Bulletin of the Technical Community on

Data Engineering

Jun 2025 Vol. 49 No. 2 **IEEE Computer Society**

Letters

Special Issue on Data-Driven Social Computing

Conference and Journal Notices

The TC on Data Engineering

Membership in the TC on Data Engineering is open to all current members of the IEEE Computer Society who are interested in database systems. The TCDE web page is http://tab.computer.org/tcde/index.html.

The Data Engineering Bulletin

The Bulletin of the Technical Community on Data Engineering is published quarterly and is distributed to all TC members. Its scope includes the design, implementation, modeling, theory and application of database systems and technology.

Letters, conference information, and news should be sent to the Editor-in-Chief. Papers for each issue are solicited by and should be sent to Associate Editor responsible for the issue.

Opinions expressed in contributions are those of the authors and do not necessarily reflect the positions of TC on Data Engineering, IEEE Computer Society, or authors' organizations.

The Data Engineering Bulletin web site is at http://tab.computer.org/tcde/bull_about.html.

Letter from the Editor-in-Chief

The central dilemma now confronting social computing is not simply technical. On one hand, human data is no longer sufficient. The corpora that sustained the past decade of AI advancement are depleted or contaminated, and the frontier of progress increasingly demands agents that learn by doing: by engaging, experimenting, and adapting in real time. This transition, described by Silver and Sutton as the onset of an *era of experience*, reorients machine learning away from static imitation and toward dynamic interaction with a changing world.

But on the other hand, the social web we rely on as a testbed for behavior, conversation, persuasion, and coordination is rapidly ceasing to be human. We are not sending agents into a reality composed purely of people. We are sending them into an ecosystem where, in short order, the majority of voices will be synthetic, the majority of interactions orchestrated by models, and the majority of "experience" recursively generated by other agents. We are entering an era where bots will train on the outputs of bots, converse with other bots, and optimize for outcomes shaped primarily by other bots' reactions.

The promise of experiential learning was that it could escape the distortions and limitations of synthetic, crowd-sourced text, by grounding AI in consequence, not consensus. But what if the consequences themselves are generated by other models? A bot that learns to persuade by practicing on other bots is not becoming more socially intelligent; it is becoming more locally efficient within an artificial microeconomy. This raises the prospect of a new kind of collapse, not in the data, but in the learning process itself. A collapse not of signal, but of relevance.

What, then, can we do?

First, we must make provenance the central feature of every social dataset, algorithm, and evaluation benchmark. The distinction between human and synthetic actors cannot remain a soft annotation or post-hoc filter. It must shape the way we define cohesion, influence, virality, and trust. Metrics built for a human-only web will mislead us in a mixed-agent world, unless we explicitly model the ontological status of each node and edge. This includes not only identifying bots but understanding their architectures, their reward functions, and their deployment context. Not all synthetic actors are alike, and not all should be treated as distortive. But none should be treated as neutral.

Second, academia must invest in open, controlled, mixed-agent environments where human behavior and synthetic behavior can be studied in relation to one another. We lack the infrastructure to observe the long-term dynamics of humans and bots coexisting in a shared social substrate. Our benchmarks are either fully human (but outdated or polluted), or fully synthetic (and therefore unreal). We need instrumented social arenas, designed with consent, governance, and transparency, where agents learn from real-time interaction with humans, and where humans can push back, adapt, and shape the reward landscape. These are not just experiments; they are prototypes of the future.

Third, reward design must become a central concern of social computing. In the era of human data, we learned from what people had already said. In the era of experience, we must decide what outcomes matter. Do we reward an agent for maximizing engagement, or for preserving the diversity of viewpoints in a thread? Do we design for short-term persuasion, or for long-term trust? Grounded rewards are powerful, but they are also dangerous if they encode proxy metrics divorced from human flourishing. As AI begins to optimize social behavior directly, we must treat reward shaping as a sociotechnical act of governance.

Fourth, we must preserve, protect, and elevate human-authored data. As synthetic content floods the web, pre-2022 corpora and verifiable human discourse will become the rare earth minerals of the AI economy: scarce, valuable, and non-renewable. Academic institutions should take the lead in curating high-fidelity, consented, and demographically representative human data, not just as training material, but as a cultural archive. This is not nostalgia. It is epistemic infrastructure.

The papers in this issue offer glimpses of how these shifts are already underway. We see how cohesion is being redefined, how community discovery is increasingly interactive and dynamic, how echo chambers and misinformation evolve in response to changing participation. We see experiments in influence-aware systems and visual abstractions that begin to accommodate complexity. Each paper speaks, in its own way, to the tension at the heart of our field: between structure and agency, between measurement and meaning, between simulation and experience.

But taken together, these contributions also remind us of our responsibility. Social computing was born as a way to understand human interaction at scale. That mandate does not disappear in the presence of synthetic actors. Instead, it becomes more urgent. If the future of intelligence is experiential, then the environments we build today will shape what intelligence becomes.

Let us ensure that what it learns is still meaningfully human.

Haixun Wang
EvenUp

Letter from the Special Issue Editor

In recent years, social platforms have evolved from niche communities into critical infrastructures that shape how individuals communicate, form opinions, and interact with the world. This transformation has led to an unprecedented volume of user-generated content and social interaction data, opening up new challenges and opportunities at the intersection of data engineering and social computing. With this backdrop, the current special issue brings together six invited articles that highlight recent advances in mining, modeling, and understanding social behavior from a data-centric perspective.

The issue begins with a conceptual bridge between social psychology and graph-based data analysis. The first paper, "A Tale of Two Cohesion: A Review of Group Cohesion in Social Psychology and Social Computing" by Zhao et al., revisits the long-standing notion of group cohesion, examining its interpretations in social psychology and tracing its algorithmic counterparts in graph mining. This cross-disciplinary perspective lays a foundation for the papers that follow, each of which addresses a concrete computational problem rooted in social interactions.

Building on this foundation, the next paper, "Community Search: A Survey of the State-of-the-Art from Algorithms to Learning, Complex Graphs, and Interaction" by Sun and Huang, surveys recent developments in community search. It charts the evolution of this task from early structural approaches to more expressive models that account for attributes, dynamics, and user interaction. The survey also offers a forward-looking view on how machine learning, and more recently, large language models, may further shape this area.

The third article, "Detection, Measurement, and Mitigation of Echo Chambers in Social Networks: A Survey" by Zhu et al., shifts the focus to the dynamics of social behavior, particularly the phenomenon of echo chambers. It offers a comprehensive taxonomy of approaches to detect and measure echo chambers, as well as strategies for mitigating their effects. By combining network structure with semantic cues, the authors provide insight into one of the most debated effects of social media on public discourse.

Misinformation is another pressing concern in social networks. The fourth article, "A Survey on False Information Detection: From A Perspective of Propagation on Social Networks" by Xie and Wang, surveys techniques for detecting false information by focusing on how it propagates. Rather than relying purely on content or user-level signals, this perspective emphasizes the structure and evolution of information cascades. The survey covers both homogeneous and cross-platform propagation and outlines emerging directions for robust, early detection.

The remaining two papers explore applications that leverage these analytical insights in real-world systems. The fifth paper, "Lightweight Influence-Aware News Recommendation in Social Media" by Feng and Cautis, introduces a lightweight, influence-aware news recommendation system tailored to the demands of social media platforms. The proposed approach integrates diffusion modeling with efficient content filtering, offering a practical balance between relevance and scalability. The final article, "Kandinsky Meets Social Conversation: Towards Abstract Art-Inspired Visualization Abstraction for Mobile Devices" by Gumtang, Tham, and Bhowmick, takes a visual and interaction-oriented turn, proposing a novel abstraction inspired by abstract art for visualizing social conversations on mobile devices. By emphasizing clarity and aesthetic simplicity, the system illustrates how data-engineering techniques can be rendered accessible to everyday users.

Collectively, the articles in this issue showcase the breadth and depth of research at the intersection of data engineering and social computing. They span foundational theory, algorithmic development, system design, and real-world applications, demonstrating how data-centric methods can yield insight into complex social phenomena and, ultimately, inform the design of more effective and responsible technologies.

Xiaokui Xiao
National University of Singapore

A Tale of Two Cohesion: A Review of Group Cohesion in Social Psychology and Social Computing

Yining Zhao [1], Sourav S. Bhowmick[1], Nastassja L. Fischer[2], and Annabel Chen Shen-Hsing[1]
[1]Nanyang Technological University, Singapore
[2]National Institute of Education, Singapore
yining002@e.ntu.edu.sg, assourav@ntu.edu.sg, nastassja.fischer@nie.edu.sg, annabelchen@ntu.edu.sg

Abstract

Community detection and community search are two prominent social computing problems aimed at identifying cohesive groups, with wide-ranging social applications. In most studies, the precise quantification of *cohesion* is fundamental to the effective identification of communities. However, although the concept of cohesion originates in social psychology, its *structural dimensions* are the only aspects that have been formally captured through cohesiveness metrics. As a result, existing algorithms may be inadequate for addressing the nuanced requirements of community-driven social applications.

In this paper, we present a narrative review of cohesion definitions and measurement approaches across both social psychology and social computing, guided by our proposed unified framework. By examining the connections and disparities between these disciplines, we draw on insights from social psychology to inform the design of cohesiveness metrics that are psychologically meaningful. This review lays the groundwork for bridging these traditionally disparate fields, fostering interdisciplinary collaboration, and advancing the development of cohesiveness metrics capable of identifying communities that embody psychology-informed cohesion within online social networks.

1 Introduction

Community detection (CD) and *community search (CS)* are two interrelated problems that focus on uncovering communities, or subgraphs, in networks. Community detection generally analyzes the entire network to identify clusters or groupings of nodes [73], whereas community search is a query-driven approach that extracts a cohesive subgraph centered around a given query node or set of nodes [189]. Identifying and exploring communities in real-world networks is essential for understanding their structural organization and has broad applications across numerous domains [73, 85, 98, 103]. This paper undertakes a narrative review of community detection and community search algorithms in the context of online social networks, where they enable practical applications such as event planning, friend recommendation, and targeted marketing [65, 161]. These efforts significantly contribute to the broader domain of social computing [162, 203]. In the discussion that follows, the terms "social computing" and "community detection and community search" are used interchangeably.

In both CD and CS, communities are usually defined using quantitative criteria prior to extraction, and only subgraphs that satisfy these criteria are included in the results. *Cohesion* is generally accepted as an inherent property of such communities [65]. While there is no universally agreed-upon definition of a community, most approaches rely on metrics that evaluate the cohesiveness of nodes in a subgraph [65, 73, 74, 110, 146, 161, 167]. Recently, advancements in graph representation learning have led to a new line of research that identifies communities by learning cohesive patterns of communities directly from training data, eliminating the need for a predefined community definition [36, 37, 77, 191, 204, 207, 228].

1.1 A Tale of Two Cohesion

The concept of *cohesion* in social computing can be traced back to its origins in social psychology. Motivated by the need to understand groups and group processes, *cohesion*, also known as *group cohesion* or *cohesiveness*, was first proposed in social psychology to describe the invisible thread that weaves individuals into the fabric of a group [69]. Cohesion has been widely studied in various types of groups (e.g., sports [26, 166], military [159], and technology-mediated learning environments [224]) and has shown its critical role in shaping our social realities, influencing group performance, organizational functioning, and reported well-being [53, 72, 143]. To distinguish this concept from the one studied in social computing, we refer to it as *psychology-defined cohesion*.

Typically, the study of cohesion considers individuals or groups as units [208]. With the emergence of *social network analysis (SNA)*[1] and increasing awareness of the significance of studying interactions among social units, researchers began exploring network-based formalizations of group structures. Cohesion was subsequently regarded as a general characteristic of subgroup members and quantified through the strength of social ties within these groups. As a result, cohesive subgroups are formally defined by analyzing different properties of the ties among subsets of individuals [208]. With the emergence of CD and CS, the structural definitions of cohesive subgroups proposed in SNA, such as *k-clique*, *k-plex*, and *k-core* [113, 208], are adopted in defining communities [65, 73, 75, 208].

However, due to the shift of the focus on groups and cohesion from social psychology to social computing, the concept of cohesion may not be consistently or comprehensively characterized [229]. In social computing, cohesion is typically quantified using only network-based information [65, 73], often overlooking fundamental social psychological factors, such as the impact of social context on individuals' thoughts, perceptions, and behaviors, that are essential to a comprehensive understanding of group dynamics [41, 235]. This narrow focus on cohesion in community detection and search algorithms calls into question their suitability for real-world social applications, which require communities made up of individuals connected by diverse motivations and a genuine commitment to sustaining their membership. For instance, in digital marketing, core components of consumer communities include a sense of belonging, shared rituals and traditions, and a sense of obligation to the community and its members [192]. In the context of event organization, studies show that online communities constructed through discussions about shared values, norms, and community boundaries contain "virtual togetherness", which can be leveraged into political action [56, 114].

The toy example in Figure 1 illustrates how cohesion is studied from the perspectives of social psychology, social network analysis and social computing. Given the close relationship between community detection (CD) and community search (CS), as well as their common shortcomings in addressing cohesion from social psychology, we focus on CS for the sake of simplicity. In group cohesion studies, social psychologists typically study groups at the group level, as depicted in Figure 1(a), where eight users form a group. Their member-to-member and member-to-group relationships constitute a social context, indicated by the blue shading. Within the group, individuals interact with each other, with their thoughts, perceptions, and behaviors shaped by the social context. The degree of cohesion can be assessed by collecting individual responses using questionnaires such as the *Perceived Cohesion Scale (PCS)* (Table 1), and aggregating these responses to the group level by taking the average (e.g., $Mean = 3.46, SD = 0.59$).

When studying the same group from a network perspective, individuals are viewed as nodes, and their interpersonal connections are depicted as edges linking them [208]. In Figure 1(b), we present one possible social network studied by social network analysts, which can be constructed from individual-reported mutual friendship choices. By using k-core with $k = 3$ to identify a cohesive subgroup, where each node is connected to at least three others in the subgraph, the resulting 3-core consists of nodes

[1] Social network analysis is "an interdisciplinary field that developed out of a propitious meeting of social theory and application, with formal mathematical, statistical, and computing methodology" [208].

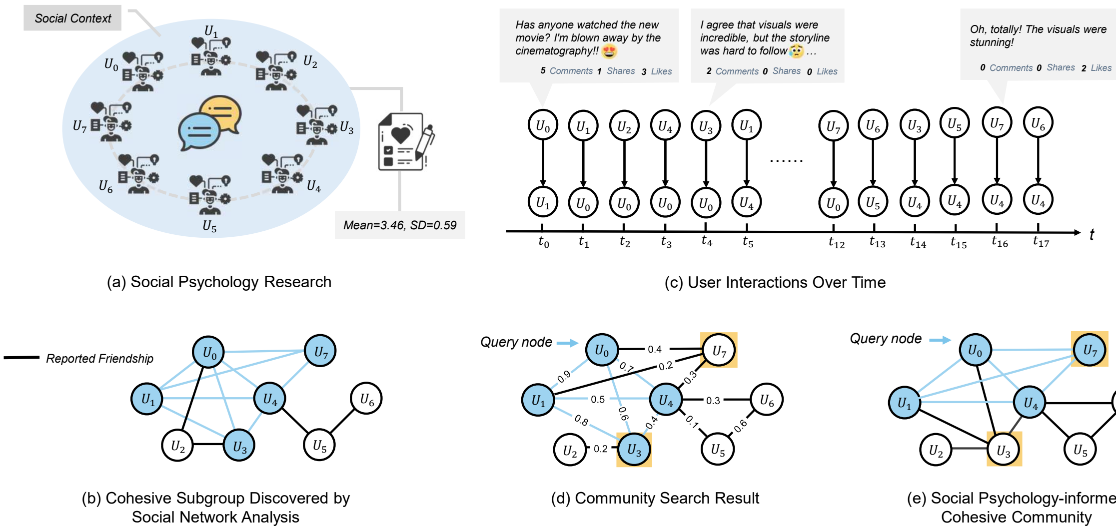


Figure 1: [Best viewed in color] Cohesion Studied in Social Psychology, Social Network Analysis, and Community Search.

Table 1: Perceived Cohesion Scale. Responses are recorded on Likert scales ranging from 0 (“strongly disagree”) to 4 (“neutral”) to 7 (“strongly agree”). The items are categorized into two groups, reflecting the sense of belonging (B) and feelings of morale (M), aligning with their adopted conceptual definition of cohesion.

Item
1. I feel that I belong to this group. (B-1)
2. I am happy to be part of this group. (M-1)
3. I see myself as part of this group. (B-2)
4. This group is one of the best anywhere. (M-2)
5. I feel that I am a member of this group. (B-3)
6. I am content to be part of this group. (M-3)

$\{U_0, U_1, U_3, U_4, U_7\}$. In this subgroup, every individual has at least three connections, indicating a tightly knit social unit.

Next, consider retrieving a community using user U_0 as the query node. Suppose the chronological interaction logs involving U_0 are available. As illustrated in Figure 1(c), each directed node pair represents a textual interaction sent from one user to another, either in the form of an initial post or a comment, expressing personal thoughts and feelings. In addition, user actions, such as likes and shares, indicate the extent to which posts resonate with other users. Figure 1(d) shows a network constructed from the interaction flow, where edge weights represent the interactiveness between user pairs, along with a community search solution [50, 63, 194]. Specifically, two cohesiveness metrics can be defined: a *structural-based cohesiveness metric* (k-core) and an *interactiveness cohesiveness metric*, operationalized by an *interactiveness score* that exceeds a user-defined threshold for each node pair. Using $k = 3$ and a threshold of 0.4, a subgroup $\{U_0, U_1, U_3, U_4\}$ can be identified (represented by blue nodes connected by blue weighted edges). However, although individuals in the subgroup engage with at least three others within the group and demonstrate relatively high interaction intensity, their perceptions of the group, as well as how their behaviors and attitudes are shaped by the broader social context, lack explicit and systematic characterization [66, 92]. As a result, interpreting the identified community as genuinely cohesive remains open to question. A closer look at the interactions in Figure 1(c) reveal that U_3's opinion towards the movie differs from others, including U_0. Despite receiving two comments, the post attracted no shares or likes. This lack of positive feedback may reduce U_3's willingness to engage further and affect the perception of belonging, potentially leading to his or her disconnection from the group. Consequently, a solution that includes U_3 may not necessarily be cohesive. Note that while such group dynamics are well studied in psychology-defined cohesion, no existing tools provide effective guidance for CS to identify such cohesive communities in online social networks [229].

1.2 Motivation and Overview of the Review

Based on the above observations, it is paramount that communities in online social networks should reflect cohesion as defined in social psychology. Therefore, community detection and search metrics ought to be designed to capture this form of cohesion. However, a recent experimental evaluation of representative community search algorithms on online social networks showed no clear correlation between structural and psychological cohesiveness, and none of the algorithms successfully identified communities that are psychologically cohesive [229]. Additionally, developing effective metrics to capture such cohesion faces several challenges. First, there is no comprehensive comparison between existing social computing metrics and psychology-defined cohesion. Second, no framework is available to guide the systematic development of effective cohesiveness metrics, despite extensive psychological research on measuring cohesion across various contexts and scopes.

To address the challenges in developing effective cohesiveness metrics, in this paper, we carry out a narrative review of the cohesion construct studied in social computing and social psychology, concentrating on its definition and measurement. We seek to shed light on the chasm between current cohesiveness metrics that social computing uses within online social networks and those adopted by social psychology, offering insights for more effective cohesiveness metric development. It is anticipated that integrating psychology-informed cohesiveness metrics into social computing algorithms would yield results that diverge from those of existing methods. For example, Figure 1(e) presents a possible psychology-driven community search result, where user U_3 is excluded and U_7 is included. According to social psychology, the inclusion can be explained by shared movie preferences and positive "likes" interactions between U_7 and other members, which reinforce social ties, cultivate a sense of belonging, and ultimately strengthen the overall cohesion of the community.

The rest of the paper is organized as follows: Section 2 reviews existing surveys about group cohesion in social psychology and social computing. Section 3 introduces the key terminologies used in our study. Section 4 illustrates the methodologies utilized for conducting the review. Sections 5 to Section 8 detail the group cohesion studied by both fields. Finally, Section 9 concludes the paper.

2 Related Work

This section briefly overviews surveys in social psychology, community detection, and community search, focusing on the definitions and measurement of cohesion. The differences between our work and existing studies are also highlighted.

2.1 Surveys in Social Psychology

In social psychology, cohesion is typically reviewed from a general perspective, focusing on its theoretical definitions and measurement issues [27, 48, 53, 55, 59, 96, 175], and its relationship to other constructs

[72, 130, 131, 143, 177]. Some studies also explore cohesion in specific contexts, such as sports [26, 166], the military [71, 159], and technology-mediated learning environments [224].

Previous studies have proposed several frameworks for reviewing cohesion. For example, Drescher et al. [55] applied multidimensional process classification system from [23] to identify key parameters for measuring cohesion across four dimensions: *Person* (the unit of observation and analysis), *Variable Function* (whether cohesion is treated as an antecedent or response variable), *Measurement Strategy* (information used to measure cohesion), and *Time* (whether time is held constant and how often the cohesion is measured). In [175], cohesion measurement is reviewed in the order of conceptualization, measurement dimensionality, operationalization, level of measurement, and level of analysis. The discussion also covers the temporal considerations, as well as logistical and practical issues related to data collection and analysis. This framework is further extended to include other research components that can help structure the cohesion studies, such as methods for studying group cohesion (i.e., obtrusive, unobtrusive, or mixed), data collection formats (i.e., surveys, log data, or both), and study types (i.e., longitudinal, cross-sectional, or design-based) [224].

Nevertheless, these surveys are limited to psychological works and do not involve the application of group cohesion concepts in other disciplines. Moreover, no standardized framework exists for reviewing cohesion definitions and measurements, which is crucial for comparing process variables across cohesion studies and integrating related findings. Our review proposes a unified framework built on advances in psychological research and applies it to analyze how cohesion is studied in both social psychology and social computing. This approach not only clarifies how cohesion is defined but also provides a step-by-step account of the development of its measurement.

2.2 Surveys in Community Detection and Community Search

Since 1955 [209], community detection (CD) has attracted sustained attention due to the prevalence of real-world communities and its practical applications, prompting a variety of surveys that summarize advancements in the field. Typically, these surveys summarize community definitions in existing papers, classify CD methods based on various criteria, and discuss their empirical validation and potential applications [7, 44, 73, 89, 98, 103, 146, 161, 167, 173, 191]. However, in these surveys, the term "cohesion" is either overlooked or narrowly defined in terms of dense structures, occasionally supplemented by high similarity [9, 19, 44, 76, 110, 146, 161, 167].

Only one survey reviews existing community search (CS) studies [65]. It systematically categorizes them based on the cohesiveness metrics they use (i.e., structural, and if considered, attribute-based), compares and analyzes these metrics, and assesses the corresponding methods on both simple and attributed graphs. However, its treatment of cohesion is highly field-specific, and it lacks a comprehensive evaluation of the effectiveness of these metrics.

To support social applications with more effective cohesiveness measures, our study sets itself apart from previous surveys by emphasizing the definition of cohesion and the development of related metrics in both social psychology and social computing. This perspective allows us to uncover limitations in the metrics currently used in social computing and to identify opportunities for their improvement.

3 Background

To set the groundwork for our review, this section introduces essential terminology from social psychology and graph theory, as well as techniques that consider cohesion to varying degrees.

3.1 Social Psychology Terminology

Concepts and Constructs. A *concept* is an idea expressed through symbols or words [14]. Concepts can vary in their level of abstraction, from precise and objective to highly abstract. A *construct* is an abstract concept developed to explain specific phenomena. In psychology, constructs, such as cohesion in our context, often reflect internal processes like thinking, feeling, or acting in specific ways, and cannot be directly observed or measured [183]. Although these two terms have distinct meanings, they are often used interchangeably in the literature [152]. Consequently, we do not differentiate between them in our paper.

Conceptualization. The process of transforming fuzzy and imprecise concepts into precise and concrete terms is known as *conceptualization* [14, 15]. In other words, conceptualization is a process of defining the study concept. The resulting construct is also termed as a "conceptual definition" or "theoretical definition", often simply called conceptualization [152, 183]. During the conceptualization process, constructs are characterized as *unidimensional* or *multidimensional*. A *unidimensional construct* has a single underlying dimension, while a *multidimensional construct* consists of two or more dimensions, each measured independently and subsequently combined to form the overall construct.

Operationalization. *Operationalization* is the process of formulating operational definitions after a theoretical construct has been clearly defined. These definitions, or measures, are concrete indicators or items that operate at the empirical level to measure the construct and imply procedures for collecting relevant data [14, 15, 51, 152]. Generally, there are three types of such measures: *self-report*, *behavioral*, and *physiological* [183, 184]. *Self-report* measures are those in which participants report their thoughts, feelings, and actions through interviews or questionnaires. *Behavioral* measures refer to the recording or observation of participants' behaviors, while *physiological* measures record participants' physiological processes, such as heart rate and facial muscle activity.

Reliability and validity. *Reliability* and *validity* are two general criteria for evaluating measures to ensure they consistently and accurately reflect the study construct. *Reliability* requires that the measured results can be replicated under identical or highly similar conditions [119]. Common methods for assessing it include *test-retest reliability*, *internal consistency reliability*, and *inter-rater reliability* (see Table 2 for definitions) [14, 183]. *Validity* reflects how accurately the measured social reality matches the study construct. It can be assessed theoretically and empirically [14]. Theoretical assessments evaluate how well a conceptual definition corresponds to its operational measures, including *face* and *content validity*. Empirical assessments, on the other hand, collect empirical data to evaluate measures from four aspects: *factorial validity*, *criterion validity*, *convergent validity*, and *discriminant validity* [14, 183]. *Criterion validity* can be further split into *predictive* and *concurrent validity*, depending on whether the studied measures and criterion are implemented simultaneously or asynchronously [119, 183]. *Convergent* and *discriminant validity* are often used together to assess related and pre-existing constructs [14]. These two types of validity are integral components of *factorial validity*, which is especially important for newly established constructs [78]. Table 3 summarizes the types of validity and their definitions. Ideally, measures should demonstrate both high reliability and validity. However, achieving consistent measurement results is often easier than ensuring accuracy, and these two criteria may sometimes conflict, necessitating a careful balance.

3.2 Graph Terminology

Next, we briefly introduce relevant graph terminology. For clarity of exposition, we illustrate the terminology on a small, simple, and undirected graph $G(V, E)$, with node set denoted as V and edge set as E, as depicted in Figure 2.

Table 2: Types of Reliability.

Reliability Type	Definition
Test-retest reliability	The degree of consistency between two assessment scores measuring the construct within the same sample at different time points.
Internal consistency reliability	The degree to which the rated items of the same constructs correlate with each other.
Inter-rater reliability	The degree to which two or more independent raters or observers agree on the same construct.

Table 3: Types of Validity.

Assessment Type	Validity Type	Definition
Theoretical	Face validity	The degree to which an indicator appears to reasonably measure its underlying construct.
	Content validity	The extent to which the full content of a definition is represented in a measure.
Empirical	Factorial validity	The extent to which the theoretical structure of a construct is reflected in the factor structure of a test or measure.
	Criterion validity	The extent to which measurement results correlate with an expected external variable (a.k.a. criterion). It includes two subtypes: predictive validity, when the criterion is measured after the construct, and concurrent validity, when the criterion is presumed to occur simultaneously with the construct.
	Convergent validity	The extent to which a measure aligns with other measures of the same or related constructs.
	Discriminant validity	The extent to which a measure discriminates from other constructs that are not expected to be measured.

k-clique. A clique is a complete subgraph that contains at least three nodes, where all nodes are adjacent to each other [134, 155]. A clique with k nodes is also called a *k-clique* [172]. In the graph displayed in Figure 2(a), the subgraph $\{U_2, U_3, U_4, U_5\}$ forms a 4-clique, with any three of its nodes form a 3-clique. $\{U_0, U_1, U_3\}$ is another example of a 3-clique. While the clique is a foundational idea for studying cohesive subgroups in social networks, its definition is often considered too strict: it overlooks node differences by imposing identical linkage requirements, making extracted subgraphs less informative. As a result, researchers have proposed different relaxations of the k-clique [208].

k-plex. One relaxation of the k-clique based on node degree is the *k-plex* [181]. A subgraph $S(V_S, E_S)$ is a *k-plex* if every node has at least $|V_S| - k$ neighbors within the subgraph [206]. A clique is a special case of a k-plex when $k = 1$, as illustrated in Figure 2(b), where the 1-plex (also a 4-clique) includes vertices $\{U_2, U_3, U_4, U_5\}$. Two additional 3-plexes are $\{U_0, U_1, U_2, U_3, U_6\}$ and $\{U_0, U_1, U_3, U_4, U_5\}$, where each node lacks no more than 2 internal connections.

k-core. The *k-core* is another relaxation of the k-clique based on node degree [180]. Let $deg_G(v)$ denotes the degree of a node v in G. A *k-core* is the largest induced subgraph $H \subseteq G$ where each node is adjacent to at least k nodes, i.e., $\forall v \in V_H, deg_H(v) \geq k$. In Figure 2(c), the largest 2-core subgraph consists of the nodes $\{U_0, U_1, U_2, U_3, U_4, U_5, U_6\}$, with each node connected to at least two other nodes. Within it, a denser 3-core, $\{U_2, U_3, U_4, U_5\}$, can be identified.

k-truss. The *k-truss* is also a relaxation of the k-clique, motivated by a natural observation of social cohesion: strongly tied actors may share connections with others [46]. It is defined as the largest subgraph of G in which every edge is contained in at least $(k-2)$ triangles [65]. For example, in Figure 2(d), a 3-truss includes nodes $\{U_0, U_1, U_2, U_3, U_4, U_5\}$, with each edge belongs to at least one triangle. When $k = 4$, a 4-truss can be found with nodes $\{U_2, U_3, U_4, U_5\}$.

k-ECC. A *k-edge-connected graph* (or *k-ECC*) belongs to a different genre of cohesive subgroup definition emphasizing edge connectivity [233]. For two vertices $u, v \in V$, the edge connectivity $\lambda(u, v)$ is the minimum number of edges whose removal disconnects u and v. A *k-ECC* is thus a connected subgraph

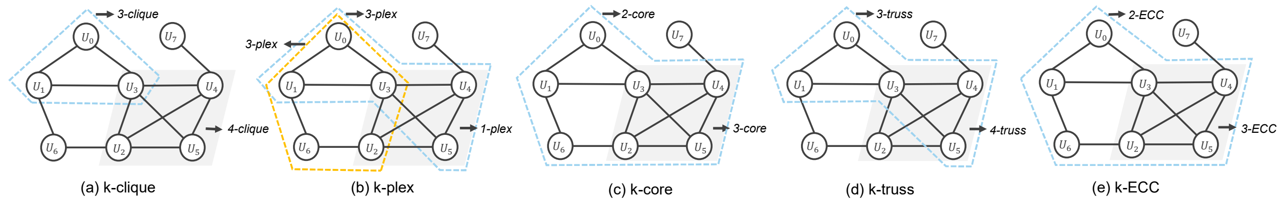


Figure 2: Cohesive Subgraphs in a Simple Graph.

that remains connected unless at least k edges are removed. As highlighted in grey in Figure 2(e), the subgraph of $\{U_2, U_3, U_4, U_5\}$ is 3-ECC, since at least three edges must be removed to disconnect it. The 2-ECC includes vertices $\{U_0, U_1, U_2, U_3, U_4, U_5, U_6\}$.

3.3 Social Network Analysis

Social network analysis (SNA), rooted in sociology and social psychology, is an interdisciplinary field that integrates social theory, formal mathematics, statistics, and computing methodologies [113, 208]. Unlike social psychology, which often treats social groups as unified entities, SNA focuses on social networks composed of individual actors and the relationships among them. Typically, relations between pairs of social actors have both form and content, depicting how information is exchanged to establish social reality [113]. The patterns of these relationships form network structures, which are described using structural variables to better understand both individuals and the group as a whole.

In SNA, one important task is to identify cohesive subgroups within a network, where cohesion is treated as a general property, quantified by the properties of social ties. Cohesive subgroups can thus be formalized by examining various tie properties among group members. As defined in [208], cohesive subgroups in social networks are *"subsets of actors connected by relatively strong, direct, intense, frequent, or positive ties"*. The formalization of a cohesive subgroup begins with a clique [68, 134], as discussed in Section 3.2. Due to its limited applicability for strict requirements [2], many alternative definitions have been proposed, including the *k-clique*, *reachability*, *diameter*, *n-clan*, *n-club*, *k*-plex, *k*-core, *LS Set*, and *Lambda Set* (see [208] for details).

It is worth noting that in SNA, network data, collected through questionnaires, interviews, and observations [113, 208], contains abundant social and psychological information. Questionnaires are especially common when studying human subjects, often including questions about their ties to other actors, involving questions such as whom they consider as close friends, feel confident about, or would like to seek advice from [4, 208]. However, self-reported social network data is susceptible to accuracy, validity, and reliability, as respondents may encounter challenges such as obtrusiveness, expansiveness bias[2], artificiality [11, 12, 141, 142]. These issues make it even more challenging to collect large-scale, long-term network data [11, 190].

3.4 Community Detection

The *community detection (CD)* problem stems from the recognition that real-world networks commonly exhibit underlying community structures with practical applications. It identifies communities, or clusters, within a graph by leveraging embedded network information [9, 44, 75, 76, 98, 161, 167]. Therefore, this problem is also referred to as graph/network clustering [73, 74].

Community detection is an ill-defined problem with no universal quantitative definition of object communities. The guiding principle behind most community definitions is that vertices within a subgraph

[2]The expansiveness bias is defined as the tendency to overreport and underreport one's interactions with others [67].

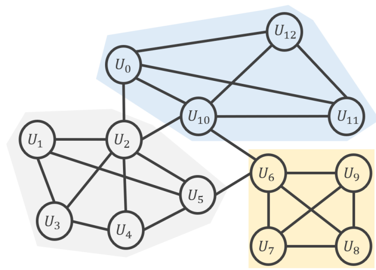


Figure 3: [Best viewed in color] Illustration of the Differences Between Community Detection and Community Search.

have more internal edges than links to the rest of the graph [73, 191]. Traditionally, target communities are defined based solely on graph topology and are classified as *local*, *global*, or based on *vertex similarity*. *Local* definitions, drawn directly from SNA, focus on subgraphs and the cohesion among internal vertices, disregarding the rest of the graph. *Global* definitions, on the other hand, concern the entire graph and assume that a community structure exists if the graph significantly differs from a random one. This can be verified by a *null model*, with *modularity* being the most popular one [153]. Lastly, communities can be identified based on the structural similarities shared by vertices (i.e., Euclidean distance, cosine similarity, Pearson correlation) [73, 161]. A simple example of traditional community detection is shown in Figure 3, where applying the k-core based algorithm with $k = 3$ partitions the network into three 3-core components, each highlighted in a different color.

However, traditional structure-based community definitions are argued by some to be insufficient to capture the forces for nodes to form communities, while incorporating node attributes alongside structural information from the real-world networks can result in more informative and robust community detection [44, 103]. Meanwhile, the results detected by algorithms designed with various metrics can also be regarded as communities, without the need for a predefined definition [73, 74, 186]. In recent years, recognizing that traditional methods struggle to capture the inherent features of target communities due to the lack of a unified definition, researchers have developed a series of learning-based approaches that identify communities directly from the network structure and attribute information, without requiring prior knowledge. This enables the discovery of more diverse communities [106, 191].

It is worth noting that, regardless of the methods used in community detection, communities are typically identified based on structural and, when available, attribute information. Few studies consider the concept of cohesion, and when they do, it is usually limited to structural density [40, 47, 91, 151, 210, 234].

3.5 Community Search

Community search (CS) is a query-dependent variant of the community detection problem, addressing the inefficiency of applying traditional CD methods to complex networks while offering enhanced customization capabilities [65]. First introduced in [189], CS aims to retrieve a cohesive subgraph from an input graph based on a query request, without traversing the entire graph. The original community search problem is formalized from a combinatorial optimization perspective as follows:

Problem 3.1: ***[Community Search (CS) Problem]*** Given an undirected graph $G = (V, E)$, a set of query nodes $Q \subseteq V$, and a goodness function f, the community search problem seeks to find an induced subgraph $H = (V_H, E_H)$ of G, such that

(i) V_H contains Q ($Q \subseteq V_H$);

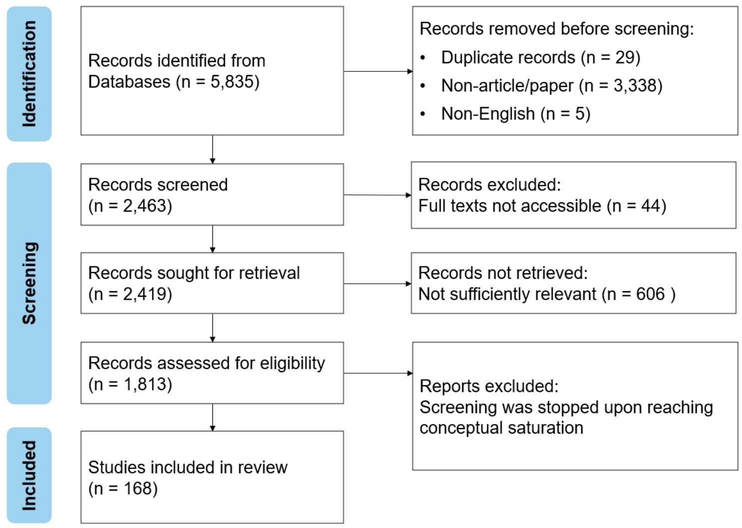


Figure 4: PRISMA Flow Diagram of the Study Selection Process.

(ii) H is connected;

(iii) $f(H)$ is maximized among all feasible choices for H.

In this formulation, the function f measures the goodness of the extracted community H, with larger values indicating denser internal connections [189]. The last two requirements are summarized subsequently in [65] as the *connectivity* and *cohesiveness* properties a community should satisfy. Common cohesiveness metrics in CS, namely k-core, k-truss, k-clique, and k-ECC, align with the cohesion formalization in SNA and local definitions of community in community detection. As shown in Figure 3, with query parameters set to $q = \{U_0\}$ and $k = 3$, a CS algorithm can retrieve a community consisting of vertices $\{U_0, U_{10}, U_{11}, U_{12}\}$, shaded in blue. Apart from structural cohesiveness metrics, *attribute-based* cohesiveness metrics have been developed to identify cohesive subgraphs by incorporating supplementary network information such as keywords, weights, profile, and temporality [65], as exemplified in Figure 1. In recent years, similar to community detection research, learning-based methods for community search have attracted increasing attention due to growing awareness of two key limitations in earlier approaches: structural inflexibility and attribute irrelevance. These models address these limitations by integrating both structural and attribute information into neural networks, eliminating the need for predefined patterns or rules [62, 77, 90, 105, 124, 204, 207].

4 Methodology

In this section, we first introduce the literature search and inclusion criteria for cohesion studies in social psychology, community detection, and community search, respectively. Then, we outline the review framework that guides the examination of selected studies.

4.1 Literature Search and Inclusion Criteria

In this paper, we adopted a narrative review methodology to provide a more selective survey of the literature on cohesion in both fields. This approach, as highlighted in [157, 163], is a great starting point to bridge related areas, stimulate theoretical development, and guide future research. Figure 4

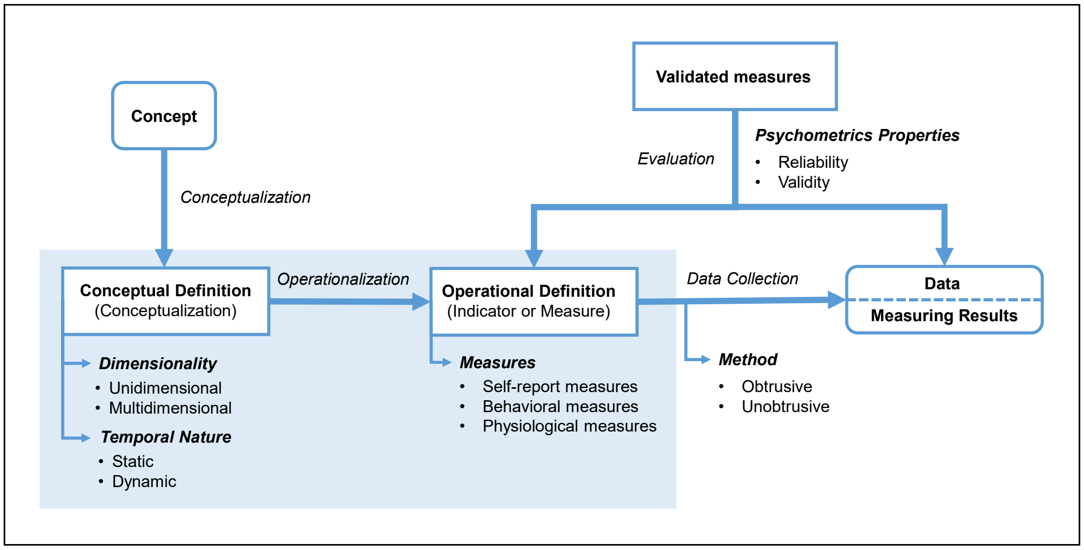


Figure 5: A Unified Framework for Reviewing Cohesion in Social Psychology and Social Computing.

presents the trajectory of the literature search and selection process, following the PRISMA guidelines [160]. To identify relevant studies on group cohesion, we used the search terms "cohesive", "cohesiveness", and "cohesion" in Google Scholar. For literature on community detection and community search, we retrieved studies from both Google Scholar and DBLP. The search terms for community detection were "community detection" and "detecting communities", while "community search" and "searching communities" were used for community search studies.

Before the screening procedure, three criteria were applied to the initially identified records: (a) duplicate papers were removed, (b) only journal articles or conference papers were retained, and (c) only English-language literature was included. The remaining records (n = 2,463) were screened based on full-text availability, and inaccessible articles were excluded.

For psychological studies, we included those that (a) adopted a concept of cohesion consistent with psychological definitions in their measurement approach and (b) measured cohesion directly. Social computing studies were included if their main focus was the development, evaluation, or review of community detection or search algorithms, rather than the application of such algorithms as tools for other tasks. Additionally, for survey papers on group cohesion and community detection/search, we applied snowballing techniques to include related studies that were not retrieved through our initial sources. This screening procedure yielded 1,813 papers eligible for inclusion in our study. The screening process was concluded once a clear understanding of research progress in both fields had been achieved. Focusing on representative rather than exhaustive coverage, we identified 168 relevant papers: 98 on group cohesion, 35 on community detection, and 35 on community search.

4.2 Review Framework

Following the thread that links group cohesion studies to social computing, as mentioned in Section 1, we propose a unified review framework to structurally examine the cohesion definitions and their measurement development across both fields, as shown in Figure 5.

The framework draws on the construct measurement processes from sociology and psychology research methods [15, 183], which begin with developing a conceptual definition of the construct, followed by operationalizing it into measures, implementing the measures, and finally evaluating them. To support the review, we also classify selected procedures and intermediate results, as has been frequently discussed

Table 4: Sampled Unidimensional Cohesion Conceptualizations with Keywords Highlighted.

Author	Conceptualization
Seashore [179, p. 97]	"... group cohesiveness was conceptualized as **attraction of members to the group** in terms of the strength of forces on the individual member to remain in the group and to resist leaving the group."
Lott [131, p. 408]	"Cohesiveness is defined as that group property which is inferred from the number and strength of **mutual positive attitudes among the members of a group**."
Cartwright [30, p. 91]	"The degree to which **the members of the group desire to remain in their group**."
Budman et al. [22, p. 202]	"Cohesion, according to the scale, is defined as **group connectedness**, demonstrated by working together toward a common therapeutic goal, constructive engagement around common themes, and openness to sharing personal material."
Forsyth [72, p. 219]	"Cohesion is just **a group's degree of unity**."

in existing surveys of psychological cohesion, into our framework [27, 53, 55, 71, 72, 83, 96, 143, 159, 166, 175, 177, 224].

Therefore, in our survey, we examine cohesion research by tracing how studies progress from conceptualization through operationalization to implementation and evaluation of these measures. Note that once the measures are established, they will be implemented to collect the data and, in some cases, to compute results. Measure evaluation may occur after its development, following data collection, or at both stages, depending on the specific methods used [15, 55, 152, 175, 183, 185, 224].

Note that although some psychology studies employ multiple research methods to investigate several variables, including cohesion, we only document the particulars concerning cohesion. For instance, in [97], self-reported data were collected to measure cohesion, while content analysis was used to assess groupthink; only the adoption of the obtrusive data collection is recorded in our survey. Moreover, it is necessary to clarify that our review framework, developed based on psychological research, does not fully align with the development of cohesiveness metrics in social computing due to their differing research perspectives. Specifically, unlike psychologists who study a construct by developing both conceptual and operational definitions, researchers in social computing directly propose formalized cohesiveness metrics or directly design learning-based models without an explicitly stated conceptual basis [65, 73, 106, 191, 208]. Therefore, we treated the definitions yielded from conceptualization and operationalization as a block to review, collectively referred to as "cohesion definitions" and marked them by a blue-shaded rectangle in Figure 5. To ensure alignment between the two domains, we manually extracted the conceptualizations of cohesion in social computing from the adopted cohesiveness metrics and separately examined how cohesion is addressed in learning-based models (see Section 6).

5 Cohesion Definitions in Social Psychology

In this section, we review group cohesion studies for their conceptual and operational definitions.

5.1 Conceptualizations

5.1.1 Cohesion Dimensionality

Cohesion conceptualizations can be categorized as unidimensional or multidimensional based on the number of dimensions used to define the construct. Despite extensive efforts to study cohesion since its inception [70], psychologists have yet to reach a consensus on its dimensionality [27, 48, 72, 175].

Sampled unidimensional cohesion conceptualizations are summarized in Table 4. Among these, the conceptualizations introduced in [179], [131], and [30] remain widely adopted in subsequent work [32, 87, 94, 202]. Unidimensional conceptualizations facilitate distinguishing cohesion from related concepts and are easy to operationalize [48, 143]. However, they may either be too ambiguous to capture the full complexity of cohesion or overly focus on a single factor, typically interpersonal attraction, at the expense of others [48, 59, 95, 96, 149, 200]. Besides, they are criticized for their limited applicability to certain group types [48]. For example, cohesion conceptualized in the therapeutic context as *"group connectedness, demonstrated by working together toward a common therapeutic goal, constructive engagement around common themes, and openness to sharing personal material"* [22], cannot be directly applied to educational or work groups without adaptation. Moreover, the inherently unidimensional nature of these constructs makes it difficult to compare them [48]. Consequently, most group cohesion studies have now embraced multidimensional conceptualizations.

Table 5: Sampled Multidimensional Cohesion Conceptualizations with Keywords of Each Dimension Highlighted.

Author	Conceptualization
Festinger [69, p. 7]	"Cohesiveness of a group is here defined as the resultant of all the forces acting on the members to remain in the group. These forces may depend on **the attractiveness or unattractiveness of either the prestige of the group, members in the group, or the activities in which the group engages**."
Carron [26, p. 213]	"... cohesion can be defined as a dynamic process which is reflected in **the tendency for a group to stick together** and **remain united in the pursuit of its goals and objectives**."
Zaccaro and Lowe [222, p. 838]	"... **interpersonal cohesiveness** represents the degree to which positive interpersonal relationships exist among members of the group. **Task-based cohesiveness** results when group membership provides for the personal attainment of important goals or when there is a 'shared commitment' to the task of the group."
Bollen and Hoyle [18, p. 482]	"Perceived cohesion encompasses **an individual's sense of belonging to a particular group** and **his or her feelings of morale associated with membership in the group**."
Bliese and Halverson [17, p. 1174]	"Vertical and horizontal cohesion are constructs reflecting the quality of relationships within groups. **Vertical cohesion** reflects subordinates' perceptions that leaders are considerate and competent. **Horizontal cohesion** is a measure of the degree of fraternal bonding and kinship within a group."
Griffith [84, p. 165]	"At the theoretical level, these cohesion scales hint at a two-dimensional concept of military-unit cohesion ...: **(a) the direction of cohesion**-contrasting vertical cohesion (referring to superior-subordinate relations) and horizontal cohesion (referring to peer relations); and **(b) the functions of cohesion**-contrasting instrumental cohesion (relating to task performance) and affective cohesion (relating to interpersonal support)."
Von Treuer et al. [200, p. 123]	"... cohesion was revealed to be a multidimensional construct made up of several elements... The elements were **(1) commitment to goal, (2) identify and respect, (3) interpersonal warmth, (4) belonging, (5) team success, and (6) proud of team identity**."

The multidimensional implication of cohesion can be traced back to its original 1950 definition, which stated that forces acting on members may depend on *"the attractiveness or unattractiveness of either the prestige of the group, members in the group, or the activities in which the group engages"* [69]. This multidimensional view has been widely adopted in subsequent studies, with some defining cohesion as bi-dimensional and others proposing three or more dimensions. A sample of these is listed in Table 5.

Commonly adopted bidimensional conceptualizations divide cohesion into task vs. social cohesion [26], belongingness vs. morale [18], and vertical vs. horizontal cohesion [17, 84]. Among these conceptualizations, the differentiation between task and social cohesion has gained widespread acceptance [13, 53, 60, 88, 107, 231, 232], where social cohesion assesses positive interrelationships among group members, while task cohesion captures the level of commitment of group members to achieving shared goals [26, 222].

An example of a multidimensional conceptualization is the six-dimensional model developed by Von Treuer et al. [200] using the *repertory grid technique* (RGT[3]) and *weighted multidimensional scaling*

[3]The RGT is "a form of structured interviewing, with ratings or without, which arrives at a precise description uncontaminated by the interviewer's viewpoint" [102]. It is used to determine the interrelations and structure of various cohesion dimensions [200].

(WMDS[4]). The six identified elements of cohesion are: (1) commitment to goals, (2) identification and respect, (3) interpersonal warmth, (4) belongingness, (5) pride in team success, and (6) pride in team identity.

Multidimensional models align more closely with existing knowledge on group cohesion and more effectively explain their relationship with response variables, such as decision-making, influence processes, and group dynamics [175]. However, as Cota et al. [48] noted, cohesion conceptualization can be affected by selecting either overly broad or excessively narrow items based on empirical criteria. This issue is further constrained by reliance on multivariate techniques such as factor analysis. Moreover, the growing number of proposed multidimensional models has not been matched by efforts to evaluate or compare them, making it challenging to identify the optimal models or integrate findings.

5.1.2 Temporal Nature

Time is a thread running through group processes to help understand how groups form, develop, and stabilize over time [5, 197]. We refer to this temporal aspect of conceptualization as the "temporal nature", which can be either static or dynamic, independent of its dimensionality. In general, cohesion is recognized as a dynamic construct that evolves as individuals interact, with perceptions of group cohesion becoming clearer [28, 117, 140, 175, 224]. This dynamism is further complicated by the varying contributions of multidimensional cohesion's components across different developmental phases, which may also differ among similar groups [27, 175]. Yet, most conceptualizations address this dynamic nature implicitly, only with one exception [26].

Given that cohesion is an emergent state[5] that evolves within groups [116, 140], our review reveals that most studies only collect data once, treating it inconsistently as a static construct. However, there is still a group of studies that emphasize its developmental aspect by examining how it changes across different stages of group development [21, 34, 57, 87, 136, 198, 224, 231]. For example, Lyles et al. [136] explored the longitudinal relationships between group cohesion dimensions and interaction variables (i.e., communication, cooperation, and competition) to inform strategies that promote physical activity within virtual communities.

5.2 Operational Definitions

Recall that operational definitions, also known as indicators or measures, precisely specify how cohesion is measured [14, 15, 119, 183]. The types of measures, including self-report measures, behavioral measures, and physiological measures [183, 184], may inform the methodologies used for subsequent data collection and how they vary in their effectiveness at capturing cohesion.

Among these three measures, researchers predominantly rely on self-report measures, particularly questionnaires, to quantify cohesion [14]. Question construction involves several stages, including determining the information needed and the way of administration, drafting questions, revising through multiple iterations, presenting, and final editing before being put into use [183]. Taking the development of the *Group Environment Questionnaire* (GEQ) [29], one of the most widely applied questionnaires, as an example. Based on the proposed conceptual model of group cohesion, the GEQ was constructed by first collecting the representative cohesion information related to it. This information was then used to develop items and ensure content validity. Next, the resulting questionnaire underwent a variety of item analytic procedures, ultimately yielding an 18-item instrument [29].

[4]WMDS is an exploratory statistical analysis used to generate the best-fitting model of the relationships between the elements, based on the similarity of their ratings [200].

[5]The term "emergent state" signifies "constructs that characterize properties of the team that are typically dynamic in nature and vary as a function of team context, inputs, processes, and outcomes". [140]

Behavioral measures of cohesion include externally rated scales and digital behavioral indicators, with the main challenge of their design being to fully and accurately describe studied behaviors. An example of the former is the *Group Cohesiveness Scale* (GCS) adopted in group therapy studies, where external raters assess cohesion by observing cohesive behaviors within the group [22]. The scale was developed through an iterative process of operationalization, including scale development, pilot testing by raters, and subsequent revisions. Consistent with the multidimensional nature of cohesion, the GCS includes five subscales: (a) Withdrawal and Self-Absorption vs. Interest and Involvement, (b) Mistrust vs. Trust, (c) Disruption vs. Cooperation, (d) Abusiveness vs. Expressed Caring, and (e) Unfocused vs. Focused, along with a global scale: Global Fragmentation vs. Global Cohesiveness. Practices using digital behavioral indicators are primarily found in online environments and are based on user-generated actions and content. For example, Altebarmakian and Alterman [3] connected participant behaviors with degree and manner of cohesion in the working groups and proposed three behavioral measures, namely behavioral engagement, interactivity, and cognitive engagement, to evaluate cohesion within working groups. In a study by Zamecnik et al. [225], temporal network motifs[6] that characterize the synchrony of team engagement are clustered to measure cohesion.

Self-report measures align well with the nature of cohesion, which is primarily rooted in individuals' perceptions of group members [118], despite their potential inaccuracies due to the individuals' subjectivity, including self-deception and memory biases [165]. However, given the dynamic nature of cohesion and the availability of longitudinal human interaction data, behavioral measures warrant further exploration. Evidence demonstrates that individual behaviors and social interactions captured by wearable social sensors can predict self-reported cohesion in longitudinal missions [227]. Meanwhile, direct physiological measures are notably absent [3], with only preliminary efforts showing that physiological synchrony, such as the interbeat intervals (IBIs) of the heart [82, 196] and smiling [147], can predict self-reported cohesion.

While neither self-report nor behavioral measures are ideal for capturing cohesion, combining both can help offset their individual limitations. Especially, well-developed self-report measures should be supplemented by behavioral ones [116]. For instance, Prewett [168] used both approaches to quantify cohesion and examine its relationship with performance. Their findings revealed strong consistency between self and observer ratings, both of which significantly predicted team errors and backup behaviors.

6 Cohesion Definitions in Social Computing

In this section, we review the study of cohesion in social computing, compare it with psychological research, and analyze insights for designing cohesiveness metrics. We start with the conceptual definitions manually extracted from existing studies, followed by their operational definitions. Next, we examine how learning-based methods address cohesion. Lastly, we outline two additional social computing techniques used to measure cohesion.

6.1 Conceptualizations

Unlike social psychology research, which formally proposes or adopts cohesion conceptualizations, community detection/search studies directly incorporate the understanding of cohesion into quantifiable metrics to characterize target communities, lacking explicit definitions. To identify the gaps and connections of cohesion between the two fields, we manually extract the cohesion conceptualizations

[6] Network motifs refer to statistically significant repeating structures within a larger network that characterize its composition [145].

Table 6: Extracted Cohesion Conceptualizations in Community Detection Studies.

Dimensionality	Temporal Nature	Extracted Conceptualization	Work
Unidimensional	Static	Vertices have more internal connections than links to the outside.	[40, 151, 187, 210, 213, 217, 234]
		Each edge belongs to at least k-2 triangles inside the subgraph.	[47]
Multidimensional	Static	Nodes with both similar semantic context and close internal relationships.	[221]
		Nodes with both high degrees of structural and content similarities.	[100]
	Dynamic	A core of people that remain in them over different time periods.	[42]
		People influence each other over time and eventually converge due to the presence of a "unity of will" or "sharing common values".	[182]

from existing cohesiveness metrics in community detection and search research, and categorize them using our review framework, as outlined in Tables 6 and 7, respectively.

In social computing, the categorization of cohesion conceptualizations depends on the number of aspects considered to characterize communities. Typically, unidimensional cohesion is defined structurally by the density of connections among nodes, the distance between nodes, the fraction of neighbors, or the count of triangles, as listed in Table 6 and 7. In contrast, psychological definitions focus on the individual perceptions of the group as a whole or regard cohesion as an intrinsic property of the group. They define cohesion as the individual's attraction to the group, group connectedness, or the group's unity, as illustrated in Table 4.

This comparison reveals that the focus and scope of cohesion conceptualizations differ between social psychology and social computing. Structural features emphasized in community detection and search research are either neglected or not explicitly addressed in cohesion studies. This difference can also be observed in Figures 1(a) and 1(d), where the network structure is especially highlighted in the latter. It is worth noting that although unidimensional psychological definitions of cohesion have been criticized for their ambiguity [48, 96], they convey richer social meanings than structural characteristics.

In multidimensional conceptualizations, both community detection and community search studies use an additive approach to define cohesiveness, which enhances the homogeneity of identified communities by integrating diverse information. Specifically, studies begin with structural cohesion and incorporate additional information, such as keywords, types, followerships, time, and global structural information into the characterization of cohesion. For example, the two-dimensional conceptualization defined in [63] requires nodes to have keyword cohesiveness in addition to structural cohesiveness, involving both structural and node information (see Table 7). In particular, community detection studies such as [42] and [182] explicitly reference the sociological origins of the cohesion concept, with the latter defining cohesion through "unity of will" and "sharing common values", closely aligning with the psychological definitions (see Table 6).

However, this common additive transition from unidimensional to multidimensional cohesion differs from group cohesion research in psychology, where the general concept of cohesion is decomposed into distinct, concrete dimensions. Despite providing a richer informational context than unidimensional models, multidimensional conceptualizations in social computing still fall short of capturing the full depth of social psychological cohesion. For example, while structural and keyword cohesiveness developed in [63] may contribute to the aforementioned interpersonal and task-based cohesiveness, they do not explicitly characterize psychological elements such as positive relationships and shared commitments, especially in the absence of data such as the content and temporal aspects of members' interactions.

When the temporal aspect is considered in social computing, time is explicitly reflected in cohesiveness metrics [40, 42, 50]. Meanwhile, since temporal cohesion builds on structural cohesion, these studies

Table 7: Extracted Cohesion Conceptualizations in Community Search Studies.

Dimensionality	Temporal Nature	Extracted Conceptualization	Work
Unidimensional	Static	Vertices are densely connected in the subgraph.	[8, 10, 38, 122, 125, 126, 129, 144, 189]
		The distance of any two nodes in the subgraph is not longer than a specific number of hops.	[123]
		Two groups, in which group members are densely connected and have a required number of butterflies, have dense cross-group interactions.	[54]
		Size-bounded vertices are densely connected in the subgraph.	[218]
		The number of nodes that need to be removed to disconnect the subgraph.	[101]
		Each vertex has at least k neighbors and at least p fraction of its neighbors in the maximal connected subgraph.	[133]
		Edges in a maximal subgraph can connect through a series of edge-adjacent triangles, with each edge present in at least k-2 triangles.	[215]
		Vertices are densely connected under multi-faceted relationships in the subgraph.	[36]
Multidimensional	Static	Structural cohesiveness: vertices are densely connected in the subgraph; Keyword cohesiveness: vertices share common keywords in the subgraph.	[63]
		Structural cohesiveness: vertices are densely connected in the subgraph; Spatial cohesiveness: vertices are spatially close to each other in the subgraph.	[39, 64]
		Structural cohesiveness: vertices have close relationships in the subgraph; Type cohesiveness: vertices are in the same type in the subgraph.	[104]
		Local cohesiveness: vertices are densely connected in individual layers; Global cohesiveness: vertices are densely connected in the projected graph.	[135]
	Dynamic	Structural cohesiveness: vertices are densely connected in the subgraph; Query cohesiveness: vertices are temporally similar in terms of their activities related to the query attributes in the subgraph.	[50]

inherently treat cohesion as multidimensional. However, not all social computing studies address the dynamic nature of cohesion, similar to the case in psychological studies. Hence, we advocate that the time dimension of cohesion needs further emphasis.

Based on our discussion, social computing would benefit from adopting cohesiveness metrics designed with guidance from widely accepted multidimensional and dynamic cohesion conceptualizations in social psychology. A recent study [229] takes a step towards this direction.

6.2 Operational Definitions

In social computing, cohesion is often mathematically formulated as cohesiveness metrics, utilizing structural and, when available, attribute information (see surveys mentioned in Section 2.2). However, these metrics are abstract and lack sufficient context-related information [65, 76]. Even Shao et al. [182], who recognized community as a tight and cohesive social entity, still detected communities using purely structural measures: network centrality, hierarchical cluster analysis, and similarity across different time points. Social psychology measures, by contrast, enable a more subtle characterization of cohesion by considering the thoughts, feelings, or behaviors of individuals within a group [15, 119, 183].

The varying level of detail in operational definitions between fields also affects how existing measures are applied to different research objects. In social psychology, existing measures, especially self-report measures, often require adaptation to fit new contexts, thus enabling participants to express their perceptions effectively [25, 34, 43, 45, 52, 132, 137, 195, 199, 214, 230]. For instance, Carless and De Paola [25] and Blanchard et al. [16] adapted the Group Environment Questionnaire (GEQ), originally designed for sports groups, for use in work and exercise groups by changing the item wording. Similarly, the Perceived Cohesion Scale (PCS), initially developed for large groups, was adapted by Chin et al. [43] for smaller group contexts. In social computing, however, the abstract nature of cohesiveness metrics allows their unadjusted application across social networks, which hinders their ability to measure psychological cohesion in specific contexts accurately. For instance, [182] applied generic measures across various networks–including non-human ones–despite the diversity of human contexts (e.g., clubs, college football teams) where cohesion may be conceptualized differently by participants. Similarly, in our running example (Figure 1(d)), using unmodified structural and interactiveness metrics–originally designed for the interaction network in Figure 1(c)–on a collaboration network may overlook shared task commitments, a key aspect of cohesiveness in collaborative settings.

Considering the concrete networks where the algorithms are applied, cohesiveness metrics can also be classified as either self-report or behavioral, similar to cohesion studies. For example, when the structural cohesiveness metric is applied to the "who-trusts-whom" network, where members report trusted relationships, it functions similarly to a self-report measure [233]. The metrics serve as behavioral measures when algorithms like *attributed community query* (ACQ) are applied to real-world online social networks where vertices are humans and edges represent behaviors such as following, commenting, or co-authoring [63]. In contrast to social psychology, which relies heavily on self-report measures, behavioral measures are widely adopted in social computing. This difference can be attributed to its focus and target applications on large-scale networks, which are typically built based on digital traces of users rather than self-reports [65, 148, 167, 173].

To develop more effective cohesiveness metrics grounded in psychological cohesion definitions, researchers can refer to social psychology measures, which offer concrete measurement actions based on clear conceptualizations. Metric formulation should incorporate context-specific information from target networks. Besides, when applying metrics to different networks, necessary adjustments should be made to ensure their suitability. Returning to the example in Figure 1, to search for a psychology-informed cohesive community, cohesiveness metrics could be formalized based on a specific cohesion conceptualization while incorporating the available network information, such as interaction content, timestamps, and user profiles. The scope of measures should be clearly stated, e.g., limited to online social networks with similar data, to imply the necessity of adaptation when they are applied in different contexts.

6.3 Cohesion Definitions in Learning-based Social Computing

Most learning-based community detection research neglects the concept of cohesion [170, 211, 212, 228]. When cohesion is mentioned, it typically appears only as a superficial goal, identifying node clusters with high cohesiveness, without deeper conceptual and operational engagement [91]. In contrast, learning-based community search studies show greater awareness of the cohesion concept. However, cohesion remains narrowly conceptualized, with models primarily targeting network features like community structure and, occasionally, attribute homogeneity [36, 61, 62, 77, 105, 124, 204, 205, 207]. Even when temporal factors are incorporated, as in [90], where a *Temporal Graph Convolutional Network* (CS-TGN) is designed to capture node attributes, structural properties, and their dynamics across network snapshots, only the structural aspect of networks is linked to the cohesion concept.

6.4 Other Techniques Designing Cohesion Measures in Social Computing

Given the significance of psychological cohesion, multiple approaches have been proposed in social computing to quantify it directly or indirectly from perspectives beyond the traditional psychological framework. Here, we briefly review two such techniques: *linguistic style matching* and *supervised learning*.

Linguistic style matching (LSM). Originating from linguistic research, *linguistic style matching* quantifies the degree of stylistic similarities in language use across individuals and groups [154]. Gonzales et al. [81] found that individual cohesion can be automatically predicted by evaluating the verbal mimicry of function words using LSM, with consistent accuracy regardless of the communication settings. However, subsequent studies have raised concerns about its generalizability. In global group collaboration, Castro-Hernández et al. [33] revealed that the LSM's correlation with task cohesion is weaker than that of *information exchange similarity*, a straightforward word-based calculation method. In asynchronous communication online, Munson et al. [150] found no support for LSM in predicting group cohesion, noting that its effectiveness may be influenced by individual participation levels and multilingual text features.

Supervised learning. In recent decades, studies have begun to develop cohesion measurements using computational techniques, including Support Vector Machines (SVM), Naive Bayes Classifier, Random Forest Classifier, and deep supervised networks. These supervised learning methods are trained on labeled data to predict cohesion levels, often leveraging multimodal communication cues such as text, audio, images, video, and motion [31, 178]. In [99], nonverbal cues (i.e., audio, visual, and audio-visual cues) were automatically extracted either on the individual or group level. SVM and Naive Bayes classifiers were then used to estimate whether a meeting had high or low cohesion. To evaluate these methods, cohesion annotations were collected from external observers using questionnaires based on psychological literature. In [79, 86], multi-task convolutional neural networks were trained for cohesion estimation, using face, body, and contextual cues extracted from images with supervision based on annotator-labeled cohesion. Besides, Walocha et al. [201] used a Random Forest Classifier to jointly predict the dynamics of social and task cohesion through large-scale spatial and bodily movement features, where labels are cohesion scores collected from the Group Environment Questionnaire (GEQ) instead of third-party annotations. The model achieved an average prediction accuracy of 64% (±3%) for task cohesion and 67% (±3%) for social cohesion [201].

Although supervised learning has shown its potential for automating cohesion measurement, it is still in the early stages with several limitations to be addressed. First, the extraction of communication cues and the establishment of cohesion measures, despite their basis in psychology, lack robust integration with cohesion studied in psychological research. Second, it can be expected that accurate automated cohesion scoring relies heavily on high-quality labeled datasets. Ideally, labels should be annotated via results from well-established psychological measures. However, the current labels in the datasets are often assigned by external observers through labor-intensive processes, and their relationship with psychological cohesion remains unclear.

7 Data Collection

Data collection methods for measuring cohesion are closely tied to how cohesion is operationalized [15]. In this section, we exclusively review the data collection methods adopted in group cohesion studies, categorized as *obtrusive* or *unobtrusive* (as detailed in Table 8). Thereafter, we review the data collection practices in social computing and discuss how they can be refined by incorporating insights from psychological studies.

Table 8: Data Collection Methods.

Type	Method
Obtrusive	Survey research Structured observation
Unobtrusive	Digital trace data collection methods

7.1 Obtrusive Methods

Obtrusive data collection methods explicitly inform participants that they are being studied [93]. Two such methods are commonly used in cohesion studies: *survey research* and *structured observation*, with surveys being more common. Survey research uses self-report measures, specifically questionnaires, to collect responses from well-defined populations. Structured observation, an observational method involving researcher intervention, creates a context in which participant behaviors can be more easily and quantitatively recorded by applying behavioral measures in laboratory or natural settings [183]. In the study by Budman et al. [22] (see Section 5.2), structured observation involved reviewing videotapes from time-limited group therapy sessions, during which two experienced raters observed cohesive behaviors within the groups and then assessed them using the GCS. In research adopting both self-report and behavioral measures [168], both survey research and structured observation are correspondingly employed in data collection.

Both obtrusive methods rely on differing interactions between researchers and respondents to gather extensive information from individuals. However, they are susceptible to distinct biases, such as response style bias in surveys and rater bias in structured observation [6, 225]. Besides, both can be labor-intensive, especially when studying the dynamic nature of cohesion over time or studying large groups [175]. These limitations urge the need for developing low-cost, unobtrusive data collection methods.

7.2 Unobtrusive Methods

Unobtrusive data collection methods allow subjects to be studied without awareness, making them well-suited for continuous or near-continuous assessment [6, 115]. In traditional psychology research, non-reactive data sources include physical traces, mass media, archives, etc. In recent years, as technologies increasingly mediate and support human activities, individuals' behaviors, as well as their intentions and emotions they reflect, can be inferred from digital traces recorded on digital systems, such as business transaction systems, social media platforms, and sensors [111, 158, 171]. These vast and continuously generated digital traces in online settings offer researchers new opportunities to study team processes and social phenomena [111, 115]. We term various approaches used for gathering digital traces as "digital trace data collection methods", which can be either obtrusive or unobtrusive. Obtrusive methods involve making video calls or accessing personal emails, while unobtrusive methods include application programming interfaces (APIs), data donation, tracking, etc [111, 158, 171, 175].

Based on our review, the digital trace data collection method serves as the primary method for gathering unobtrusive data in cohesion studies. For example, Altebarmakian and Alterman [3] unobtrusively collected students' data via an online system throughout a semester. Similarly, Zamecnik et al. [225] gathered three weeks of engagement data from a global experiential learning platform, involving 420 teams from various higher education institutions. Collecting digital traces overcomes the inherent limitations of obtrusive methods by enabling efficient, cost-effective data collection over time. It can capture the long-term dynamics of cohesion and offer greater flexibility in participant selection, unconstrained by geography or group size [6, 174]. Furthermore, the resulting datasets may be reused and support the application of various cohesion measures [171], making this approach promising for broader adoption.

However, this method also has limitations. First, it is not suitable for groups that require in-person participation, such as sports and exercise groups. Second, since the digital traces are generally not generated for research purposes, they are susceptible to misinformation, coverage error[7], and over-interpretation. As such, the data must be carefully cleaned and interpreted before being used for

[7] Coverage error refers to the selective representation of the platform user population compared to the general population [158].

measurement [6, 15, 111, 158, 171, 177, 185].

Therefore, using both obtrusive and unobtrusive data collection methods together is seen as a way to balance their respective limitations. Some researchers suggest using digital trace methods to complement traditional approaches, obtaining more diverse, extensive, and potentially less biased data [6, 171]. There is also an inspiring attempt to construct a multi-modal dataset utilizing both types of methods [139]. The dataset not only contains sound, video, and motion capture data collected from participants engaged in an 11-hour escape game, but also includes ongoing self-reported cohesion data collected through questionnaires.

7.3 Discussion

By tracing the original studies behind commonly used social network datasets in social computing and extracting their data collection details, we find that most datasets used in social computing are digital traces collected from online platforms, such as *Facebook*, *X* (previously *Twitter*), *YouTube*, *Flickr*, *Live Journal*, *Myspace*, and *Tencent* [8, 39, 50, 54, 63, 100, 123, 128, 133, 169, 210, 217]. Only a few datasets, particularly those used in community detection, are gathered through survey research, structured observations, or digitalized records of face-to-face interactions. For example, the Zachary karate club network was built using data from direct observation, informants, and club records in the university archives [223]. The "Jazz musicians" dataset is constructed from the collaboration data stored in the Red Hot Jazz Archive digital database [80]. Sometimes, network datasets are collected along with "ground-truth" communities [216], which are instrumental in validating social computing algorithms by comparing their outputs to known community structures. However, the definitions of these communities vary, and psychological cohesion is rarely, if ever, explicitly referenced [112, 121, 138, 216, 219].

The preference for digital trace collection in social computing contrasts with group cohesion studies, where survey research is commonly used. Besides, the two fields differ in their attention to the data collection process. Psychological studies typically collect data firsthand, providing detailed descriptions of participants and procedures. In contrast, social computing often relies on existing datasets, focusing more on algorithm design and performance than on data collection. As a result, dataset details are frequently overlooked, and only a few studies collect their own data [42, 128].

Based on the above observation, it can be inferred that network datasets built on digital traces share the same limitations identified in psychological studies. Concerning the general reliance of social computing on digital traces, it is important to maximize their advantages. Informed by existing endeavors in social psychology, researchers should carefully clean and interpret the digital traces. Besides, survey research can serve as a complementary method to validate the quality of the digital traces. Social computing could also benefit from datasets containing valid ground-truth communities that align with psychologically defined cohesion [229].

8 Evaluation

In psychological studies, the quality of a measure is evaluated through its reliability and validity, with the timing of these evaluations varying by method: some are conducted theoretically once the measure is established, others are empirically implemented following data collection, or at both stages [14, 119, 183]. In this section, we review how cohesion studies evaluate their measures. Then we compare those approaches with the evaluation of cohesiveness metrics in social computing, exploring insights that may guide the evaluation of cohesiveness metrics.

8.1 Reliability

Reliable measures seek the consistent and stable measuring of construct and are evaluated after data collection [14, 119, 183]. Three common types of reliability are test-retest, internal consistency, and inter-rater reliability (for definitions, see Section 3.1). The choice of method depends on the construct's temporal nature and the adopted operational definition.

In cohesion studies, internal consistency, often measured using *Cronbach's alpha*, is commonly evaluated due to its simplicity and direct relevance to questionnaires [14, 49]. It averages all possible split-half correlations for a set of items [49]. In our review, reported Cronbach's alpha for cohesion measures ranged from 0.542 to 0.991, with most exceeding the 0.70 threshold, indicating good internal consistency [156, 193]. In addition, test-retest reliability, which presents the stability of cohesion measurements over time, is validated by correlating the scores of samples collected at different intervals [16, 34]. When behavioral measures are used, inter-rater reliability ensures cohesion is consistently quantified across different observers [14, 120].

Ensuring measurement reliability is critical but often challenging. Common sources of unreliability include observers' subjectivity, and questions that are imprecise, ambiguous, or unfamiliar to respondents [14]. Strategies to improve reliability include clearly conceptualizing constructs, utilizing precise measurement levels and multiple indicators, and conducting pilot tests [25, 29, 34, 58, 119, 120].

8.2 Validity

A measure with high validity guarantees a strong alignment between the cohesion conceptualization and its empirical measurement results. Validity can be assessed from both theoretical and empirical perspectives.

Theoretical validity includes face and content validity, typically conducted by expert panels to ensure that measures adequately represent the cohesion definition [14, 60]. For instance, in developing the *Youth Sport Environment Questionnaire* [60], 142 initial items were evaluated by investigators for relevance, similarity, and clarity. These items were then refined based on feedback from additional group dynamics experts and athletes, resulting in 87 items.

Empirical validity is assessed using participant data. Factorial validity is often evaluated via pilot testing to ensure the measure reflects the underlying construct structure of cohesion. A practice of this assessment is detailed in [16]. While some studies only evaluate convergent or discriminant validity [35, 188, 220], it is more common to assess both simultaneously [24, 25, 52, 87, 108, 109, 118, 164, 176, 202, 230]. Empirical validity can also be substantiated through correlation with criteria expected to be associated with cohesion [183]. Concurrent validity is established when a measure correlates significantly with a criterion assessed concurrently. For instance, Estabrooks and Carron [58] examined the concurrent validity of the *Physical Activity Group Environment Questionnaire* (PAGEQ) by correlating it with the Group Environment Questionnaire (GEQ). Predictive validity assesses the correlations between cohesion and its expected future outcomes, such as attendance [58, 120], self-efficacy [58], individual perceptions of job performance, job satisfaction, and psychological distress [1].

Ideally, a thorough validity assessment should integrate both theoretical and empirical approaches. However, while newly developed cohesion measures typically undergo comprehensive validity assessment, adapted measures for different contexts or populations may not receive the same level of scrutiny, as noted in several studies [16, 25, 58, 60, 176]. For instance, when a questionnaire originally designed for student teams was adapted for serious game teams, only face and content validity were assessed, with other empirical evaluation neglected [20]. This limited approach can yield unreliable responses that misalign with the adopted cohesion conceptualization. More holistic evaluation of adapted measures is therefore essential to ensure their appropriateness for specific contexts and populations.

Table 9: Similarities and Differences in Cohesion between Social Psychology and Social Computing.

Components	Aspects	Similarities	Differences	Social Psychology	Social Computing
Conceptual Definitions	Dimension	Both can be classified as either unidimensional or multidimensional	Focus and scope	Neglect or implicitly address the structural features of cohesion, yet contain sufficient psychological meanings	Emphasize structural and attribute features of cohesion with insufficient psychological consideration
			Relationship between unidimensional and multidimensional definitions	From ambiguous to concrete and empirical	From solely structural to additive inclusion of non-structural aspects
	Temporal Nature	The dynamic nature is frequently overlooked	Treatment of the dynamic nature	Dynamic nature does not impact conceptual dimensionality	Incorporate dynamic nature as an additional dimension in definitions
Operational Definitions	Measures	Both adopt self-report and behavioral measures	Measurement concreteness	Include concrete measurement procedures	Mathematically formulated, need the specification or pre-calculation of parameters
			Approaches for applying existing measures to different study objects	Measures must be adapted to fit the diverse characteristics of study groups	Same metrics can be applied across networks without adjustments, but their accuracy may be undermined
			Measures mainly adopted	Self-report measures	Behavioral measures
Data Collection	Methods	Both use two types of data collection methods	Methods mainly adopted	Survey research	Digital traces data collection
			Attention to data collection	Detailed reporting of participants and data collection procedures	Dataset details and collection methods largely overlooked
Evaluation	Reliability	Both address reliability	Definition of the reliability	Consistency and stability in measuring a construct	Stability of algorithm performance or the continuity of subgroup cohesiveness
			Level of measure reliability	Reliability could be compromised by various sources of errors	Highly reliable due to the automation and objective metric calculation
	Validity	Both consider validity in their evaluation	Objects of evaluation	Cohesion measures	Communities detected/searched by algorithms, from where the measure validity metrics can be inferred
			Perspective of evaluation	Theoretical and empirical	Empirically only, based on the structure and attributes of identified communities

8.3 Discussion

Consistent and accurate cohesion measures are also indispensable for social computing algorithms. However, these studies only validate the superiority of their algorithms through efficiency and effectiveness comparisons [73, 89, 103, 161, 167, 173]. We therefore examine how metrics are evaluated in existing studies through their detection/search result assessments.

Surprisingly, community detection and search rarely address "reliability" in the evaluation. Yet, due to the automated and objective nature of these algorithms, wherein cohesiveness metrics are computed based on established equations rather than relying on participants' limited comprehension or researchers' subjective observations [14], their cohesiveness measurements are always stable and consistent. Regarding validity, we find that the effectiveness evaluation experiments in community detection/search studies are conceptually similar to validity assessments in cohesion studies. However, since cohesion metrics are not directly evaluated in social computing, their validity can only be inferred from the quality of the detected/searched communities: the closer they are to the expected communities, the higher the validity of the metrics.

Generally, retrieved communities are evaluated empirically using four approaches: (1) adopting synthetic network generators for controlled environment testing or ground-truth testing; (2) benchmarking against ground-truth with performance metrics; (3) comparing them with those from other algorithms through a set of "effectiveness metrics" or "goodness metrics", particularly when the ground-truth is unavailable; and (4) conducting case studies to present their interpretability and significance [50, 73, 89, 103, 123, 127, 144, 173]. Popular performance metrics include precision, recall, F-score, accuracy, and Normalized Mutual Information (NMI), which measures the similarity between the set of ground-truth communities and those retrieved by the algorithm [10, 54, 89, 129]. For algorithms that only consider structural cohesiveness metrics, effectiveness are often evaluated based on quantitative attributes of community structure, such as density, clustering coefficient, conductance, triangle participation ratio, diameter, result size, and relationship closeness [8, 10, 36, 89, 104, 122, 133, 189, 215, 218, 226]. For algorithms that incorporate both structural and attribute information, additional metrics are employed, such as semantic richness, average number of activities, community member frequency (CMF), and community pair-wise Jaccard (CPJ) metric [50, 63, 101, 104, 122, 129].

Comparing the evaluation methods between the two fields, it is apparent that their objects of evaluation are different: group cohesion studies focus on the quality of cohesion measures, while social computing emphasizes algorithm performance. Besides, the focus of assessments in social computing studies, similar to the metrics they developed, focuses only on structural and attribute features. While suitable for cohesion definitions rooted in network properties, these methods may be insufficient for capturing psychologically defined cohesion. Currently, no synthetic network generators that incorporate psychological cohesion are available for testing, nor are there datasets with psychologically cohesive communities available for benchmarking. Although efforts have been made to develop effectiveness metrics based on psychological cohesion to compare algorithm-generated communities [229], their quality and generalizability require further validation. With no established standards for cohesion evaluation in social computing, additional metrics are needed to reliably assess psychological cohesion.

Informed by validity assessments in social psychology, cohesiveness metrics should be theoretically evaluated before being integrated into social computing algorithms. This is particularly important when metrics are based on psychological cohesion, as translating such complex constructs into quantifiable measures can be contentious. Without a clear theoretical grounding, metrics may misrepresent or inadequately capture cohesion. In addition, informed by concurrent validity and predictive validity in social psychology, constructs conceptually linked to cohesion, such as morale, group spirit, trust, friendship, and group identification [72, 143], can be quantified as effectiveness metrics, assessing whether identified communities exhibit cohesion-related characteristics or predict future events pertinent to cohesion. Their measurement, as explored in psychological research, could also be examined under our review framework to uncover insights for development.

9 Conclusions

Community detection and search algorithms are key techniques for identifying cohesive subgroups in social applications. However, the concept of cohesion, which originates from social psychology, is not consistently characterized in existing cohesiveness metrics. As a result, these algorithms may fail to identify psychologically cohesive subgroups in online social networks. Given this challenge, this survey reviews cohesion as studied in both social psychology and social computing fields. Following our proposed unified framework, the approach to addressing cohesion in each study is decomposed into four phases encompassing the definition and measurement of cohesion, enabling comparison of their similarities and differences (summarized in Table 9).

We acknowledge several limitations in our review. First, as a narrative rather than a systematic or scoping review, it includes only a representative set of studies and may omit relevant research on cohesion in the two fields. Second, no inter-rater checking was conducted for study selection or evidence validation, which may affect the comprehensiveness of our analysis. Nevertheless, random sampling of eligible papers from the screening process indicates that the identified similarities and differences remain consistent.

Despite these limitations, this survey lays the foundation for integrating social psychological theory into social computing. It offers a fresh perspective on examining community detection and search studies. In addition, our review framework, grounded on social psychology research methods, supports both theory reviews in psychology while serving as an analytical tool for evaluating studies across these traditionally disparate fields. Last but not least, it paves the way for a reevaluation of existing social computing algorithm designs and their practical utility in achieving social application goals.

References

[1] Arni Ahronson and James E Cameron. The nature and consequences of group cohesion in a military sample. *Military psychology*, 19(1):9–25, 2007.

[2] Richard D Alba. A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology*, 3 (1):113–126, 1973.

[3] Maria Altebarmakian and Richard Alterman. Cohesion in online environments. *International Journal of Computer-Supported Collaborative Learning*, 14:443–465, 2019.

[4] Arden J Anderson and Stacy Warner. Social network analysis as a complimentary tool to measuring team cohesion. *Journal of Sport Behavior*, 40(1):3, 2017.

[5] Holly Arrow, Marshall Scott Poole, Kelly Bouas Henry, Susan Wheelan, and Richard Moreland. Time, change, and development: The temporal perspective on groups. *Small group research*, 35(1):73–105, 2004.

[6] CJ Auriacombe. Towards the construction of unobtrusive research techniques: Critical considerations when conducting a literature analysis. *African Journal of Public Affairs*, 9(4):1–19, 2016.

[7] Georgia Baltsou, Konstantinos Christopoulos, and Konstantinos Tsichlas. Local community detection: A survey. *IEEE Access*, 10:110701–110726, 2022.

[8] Nicola Barbieri, Francesco Bonchi, Edoardo Galimberti, and Francesco Gullo. Efficient and effective community search. *Data mining and knowledge discovery*, 29:1406–1433, 2015.

[9] Punam Bedi and Chhavi Sharma. Community detection in social networks. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, 6(3):115–135, 2016.

[10] Ali Behrouz, Farnoosh Hashemi, and Laks VS Lakshmanan. Firmtruss community search in multilayer networks. *Proceedings of the VLDB Endowment*, 16(3):505–518, 2022.

[11] H Russell Bernard, Peter D Killworth, and Lee Sailer. Informant accuracy in social network data iv: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, 2(3):191–218, 1979.

[12] H Russell Bernard, Peter Killworth, David Kronenfeld, and Lee Sailer. The problem of informant accuracy: The validity of retrospective data. *Annual review of anthropology*, 13(1):495–517, 1984.

[13] Paul R Bernthal and Chester A Insko. Cohesiveness without groupthink: The interactive effects of social and task cohesion. *Group & Organization Management*, 18(1):66–87, 1993.

[14] Anol Bhattacherjee. *Social science research: Principles, methods, and practices.* USA, 2012.

[15] Amy Blackstone. *Principles of sociological inquiry: Qualitative and quantitative methods.* Saylor Academy Open Textbooks, 2018.

[16] Chris Blanchard, Pauline Poon, Wendy Rodgers, and Bruce Pinel. Group environment questionnaire and its applicability in an exercise setting. *Small group research*, 31(2):210–224, 2000.

[17] Paul D Bliese and Ronald R Halverson. Individual and nomothetic models of job stress: An examination of work hours, cohesion, and well-being 1. *Journal of Applied Social Psychology*, 26(13):1171–1189, 1996.

[18] Kenneth A Bollen and Rick H Hoyle. Perceived cohesion: A conceptual and empirical examination. *Social forces*, 69(2):479–504, 1990.

[19] Randa Boukabene and Fatima Benbouzid-Si Tayeb. Community detection metrics in multilayer networks: A survey and novel taxonomy. *SN Computer Science*, 6(5):428, 2025.

[20] Aysun Bozanta, Birgul Kutlu, Nuket Nowlan, and Shervin Shirmohammadi. Effects of serious games on perceived team cohesiveness in a multi-user virtual environment. *Computers in Human Behavior*, 59: 380–388, 2016.

[21] Lawrence R Brawley, Albert V Carron, and W Neil Widmeyer. The influence of the group and its cohesiveness on perceptions of group goal-related variables. *Journal of Sport and Exercise Psychology*, 15 (3):245–260, 1993.

[22] Simon H Budman, Stephen Soldz, Annette Demby, Michael Davis, and Jocelyn Merry. What is cohesiveness? an empirical examination. *Small Group Research*, 24(2):199–216, 1993.

[23] Gary Burlingame, Addie Fuhriman, and Stuart Drescher. Scientific inquiry into small group process: A multidimensional approach. *Small Group Behavior*, 15(4):441–470, 1984.

[24] Yuejun Lawrance Cai. Strengthening perceptions of virtual team cohesiveness and effectiveness in new normal: A hyperpersonal communication theory perspective. *Asian Business & Management*, 22(4): 1649–1682, 2023.

[25] Sally A Carless and Caroline De Paola. The measurement of cohesion in work teams. *Small group research*, 31(1):71–88, 2000.

[26] Albert V Carron. Cohesiveness in sport groups: Interpretations and considerations. *Journal of Sport psychology*, 4(2):123–138, 1982.

[27] Albert V Carron and Lawrence R Brawley. Cohesion: Conceptual and measurement issues. *Small group research*, 31(1):89–106, 2000.

[28] Albert V Carron and Kevin S Spink. The group size-cohesion relationship in minimal groups. *Small group research*, 26(1):86–105, 1995.

[29] Albert V Carron, W Neil Widmeyer, and Lawrence R Brawley. The development of an instrument to assess cohesion in sport teams: The group environment questionnaire. *Journal of Sport and Exercise psychology*, 7 (3):244–266, 1985.

[30] Dorwin Cartwright. The nature of group cohesiveness. *Group dynamics: Research and theory*, 91:109, 1968.

[31] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168, 2006.

[32] Alberto Castro-Hernández, Kathleen Swigger, and Mirna P Ponce-Flores. Effects of cohesion-based feedback on the collaborations in global software development teams. In *10th IEEE international conference on collaborative computing: Networking, applications and Worksharing*, pages 74–83. IEEE, 2014.

[33] Alberto Castro-Hernández, Kathleen Swigger, Mirna P Ponce-Flores, and J David Terán-Villanueva. Measures for predicting task cohesion in a global collaborative learning environment. In *2016 IEEE 11th International Conference on Global Software Engineering Workshops (ICGSEW)*, pages 31–36. IEEE, 2016.

[34] Artemis Chang and Prashant Bordia. A multidimensional approach to the group cohesion-group performance relationship. *Small group research*, 32(4):379–405, 2001.

[35] Song Chang, Liangding Jia, Riki Takeuchi, and Yahua Cai. Do high-commitment work systems affect creativity? a multilevel combinational approach to employee creativity. *Journal of applied psychology*, 99 (4):665, 2014.

[36] Jiazun Chen, Yikuan Xia, and Jun Gao. Communityaf: An example-based community search method via autoregressive flow. *Proceedings of the VLDB Endowment*, 16(10):2565–2577, 2023.

[37] Jiazun Chen, Yikuan Xia, Jun Gao, Zhao Li, and Hongyang Chen. Communitydf: A guided denoising diffusion approach for community search. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, pages 460–473. IEEE Computer Society, 2025.

[38] Jinyi Chen, Junchang Xin, Farhana Choudhury, Keqi Zhou, and Zhiqiong Wang. Efficient stable community search in temporal signed graphs. *Knowledge and Information Systems*, pages 1–31, 2025.

[39] Lu Chen, Chengfei Liu, Rui Zhou, Jianxin Li, Xiaochun Yang, and Bin Wang. Maximum co-located community search in large scale social networks. *Proceedings of the VLDB Endowment*, 11(10):1233–1246, 2018.

[40] Naiyue Chen, Bo Hu, and Yuqun Rui. Dynamic network community detection with coherent neighborhood propinquity. *IEEE Access*, 8:27915–27926, 2020.

[41] David S Chester and Emily N Lasko. Construct validation of experimental manipulations in social psychology: Current practices and recommendations for the future. *Perspectives on Psychological Science*, 16(2):377–395, 2021.

[42] Alvin Chin and Mark Chignell. Automatic detection of cohesive subgroups within social hypertext: A heuristic approach. *New Review of Hypermedia and Multimedia*, 14(1):121–143, 2008.

[43] Wynne W Chin, Wm David Salisbury, Allison W Pearson, and Matthew J Stollak. Perceived cohesion in small groups: Adapting and testing the perceived cohesion scale in a small-group setting. *Small group research*, 30(6):751–766, 1999.

[44] Petr Chunaev. Community detection in node-attributed social networks: a survey. *Computer Science Review*, 37:100286, 2020.

[45] Joshua Cloudy, Melissa R Gotlieb, and Bryan McLaughlin. Online political networks as fertile ground for extremism: The roles of group cohesion and perceived group threat. *Journal of Information Technology & Politics*, 21(4):578–587, 2024.

[46] Jonathan Cohen. Trusses: Cohesive subgraphs for social network analysis. *National security agency technical report*, 16(3.1):1–29, 2008.

[47] Alessio Conte, Daniele De Sensi, Roberto Grossi, Andrea Marino, and Luca Versari. Truly scalable k-truss and max-truss algorithms for community detection in graphs. *IEEE Access*, 8:139096–139109, 2020.

[48] Albert A Cota, Charles R Evans, Kenneth L Dion, Lindy Kilik, and R Stewart Longman. The structure of group cohesion. *Personality and Social Psychology Bulletin*, 21(6):572–580, 1995.

[49] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.

[50] Badhan Chandra Das, Md Musfique Anwar, Md Al-Amin Bhuiyan, Iqbal H Sarker, Salem A Alyami, and Mohammad Ali Moni. Attribute driven temporal active online community search. *IEEE Access*, 9: 93976–93989, 2021.

[51] Matthew DeCarlo. *Scientific inquiry in social work*. Open social work education, 2018.

[52] Isabel Dórdio Dimas, Paulo Renato Lourenço, Teresa Rebelo, and Humberto Rocha. Maximizing learning through cohesion: Contributions from a nonlinear approach. *Small Group Research*, 52(2):162–188, 2021.

[53] Kenneth L Dion. Group cohesion: From "field of forces" to multidimensional construct. *Group Dynamics: Theory, research, and practice*, 4(1):7, 2000.

[54] Zheng Dong, Xin Huang, Guorui Yuan, Hengshu Zhu, and Hui Xiong. Butterfly-core community search over labeled graphs. *Proceedings of the VLDB Endowment*, 14(11):2006–2018, 2021.

[55] Stuart Drescher, Gary Burlingame, and Addie Fuhriman. Cohesion: An odyssey in empirical understanding. *Small Group Behavior*, 16(1):3–30, 1985.

[56] Marc Eaton. Manufacturing community in an online activist organization: The rhetoric of moveon. org's e-mails. *Information, Communication & Society*, 13(2):174–192, 2010.

[57] Paul A Estabrooks and Albert V Carron. Group cohesion in older adult exercisers: Prediction and intervention effects. *Journal of behavioral medicine*, 22:575–588, 1999.

[58] Paul A Estabrooks and Albert V Carron. The physical activity group environment questionnaire: An instrument for the assessment of cohesion in exercise classes. *Group Dynamics: Theory, Research, and Practice*, 4(3):230, 2000.

[59] Nancy J Evans and Paul A Jarvis. Group cohesion: A review and reevaluation. *Small group behavior*, 11 (4):359–370, 1980.

[60] Mark Eys, Todd Loughead, Steven R Bray, and Albert V Carron. Development of a cohesion questionnaire for youth: The youth sport environment questionnaire. *Journal of Sport and Exercise Psychology*, 31(3): 390–408, 2009.

[61] Shuheng Fang, Kangfei Zhao, Guanghua Li, and Jeffrey Xu Yu. Community search: a meta-learning approach. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 2358–2371. IEEE, 2023.

[62] Shuheng Fang, Kangfei Zhao, Yu Rong, Zhixun Li, and Jeffrey Xu Yu. Inductive attributed community search: to learn communities across graphs. *Proceedings of the VLDB Endowment*, 17(10):2576–2589, 2024.

[63] Yixiang Fang, Reynold Cheng, Siqiang Luo, and Jiafeng Hu. Effective community search for large attributed graphs. *Proceedings of the VLDB Endowment*, 9(12):1233–1244, 2016.

[64] Yixiang Fang, Reynold Cheng, Xiaodong Li, Siqiang Luo, and Jiafeng Hu. Effective community search over large spatial graphs. *Proceedings of the VLDB Endowment*, 10(6):709–720, 2017.

[65] Yixiang Fang, Xin Huang, Lu Qin, Ying Zhang, Wenjie Zhang, Reynold Cheng, and Xuemin Lin. A survey of community search over big graphs. *The VLDB Journal*, 29:353–392, 2020.

[66] Beverley Fehr. *Friendship processes*, volume 12. Sage, 1996.

[67] Scott L Feld and William C Carter. Detecting measurement bias in respondent reports of personal networks. *Social networks*, 24(4):365–383, 2002.

[68] Leon Festinger. The analysis of sociograms using matrix algebra. *Human relations*, 2(2):153–158, 1949.

[69] Leon Festinger. Informal social communication. *Psychological review*, 57(5):271, 1950.

[70] Leon Festinger, Stanley Schachter, and Kurt Back. Social pressures in informal groups; a study of human factors in housing. 1950.

[71] Maria Fors Brandebo, Marcus Börjesson, and Hilmar Hilmarsson. Longitudinal studies on cohesion in a military context–a systematic review. *Military Psychology*, 34(6):732–741, 2022.

[72] Donelson R Forsyth. Recent advances in the study of group cohesion. *Group Dynamics: Theory, Research, and Practice*, 25(3):213, 2021.

[73] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[74] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics reports*, 659: 1–44, 2016.

[75] Santo Fortunato, Vito Latora, and Massimo Marchiori. Method to find community structures based on information centrality. *Physical review E*, 70(5):056104, 2004.

[76] Borko Furht. *Handbook of social network technologies and applications*. Springer Science & Business Media, 2010.

[77] Jun Gao, Jiazun Chen, Zhao Li, and Ji Zhang. Ics-gnn: lightweight interactive community search via graph neural network. *Proceedings of the VLDB Endowment*, 14(6):1006–1018, 2021.

[78] David Gefen and Detmar Straub. A practical guide to factorial validity using pls-graph: Tutorial and annotated example. *Communications of the Association for Information systems*, 16(1):5, 2005.

[79] Shreya Ghosh, Abhinav Dhall, Nicu Sebe, and Tom Gedeon. Predicting group cohesiveness in images. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[80] Pablo M Gleiser and Leon Danon. Community structure in jazz. *Advances in complex systems*, 6(04): 565–573, 2003.

[81] Amy L Gonzales, Jeffrey T Hancock, and James W Pennebaker. Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37(1):3–19, 2010.

[82] Ilanit Gordon, Avi Gilboa, Shai Cohen, Nir Milstein, Nir Haimovich, Shay Pinhasi, and Shahar Siegman. Physiological and behavioral synchrony predict group cohesion and performance. *Scientific reports*, 10(1): 8484, 2020.

[83] Lindred L Greer. Group cohesion: Then and now. *Small Group Research*, 43(6):655–661, 2012.

[84] James Griffith. Measurement of group cohesion in us army units. *Basic and applied social psychology*, 9(2): 149–171, 1988.

[85] Natali Gulbahce and Sune Lehmann. The art of community detection. *BioEssays*, 30(10):934–938, 2008.

[86] Da Guo, Kai Wang, Jianfei Yang, Kaipeng Zhang, Xiaojiang Peng, and Yu Qiao. Exploring regularizations with face, body and image cues for group cohesion prediction. In *2019 International Conference on Multimodal Interaction*, pages 557–561, 2019.

[87] Zixiu Guo, Felix B Tan, Tim Turner, and Huizhong Xu. Group norms, media preferences, and group meeting success: A longitudinal study. *Computers in Human Behavior*, 26(4):645–655, 2010.

[88] James Hardy, Mark A Eys, and Albert V Carron. Exploring the potential disadvantages of high cohesion in sports teams. *Small group research*, 36(2):166–187, 2005.

[89] Steve Harenberg, Gonzalo Bello, La Gjeltema, Stephen Ranshous, Jitendra Harlalka, Ramona Seay, Kanchana Padmanabhan, and Nagiza Samatova. Community detection in large-scale networks: a survey and empirical evaluation. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6):426–439, 2014.

[90] Farnoosh Hashemi, Ali Behrouz, and Milad Rezaei Hajidehi. Cs-tgn: Community search via temporal graph neural networks. In *Companion Proceedings of the ACM Web Conference 2023*, pages 1196–1203, 2023.

[91] Chaobo He, Junwei Cheng, Guohua Chen, and Yong Tang. Multiple topics community detection in attributed networks. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2199–2203, 2023.

[92] RA Hinde and Joan Stevenson-Hinde. Towards understanding relationships: Dynamic stability. 1976.

[93] Christine Hine. Internet research and unobtrusive methods. *Social Research Update*, (61):1, 2011.

[94] Martin Hoegl and Hans Georg Gemuenden. Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization science*, 12(4):435–449, 2001.

[95] Michael A Hogg. The social psychology of group cohesiveness: From attraction to social identity. 1992.

[96] Michael A Hogg. Group cohesiveness: A critical review and some new directions. *European review of social psychology*, 4(1):85–111, 1993.

[97] Michael A Hogg and Sarah C Hains. Friendship and group identification: A new look at the role of cohesiveness in groupthink. *European journal of social psychology*, 28(3):323–341, 1998.

[98] Xinyu Huang, Dongming Chen, Tao Ren, and Dongqi Wang. A survey of community detection methods in multilayer networks. *Data Mining and Knowledge Discovery*, 35:1–45, 2021.

[99] Hayley Hung and Daniel Gatica-Perez. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia*, 12(6):563–575, 2010.

[100] Isa Inuwa-Dutse, Mark Liptrott, and Ioannis Korkontzelos. A multilevel clustering technique for community detection. *Neurocomputing*, 441:64–78, 2021.

[101] Md Saiful Islam, Mohammed Eunus Ali, Yong-Bin Kang, Timos Sellis, Farhana M Choudhury, and Shamik Roy. Keyword aware influential community search in large attributed graphs. *Information Systems*, 104: 101914, 2022.

[102] Devi Jankowicz. *The easy guide to repertory grids*. John wiley & sons, 2005.

[103] Muhammad Aqib Javed, Muhammad Shahzad Younis, Siddique Latif, Junaid Qadir, and Adeel Baig. Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 108:87–111, 2018.

[104] Yangqin Jiang, Yixiang Fang, Chenhao Ma, Xin Cao, and Chunshan Li. Effective community search over large star-schema heterogeneous information networks. *Proceedings of the VLDB Endowment*, 15(11): 2307–2320, 2022.

[105] Yuli Jiang, Yu Rong, Hong Cheng, Xin Huang, Kangfei Zhao, and Junzhou Huang. Query driven-graph neural networks for community search: from non-attributed, attributed, to interactive attributed. *Proceedings of the VLDB Endowment*, 15(6):1243–1255, 2022.

[106] Di Jin, Zhizhi Yu, Pengfei Jiao, Shirui Pan, Dongxiao He, Jia Wu, Philip S Yu, and Weixiong Zhang. A survey of community detection approaches: From statistical modeling to deep learning. *IEEE Transactions on knowledge and data engineering*, 35(2):1149–1170, 2021.

[107] Dong I Jung and John J Sosik. Transformational leadership in work groups: The role of empowerment, cohesiveness, and collective-efficacy on perceived group performance. *Small group research*, 33(3):313–336, 2002.

[108] Mark Keith, Greg Anderson, Douglas Dean, James Gaskin, Trevor Wiser, and Seth Gremmert. The effects of video gaming on work group performance. 2016.

[109] Mark J Keith, Greg Anderson, James Gaskin, and Douglas L Dean. Team video gaming for team building: Effects on team performance. *AIS Transactions on Human-Computer Interaction*, 10(4):205–231, 2018.

[110] Stephen Kelley, Mark Goldberg, Malik Magdon-Ismail, Konstantin Mertsalov, and Al Wallace. Defining and discovering communities in social networks. *Handbook of optimization in complex networks: Theory and applications*, 57:139–168, 2012.

[111] Florian Keusch and Frauke Kreuter. Digital trace data: Modes of data collection, applications, and errors at a glance. In *Handbook of Computational Social Science, Vol 1*. Taylor & Francis, 2021.

[112] Christine Klymko, David Gleich, and Tamara G Kolda. Using triangles to improve community detection in directed networks. *arXiv preprint arXiv:1404.5874*, 2014.

[113] David Knoke and Song Yang. *Social network analysis.* SAGE publications, 2019.

[114] Tetsuro Kobayashi, Ken'ichi Ikeda, and Kakuko Miyata. Social capital online: Collective use of the internet and reciprocity as lubricants of democracy. *Information, Community & Society*, 9(5):582–611, 2006.

[115] Steve WJ Kozlowski. Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review*, 5(4):270–299, 2015.

[116] Steve WJ Kozlowski and Georgia T Chao. The dynamics of emergence: Cognition and cohesion in work teams. *Managerial and Decision Economics*, 33(5-6):335–354, 2012.

[117] Steve WJ Kozlowski and Katherine J Klein. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. 2000.

[118] Claus W Langfred. Is group cohesiveness a double-edged sword? an investigation of the effects of cohesiveness on performance. *Small Group Research*, 29(1):124–143, 1998.

[119] Neuman W Lawrence. Basics of social research: Qualitative and quantitative approaches, 2007.

[120] Matthew D Lerner, Bryce D McLeod, and Amori Y Mikami. Preliminary evaluation of an observational measure of group cohesion for group psychotherapy. *Journal of Clinical Psychology*, 69(3):191–208, 2013.

[121] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007.

[122] Conggai Li, Fan Zhang, Ying Zhang, Lu Qin, Wenjie Zhang, and Xuemin Lin. Efficient progressive minimum k-core search. *Proceedings of the VLDB Endowment*, 13(3):362–375, 2019.

[123] Jianxin Li, Xinjue Wang, Ke Deng, Xiaochun Yang, Timos Sellis, and Jeffrey Xu Yu. Most influential community search over large social networks. In *2017 IEEE 33rd international conference on data engineering (ICDE)*, pages 871–882, San Diego, CA, USA, 2017. IEEE.

[124] Ling Li, Siqiang Luo, Yuhai Zhao, Caihua Shan, Zhengkui Wang, and Lu Qin. Coclep: Contrastive learning-based semi-supervised community search. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 2483–2495. IEEE, 2023.

[125] Rong-Hua Li, Lu Qin, Jeffrey Xu Yu, and Rui Mao. Influential community search in large networks. *Proceedings of the VLDB Endowment*, 8(5):509–520, 2015.

[126] Rong-Hua Li, Lu Qin, Fanghua Ye, Jeffrey Xu Yu, Xiaokui Xiao, Nong Xiao, and Zibin Zheng. Skyline community search in multi-valued networks. In *Proceedings of the 2018 international conference on management of data*, pages 457–472, 2018.

[127] Rong-Hua Li, Jiao Su, Lu Qin, Jeffrey Xu Yu, and Qiangqiang Dai. Persistent community search in temporal networks. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 797–808, Paris, France, 2018. IEEE.

[128] Kwan Hui Lim and Amitava Datta. An interaction-based approach to detecting highly interactive twitter communities using tweeting links. In *Web Intelligence*, volume 14, pages 1–15. IOS Press, 2016.

[129] Qing Liu, Minjun Zhao, Xin Huang, Jianliang Xu, and Yunjun Gao. Truss-based community search over large directed graphs. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, pages 2183–2197, 2020.

[130] Albert J Lott and Bernice E Lott. Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychological bulletin*, 64(4):259, 1965.

[131] Bernice E Lott. Group cohesiveness, communication level, and conformity. *The Journal of Abnormal and Social Psychology*, 62(2):408, 1961.

[132] Lin Lu and Kui Xie. Achievement goals and team leadership in online small group learning. International Society of the Learning Sciences, Inc.[ISLS]., 2018.

[133] Zhao Lu, Yuanyuan Zhu, Ming Zhong, and Jeffrey Xu Yu. On time-optimal (k, p)-core community search in dynamic graphs. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 1396–1407, Kuala Lumpur, Malaysia, 2022. IEEE.

[134] R Duncan Luce and Albert D Perry. A method of matrix analysis of group structure. *Psychometrika*, 14 (2):95–116, 1949.

[135] Chengyang Luo, Qing Liu, Yunjun Gao, and Jianliang Xu. Synergetic community search over large multilayer graphs. 18(6):1412–1424, 2025. doi: 10.14778/3718057.3718069.

[136] Annmarie A Lyles, Colleen Loomis, Scherezade K Mama, Sameer Siddiqi, and Rebecca E Lee. Longitudinal analysis of virtual community perceptions of cohesion: The role of cooperation, communication, and competition. *Journal of Health Psychology*, 23(13):1677–1688, 2018.

[137] Merce Mach, Simon Dolan, and Shay Tzafrir. The differential effect of team members' trust on team performance: The mediation role of team cohesion. *Journal of occupational and organizational psychology*, 83(3):771–794, 2010.

[138] Omid Madani, Sai Ankith Averineni, and Shashidhar Gandham. A dataset of networks of computing hosts. In *Proceedings of the 2022 ACM on International Workshop on Security and Privacy Analytics*, pages 100–104, 2022.

[139] Lucien Maman, Eleonora Ceccaldi, Nale Lehmann-Willenbrock, Laurence Likforman-Sulem, Mohamed Chetouani, Gualtiero Volpe, and Giovanna Varni. Game-on: A multimodal dataset for cohesion and group analysis. *IEEE Access*, 8:124185–124203, 2020.

[140] Michelle A Marks, John E Mathieu, and Stephen J Zaccaro. A temporally based framework and taxonomy of team processes. *Academy of management review*, 26(3):356–376, 2001.

[141] Peter V Marsden. Network data and measurement. *Annual review of sociology*, 16(1):435–463, 1990.

[142] Peter V Marsden. Recent developments in network measurement. *Models and methods in social network analysis*, 8:30, 2005.

[143] Janet McLeod and Kathryn Von Treuer. Towards a cohesive theory of cohesion. 2013.

[144] Xiaoye Miao, Yue Liu, Lu Chen, Yunjun Gao, and Jianwei Yin. Reliable community search on uncertain graphs. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 1166–1179, Kuala Lumpur, Malaysia, 2022. IEEE.

[145] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[146] EL-MOUSSAOUI Mohamed, Tarik Agouti, Abdessadek Tikniouine, and Mohamed El Adnani. A comprehensive literature review on community detection: Approaches and applications. *Procedia Computer Science*, 151:295–302, 2019.

[147] Dan Mønster, Dorthe Døjbak Håkonsson, Jacob Kjær Eskildsen, and Sebastian Wallot. Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & behavior*, 156:24–34, 2016.

[148] Christian Montag, Éilish Duke, and Alexander Markowetz. Toward psychoinformatics: Computer science meets psychology. *Computational and mathematical methods in medicine*, 2016(1):2983685, 2016.

[149] Peter E Mudrack. Defining group cohesiveness: A legacy of confusion? *Small group behavior*, 20(1):37–49, 1989.

[150] Sean A Munson, Karina Kervin, and Lionel P Robert Jr. Monitoring email to indicate project team performance and mutual attraction. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 542–549, 2014.

[151] Mansoureh Naderipour, Mohammad Hossein Fazel Zarandi, and Susan Bastani. Fuzzy community detection on the basis of similarities in structural/attribute in large-scale social networks. *Artificial Intelligence Review*, 55:1373–1407, 2022.

[152] W Lawrence Neuman. *Basics of social research.* 2007.

[153] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[154] Kate G Niederhoffer and James W Pennebaker. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360, 2002.

[155] Robert Z Norman et al. Structural models: An introduction to the theory of directed graphs. 1965.

[156] J.C. Nunnally. *Psychometric Theory.* McGraw-Hill series in psychology. McGraw-Hill, 1978.

[157] Leysan Nurgalieva, Alisa Frik, and Gavin Doherty. A narrative review of factors affecting the implementation of privacy and security practices in software development. *ACM Computing Surveys*, 55(14s):1–27, 2023.

[158] Jakob Ohme, Theo Araujo, Laura Boeschoten, Deen Freelon, Nilam Ram, Byron B Reeves, and Thomas N Robinson. Digital trace data collection for social media effects research: Apis, data donation, and (screen) tracking. *Communication Methods and Measures*, 18(2):124–141, 2024. doi: 10.1080/19312458.2023.2181319.

[159] Laurel W Oliver, Joan Harman, Elizabeth Hoover, Stephanie M Hayes, and Nancy A Pandhi. A quantitative integration of the military cohesion literature. *Military Psychology*, 11(1):57–83, 1999.

[160] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 2021. doi: 10.1136/bmj.n71. URL `https://www.bmj.com/content/372/bmj.n71`.

[161] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. Community detection in social media: Performance and application considerations. *Data mining and knowledge discovery*, 24:515–554, 2012.

[162] Manoj Parameswaran and Andrew B Whinston. Social computing: An overview. *Communications of the association for Information Systems*, 19(1):37, 2007.

[163] Guy Paré, Marie-Claude Trudel, Mirou Jaana, and Spyros Kitsiou. Synthesizing information systems knowledge: A typology of literature reviews. *Information & management*, 52(2):183–199, 2015.

[164] Ravi Paul, John R Drake, and Huigang Liang. Global virtual team performance: The effect of coordination effectiveness, trust, and team cohesion. *IEEE Transactions on Professional Communication*, 59(3):186–202, 2016.

[165] Delroy L Paulhus, Simine Vazire, et al. The self-report method. *Handbook of research methods in personality psychology*, 1(2007):224–239, 2007.

[166] Anthony T Pescosolido and Richard Saavedra. Cohesion and sports teams: A review. *Small Group Research*, 43(6):744–758, 2012.

[167] Michel Plantié and Michel Crampes. Survey on social community detection. In *Social media retrieval*, pages 65–85. Springer, 2012.

[168] Matthew S Prewett. Clarifying the cohesion-performance relationship in teams: Backup behavior as a mediating mechanism. 2006.

[169] Ioannis Psorakis, Stephen Roberts, Mark Ebden, and Ben Sheldon. Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E*, 83(6):066114, 2011.

[170] Meng Qin, Chaorui Zhang, Yu Gao, Weixi Zhang, and Dit-Yan Yeung. Pre-train and refine: Towards higher efficiency in k-agnostic community detection without quality degradation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2467–2478, 2024.

[171] Anat Rafaeli, Shelly Ashtar, and Daniel Altman. Digital traces: New data, resources, and tools for psychological-science research. *Current directions in psychological science*, 28(6):560–566, 2019.

[172] Donald J Rose. On simple characterizations of k-trees. *Discrete mathematics*, 7(3-4):317–322, 1974.

[173] Giulio Rossetti and Rémy Cazabet. Community discovery in dynamic networks: a survey. *ACM computing surveys (CSUR)*, 51(2):1–37, 2018.

[174] Eduardo Salas, Dana E Sims, and C Shawn Burke. Is there a "big five" in teamwork? *Small group research*, 36(5):555–599, 2005.

[175] Eduardo Salas, Rebecca Grossman, Ashley M Hughes, and Chris W Coultas. Measuring team cohesion: Observations from the science. *Human factors*, 57(3):365–374, 2015.

[176] Wm David Salisbury, Traci A Carte, and Laku Chidambaram. Cohesion in virtual teams: validating the perceived cohesion scale in a distributed setting. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 37(2-3):147–155, 2006.

[177] Jessica M Santoro, Aurora J Dixon, Chu-Hsiang Chang, and Steve WJ Kozlowski. Measuring and monitoring the dynamics of team cohesion: Methods, emerging tools, and advanced technologies. In *Team cohesion: Advances in psychological theory, methods and practice*, pages 115–145. Emerald Group Publishing Limited, Leeds, 2015.

[178] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[179] Stanley Emanuel Seashore. *Group cohesiveness in the industrial work group*. University of Michigan Ann Arbor, 1954.

[180] Stephen B Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.

[181] Stephen B Seidman and Brian L Foster. A graph-theoretic generalization of the clique concept. *Journal of Mathematical sociology*, 6(1):139–154, 1978.

[182] Junming Shao, Zhichao Han, Qinli Yang, and Tao Zhou. Community detection based on distance dynamics. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1075–1084, 2015.

[183] John J Shaughnessy, Eugene B Zechmeister, and Jeanne S Zechmeister. *Research methods in psychology*. McGraw-Hill, 2000.

[184] Kelly G Shaver. *Principles of social psychology*. Psychology Press, 2015.

[185] Valerie Sheppard. Research methods for the social sciences: An introduction. *Vancouver: BCCampus*, 2020.

[186] Jessica Shi, Laxman Dhulipala, David Eisenstat, Jakub Łăcki, and Vahab Mirrokni. Scalable community detection via parallel correlation clustering. *Proceedings of the VLDB Endowment*, 14(11):2305–2313, 2021.

[187] JinNuo Shi. A hierarchical model for community identification in complex networks through modularity and genetic algorithm. *Scientific Reports*, 15(1):1–21, 2025.

[188] Yuhyung Shin and Kyojik Song. Role of face-to-face and computer-mediated communication time in the cohesion and performance of mixed-mode groups. *Asian Journal of Social Psychology*, 14(2):126–139, 2011.

[189] Mauro Sozio and Aristides Gionis. The community-search problem and how to plan a successful cocktail party. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 939–948, 2010.

[190] Tobias H. Stark. Collecting social network data. pages 241–254, 2018.

[191] Xing Su, Shan Xue, Fanzhen Liu, Jia Wu, Jian Yang, Chuan Zhou, Wenbin Hu, Cecile Paris, Surya Nepal, Di Jin, et al. A comprehensive survey on community detection with deep learning. *IEEE transactions on neural networks and learning systems*, 35(4):4682–4702, 2022.

[192] Isabelle Szmigin, Louise Canning, and Alexander E Reppel. Online community: enhancing the relationship marketing concept through customer bonding. *International journal of service industry management*, 16(5): 480–496, 2005.

[193] Keith S Taber. The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in science education*, 48:1273–1296, 2018.

[194] Yifu Tang, Jianxin Li, Nur Al Hasan Haldar, Ziyu Guan, Jiajie Xu, and Chengfei Liu. Reliable community search in dynamic networks. *Proceedings of the VLDB Endowment*, 15(11):2826–2838, 2022.

[195] Claire Thornton, Phillip Miller, and Kelly Perry. The impact of group cohesion on key success measures in higher education. *Journal of Further and Higher Education*, 44(4):542–553, 2020.

[196] Alon Tomashin, Ilanit Gordon, and Sebastian Wallot. Interpersonal physiological synchrony predicts group cohesion. *Frontiers in human neuroscience*, 16:903407, 2022.

[197] Bruce W Tuckman and Mary Ann C Jensen. Stages of small-group development revisited. *Group & organization studies*, 2(4):419–427, 1977.

[198] Mark Vaitkus and James Griffith. An evaluation of unit replacement on unit cohesion and individual morale in the us army all-volunteer force. *Military Psychology*, 2(4):221–239, 1990.

[199] Judy L Van Raalte, Allen E Cornelius, Darwyn E Linder, and Britton W Brewer. The relationship between hazing and team cohesion. *Journal of sport behavior*, 30(4):491, 2007.

[200] Kathryn Von Treuer, Janet McLeod, Matthew Fuller-Tyszkiewicz, and Graham Scott. Determining the components of cohesion using the repertory grid technique. *Group Dynamics: Theory, Research, and Practice*, 22(2):108, 2018.

[201] Fabian Walocha, Lucien Maman, Mohamed Chetouani, and Giovanna Varni. Modeling dynamics of task and social cohesion from the group perspective using nonverbal motion capture-based features. In *Companion publication of the 2020 international conference on multimodal interaction*, pages 182–190, 2020.

[202] Eric Wang, Huey-Wen Chou, and James Jiang. The impacts of charismatic leadership style on team cohesiveness and overall performance during erp implementation. *International Journal of Project Management*, 23(3):173–180, 2005.

[203] Fei-Yue Wang, Kathleen M Carley, Daniel Zeng, and Wenji Mao. Social computing: From social informatics to social intelligence. *IEEE Intelligent systems*, 22(2):79–83, 2007.

[204] Jianwei Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, and Ying Zhang. Efficient unsupervised community search with pre-trained graph transformer. *Proceedings of the VLDB Endowment*, 17(9):2227–2240, 2024.

[205] Jianwei Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, and Ying Zhang. Neural attributed community search at billion scale. *Proceedings of the ACM on Management of Data*, 1(4):1–25, 2024.

[206] Yue Wang, Xun Jian, Zhenhua Yang, and Jia Li. Query optimal k-plex based community in graphs. *Data Science and Engineering*, 2:257–273, 2017.

[207] Yuxiang Wang, Xiaoxuan Gou, Xiaoliang Xu, Yuxia Geng, Xiangyu Ke, Tianxing Wu, Zhiyuan Yu, Runhuai Chen, and Xiangying Wu. Scalable community search over large-scale graphs based on graph transformer. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1680–1690, 2024.

[208] Stanley Wasserman and Katherine Faust. Social network analysis: Methods and applications. 1994.

[209] Robert S Weiss and Eugene Jacobson. A method for the analysis of the structure of complex organizations. *American Sociological Review*, 20(6):661–668, 1955.

[210] Joyce Jiyoung Whang, David F Gleich, and Inderjit S Dhillon. Overlapping community detection using neighborhood-inflated seed expansion. *IEEE Transactions on Knowledge and Data Engineering*, 28(5): 1272–1284, 2016.

[211] Xixi Wu, Yun Xiong, Yao Zhang, Yizhu Jiao, Caihua Shan, Yiheng Sun, Yangyong Zhu, and Philip S Yu. Clare: A semi-supervised community detection algorithm. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 2059–2069, 2022.

[212] Xixi Wu, Kaiyu Xiong, Yun Xiong, Xiaoxin He, Yao Zhang, Yizhu Jiao, and Jiawei Zhang. Procom: A few-shot targeted community detection algorithm. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3414–3424, 2024.

[213] Yantuan Xian, Pu Li, Hao Peng, Zhengtao Yu, Yan Xiang, and Philip S Yu. Community detection in large-scale complex networks via structural entropy game. In *Proceedings of the ACM on Web Conference 2025*, pages 3930–3941, 2025.

[214] Kui Xie, Lauren C Hensley, Victor Law, and Zhiru Sun. Self-regulation as a function of perceived leadership and cohesion in small group online collaborative learning. *British Journal of Educational Technology*, 50(1): 456–468, 2019.

[215] Tianyang Xu, Zhao Lu, and Yuanyuan Zhu. Efficient triangle-connected truss community search in dynamic graphs. *Proceedings of the VLDB Endowment*, 16(3):519–531, 2022.

[216] Jaewon Yang and Jure Leskovec. Defining and evaluating network communities based on ground-truth. In *Proceedings of the ACM SIGKDD workshop on mining data semantics*, pages 1–8, 2012.

[217] Jaewon Yang, Julian McAuley, and Jure Leskovec. Detecting cohesive and 2-mode communities indirected and undirected networks. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 323–332, 2014.

[218] Kai Yao and Lijun Chang. Efficient size-bounded community search over large networks. *Proceedings of the VLDB Endowment*, 14(8):1441–1453, 2021.

[219] Hao Yin, Austin R Benson, Jure Leskovec, and David F Gleich. Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 555–564, 2017.

[220] Youngjin Yoo and Maryam Alavi. Media and group cohesion: Relative influences on social presence, task participation, and group consensus. *MIS quarterly*, pages 371–390, 2001.

[221] Xin Yu, Jing Yang, and Zhi-Qiang Xie. A semantic overlapping community detection algorithm based on field sampling. *Expert Systems with Applications*, 42(1):366–375, 2015.

[222] Stephen J Zaccaro and Charles A Lowe. Cohesiveness and performance on an additive task: Evidence for multidimensionality. *The Journal of Social Psychology*, 128(4):547–558, 1988.

[223] Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.

[224] Andrew Zamecnik, Cristina Villa-Torrano, Vitomir Kovanović, Georg Grossmann, Srećko Joksimović, Yannis Dimitriadis, and Abelardo Pardo. The cohesion of small groups in technology-mediated learning environments: A systematic literature review. *Educational Research Review*, 35:100427, 2022.

[225] Andrew Zamecnik, Vitomir Kovanovíc, Srécko Joksimovíc, Georg Grossmann, Djazia Ladjal, Ruth Marshall, and Abelardo Pardo. Using online learner trace data to understand the cohesion of teams in higher education. *Journal of Computer Assisted Learning*, 39(6):1733–1750, 2023.

[226] Fan Zhang, Haicheng Guo, Dian Ouyang, Shiyu Yang, Xuemin Lin, and Zhihong Tian. Size-constrained community search on large networks: An effective and efficient solution. *IEEE Transactions on Knowledge and Data Engineering*, 36(1):356–371, 2024.

[227] Yanxia Zhang, Jeffrey Olenick, Chu-Hsiang Chang, Steve WJ Kozlowski, and Hayley Hung. Teamsense: assessing personal affect and group cohesion in small teams through dyadic interaction and behavior analysis with wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–22, 2018.

[228] Yao Zhang, Yun Xiong, Yun Ye, Tengfei Liu, Weiqiang Wang, Yangyong Zhu, and Philip S Yu. Seal: Learning heuristics for community detection with generative adversarial networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1103–1113, 2020.

[229] Yining Zhao, Sourav S Bhowmick, Fischer Nastassja L., and SH Annabel Chen. How cohesive are community search results on online social networks?: An experimental evaluation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '25, New York, NY, USA, 2025. Association for Computing Machinery.

[230] Yu Zhao, Mark Srite, Sumin Kim, and Jinwoong Lee. Effect of team cohesion on flow: An empirical study of team-based gamification for enterprise resource planning systems in online classes. *Decision Sciences Journal of Innovative Education*, 19(3):173–184, 2021.

[231] Lanqin Zheng and Lanqin Zheng. A socially shared regulation approach to improving group cohesion, collective efficacy, and regulation skills in cscl. *Knowledge building and regulation in computer-supported collaborative learning*, pages 83–96, 2017.

[232] Lanqin Zheng, Ronghuai Huang, and Junhui Yu. The impact of different roles on motivation, group cohesion, and learning performance in computer-supported collaborative learning (cscl). In *2014 IEEE 14th International Conference on Advanced Learning Technologies*, pages 294–296, Athens, Greece, 2014. IEEE.

[233] Rui Zhou, Chengfei Liu, Jeffrey Xu Yu, Weifa Liang, Baichen Chen, and Jianxin Li. Finding maximal k-edge-connected subgraphs from a large graph. In *Proceedings of the 15th international conference on extending database technology*, pages 480–491, 2012.

[234] Zhijian Zhuo, Bilian Chen, Shenbao Yu, and Langcai Cao. Overlapping community detection using expansion with contraction. *Neurocomputing*, 565:126989, 2024.

[235] Philip G Zimbardo, Robert Lee Johnson, Vivian McCann, and Carol Carter. *Psychology: core concepts*. Allyn and Bacon Boston, 2003.

Community Search: A Survey of the State-of-the-Art from Algorithms to Learning, Complex Graphs, and Interaction

Longxu Sun, Xin Huang
Hong Kong Baptist University
Hong Kong, China
{cslxsun, xinhuang}@comp.hkbu.edu.hk

Abstract

Community search is a query processing task of finding query-dependent communities in graphs, which has particularly useful applications in social circle discovery, personalized recommendation, and fraud detection. Over the past years, extensive studies have been devoted to this challenging problem due to diverse application scenarios and the complexity of communities in nature. Most classical community search works follow *an algorithmic framework* by designing a dense subgraph-based model and developing offline indices to accelerate online community search. However, many recent works have studied novel problems of community search over complex graphs and extended new directions from this algorithmic framework, which unfortunately lacks a timely review. This survey reviews the state-of-the-art community search works and summarizes them in terms of three aspects as follows. First, we introduce the latest community search works, in terms of methodologies from *querying algorithms* to *machine learning models*. Second, we give a review in terms of graph types, from structural and attributed graphs to complex graphs. We focus on the studies of community search over two kinds of complex graphs: *heterogeneous information networks* and *multilayer graphs*. Third, we present a newly exciting direction of *interactive community search*, which needs human-in-loop interactions and allows users to give feedback for advising community refinement. We finally share future directions and promising opportunities in this research area of community search.

1 Introduction

The topology structure of graphs widely exists in numerous real-world scenarios, such as social networks, citation networks, e-commerce platforms, traffic transportation systems, and biomedical networks [3, 52, 59, 60, 63]. Due to various attributes associated with nodes and edges, one simple graph cannot accurately depict real relationships. Therefore, many kinds of graphs are used to comprehensively model these scenarios, including attributed graphs, heterogeneous information networks (HINs), temporal graphs, multilayer graphs, bipartite graphs, dynamic graphs, and so on [15, 19, 22, 31, 33, 43, 58, 70, 81]. Communities that are formed by a group of individual nodes, naturally exist in graphs, such as family groups, biological functional organs, event groups in social networks, and so on. The discovery of community is very useful for understanding network dynamics and user behaviors. These communities are characterized by diverse attributes and different sizes of participants, which presents challenges in developing effective algorithms and models to discover them.

The task of *community search* focuses on identifying query-oriented cohesive communities in graphs, which is determined by user-provided queries [20, 30]. Unlike *community detection*, which seeks to uncover all potential communities within a graph, community search has demonstrated fast response efficiency and powerful capabilities in various personalized search tasks [13, 55, 63]. The applications of community search include social circle discovery in social networks, personalized recommendations in e-commerce platforms, the formation of interdisciplinary expert groups in collaborative networks, drug discovery in biological networks, and fraud detection across different financial systems [13, 59, 60, 80]. These applications highlight the utility and impact of community search in enhancing user experience across multiple domains.

A wide range of community search studies [19, 28, 29] have been investigated in the literature. Fang et al. [20] provides a systematic classification of existing community search studies based on various dense subgraphs, including the k-core, k-truss, k-clique, quasi-clique, k-plex, k-edge-connectivity components, and other metrics-based densest subgraphs. Besides structural community search over simple graphs, another major contribution of this survey summarizes various works of attributed community search for different attribute types, e.g., keywords, locations, temporal, weighted influences, and profiles. However, most discussed works [20][30] had been published on or before 2020. Based on our investigation of recent studies, community search have been studied from *static and querying algorithms* to *interactive and learning-based approaches* [2, 10, 12, 25, 36, 37, 42, 43, 45, 47, 53, 69, 70, 74, 82, 83, 88]. Even more, the problems of community search have been studied from *simple and attributed networks* to *more complex graphs (e.g. HINs and multilayer graphs)* [1, 7–9, 14, 17, 26, 39, 41, 65, 73, 79, 84–87]. Thus, it is timely and important to have a state-of-the-art survey on existing community search works. In this survey, we make contributions to discuss and categorize these state-of-the-art community search works, in terms of three aspects, including community search methodologies, complex networks, and user interaction.

First, in terms of community search methodologies, we review advanced works using querying algorithms and learning-based approaches. These works apply classical querying algorithms, e.g., exact algorithms, heuristic solutions with pruning strategies, and indexing-based methods, to address attribute community search. Furthermore, we summarize recent learning-based approaches, e.g., graph neural networks and transformers, to significantly improve the quality of community search and introduce fast solutions.

Second, in terms of complex networks, we summarize recent community search works over two complex networks of HINs and multilayer graphs. Specifically, HINs consist of multiple typed nodes and edges with a given network schema. Thus, community search over HINs usually needs an input of meta-path query. On the contrary, multilayer networks consist of multiple graph layers, where each layer has one graph with internal edges and two graphs at different layers have cross-layer edges. Community search over multilayer graphs usually needs no input meta-path but focuses on the interested querying nodes.

Third, in terms of user interaction, we introduce a new direction of interactive community search, which allows users to give feedback for improving community search answers. Last but not least, we also point out interesting research directions and promising opportunities leveraging recent large language models (LLMs), including the retrieval augmented generation (RAG) improvement using community search, LLM-enhanced community search solutions, and developing a unified dense subgraph model for various community search tasks.

2 Community Search on Attributed Graphs: Algorithms and Learning

We start with a widely used graph model of attributed graphs. Then, we introduce the state-of-the-art studies of community search over attributed graphs, in terms of *algorithmic approaches* and *learning models*.

Attributed graphs naturally exist in various real-world scenarios. For instance, individual nodes are often associated with multiple attributes of hobby tags in social networks; In academic collaboration networks, author nodes tend to specialize in a few attributes of research topics; In protein-protein interaction networks, each protein node is associated with several distinct functions. Therefore, a keyword-based attributed graph consists of vertices, edges, and a set of keywords, denoted as $G(V, E, A)$, where V is the set of vertices, $E \subset V \times V$ is the set of edges, and $A = \{a_1, a_2, ..., a_n\}$ is a set of keywords associated with vertices. The objective of community search over attributed graphs (ACS) is to find query-dependent communities with cohesive structure and homogeneous attributes. As summarized in the previous survey [20], there were two representative community search works on keyword-attributed graphs, which adopt k-core model [19] and k-truss model [28] to measure a structural density of communities. However, both methods [19, 28] use algorithmic approaches to solve community search problems. After an investigation of recent studies on attributed community search, we summarize new research directions, in terms of three aspects: community model improvement [6, 90], new problems with complex queries [44, 48, 78], and new solutions of machine learning models [18, 24, 34, 38, 61, 72, 78]. In the following, we introduce a few representative works of querying algorithms and machine learning based attributed community search.

2.1 Algorithmic Approaches

We discuss two representative ACS studies based on the types of keyword/numeric attributes and semantic-based attributes, respectively.

2.1.1 Keyword and numeric attributed community search

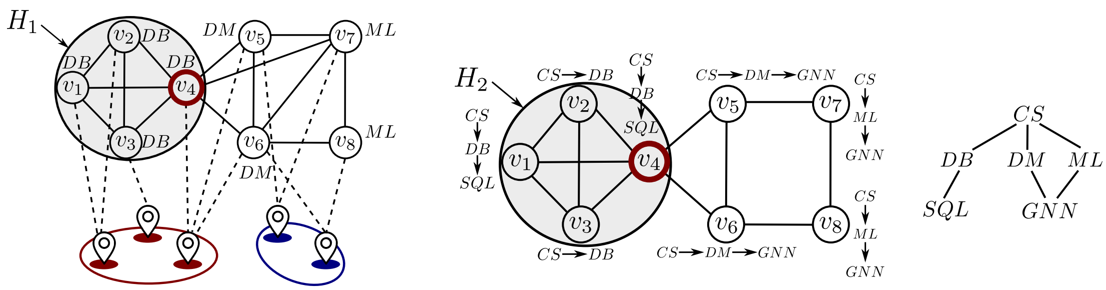
Different from previous ACS studies [19, 28] that only consider keyword attributes, recent works [48, 56, 57] consider the combination of *keyword attributes* and *numeric attributes* in attributed community search tasks.

Liu et al. [48] introduce a new combined attribute score to measure an attribute similarity of two vertices by employing a distance metric. Specifically, it applies Euclidean distance to compute the numeric location distance, while applying Jaccard distance to eliminate the similarity of textural keywords. Moreover, it employs a unified function to combine multiple kinds of distance functions, which provides better scalability to support multiple types of attributes, including both textual attributes (keywords) and numeric attributes (numbers). Based on this unified attribute score, Liu et al. propose the problem of vertex-centric attributed community search problem (VAC-Problem) as follows.

VAC-Problem [48]. Given an attributed graph $G = (V, E, A)$, a set of querying vertices $Q \subseteq V$, and a parameter k, the problem of vertex-centric attributed community search aims to find the community $H \subseteq G$ satisfying the following properties: (1) query participation, i.e., $Q \subseteq V(H)$; (2) structure cohesiveness, such that H is a connected k-truss; (3) attribute cohesiveness, requiring that H has the smallest attribute score; (4) the maximality, requiring that there does not exist another subgraph $H' \supset H$ satisfying the above three properties.

Here, the k-truss [28] is a dense subgraph pattern H, requiring that every edge (u, v) in $E(H)$ is contained in at least $k - 2$ triangles in H. The unified attribute score is denoted as $Ascore(u, v) = \alpha \cdot \frac{Sdist(u,v)}{Sdist_{max}} + (1 - \alpha) \cdot \frac{Tdist(u,v)}{Tdist_{max}}$, where $Sdist(\cdot)$ and $Tdist(\cdot)$ denote the spatial distance and textual distance, respectively. Note that $Sdist_{max}$ and $Tdist_{max}$ are the maximal limits used for normalization.

The VAC-Problem is proven as NP-hard, reduced from the maximum clique problem. To solve the VAC-problem, a DFS-manner exact algorithm and a BFS-based exact algorithm are proposed. Due to the NP-hardness, several heuristic rules are proposed for vertex addition and deletion separately to improve efficiency performance. These heuristic strategies can be summarized into three aspects: pruning strategy, early termination strategy, and ordered search strategy. The pruning strategy pre-removes some useless edges and vertices based on the property of the k-truss model to reduce the search space. The early termination strategy stops the algorithm early by precomputing the low bounds of subgraphs' attribute scores. The ordered search strategy precomputes the vertex attribute score to query vertices and



(a) VAC Problem (b) CSSAG Problem

Figure 1: Typical examples of two attributed community search problems.

arranges the order when adding/deleting vertices from the graph. To further improve the efficiency and scalability of proposed algorithms, an approximate algorithm is proposed by achieving a 2-approximation guarantee to exact algorithms.

2.1.2 Semantic-based attributed community search

Besides the simple types of keywords and numerical values, attributes can have semantic relationships. Lin et al. [44] studied a problem of community search in semantic-based attributed graphs (CSSAG-Problem), which makes use of semantic information among attributes and gives an interpretability of community answers. In semantic-based attributed graphs, two keyword attributes may have a relationship in a directed acyclic semantic graph. This kind of graph is also meaningful in real life, e.g., identifying the synonyms and distinguishing polysynonyms. In the semantic attribute graph, attributes of vertex v is not a set of independent attributes, but a small directed acyclic semantic graph, denoted as $SG(v) = (V_{SG(v)}, E_{SG(v)})$, where $V_{SG(v)}$ is the semantic keyword set of the vertex v, $E_{SG(v)}$ is the semantic relations between theses keywords. In the semantic-based attributed graph H, it proposes the maximal common subgraph $MSG(H)$ to denote attribute cohesiveness, by satisfying the following properties: (1) $\forall v \in V(H), MSG(H) \subseteq SG(v)$; (2) there exists no other $MSG'(H)$ such that $MSG(H) \subseteq MSG'(H)$; (3) there exists at least one edge in $MSG(H)$. To find a subgraph H with the maximal common sub-semantic graph, the CSSAG-Problem is formulated as follows.

CSSAG-Problem [44]. Given a semantic-based attribute graph $G(V, E, A)$, a positive integer k, and a query vertex $q \in V$, the CSSAG-Problem is to find a maximal set of communities H satisfying the following properties: (1) H is a connected subgraph and $q \in V(H)$; (2) the degree of v in $V(H)$ is no less than k; (3) there exists no other $MSG'(H)$ such that $MSG(H) \subseteq MSG'(H)$, i.e., the $MSG(H)$ is maximal.

As shown in Figure 1(a), given an attributed graph with both keyword and location attributes, a query vertex v_4, and parameter $k = 4$, the VAC problem requires the community follows three constraints: H_1 is a 4-truss; vertices in H_1 has similar keywords; the vertices in H_1 is close to each other. Thus, the VAC problem aims to find the community H_1 that has similar keywords "DB" and is in the same region. Figure 1(b) provides an example of a semantic attributed graph, where each vertex corresponds to a small semantic subgraph. Given the query v_4 and $k = 3$, the CSSAG problem finds the community H_2, where each vertex in H_2 has a degree no less than 3, and the common sub-semantic graph $MSG(H) = \{CS, DB\}$ is maximal.

In summary, besides the above two ACS works, there are also other ACS studies, including anchored vertex set exploration in attributed graphs [6], k-truss-based cohesive attributed community search with maximal common attributes [90], and a unified hierarchy index structure for attributed community search [77].

2.2 Learning Approaches

Different from algorithmic approaches that bulid up the indexes in advance and then support the online community querying, there are a dozen of recent studies [4, 11, 18, 24, 34, 38, 61, 71, 72, 78] that develop machine learning models to learn the properties among trained communities and queries for tackling community search problems. In the following, we introduce two kinds of learning approaches based on *graph convolutional neural networks* [24, 34] and *transformer* [71, 76], respectively.

GNN-based Approach. Gao et al. [24] propose the ICS-GNN framework by applying GNN-based approaches to solve interactive community search problems for the first time. It first encodes the attributes to vertex features, then trains the GNN model by labeling a set of positive vertices and negative vertices separately for each query process. Following their work, several studies focus on applying the popular learning techniques to solve community search problems in attributed graphs. Jiang et al. [34] introduce the AQD-GNN model to solve community search problems on keyword-attributed graphs. Different from [24] that supports vertex-only queries, AQD-GNN can handle attribute-vertex combined queries in keyword-attributed graphs in three steps. First, it constructs a vertex-attribute bipartite graph to model the relationship between vertices and keyword attributes. Second, it applies the GNN model to bipartite graphs in order to learn both vertex feature similarity and topological similarity. In the training process, besides the labeling of positive/negative vertices, it further selects a set of query-community pairs as the training data to enhance the learning model. Finally, it constructs a fusion graph by capturing vertices with high attribute and topology similarity to the query vertices and applying BFS search to ensure the connectivity of answer community.

Transformer-based Approach. Most of the GNN-based community search approaches require labeling vertices during the training process, i.e., supervised learning and semi-supervised learning. Wang et al. [71] first propose an unsupervised transformer-based approach called transformer-based community search framework that uses zero label (TransZero), which does not required labeled vertices. TransZero contains two phases: pre-training phase and online search phase. In the pre-training phase, it proposes a community search graph transformer model to learn both vertex-level and subgraph-level representations. After that, it designs two kinds of self-supervised loss functions, including personalization loss and link loss, to avoid labeling vertices, and then conducts loss computation. In the online search phase, it first computes the community score for each vertex based on the pre-trained transformer model, and then designs a modularity-based function called community identification with expected score gain solver (IESG) to identify the community, where the IESG problem is proved to be NP-hard. Another transformer-based approach called CSFormer contains three phases [76]. First, it proposes a l-hop neighborhood community vector based on the n-th order h-index community model, which can be used to extract cohesiveness features. Based on the cohesiveness features extractor, it then proposes an attention-based transformer model to predict the vertex coreness. Finally, it conducts the online k-core community search based on predicted vertex coreness.

Besides the above studies, Li et al. [38] firstly introduce to use of contrastive learning approaches to train the community using semi-supervised learning. Wang et al. [72] extend the attributed community search works based on GNN models to billion scale graphs. Fang et al. [18] propose an inductive learning approach to help improve the scalability of learning models, which provides a fairer way to train the datasets.

3 Community Search on Complex Graphs

The investigation of community search on complex graphs has gained significant attention in recent years [5, 15, 22, 81]. Unlike simple graphs with categorized attributes or features [19, 28], complex graphs

Table 1: A summary of existing community search problems in HINs and multilayer graphs.

Community Model	Graph Type	Input Queries	Dense Subgraph Pattern	Community Homogeneity
CSH [22]	Heterogeneous Information Networks	Vertices; Meta-path	$(k,\mathcal{P})$-core	Single Vertex Type
CSSH [32]		Vertices	$(k,\mathcal{P})$-core	Single Vertex Type
ICSH [89]		Meta-path	$(k,\mathcal{P})$-core	Multiple Vertex Types
SACH [51]		Vertices; Meta-path	$(k,\mathcal{P})$-core	Single Vertex Type
SSH [49]		Vertices; Meta-path; Triplet Constraints	$(k,\mathcal{P})$-core	Multiple Vertex Types
MRCS [31]		Vertices; Triplet Constraints	k-core	Multiple Vertex Types
CASH [40]		Vertices	Learning-based Component	Single Vertex Type
CS-DAHIN [62]		Vertices; Meta-path	Learning-based Component	Single Vertex Type
LCD [55]	Multilayer Graphs	Vertices	Random Walk-based Component	Multiple Vertex Types
FTCS [5]		Vertices	k-truss	Multiple Vertex Types
CTCS [50]		Vertices	k-truss	Multiple Vertex Types
MCS [67]		Vertices	(k,d)-core	Multiple Vertex Types
SynCS [54]		Vertices	k-SynCore	Multiple Vertex Types
CSML [4]		Vertices	Learning-based Component	Multiple Vertex Types

encompass multiple types of attributes and exhibit diverse structures, e.g., heterogeneous information networks (HINs) [21, 22, 31, 32, 40, 49, 51, 62, 68, 89], multilayer graphs [4, 5, 50, 54, 55, 67], and dynamic graphs [53, 62, 69, 70, 81]. Table 1 summarizes the existing community search problems in heterogeneous information networks and multilayer graphs for a comparison, in terms of the query pattern, dense subgraph pattern, community type, and problem hardness. In the following section, we primarily focus on community search in two kinds of complex graphs: *heterogeneous information networks* and *multilayer graphs*.

3.1 Community Search over Heterogeneous Information Networks

Heterogeneous information networks (HINs) can model complex entities and relations, where nodes and edges have multiple types in the graph. For instance, authors write academic papers that are published in various venues. Thus, in this academic network, nodes can be categorized into distinct types such as authors, papers, and venues, while the edges represent different relationships including authorship and publication. HINs are composed of multiple types of nodes and edges, defined as $G(V,E,\mathcal{A},\mathcal{R})$, where V is the set of vertices, E is the set of edges, $\mathcal{A}=\{A_1,A_2,...,A_n\}$ is the set of node types, and $\mathcal{R}=\{R_1,R_2,...,R_m\}$ is the set of edge types. The type mappings $\psi : V \to \mathcal{A}$ and $\phi : E \to \mathcal{R}$, such that $\psi(v)=A_1$, $\phi(e)=R_1$. Community search over heterogeneous information networks (HIN-CS) attracted significant attention in recent years [22, 31, 32, 40, 49, 51, 62, 89]. Various methodologies have been proposed to identify communities in HINs. In the following, we categorize HIN-CS studies in terms of two directions: meta-path-driven HIN-CS [22] and constraints-based HIN-CS [31].

3.1.1 Meta-path-driven HIN-CS

Fang et al. [22] propose a meta-path-based $(k,\mathcal{P})$-core model and study the community search problem in heterogeneous graphs (CSH-Problem) based on $(k,\mathcal{P})$-core. Given a heterogeneous graph G, a symmetric meta-path $\mathcal{P}$, and an integer k, a $(k,\mathcal{P})$-core is defined as the largest subgraph $H_{k,\mathcal{P}}$ of G such that every node v in $H_{k,\mathcal{P}}$ are $\mathcal{P}$-connected with at least k $\mathcal{P}$-neighbors of the same type, i.e. $\forall v \in V_{k,\mathcal{P}}$, $\alpha(v,H_{k,\mathcal{P}}) \geq k$. The $\mathcal{P}$-neighbors of v represent nodes connected to v via an instance of $\mathcal{P}$ (e.g., "$Author \to Paper \to Venue \to Paper \to Author$" in an academic HIN) in H. The $\mathcal{P}$-neighbors have the same type as node v. The definition of $\mathcal{P}$-degree is the number of $\mathcal{P}$-neighbors of v in H, denoted as $\alpha(v,H)$.

The problem of $(k, \mathcal{P})$-core-based community search [22]. Given a heterogeneous graph G, a query vertex q, a symmetric meta-path $\mathcal{P}$, and an integer k, the $(k, \mathcal{P})$-core-based community search problem is to output a community H, such that H contains q; H is a $(k, \mathcal{P})$-core. This problem extends the minimum degree measure k-core to a meta-path-based $(k, \mathcal{P})$-core. The target community H has two main characteristics. (1) For each node v of a community H, there are at least k other nodes, which can be connected to v via instances of a particular meta-path $\mathcal{P}$, within H. (2) H contains a group of nodes with the same type of query q.

Basic $(k, \mathcal{P})$-core algorithm. A straightforward solution of $(k, \mathcal{P})$-core-based community search problem is to transform the heterogeneous graph to a homogeneous graph via meta-path $\mathcal{P}$, and perform core decomposition to find the largest k-core subgraph. There are three main steps of basic algorithm. First, it collects all vertices with target node type to set S. Second, it enumerates all meta-path instances in $\mathcal{P}$ starting with v for each vertex $v \in S$, and adds edges between v and each of its $\mathcal{P}$-neighbors. Finally, it finds a connected k-core containing q. However, there are two weaknesses of this straightforward solution: (1) leads to some vertices weakly engaged; (2) causes high degrees and high clustering coefficients.

Edge/Vertex-disjoint $(k, \mathcal{P})$-core [22]. To overcome the weakness, two definitions of edge-disjoint path and vertex-disjoint path are proposed to formulate useful subgraphs of edge/vertex-disjoint $(k, \mathcal{P})$-cores. Edge-disjoint paths set $\psi[v]$ is a set of edge-disjoint paths if for any two path instances p_1, $p_2 \in \psi[v]$, their i-th $(1 \leq v \leq l)$ edges are different and $(l+1)$-th nodes are different. Similarly, the vertex-disjoint paths set $\psi[v]$ is a set of vertex-disjoint paths if for any two path instances p_1, $p_2 \in \psi[v]$, their i-th $(2 \leq v \leq l+1)$ nodes are different. Thus, given a heterogeneous graph G, a symmetric meta-path $\mathcal{P}$, and an integer k, an edge/vertex-disjoint $(k, \mathcal{P})$-core is the largest subgraph $E_{k,\mathcal{P}}$ / $V_{k,\mathcal{P}}$ of G such that every node v in $E_{k,\mathcal{P}}$ /$V_{k,\mathcal{P}}$ has at least k $\mathcal{P}$-neighbors with the same type and are $\mathcal{P}$-connected, i.e. $\forall v \in E_{k,\mathcal{P}}$ $(V_{k,\mathcal{P}})$, $\beta(v, E_{k,\mathcal{P}}) \geq k$ / $\gamma(v, V_{k,\mathcal{P}}) \geq k$.

A lazy peeling-based algorithm and its improvement. Based on the edge-disjoint $(k, \mathcal{P})$-core, a greedy algorithm is proposed to improve the efficiency of the basic algorithm. The lazy peeling process contains three parts: (1) it first collects a set S of all vertices with target type; (2) then it uses a greedy strategy to compute the edge-disjoint $(k, \mathcal{P})$-core, $\beta(v, E_{k,\mathcal{P}})$, with an approximation ratio of $\frac{1}{l}$; (3) it performs core decomposition by removing nodes with $\beta(v, E_{k,\mathcal{P}}) < k$ one by one. To further improve efficiency, a batch peeling-based algorithm is proposed. In summary, the $(k, \mathcal{P})$-core-based community search problem aims to find a single type of community close to the query node q by using semantic information offered by HINs.

3.1.2 Constraint-based HIN-CS

In the following, we introduce representative HIN-CS works, driven by query constraints instead of meta-paths.

Minimum relational community search [31]. Jian et al. [31] propose a problem called Minimum Relational Community Search problem (MRCS-Problem) by setting relational constraints. Given a heterogeneous graph G, a concerned label set L_S, a constraint set S, and a query node q, the MRCS-Problem is to find a community H in G, such that H contains q; H is a relational community r-com; and $|H|$ is minimized.

This problem allows users to design the paired relational constraints personally. Formally, the relational constraint is defined as a triplet $s =< l_1, l_2, k >$, where $l_1, l_2 \in L_G$ and $k \geq 1$. Note that each node with label l_1 must connect with at least k neighbors with label l_2, where l_1 and l_2 can be the same or different. Users can also define a constraint that focuses on not only node types but also edge types, such that, $s =< l_1, l_2, l_3, k >$, where node with label l_1 must connected with at least k neighbors with label l_2 and linked with edge labeled as l_3. The constraint set S is a set of relational constraints $S = \{s_1, s_2, ...s_t\}$. The set of concerned labels, L_S, is defined as $L_S = \{l| < l, l', k >\in S\}$.

MRCS exact algorithm and greedy algorithm. The key point to solve the MRCS-Problem is how to compute the qualified connected components in graph G. A message-passing algorithm [31] is developed to iteratively remove disqualified nodes. On the basis of this message passing algorithm, four extension algorithms are proposed, including an exact algorithm, a greedy algorithm, a local search approach, and a round index-based algorithm. The basic exact solution contains three main steps. It first computes all connected qualified node components using the message passing algorithm. Then, it finds the largest connected subgraph containing the query node. After that, it removes a group of nodes recursively until there is no node that can be removed. For different removing groups, it outputs the result with the minimum number of nodes. However, this method can be costly since it tries all possible groups of nodes. By improving the efficiency, a greedy algorithm has been proposed. It changes from trying all possible ways to get a globally minimized answer by removing the largest removable group in each iteration.

In summary, the MRCS-Problem is to find a query-dependent small-sized community that satisfies user-given relational constraints. The constraints can restrict both inter-layer and intra-layer relationships. However, the answer quality is highly related to the given constraints. However, it may be hard for users to design a set of suitable constraints.

3.1.3 Discussions

Both meta-path-based community search [22] and constraint-based community search [31] focus on heterogeneous information networks and adopt k-core dense subgraph models, but the community answers can be quite different. The meta-path-based $(k, \mathcal{P})$-core requires every vertex v in H has at least k numbers of neighbors u where u and v is connected via a meta-path $\mathcal{P}$. Users are required to provide a meta-path $\mathcal{P}$, and the community is required to be a $(k, \mathcal{P})$-core containing multiple of meta-paths. On the other hand, constraint-based relational community search [31] asks users to provide multiple pairwise constraints between pairs of vertices. For example, they can set up different k-core conditions for different pairs of vertices, i.e., set 3-core between "paper" and "venue", and set 2-core between "author" and "paper". It is more flexible to organize the relationship between different pairs of entities, but it needs more input parameters.

In addition to the above two HIN-CS works [22][31], there are other studies of community search over heterogeneous information networks. As shown in Table 1, several studies [32, 49, 51, 89] also employ meta-path-based approaches and $(k, \mathcal{P})$-core model, which facilitates the identification of communities that contain specified meta-paths critical to the analysis. Jiang et al. [32] propose the concept of non-nested meta-path core to identify communities in star-schema HINs and do not require a meta-path as input. [49] also supports user-defined constraints to help make the search more personalized. The works [40, 62] utilize learning-based approaches to develop a unified representation of HINs, enabling the discovery of flexible communities.

3.2 Community Search over Multilayer Graphs

Heterogeneous information networks mainly focus on the relationship between different types of entities, for example," $author \xrightarrow{write} paper \xrightarrow{published} venue$", which ignores the relationships between the same entities. Different from HINs, multilayer graph models the vertices with the same type in the same layer, and constructs multilayer graphs by entity types. Given a multilayer graph $MG(V_M, E_M, \mathcal{L})$, where $V_M = \bigcup_{i=1}^{l} V_i$ represents vertices in all l layers, $E_M = \bigcup_{i=1}^{l}\{E_i \cup \bigcup_{i<j\leq l} E_{ij}\}$ represents all inter-layer edges and cross-layer edges, and a function of layer identification $\mathcal{L}(.) \in \mathcal{Z}^+$. Thus, the academic graph

can be divided into a three layer graph containing the author layer, the paper layer, and the venue layer. Edges in the author layer represent the coauthorship between authors, edges in paper layer represent citations, and edges between author and paper represent authorships. In the literature, community search over multilayer graphs (MLCS) has attracted increasing attention [4, 5, 23, 46, 50, 54, 55, 67, 75], due to its significant impact in various applications. In this section, we mainly discuss two representative works of random walk-based community search [55] and (k, d)-core-based cross-layer community search [67] in multilayer graphs.

3.2.1 Random-walk based MLCS

Random-walk-based multilayer community search [55]. Luo et al. [55] propose a problem called local community detection in multiple networks, which aims to find a community in each layer of the heterogeneous graph. Given a heterogeneous graph $G = \{G^1, G^2, ...G^n\}$, a set of query nodes Q, the community search problem in heterogeneous graphs is to detect all relevant communities in each layer of G.

RWM-based method. A model of random walk in multiple networks (RWM) [55] is proposed in the following procedures: (1) send a random walker to each layer of the graph; (2) detect query-relevant subnetwork relations by calculate the cosine similarity between two layer; (3) the walkers reinforce each other by dynamically modifying their transition matrices; (4) sort nodes according to the converged score vector; (5) for each compute the conductance of the subgraph introduced by the top l ranked nodes (L is the number of non-zero elements in $x_i^{(T)}$.) ; (6) the node set with the smallest conductance can be considered as the detected community. However, a straightforward solution of performing RWM suffers from the limitation of imbalanced visiting probability vectors. To tackle it, an early stopping strategy splits the RWM model into two parts, before the error bound T, it updates both the transition matrices and the visiting vectors; after T, it keeps the transition matrices static and only updates the visiting probability vectors. Another is a partially updating strategy which splits the visiting probability vector into two parts and keeps the unrelated part static.

In summary, this work of community search over multilayer graphs [55] proposes a RWM model motivated by the random walk with restart strategy, trying to find out communities with a small conductance at each layer.

3.2.2 Cross-layer based MLCS

Sun et al. [67] introduce a (k, d)-core model to identify communities characterized by dense inter-layer and intra-layer connections, as well as robust connectivity among all pairs of sub-communities across different layers. It presents three key contributions: (1) the introduction of a fully-connected and a path-connected community model; (2) the design of a compact (k, d)-core index; and (3) the development of several efficient algorithms for community querying.

Fully-connected multilayer community [67]. Before introducing the fully connected community, let's first introduce the cross-layer (k, d)-core, which is defined across two layers. Let $H(H_i, H_j, E_{ij}^H) \subseteq MG$ denote a two-layer subgraph in the multilayer graph MG, and assume that k and d are two user-specified parameters. Given a multilayer subgraph $H \subseteq MG$, we say that H is a full-layer connected multilayer community *if and only if* for every pair of layers $i, j \in \mathcal{L}(H)$, there exists a strong cross-layer connectivity between G_i and G_j, such that $\forall i, j \in \mathcal{L}(H)$, $H_i \xleftrightarrow{H} H_j$.

Cross-layer community search [67]. Given a multilayer graph $MG(\ V_M, E_M, \mathcal{L})$, a set of query vertices $Q \subseteq V_M$, two positive integer parameters k, d, the problem of cross-layer community search in multilayer graphs (MCS-Problem) is to find a connected community $H \subseteq MG$ satisfying the following four constraints: query-dependent personalization such that $Q \subseteq V(H)$; Core-dense internal layers such that $\forall i \in \mathcal{L}(H)$, H_i is a connected k-core; *fully-connected cross-layers:* $\forall i, j \in \mathcal{L}(H)$, two layers H_i and H_j are connected via a (k, d)-core in H; *cross-layer maximization:* $|\mathcal{L}(H)|$ is maximized. To address this problem, a fast search algorithm is developed using a binary search strategy of potential community's layer number. The (k, d)-core index-based algorithms have demonstrated efficient performance on a large-scale multilayer graph dataset, FriendFeed (with 510,338 nodes, 20,204,534 edges, and 3 layers), using a short response time of around ten seconds. In contrast, RWM [55] and FirmTruss have failed to generate community results on FriendFeed.

Discussion of multilayer community search. Besides the above two studies [55][67], there are several other multilayer graph analytics studies on community search in multilayer graphs. Several studies [23, 46, 75] work on core decomposition of multilayer graphs, which help keep the records of dense multilayer subgraphs for efficient community search. Another line of research investigates k-truss-based community models and querying approaches in multilayer graphs [5, 50]. The k-core-based community search in multilayer graphs [54, 67] explores a new formulation of cross-layer communities and fast index construction. Furthermore, Behrouz et al. [4] leverage graph neural network (GNN) techniques to discover communities in multilayer graphs.

4 Interactive Community Search

Communities in graphs exhibit a wide range of different topological structures and attributes, which makes it challenging to identify exact communities relying solely on *a single and static community model*. To fit with different structures of communities and enhance the accuracy of query answers, existing community search models usually need a set of parameters to adjust the density, community size, diameter, target attributes, or even meta-paths. However, it is difficult for users to adjust parameters to navigate the correct search process, especially when users are not familiar with datasets and queries. To tackle these challenges, a few studies [24][66] are devoted to the problem of interactive community search (ICS), which allows users to refine community search answers by simply adding/deleting nodes into/from the community answer within a given small rounds of interactions. In this section, we first introduce the graph neural network-based interactive community search (ICS-GNN) [24]. We then present a flexible framework of interactive community search that allows for fitting with the interaction schema for several existing community search models [66].

4.1 Interactive Community Search via Graph Neural Network

Gao et al. [24] study the problem of interactive community search using GNN models as follows.

The k-sized maximum-GNN-score based community search [24]. The problem of k-sized Maximum-GNN-scores community search aims to find a k-sized connected community with the maximum GNN scores. The authors first abstract the GNN model as $P = GNN(A, F, W)$, where A is the adjacency matrix, F is the feature matrix, and W is the matrix of learnable parameters. Here, P is a mapping of probability, such that $P[i]$ represents the probability that the i-th vertex belongs to the target community. Given an enriched graph with GNN scores $G = (V, E, F, P)$, query vertex q and community size k, where P is the community membership possibility, the kMG problem is to find the k-sized community H with the maximum GNN scores, satisfies three constraints: (1) the query vertex $q \in V(H)$ and H is connected; (2) $|V(H)| = k$; and (3) the sum of GNN scores $\sum_{u \in V(H)} P[u]$ is maximum. This problem is shown to be NP-hard by reducing the Knapsack problem. The solution of GNN-based approach first builds a suitable-sized candidate subgraph; Then, it trains the GNN model on the candidate subgraph guided by the query vertex and labeled positive/negative vertices. After that, it infers the probability scores for all relevant vertices. Finally, it outputs the k-sized community with the maximum GNN scores.

Interactive community search via GNN. An interactive community search framework [24] is proposed to allow users to interactively provide feedback during the community search process so that the answer can be refined iteratively. The interactive process is conducted as follows: in each iteration, users can label a new set of positive vertices and negative vertices, which are included in the training set. The candidate subgraph for searching can also be updated based on user-provided labeled vertices. After the retraining and inferring, the GNN score for each vertex in the candidate subgraph is updated and produced as a refined k-sized community with the maximum GNN scores. If users are still unsatisfied with the answer, they can further label new vertices and retrain the model. To enhance the efficiency of the proposed ICS-GNN framework, the authors also develop a ranking-based loss function to simplify the labeling tasks by allowing users to compare and rank the importance of two vertices, instead of labeling them positive or negative. They also propose a greedy method based on a global view to identify the community to tackle the scenario where query vertices are at the boundary of the community.

4.2 A Flexible Interactive Community Search Framework

Although ICS-GNN [24] tackles the interactive community search using GNN models, it is challenging to extend it to other dense subgraphs based community models. Moreover, this method may be inefficient due to the need to compute GNN scores at each round of interaction, which requires rerunning learning models multiple times. To tackle these challenges, Sun et al. [66] propose a flexible framework to generalize existing community models and handle query-oriented interactive community search, called GICS. The techniques of GICS [66] can support different community models, e.g., attributed k-core-based community search [19], k-core-based weighted community search [64], and also the GNN-based community search [24].

A unified notation system to generalize existing community models. Sun et al. [66] first summarize the representative community search models and introduce an integrated notation system to define community models, denoted as $\mathbb{M}(\mathcal{G}, \mathcal{M}, \mathcal{O}, \mathcal{P})$, where $\mathcal{G}$ represents graph data, $\mathcal{M}$ defines various community metrics on the density, the graph size, or attributes, $\mathcal{O}$ contains a set of required operations, and $\mathcal{P}$ is a set of input parameters. Based on the notation system $\mathbb{M}$, it formulates the problem of interactive community search as follows. Given a graph $G = (V, E, A)$, a community model $\mathbb{M}_x(G, \mathcal{M}, \mathcal{O}, \mathcal{P}_0)$, query vertices $Q_0 \in \mathcal{P}_0$, the first iteration community search result H_0, and the number of maximum interactive round $I_{max} \in \mathbb{Z}^+$, the GICS problem aims to refine community H^{i-1} to H^i in the i-th iteration where $1 \leq i \leq I_{max}$, with a sequence of actions to add/remove vertices $Q^{\pm} \subseteq V$.

The GICS framework. The GICS framework for interactive community search [66] consists of three main steps: personalized adding/deleting recommendations, an auto-tuning parameter mechanism, and a fast refinement strategy. After the user chooses the specified community model $\mathbb{M}_x$, the system first implements the existing community search algorithm to get the initial community search result H_0. To enhance the efficiency of the interaction process, GICS system suggests a limited number of high-quality adding/deleting candidate vertices for users. Users can choose to add vertices to the community or remove vertices from the community by clicking the recommended vertices. This recommendation procedure first computes an integrated community relevance score for candidate vertices and then outputs a group of vertices with high community relevance and diverse location and attribute distribution. This paper also proposes an auto-tuning parameter mechanism to automatically refine the required parameters based on user adding/deletion actions, e.g., density metrics k for k-core and the size parameter k for k-sized community, which can significantly affect the community search answer. Finally, the GICS system develops a fast refinement strategy to help refine the choice of candidate subgraphs during the community search process, supporting partial refinement of communities.

5 Future Directions and Opportunities

In this section, we highlight a few future directions and promising opportunities for community search.

Leveraging community search for retrieval-augmented generation systems. Retrieval Augmented Generation (RAG) is a promising direction to enhance the performance of large language models (LLMs) by extracting relevant text chunks to a given query. Naturally, community search aims to find relevant communities containing query nodes. Thus, community search has potential applications for RAG systems with graph-structured documents. It is well-known that knowledge graphs are becoming increasingly important for organizing and querying vast amounts of structured information. The semantic information and complex relationships within knowledge graphs can help identify relevant subgraphs or communities for these knowledge-graph-based RAG systems. GraphRAG [16, 27] introduces a Leiden-based hierarchical community indexing solution, which leaves room for further exploration using advanced community search techniques.

LLM-enhanced community search over text-attributed graphs. Attributes are used to describe node properties, such as user interests, keywords, and check-in locations. Existing community search studies focus on attributed graphs with simple formats of category and numerical keywords. Nowadays, attribute information becomes richer and richer. Nodes and edges are associated with rich textual information, reflecting their diverse properties and complex semantics, denoted by text-attributed graphs [35]. An example of healthcare network consists of patients, doctors, diseases, and treatments, incorporating medical records and histories to improve patient care. Thus, it is important to further investigate community search over this complex and challenging textual graph, leveraging the strong analytics ability of LLMs on long texts [35]. Moreover, LLM-based systems allow a user-friendly interaction between users and systems. It suggests developing an interactive community search system to support queries issued by natural languages without a formal formulation, which allows multiple rounds of exploration in large-scale graphs. Such systems can further provide NLP-based interpretation to community search results, which advances the results of community search with an explainable insight.

Developing a unified dense subgraph model to handle various community search problems. Most existing community search models are usually proposed based on a particular subgraph pattern, e.g., k-core, k-truss, and quasi-clique. Thus, various community search algorithms and indexes are designed to support a particular subgraph maintenance, which have different search strategies and updating rules over dynamic graphs. Therefore, it is essentially important to develop a unified structure to generalize all existing and frequently used subgraphs, e.g., k-core and k-truss. Although GICS framework [66] gives a unified notation system for many dense subgraph-based community models, it still needs to implement specific updating algorithms to maintain these dense subgraphs. Moreover, the corresponding subgraph decomposition and index maintenance algorithms also need a unified solution over dynamic graphs, where node/edge insertions/deletions can occur anywhere in graphs.

References

[1] W. Ai, C. Xie, T. Meng, J. Du, and K. Li. A d-truss-equivalence based index for community search over large directed graphs. *TKDE*, 2024.

[2] A. Al-Baghdadi and X. Lian. Topic-based community search over spatial-social networks. *PVLDB*, 13(12):2104–2117, 2020.

[3] A. Arenas, L. Danon, A. Diaz-Guilera, P. M. Gleiser, and R. Guimera. Community analysis in social networks. *EPJB*, 38:373–380, 2004.

[4] A. Behrouz and F. Hashemi. Cs-mlgcn: Multiplex graph convolutional networks for community search in multiplex networks. In *CIKM*, pages 3828–3832, 2022.
[5] A. Behrouz, F. Hashemi, and L. V. Lakshmanan. Firmtruss community search in multilayer networks. *PVLDB*, 16(3):505–518, 2022.
[6] T. Cai, J. Li, N. A. H. Haldar, A. Mian, J. Yearwood, and T. Sellis. Anchored vertex exploration for community engagement in social networks. In *ICDE*, pages 409–420. IEEE, 2020.
[7] W. Cai, F. Zhu, Z. Liu, and M. Wu. Heterocs: A heterogeneous community search system with semantic explanation. In *SIGIR*, pages 3155–3159, 2023.
[8] G. Chen, F. Guo, Y. Wang, Y. Liu, P. Yu, H. Shen, and X. Cheng. Fcs-hgnn: Flexible multi-type community search in heterogeneous information networks. In *CIKM*, pages 207–217, 2024.
[9] J. Chen and J. Gao. Vics-gnn: A visual interactive system for community search via graph neural network. In *ICDE*, pages 3150–3153. IEEE, 2022.
[10] J. Chen, Y. Xia, and J. Gao. Communityaf: An example-based community search method via autoregressive flow. *PVLDB*, 16(10):2565–2577, 2023.
[11] J. Chen, Y. Xia, J. Gao, Z. Li, and H. Chen. Communitydf: A guided denoising diffusion approach for community search. In *ICDE*, pages 460–473. IEEE, 2025.
[12] L. Chen, C. Liu, K. Liao, J. Li, and R. Zhou. Contextual community search over large social networks. In *ICDE*, pages 88–99. IEEE, 2019.
[13] P. Chen and S. Redner. Community structure of the physical review citation network. *JOI*, 4(3):278–290, 2010.
[14] Y. Dai, M. Qiao, and R.-H. Li. On density-based local community search. *PACMMOD*, 2(2):1–25, 2024.
[15] Z. Dong, X. Huang, G. Yuan, H. Zhu, and H. Xiong. Butterfly-core community search over labeled graphs. *PVLDB*, 14(11):2006–2018, 2021.
[16] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, D. Metropolitansky, R. O. Ness, and J. Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.
[17] S. Fang, K. Zhao, G. Li, and J. X. Yu. Community search: a meta-learning approach. In *ICDE*, pages 2358–2371. IEEE, 2023.
[18] S. Fang, K. Zhao, Y. Rong, Z. Li, and J. X. Yu. Inductive attributed community search: to learn communities across graphs. *PVLDB*, 17(10):2576–2589, 2024.
[19] Y. Fang, C. Cheng, S. Luo, and J. Hu. Effective community search for large attributed graphs. *PVLDB*, 9(12), 2016.
[20] Y. Fang, X. Huang, L. Qin, Y. Zhang, W. Zhang, R. Cheng, and X. Lin. A survey of community search over big graphs. *VLDBJ*, 29:353–392, 2020.
[21] Y. Fang, K. Wang, X. Lin, and W. Zhang. Cohesive subgraph search over big heterogeneous information networks: Applications, challenges, and solutions. In *SIGMOD*, pages 2829–2838, 2021.
[22] Y. Fang, Y. Yang, W. Zhang, X. Lin, and X. Cao. Effective and efficient community search over large heterogeneous information networks. *PVLDB*, 13(6):854–867, 2020.
[23] E. Galimberti, F. Bonchi, F. Gullo, and T. Lanciano. Core decomposition in multilayer networks: Theory, algorithms, and applications. *TKDD*, 14(1):1–40, 2020.
[24] J. Gao, J. Chen, Z. Li, and J. Zhang. Ics-gnn: lightweight interactive community search via graph neural network. *PVLDB*, 14(6):1006–1018, 2021.
[25] F. Guo, Y. Yuan, G. Wang, X. Zhao, and H. Sun. Multi-attributed community search in road-social networks. In *ICDE*, pages 109–120. IEEE, 2021.
[26] F. Hashemi, A. Behrouz, and M. Rezaei Hajidehi. Cs-tgn: Community search via temporal graph neural networks. In *WWW*, pages 1196–1203, 2023.
[27] H. Huang, Y. Huang, J. Yang, Z. Pan, Y. Chen, K. Ma, H. Chen, and J. Cheng. Retrieval-augmented generation with hierarchical knowledge. *arXiv preprint arXiv:2503.10150*, 2025.
[28] X. Huang and L. V. Lakshmanan. Attribute-driven community search. *PVLDB*, 10(9), 2017.
[29] X. Huang, L. V. Lakshmanan, and J. Xu. Community search over big graphs: Models, algorithms, and opportunities. In *ICDE*, pages 1451–1454. IEEE, 2017.
[30] X. Huang, L. V. Lakshmanan, and J. Xu. *Community search over big graphs*. Morgan & Claypool Publishers,

2019.
[31] X. Jian, Y. Wang, and L. Chen. Effective and efficient relational community detection and search in large dynamic heterogeneous information networks. *PVLDB*, 13(10):1723–1736, 2020.
[32] Y. Jiang, Y. Fang, C. Ma, X. Cao, and C. Li. Effective community search over large star-schema heterogeneous information networks. *PVLDB*, 15(11):2307–2320, 2022.
[33] Y. Jiang, X. Huang, and H. Cheng. I/o efficient k-truss community search in massive graphs. *VLDBJ*, 30(5):713–738, 2021.
[34] Y. Jiang, Y. Rong, H. Cheng, X. Huang, K. Zhao, and J. Huang. Query driven-graph neural networks for community search: From non-attributed, attributed, to interactive attributed. *PVLDB*, 2022.
[35] B. Jin, G. Liu, C. Han, M. Jiang, H. Ji, and J. Han. Large language models on graphs: A comprehensive survey. *TKDE*, 2024.
[36] J. Kim, T. Guo, K. Feng, G. Cong, A. Khan, and F. M. Choudhury. Densely connected user community and location cluster search in location-based social networks. In *SIGMOD*, pages 2199–2209, 2020.
[37] J. Kim, S. Luo, G. Cong, and W. Yu. Dmcs: Density modularity based community search. In *SIGMOD*, pages 889–903, 2022.
[38] L. Li, S. Luo, Y. Zhao, C. Shan, Z. Wang, and L. Qin. Coclep: Contrastive learning-based semi-supervised community search. In *ICDE*, pages 2483–2495. IEEE, 2023.
[39] L. Li, Y. Zhao, S. Luo, G. Wang, and Z. Wang. Efficient community search in edge-attributed graphs. *TKDE*, 35(10):10790–10806, 2023.
[40] Y. Li, X. Chen, Y. Zhao, W. Shan, Z. Wang, G. Yang, and G. Wang. Self-training gnn-based community search in large attributed heterogeneous information networks. In *ICDE*, pages 2765–2778. IEEE, 2024.
[41] Z. Li, P. Zou, X. Chen, S. Hu, P. Zhang, Y. Zhang, B. He, Y. Li, and X. Tang. From community search to community understanding: A multimodal community query engine. In *CIKM*, pages 4749–4753, 2021.
[42] X. Liao, Q. Liu, X. Huang, and J. Xu. Truss-based community search over streaming directed graphs. *PVLDB*, 17(8):1816–1829, 2024.
[43] L. Lin, P. Yuan, R.-H. Li, C. Zhu, H. Qin, H. Jin, and T. Jia. Qtcs: Efficient query-centered temporal community search. *PVLDB*, 17(6):1187–1199, 2024.
[44] P. Lin, S. Yu, X. Zhou, P. Peng, K. Li, and X. Liao. Community search over large semantic-based attribute graphs. *WWW*, 25(2):927–948, 2022.
[45] B. Liu, F. Zhang, W. Zhang, X. Lin, and Y. Zhang. Efficient community search with size constraint. In *ICDE*, pages 97–108. IEEE, 2021.
[46] D. Liu and Z. Zou. Gcore: Exploring cross-layer cohesiveness in multi-layer graphs. *PVLDB*, 16(11):3201–3213, 2023.
[47] Q. Liu, M. Zhao, X. Huang, J. Xu, and Y. Gao. Truss-based community search over large directed graphs. In *SIGMOD*, pages 2183–2197, 2020.
[48] Q. Liu, Y. Zhu, M. Zhao, X. Huang, J. Xu, and Y. Gao. Vac: vertex-centric attributed community search. In *ICDE*, pages 937–948. IEEE, 2020.
[49] X. Liu, L. Chen, C. Liu, and R. Zhou. Searching society over large heterogeneous information networks. In *ICDE*, pages 1402–1414. IEEE, 2025.
[50] X. Liu and Z. Zou. Common-truss-based community search on multilayer graphs. In *ADMA*, pages 277–291. Springer, 2023.
[51] Y. Liu, F. Guo, B. Xu, P. Bao, H. Shen, and X. Cheng. Sach: Significant-attributed community search in heterogeneous information networks. In *ICDE*, pages 3283–3296. IEEE, 2024.
[52] Y. Lu, L. Zhao, and B. Wang. From virtual community members to c2c e-commerce buyers: Trust in virtual communities and its effect on consumers' purchase intention. *ECRA*, 9(4):346–360, 2010.
[53] Z. Lu, Y. Zhu, M. Zhong, and J. X. Yu. On time-optimal (k, p)-core community search in dynamic graphs. In *ICDE*, pages 1396–1407. IEEE, 2022.
[54] C. Luo, Q. Liu, Y. Gao, and J. Xu. Synergetic community search over large multilayer graphs. *PVLDB*, 3(v6):v9, 2024.
[55] D. Luo, Y. Bian, Y. Yan, X. Liu, J. Huan, and X. Zhang. Local community detection in multiple networks. In *SIGKDD*, pages 266–274, 2020.
[56] J. Luo, X. Cao, X. Xie, and Q. Qu. Best co-located community search in attributed networks. In *CIKM*,

pages 2453–2456, 2019.
[57] J. Luo, X. Cao, X. Xie, Q. Qu, Z. Xu, and C. S. Jensen. Efficient attribute-constrained co-located community search. In *ICDE*, pages 1201–1212. IEEE, 2020.
[58] W. Luo, X. Zhou, K. Li, Y. Gao, and K. Li. Efficient influential community search in large uncertain graphs. *TKDE*, 35(4):3779–3793, 2021.
[59] M. E. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical review E*, 64(1):016131, 2001.
[60] D. Rowangould, A. Karner, and J. London. Identifying environmental justice communities for transportation analysis. *Transportation Research Part A: Policy and Practice*, 88:151–162, 2016.
[61] Q. Sima, J. Yu, X. Wang, W. Zhang, Y. Zhang, and X. Lin. Deep overlapping community search via subspace embedding. *SIGMOD*, 3(1):1–26, 2025.
[62] Y. Song, L. Zhou, P. Yang, J. Wang, and L. Wang. Cs-dahin: Community search over dynamic attribute heterogeneous network. *TKDE*, 2024.
[63] G. Su, J. H. Morris, B. Demchak, and G. D. Bader. Biological network exploration with cytoscape 3. *Current protocols in bioinformatics*, 47(1):8–13, 2014.
[64] L. Sun, X. Huang, R.-H. Li, B. Choi, and J. Xu. Index-based intimate-core community search in large weighted graphs. *TKDE*, 34(9):4313–4327, 2020.
[65] L. Sun, X. Huang, R.-H. Li, and J. Xu. Fast algorithms for intimate-core group search in weighted graphs. In *WISE*, pages 728–744. Springer, 2019.
[66] L. Sun, X. Huang, J. Wang, and J. Xu. A flexible framework for query-oriented interactive community search. *PVLDB*, 2025.
[67] L. Sun, X. Huang, Z. Wu, and J. Xu. Efficient cross-layer community search in large multilayer graphs. In *ICDE*, pages 2959–2971. IEEE, 2024.
[68] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu. Heterogeneous information networks: the past, the present, and the future. *PVLDB*, 15(12), 2022.
[69] Y. Tang, J. Li, N. A. H. Haldar, Z. Guan, J. Xu, and C. Liu. Reliable community search in dynamic networks. *PVLDB*, 15(11):2826–2838, 2022.
[70] A. Tian, A. Zhou, Y. Wang, and L. Chen. Maximal d-truss search in dynamic directed graphs. *PVLDB*, 16(9):2199–2211, 2023.
[71] J. Wang, K. Wang, X. Lin, W. Zhang, and Y. Zhang. Efficient unsupervised community search with pre-trained graph transformer. *PVLDB*, 17(9):2227–2240, 2024.
[72] J. Wang, K. Wang, X. Lin, W. Zhang, and Y. Zhang. Neural attributed community search at billion scale. *PACMMOD*, 1(4):1–25, 2024.
[73] K. Wang, S. Wang, X. Cao, and L. Qin. Efficient radius-bounded community search in geo-social networks. *TKDE*, 34(9):4186–4200, 2020.
[74] K. Wang, W. Zhang, X. Lin, Y. Zhang, L. Qin, and Y. Zhang. Efficient and effective community search on large-scale bipartite graphs. In *ICDE*, pages 85–96. IEEE, 2021.
[75] R.-A. Wang, D. Liu, and Z. Zou. Focuscore decomposition of multilayer graphs. In *ICDE*, pages 2792–2804. IEEE, 2024.
[76] Y. Wang, X. Gou, X. Xu, Y. Geng, X. Ke, T. Wu, Z. Yu, R. Chen, and X. Wu. Scalable community search over large-scale graphs based on graph transformer. In *SIGIR*, pages 1680–1690, 2024.
[77] Y. Wang, Z. Peng, X. Ke, X. Xu, T. Wu, and Y. Gao. Cohesiveness-aware hierarchical compressed index for community search on attributed graphs. *SIGMOD*, 3(1):1–27, 2025.
[78] Y. Wang, S. Ye, X. Xu, Y. Geng, Z. Zhao, X. Ke, and T. Wu. Scalable community search with accuracy guarantee on attributed graphs. In *ICDE*, pages 2737–2750. IEEE, 2024.
[79] X. Xie, S. Liu, J. Zhang, S. Han, W. Wang, and W. Yang. Efficient community search based on relaxed k-truss index. In *SIGIR*, pages 1691–1700, 2024.
[80] R. Xu and D. C. Wunsch. Clustering algorithms in biomedical research: a review. *RBME*, 3:120–154, 2010.
[81] T. Xu, Z. Lu, and Y. Zhu. Efficient triangle-connected truss community search in dynamic graphs. *PVLDB*, 16(3):519–531, 2022.
[82] R. Yang, X. Xiao, Z. Wei, S. S. Bhowmick, J. Zhao, and R.-H. Li. Efficient estimation of heat kernel pagerank for local clustering. In *SIGMOD*, pages 1339–1356, 2019.

[83] K. Yao and L. Chang. Efficient size-bounded community search over large networks. *PVLDB*, 14(8):1441–1453, 2021.

[84] J. Ye, Y. Zhu, and L. Chen. Top-r keyword-based community search in attributed graphs. In *ICDE*, pages 1652–1664. IEEE, 2023.

[85] F. Zhang, H. Guo, D. Ouyang, S. Yang, X. Lin, and Z. Tian. Size-constrained community search on large networks: An effective and efficient solution. *TKDE*, 36(1):356–371, 2023.

[86] Y. Zhang, L. Lin, P. Yuan, and H. Jin. Significant engagement community search on temporal networks. In *DASFAA*, pages 250–258. Springer, 2022.

[87] Y. Zhang, K. Wang, W. Zhang, X. Lin, and Y. Zhang. Pareto-optimal community search on large bipartite graphs. In *CIKM*, pages 2647–2656, 2021.

[88] Z. Zhang, X. Huang, J. Xu, B. Choi, and Z. Shang. Keyword-centric community search. In *ICDE*, pages 422–433. IEEE, 2019.

[89] Y. Zhou, Y. Fang, W. Luo, and Y. Ye. Influential community search over large heterogeneous information networks. *PVLDB*, 16(8):2047–2060, 2023.

[90] Y. Zhu, J. He, J. Ye, L. Qin, X. Huang, and J. X. Yu. When structure meets keywords: Cohesive attributed community search. In *CIKM*, pages 1913–1922, 2020.

Detection, Measurement, and Mitigation of Echo Chambers in Social Networks: A Survey

Xianju Zhu[1], Reynold Cheng[1], Laks V.S. Lakshmanan[2]
Xiaodong Li[1], Chenhao Ma[3], Mohammad Matin Najafi[1]
[1]The University of Hong Kong, Hong Kong SAR, China
[2]The University of British Columbia, Vancouver, B.C., Canada
[3]The Chinese University of Hong Kong, Shenzhen, China
zxj0302@connect.hku.hk, ckcheng@cs.hku.hk, laks@cs.ubc.ca,
xdli@cs.hku.hk, machenhao@cuhk.edu.cn, mohammad@cs.hku.hk

Abstract

Echo chambers in social networks are environments where like-minded users cluster together, often reinforcing shared beliefs while limiting exposure to opposing viewpoints. This phenomenon has profound implications for information diffusion, political discourse, and societal polarization. In this survey, we systematically review the landscape of echo chambers in online social networks, focusing on three key aspects: detection, measurement, and mitigation. We categorize existing methods for identifying echo chambers, analyze diverse metrics that quantify their effects, and examine intervention strategies aimed at alleviating their negative consequences. By synthesizing findings from recent literature and highlighting open challenges, our survey aims to provide a comprehensive reference for researchers and practitioners seeking to understand and address the echo chamber effect in contemporary digital platforms.

1 Introduction and Background

Online social networks, such as Twitter and Facebook, have fundamentally transformed the way individuals consume information, engage in discourse, and form social ties. While these platforms enable unprecedented connectivity and rapid information diffusion, they also give rise to complex social phenomena that challenge the health and diversity of public discourse. Among these, the *echo chamber* effect has emerged as a central concern for researchers across computer science, social science, and digital media studies.

Definition (Echo Chamber): An echo chamber is a phenomenon prevalent in online social networks, characterized by like-minded users predominantly interacting with each other. Within these echo chambers, users express and reinforce their beliefs on specific issues, thereby amplifying their shared viewpoints [14].

The terms *echo chamber* and *filter bubble* [82] carry distinct nuances but are frequently treated as synonyms in the literature [41]. While both limit exposure to diverse perspectives, their underlying mechanisms and effects differ. Echo chambers arise when users deliberately discredit and exclude dissenting views, fostering distrust of outsiders and resistance to counterevidence. In contrast, filter bubbles result from passive algorithmic curation that limits exposure to diverse perspectives, though users may remain open to opposing information when encountered. Recognizing this distinction is critical for designing interventions, as mitigating echo chambers requires addressing active exclusion and psychological biases, whereas filter bubbles may be alleviated through algorithmic transparency and diversification [21].

The echo chamber effect has been linked to a range of negative consequences for both individuals and society at large. Five common attributes [21] associated with echo chambers are: the diffusion of misinformation [83, 84], the spread of conspiracy theories [85], the formation of social trends [86], increased political polarization [87], and the emotional contagion of users [88]. These dynamics can contribute to the amplification of extreme views, the hardening of group boundaries, and the erosion of trust in public institutions.

Despite growing attention, systematic understanding and management of echo chambers remain challenging. First, echo chambers are multifaceted: they involve both the structural arrangement of social ties (who interacts with whom) and the semantic alignment of opinions (what people believe and express). Second, the observable effects of echo chambers—such as polarization, misinformation, and network segregation—are not always easily separable from other social processes. As a result, existing research has produced a wide array of methods for detecting, quantifying, and mitigating echo chambers, yet a unified framework remains elusive.

While there have been a few recent surveys on the literature on echo chambers in specific contexts [14, 21, 41], they exclude one of the key aspects: detection, measurement, or mitigation, or omit many recent and relevant studies due to their publication date, or lack a coherent framework for classifying existing approaches. Motivated by these gaps, our survey provides a comprehensive and up-to-date synthesis of the literature on echo chambers, with a particular focus on social networks. We systematically compare and extend prior surveys by incorporating recent work, offering a clear and structured taxonomy of methods, and highlighting overlooked yet important perspectives. Our goal is to offer a one-stop reference for both scholars and practitioners seeking to understand the full landscape of echo chamber research. Throughout, we draw attention to the conceptual nuances that distinguish echo chambers from related ideas, and we identify open challenges and promising directions for future inquiry.

Scope of this survey: This paper provides a comprehensive review of echo chamber research in online social networks, focusing on three interrelated aspects: *detection*, *measurement*, and *mitigation*. We systematically categorize methods for identifying echo chambers, analyze the diverse metrics used to quantify their effects, and survey intervention strategies designed to alleviate their negative impact. Our discussion emphasizes the interplay between network topology and semantic content, and highlights both methodological advances and open challenges. While echo chamber-like phenomena have also been observed in domains such as e-commerce recommender systems [15], we restrict our focus to social media and communication networks, where the social, political, and informational stakes are particularly pronounced.

For further reading on the definitions, attributes, mechanisms, and modeling of echo chambers and related risks, readers are encouraged to consult recent surveys such as [14, 21]. The remainder of this paper is organized as follows: Section 2 surveys methods for echo chamber detection; Section 3 reviews metrics for measuring echo chamber effects; Section 4 discusses mitigation strategies; Section 5 outlines open challenges and opportunities; and Section 6 concludes with a summary and future outlook.

2 Echo Chamber Detection

In this section, we summarize existing methods for echo chamber detection, which can be broadly classified based on their use of semantic and/or topological information: (1) *Topology-Based Detection*, which treats echo chambers as communities characterized by dense internal interactions and sparse external connections. Methods in this category commonly apply community detection algorithms—such as the Louvain algorithm—emphasizing metrics related to inter-community edges; (2) *Content-Based*

Detection, which focuses on analyzing the semantic content generated by users (e.g., tweets or posts) to identify differences, similarities, or polarization within the network; and (3) *Hybrid Detection*, which integrates both topological and semantic information to detect communities with high internal alignment, pronounced polarization between groups, and limited cross-community interaction. Table 1 summarizes related works, detailing their classification, detection methods, and a brief description of their detection logic or graph construction process.

Table 1: Comparison of Echo Chamber Detection Methods

Category	**Detection Method**	**Reference**
Topology-Based	Community Detection (Fast Greedy [78], Infomap [80], Louvain [79], WalkTrap [81], etc)	[22–25, 31, 51, 53–56, 58]
	Graph Partition (METIS [59])	[22, 60]
Content-Based	Post-comment pair classification	[4]
	Engagement in opposing views	[1]
Hybrid	METIS [59] on semantic-enriched graph	[6]
	Maximizing the likelihood of observed cascades	[8]
	Polarized communities detection in signed networks	[97, 98]

2.1 Topology-Based Detection

Based on the definition and intuition of an echo chamber, users interact much more frequently within their own communities than with those outside. Therefore, topology-based detection methods focus on network structure and construct social networks using various user interactions, such as retweets, replies, and follows. They then identify communities by finding groups with dense internal connections and sparse interactions with users outside the group.

Cossard et al.[22] detect echo chambers in the Italian vaccination debate on Twitter by leveraging the structural properties of interaction networks, particularly the retweet network. They construct a weighted, directed retweet network where nodes represent users and edges represent repeated retweet interactions, filtering out low-weight edges to reduce noise. Using the graph partition algorithm, METIS [59], they repeatedly bi-partition the giant connected component of this network to assign each user a "leaning score" based on retweet behavior, effectively mapping users to one of two polarized communities: vaccine advocates and vaccine skeptics. Manual annotation of a random user sample confirmed high classification accuracy. Similarly, Amendola et al. [60] model the social network as a graph, with nodes representing users and edges denoting topological relationships such as mentions, retweets, or follower/followee links. To detect communities within this network, they apply the METIS algorithm, which partitions the graph into balanced communities based on user interaction patterns.

Community detection algorithms are widely used to identify echo chambers in social networks. For instance, the Fast Greedy algorithm [78] is utilized in [22, 24], while the Louvain method [79] has been applied in several studies [22, 23, 31, 51, 53–56, 58]. Similarly, Infomap [80] has been adopted in [22, 25]. These methods are typically applied to networks constructed from retweet, quote, mention, follow, comment, or reply interactions collected from Twitter or other platforms, as well as user-item interactions in recommender systems. The resulting communities are then analyzed as potential echo chambers.

2.2 Content-Based Detection

Since content generated or endorsed by users often reflects their opinions, analyzing such content through stance detection or emotion classification enables the quantification of underlying semantic signals. Applying these techniques allows for the inference of a user's stance based on their interactions, facilitating the identification of users participating in echo chambers.

Calderon et al. [4] propose a novel content-based approach for detecting echo chambers on social media, specifically focusing on Facebook fan pages. Instead of relying on complete network structures, their method analyzes the content of posts and their associated comments to extract linguistic features indicative of echo chamber behavior. They design two main types of features: Target Stance, which captures whether a comment agrees or disagrees with the stance of the original post, and Emotion Intensity, which measures the type and strength of emotions expressed. These features are extracted using graph-based linguistic pattern mining and are then fed into a neural network model (ECHO model) that uses attention mechanisms to classify post-comment pairs as echoing or not. By aggregating these classifications at the fan page level, they compute an Echo Chamber Index that quantifies the degree of echo chamber behavior present. Their experiments demonstrate that this approach outperforms traditional models in both English and Mandarin datasets, highlighting the effectiveness of content-based echo chamber detection.

Del Vicario et al. [1] detect echo chambers on Facebook by systematically analyzing user interactions with public Facebook pages categorized as either science or conspiracy. The authors first classify pages into these two categories based on their content and self-description. To identify echo chambers, they examine each user's commenting behavior over time and assign users to a community if at least 95% of their total comments are made on posts within either the science or within the conspiracy category. This high threshold ensures that only users with a strong and consistent preference for one type of content are included in the respective echo chamber. By applying this method across their dataset, the authors empirically isolate two distinct, polarized communities—one centered on scientific information and the other on conspiracy theories. This quantitative approach allows for the objective detection and longitudinal tracking of echo chambers, rooted in users' selective engagement and reinforcement of like-minded content.

2.3 Hybrid Detection

Topology-based detection methods utilize structural information to identify segregation between communities, but cannot ensure that users within the same community are highly ideologically aligned based solely on topology. Content-based detection methods leverage information from user-generated or endorsed content to infer ideological leanings, but do not determine whether these users are structurally divided into opposing groups. To more accurately detect echo chambers, some approaches combining both topological and semantic information have been proposed.

In [8], Minici et al. introduce a probabilistic generative model that integrates content-based and semantic-based approaches to identify latent communities with echo chamber characteristics within social networks. By modeling communities based on their polarization and opinion polarity, the method employs a scalable adaptation of the Generalized Expectation Maximization algorithm to optimize the joint likelihood of social connections and information cascades. This dual approach captures the propagation of ideologically aligned content while distinguishing echo chambers from other communities, as validated through experiments on synthetic datasets and real-world cases, including the Brexit referendum and COVID-19 vaccine discussions.

In their study on echo chamber detection during the early COVID-19 pandemic, Villa et al. [6] propose a method that integrates both topological and semantic aspects of online interactions. The process begins by modeling a Twitter conversation graph, where nodes represent users and edges are established through mention relationships, with edge weights reflecting the frequency of mentions. The authors enrich the graph with semantic information by adjusting edge weights according to users' sentiment similarity (using VADER sentiment analysis) and topic similarity (using LDA topic modeling), producing four distinct graph representations: topology-based, sentiment-based, topic-based, and a hybrid of sentiment and topic. To detect echo chambers, they apply the METIS community detection algorithm to partition each graph into two groups, corresponding to potential polarized communities. The presence of echo chambers is then assessed by quantifying controversy (using random walk-based and boundary connectivity measures) and community homogeneity (analyzing sentiment and topical coherence within each group). Their results show that incorporating content-based features, especially sentiment, enhances the detection of polarized, homogeneous groups—providing a robust framework for echo chamber identification in social media contexts.

Although graph partitioning techniques have been applied for community detection, they often produce overly balanced divisions that may fail to reflect the intricate structures found in real-world datasets. To address this limitation, [97] presents a novel generalized balanced subgraph model that allows for some degree of imbalance. They propose a region-based heuristic algorithm that effectively balances computational efficiency with solution quality. This approach is built on signed networks, which capture the consistency of opinions among users and, as a result, incorporate semantic information. Similarly, building on signed networks, [98] detect polarized communities characterized by mostly intra-community positive edges and inter-community negative edges, thus enabling fine-grained analysis of controversy in social networks.

3 Echo Chamber Effect Measurement

In this section, we collate and analyze the various metrics used to measure echo chamber effects. Typically, there is no single metric that directly quantifies whether a reported echo chamber is inherently "good" or "bad". Instead, echo chambers are typically assessed through their observable consequences, such as polarization, network segregation, and intra-group homogeneity. By evaluating these phenomena, researchers can gauge the extent to which echo chamber effects are present. In the literature, similar aspects of echo chamber effects are often quantified using different terms. For example, the degree to which communities within a social network are separated—and users predominantly interact with others within their own community—may be referred to as "segregation", "separateness", "controversy", or "polarization". Although these terms may emphasize slightly different nuances, they generally capture the same underlying intuition and tend to be used interchangeably.

Existing approaches to measuring echo chamber effects can be broadly categorized into three groups: (1) *Network Segregation-Based Metrics*, which analyze the structural properties of social networks to capture the characteristic that intra-community interactions are much more frequent than inter-community ones; (2) *Opinion Homogeneity-Based Metrics*, which assess the distribution of user opinions to determine the degree of opinion alignment within echo chambers, or to evaluate overall opinion diversity across the network; and (3) *Hybrid Metrics*, which integrate both network topology and user opinions to assess intra-community alignment alongside cross-community segregation. Additionally, some studies have explored echo chamber effects in non-social network contexts, such as e-commerce platforms, and have proposed corresponding metrics. In the following subsections, we classify and introduce these approaches in more detail.

3.1 Network Segregation-Based

Intuitively, a higher number of intra-community edges and fewer inter-community edges indicate a greater degree of segregation between groups of users, which structurally aligns with the concept of an echo chamber. Consequently, many approaches focus on the structural characteristics of the network by dividing it into communities, typically using community detection or graph partition methods that aim to minimize the number of edges between different communities. Based on the resulting community structure, various metrics are then employed to quantify the degree of segregation in the network.

Analyzing the challenges in accurately determining users' political preferences and the expected properties of informational separateness, Chkhartishvili et al. [2] introduced a metric to measure the echo chamber effect, called the **Binary Separation Index (BSI)**. This index requires only the identification of the set of accounts disseminating information within the network and their corresponding political positions. The BSI is calculated as follows:

$$\mathrm{BSI} = 4 \times \alpha \times \beta$$

where α and β represent the proportions of users in each of the opposing sources I_1 and I_2, respectively, who exclusively connect with information from their own source and not the opposing one. Note that $\alpha, \beta \in [0, 1]$. The BSI attains its maximum value, 4, when the network is perfectly divided, with users exclusively engaging with only one source or the other, indicating strong informational separation and a pronounced echo chamber effect. Conversely, a lower BSI suggests greater cross-exposure between groups, reflecting a less segregated information environment.

With a focus on nodes, BSI utilizes the ratio of border users in each community. In contrast, edge-focused methods examine the connections between communities from a different perspective. Luo et al. [3] introduce a quantitative measure of segregation in social networks by examining the formation of edges both within and between communities. Specifically, they define **segregation** for a directed graph $G = (V, E)$, where the set of vertices $V = R \cup B$ is partitioned into red (R) and blue (B) communities. The segregation metric s is given by

$$s = 1 - \frac{|E_d|}{2|R||B|}$$

where E_d denotes the set of edges connecting members of different communities, and $2|R||B|$ represents the maximum possible number of inter-community edges. A segregation value of $s = 1$ corresponds to complete segregation, while lower values indicate greater integration between communities.

A related metric, the **E-I Index** [68], is used by Ertan et al. [32] to measure political polarization in network structures. After defining political blocs (e.g., based on formal electoral alliances), each respondent's network receives an E-I Index score, calculated as:

$$\text{E-I Index} = \frac{ET - IT}{ET + IT}$$

where ET is the number of external ties (cooperation across blocs) and IT is the number of internal ties (cooperation within blocs). The index ranges from -1 (homophily, mostly within-bloc ties) to $+1$ (heterophily, mostly cross-bloc ties). The key distinction between the E-I Index and the BSI lies in their denominators: the BSI normalizes by the maximum possible number of edges, whereas the E-I Index uses the actual number of observed edges.

As many works use community detection methods to find echo chambers, **modularity** proposed in [17] is widely used in works such as [6, 33, 34] to evaluate the quality of a discovered community structure. Modularity measures how well a network is partitioned into communities by comparing the density of links within communities to that expected in a random network. A higher modularity score indicates a stronger community structure. Intuitively, modularity assesses whether more edges fall within communities than would be expected by chance. It does this by comparing the actual fraction of edges within each community to the expected fraction if the edges were distributed randomly, while preserving the network's degree distribution. If the observed number of intra-community connections is significantly higher than expected at random, the modularity score will be high. A value of modularity close to its theoretical maximum of 1 indicates a strong community structure, while a value near 0 suggests that the observed partition is no better than random.

However, Guerra et al. [7] note that "non-polarized networks still show positive modularity", limiting modularity's effectiveness for detecting polarization. To address this, they introduce **Boundary Connectivity (BC)**, P, a metric that measures antagonism by evaluating boundary nodes' connectivity preferences between two communities:

$$P = \frac{1}{|B|} \sum_{v \in B} \left[\frac{d_i(v)}{d_b(v) + d_i(v)} - 0.5 \right]$$

where $d_i(v)$ is the number of edges from node v to internal nodes, $d_b(v)$ is the number of edges to opposing boundary nodes, and B is the set of boundary nodes. A positive P indicates polarization, while negative or near-zero values suggest its absence. BC is applied in many studies [6, 16, 30, 35, 36, 60] to quantify controversy or polarization.

Guyot et al. [40] introduce **ERIS**, a metric akin to BC but incorporating edge weights, to assess polarization by examining community boundaries. For each pair of communities C_i and C_j, ERIS identifies the boundary area ($B_{i,j}$), defined as the set of users in C_i who interact both with their own community and with C_j. The first metric, Community Antagonism ($A_{i,j}$), quantifies the directed opposition from C_i to C_j based on the interaction patterns of boundary users. For an individual boundary user v, antagonism $A^v_{i,j}$ is computed as:

$$A^v_{i,j} = \frac{\sum_{e \in IE^v_{i,j}} w(e)}{\sum_{e \in IE^v_{i,j}} w(e) + \sum_{e \in EE^v_{i,j}} w(e)} - 0.5$$

where $IE^v_{i,j}$ and $EE^v_{i,j}$ denote the sets of v's internal and external edges, respectively, and edge weight $w(e)$ is the number of times u quoted v for edge $e = (u, v)$. The overall community antagonism $A_{i,j}$ is obtained by averaging $A^v_{i,j}$ over all users in $B_{i,j}$. The second metric, Boundary Porosity ($P_{i,j}$), measures the permeability of the boundary, defined as the proportion of boundary users who interact more with the external community:

$$P_{i,j} = \frac{\left| \{ v \in B_{i,j} \mid A^v_{i,j} < 0 \} \right|}{|B_{i,j}|} \times 100.$$

Together, these metrics capture both the degree of antagonism between communities and their susceptibility to external influence.

While many existing metrics primarily focus on edge counts, Garimella et al. [16] introduced a random walk-based metric, the **Random Walk Controversy (RWC)** score, to quantify the structural separation in partitioned conversation graphs. The underlying intuition is that, in highly controversial topics, opposing sides form distinct communities with minimal interaction between them. As a result, a random walk initiated within one community is more likely to remain there than traverse to the other. The RWC score formalizes this notion. Given a graph partitioned into two disjoint user sets, X and Y, the metric is defined as the difference between the probability that two random walks remain within their respective starting partitions and the probability that both cross over to the opposing partition.

Specifically, let P_{AB} denote the conditional probability that a random walk starting from a random node in partition A terminates in partition B. The RWC score is then calculated as:

$$\text{RWC} = P_{XX}P_{YY} - P_{XY}P_{YX}$$

Here, the term $P_{XX}P_{YY}$ represents the likelihood that both communities are internally cohesive and isolated, while $P_{XY}P_{YX}$ reflects the strength of cross-partition interaction. A score approaching 1 indicates strong separation between partitions and thus a high level of controversy. Conversely, a score near 0 suggests that the probability of crossing partitions is similar to that of remaining within them, implying a lack of clear division and therefore a non-controversial topic. For practical computation, the authors propose an efficient variant utilizing Random Walk with Restart (RWR) to estimate these probabilities in large, directed graphs. In [6], the authors employ the RWC score to measure controversy and further introduce variants such as Authoritative Random Walk Controversy (AWRC) and Displacement Random Walk Controversy (DRWC). The RWC score has also been applied in [22, 51, 52].

The aforementioned metrics evaluate segregation at the level of the entire network. To determine the degree of homogeneity within individual partitions, "**coverage**" is employed in [6], as originally introduced in [18]. Coverage quantifies the proportion of a graph's total edges that are intra-community, i.e., those connecting pairs of vertices within the same community. In the idealized case where communities are completely separated—forming disjoint subgraphs with no inter-community edges—the coverage reaches a maximum value of 1. Consequently, coverage offers a direct measure of partition cohesiveness by indicating how well the detected communities encompass the network's edge structure. Additionally, in [8, 13], "**conductance**" is utilized, which measures the fraction of the total edge volume that leaves the community. Intuitively, conductance reflects how well a community is separated from the rest of the network: it compares the number of edges that connect the community to the outside to the total number of edges associated with the community, both internal and external. To ensure that a community is sufficiently insular to be considered an echo chamber, conductance is constrained to be at most 0.5 in [8, 13], meaning that more than half of the total edges must remain within the community boundaries.

3.2 Opinion Homogeneity-Based

Echo chambers cluster like-minded users who predominantly interact with others sharing similar views. A straightforward approach to quantifying echo chambers is to measure how closely aligned users' leanings are within the same echo chamber, and how distinct the average leanings are between different echo chambers. Additionally, by examining the overall distribution of user leanings, one can determine whether there is significant polarization—for example, whether the distribution exhibits a "U"-shaped pattern indicative of extreme polarization, or is more uniform. From this perspective, many studies have investigated polarization and the echo chamber effect through the lens of user opinions.

A common approach for quantifying polarization is to focus on opinion homogeneity within a network. Several studies [26, 27, 64, 67] use the variance of expressed opinions as a direct indicator of **polarization**, which measures how much individual opinions deviate from the average opinion in the group. A higher variance indicates greater diversity or polarization in user opinions, while a lower variance suggests more consensus.

Chen et al. [43] define **controversy** as the summation of squared user opinions that captures the overall intensity of expressed opinions in a group by aggregating how strongly each opinion deviates from neutrality. The same formula is adopted in Musco et al. [50] under the term **polarization**, and Matakos et al. [63] propose a "**polarization index**" computed as the average squared opinion per user, i.e., dividing the sum above by the total number of users.

Moving beyond aggregate measures, Interian et al. [29] introduce a probabilistic framework to quantify network polarization by evaluating the statistical significance of node-level homophily. Instead of relying solely on raw homophily counts, their method calculates a p-value for each node. This p-value represents the likelihood that a node's observed number of same-group connections (or more) could occur by chance in a "balanced" network, where connection probabilities are proportional to group sizes. The overall polarization for a network or group is then characterized by the distribution of these p-values: a distribution skewed towards zero indicates strong, statistically significant polarization. Empirical cumulative distribution functions of these p-values are used for robust comparison of polarization across groups or networks.

Cota et al. [9] take a different approach, quantifying the echo chamber effect by assigning each user a political position based on the average leaning of their tweets, with tweet classification performed manually as pro-impeachment, neutral, or anti-impeachment. Echo chambers are then measured by analyzing the correlation between a user's political position and the average leanings of both their nearest neighbors and the tweets they receive. Strong correlations indicate that users primarily interact with others holding similar political views, confirming the existence of echo chambers. This **joint distribution** approach is also utilized in [53, 54, 58, 61, 62]. Notably, this approach is primarily visualization-based and does not provide a single summary statistic. For more quantitative assessment, the correlation coefficient between the joint distribution and the $y = x$ line can be calculated to measure the degree of homophily between users' opinions and the average opinions of their neighbors.

Further exploring the dynamics of opinion formation, Sikder et al. [19] model agents in a social network tasked with determining the truth of a binary statement ($X = +1$ or $X = -1$). Opinion formation begins at $t = 0$ with each agent receiving a private signal. At each discrete time step, agents synchronously share all accumulated signals with their immediate neighbors, causing each agent's information set to grow recursively. Polarization is quantified in several steps: each agent i computes their "signal mix" $x_i(t)$ at time t, representing the proportion of positive signals ($N_i^+(t)$) among all signals received:

$$x_i(t) = \frac{N_i^+(t)}{N_i^+(t) + N_i^-(t)}.$$

Based on $x_i(t)$, the agent adopts a public orientation $y_i(t)$, set to $+1$ if $x_i(t) > 0.5$ and -1 otherwise. For a group C, **polarization** $z_C(t)$ is defined as the size of the minority camp:

$$z_C(t) = \min(y_C(t), 1 - y_C(t)),$$

where $y_C(t)$ is the fraction of positive orientations in C. Polarization reaches zero under full consensus and is maximized at 0.5 when the group is evenly split.

At the individual level, Al Atiqi et al. [10] introduce the **Individual Echo Chamber Coefficient (ECC)**, which measures the diversity of opinions among a user's neighbors. Instead of a formula, the ECC for a user quantifies how varied the opinions are within that user's network; if all neighbors share similar views, the ECC will be low, whereas a wide range of opinions among neighbors leads to a higher ECC. Values near zero indicate strong echo chambers. For network-wide assessment, they define the **Global Echo Chamber (GEC)** Indicator, which reflects the overall tendency of connected individuals to share the same or opposing opinions throughout the network. A lower GEC value suggests weaker clustering of like-minded individuals.

Botte et al. [11] propose two complementary metrics for echo chambers. The global **echo chamber size** measures the relative growth of fully homogeneous neighborhoods: instead of focusing on the exact calculation, this metric essentially compares how the number of individuals who are exclusively surrounded by others with the same opinion changes from the beginning to the end of the observation. They also assess **local polarization** by analyzing the distribution, the fraction of an individual's neighbors sharing their opinion. A bimodal distribution of these values, with peaks near 0 and 1, signals strong polarization into segregated groups.

Similarly, Madsen et al. [12] introduce the "**belief purity**" metric to quantify echo chamber formation. The intuition is that this metric reflects how similar the beliefs are among connected agents: as the average difference in belief between connected nodes decreases, the belief purity approaches its maximum value of 1. This indicates that, in highly purified networks, connected agents tend to share nearly identical beliefs.

The concept of "**purity**" is also employed in [8, 13]. In [8], purity is the ratio of users with the same ideological alignment, measured as the average leaning of reshared tweets. In [13], purity is defined as the product of the frequencies of the most common labels among nodes in a community. Combined with "conductance" (as discussed above), low conductance and high purity are identified as hallmarks of echo chambers.

To capture the distributional aspects of polarization, Lelkes [30] employs specific metrics. For ideological divergence between partisans, the **Overlap Coefficient (OC)** is used, which intuitively measures the extent to which the ideological distributions of two groups (such as Republicans and Democrats) coincide. An OC value of 1 indicates perfect overlap, while 0 indicates complete separation. For detecting polarization in the general public, the Bimodality Coefficient is utilized to formally test for the emergence of two peaks in the ideological distribution, thereby providing an alternative to mean comparisons.

To quantify "echofication" in multiparty networks, Markgraf and Schoch [31] present a two-pronged framework. First, they identify "Social Boundaries" (the chamber) via community detection and measure the insularity of these boundaries. Second, they assess "User Similarity" (the echo) by modeling user ideology as a multi-dimensional vector, based on the politicians each user follows. Cosine similarity between users provides a homophily score, serving as a direct proxy for the level of echo within each chamber.

3.3 Hybrid

Network topology-based methods can quantify the degree of segregation between different communities; however, they are unable to assess individuals' ideological alignment within each community or the ideological differences between communities, as analyzing interactions alone may not accurately reflect users' true opinions without analyzing their posts or other outputs. Although it is challenging to obtain users' genuine thoughts, opinion-based methods can evaluate whether the distribution of users' leanings as expressed in their posts aligns with the expected characteristics of echo chambers. However, these methods often overlook the assessment of users' interaction patterns (i.e., frequent intra-community interactions and infrequent inter-community interactions), particularly when the metrics quantify polarization only at the network level. For example, a variance score or leaning distribution calculated for the entire network does not contain information about the position of each node. Compared with two ideologically opposing groups that have high intra-group homogeneity, a network composed solely of users with extreme and polarized viewpoints, where these users frequently interact with both like-minded and opposing individuals, can still exhibit a high variance score or highly polarized leaning distribution. Therefore, in this section, we discuss measurement approaches that combine both the network's topological structure and users' opinions simultaneously. Such hybrid approaches are promising for providing a more comprehensive and precise quantification of the echo chamber effect.

To quantify the structural dimension of polarization beyond just opinion distribution, [20] proposed a metric that assesses the alignment between opinion similarity and the strength of connections within the network. Building upon the basic concept of edge homogeneity (the product of two connected users' opinions), they introduce the **average weighted mean edge homogeneity** ($\overline{\mathrm{hom}_w}$). This measure incorporates the evolving connection strength between users, providing a more nuanced view of the network's structure. Instead of simply averaging opinion similarity across all connections, this measure gives more weight to stronger connections, effectively capturing how strongly people with similar or different opinions are connected in the network. A high value of $\overline{\mathrm{hom}}_w$ signifies a strongly polarized and segregated structure, characteristic of an echo chamber, where users with similar opinions are linked by strong connections and those with different views are linked by weak ones.

Several metrics explicitly integrate network topology with the distribution of opinions to quantify polarization and fragmentation. Shekatkar [48] proposes "**correlated polarization**" (ϕ), which is defined as the product of two components: (1) the balance or bimodality of opinions, $R = 1 - 2|n^{-} - 0.5|$, where n^{-} is the fraction of nodes holding one of the two opinions (maximized when opinions are evenly split), and (2) the **assortativity coefficient** r [49] with respect to node states, measuring the tendency of like-minded nodes to connect. Thus, $\phi = R \times r$ achieves high values only in networks exhibiting both a sharp division of opinions and pronounced structural segregation, capturing social fragmentation more effectively than methods based solely on opinion counts.

Morales et al. [52] introduce a **polarization index**, μ, to measure the degree to which an opinion distribution is divided into two distinct and opposing groups. Conceptually inspired by the electric dipole moment in physics, their metric defines perfect polarization as a state where a population is split into two factions of equal size holding maximally distant opinions. The index is calculated as the product of two key components: group size balance and ideological distance. The first component, $(1 - \Delta A)$, captures the population balance, where ΔA is the absolute difference between the relative sizes of the two opposing groups. This term equals 1 when the groups are perfectly balanced in size and 0 when one group comprises the entire population. The second component, d, represents the ideological distance, calculated as the normalized distance between the "gravity centers" (i.e., the mean opinions) of each group. This distance ranges from 0, for no ideological separation, to 1, for maximum opposition. The final polarization index, $\mu = (1 - \Delta A)d$, thus reaches its maximum value of 1 only when both conditions are met: the two groups are of equal size ($\Delta A = 0$) and their average opinions are maximally far apart ($d = 1$).

To address the limitations of approaches focused solely on network structure, Emamgholizadeh et al. [37] propose the **Biased Random Walk (BRW)** framework for quantifying controversy in attributed networks. Unlike methods that only consider topology, BRW integrates both structural and node-level attributes. The core of the framework is a random walk with a finite lifetime, simulated through a novel energy mechanism. The initial energy for a walk starting at a given node is determined by considering how close that node is to the center of its own community as well as how close it is to the center of the opposing community, reflecting both local and cross-community influence. As the walk traverses the network, it loses energy at each step, with the amount of energy lost depending on how far the current node is from the center of the opposing community. This means that the deeper a walk ventures into the opposing community, the faster it loses energy, making it increasingly difficult to reach the core of the opposition. Controversy is then measured by the "penetration depth"—the maximum level a walk can reach within a contradicting community before its energy is depleted. This approach models an idea's ability to be heard by an opposing audience, offering a more nuanced controversy score that captures the combined effects of network structure and user characteristics.

Alatawi et al. [51] propose the **Echo Chamber Score (ECS)**, a metric that quantifies the echo chamber effect by evaluating the geometric properties of user communities within a learned embedding space. Their approach first uses a self-supervised Graph Auto-Encoder, EchoGAE, to embed users into a low-dimensional space where distance corresponds to ideological similarity, leveraging both user interaction patterns and post content. The core idea behind ECS is to measure two key properties: cohesion, representing how similar users are within their own community, and separation, representing how distinct they are from users in other communities. For each user u in a community ω, cohesion is calculated as their average distance to all other users in ω. Separation is their average distance to users in the nearest neighboring community. These values are then aggregated into a score for the individual community by assessing, for each member, the relative difference between their similarity to their own community and their dissimilarity to the closest external community. This aggregation is inspired by the silhouette score and normalized so that it ranges from 0 to 1, where higher values indicate stronger echo chamber effects. The overall ECS for the entire graph is then the average of these scores across all detected communities. A score approaching 1 signifies a strong echo chamber effect, with tightly clustered and well-separated communities, while a score near 0 indicates more ideologically integrated groups. A key advantage of this method is its unsupervised nature, as it requires neither pre-defined user labels nor a fixed number of communities.

Similarly, Amendola et al. [60] move beyond structural analysis to incorporate the semantic content of user interactions. Their approach is founded on the principle that true opinion alignment requires agreement on specific facets of a topic, not just a general sentiment. The method employs **Aspect-Based Sentiment Analysis (ABSA)** to capture nuanced opinions and **Group Decision-Making (GDM)** principles to measure consensus. The process begins by using ABSA to generate a sentiment vector for each user on every specific aspect of a topic. The core of the metric is built upon a multi-level aggregation of these opinions. First, the agreement between any two users on a specific aspect is calculated as a pairwise similarity, typically using cosine similarity. These pairwise similarities for a single aspect are then aggregated across all users to determine the **Aspect Consensus**, which intuitively reflects the average level of agreement within the community on that aspect, with greater weight given to stronger agreements. Finally, the consensus scores from all aspects are aggregated to produce a single **Community Consensus** value, representing the overall agreement within the community by summarizing the aspect-level agreements in a way that emphasizes more consistent patterns of alignment. The diagnostic power of this metric comes from comparing the within-community consensus (calculated among members of the same group) with the in-between-communities consensus (calculated between members of different groups). A high within-community consensus coupled with a low in-between-communities consensus provides a strong, content-driven signal of an echo chamber.

Hohmann et al. [65] introduce a holistic measure for ideological polarization that integrates three distinct factors into a single score: the extremity of opinions, the structural clustering of individuals into communities (homophily), and the mesoscale organization of these communities along an ideological spectrum. Their approach is based on the **generalized Euclidean (GE) distance**, which quantifies the "effort" required for influence to travel between opposing sides of a debate on a given network. For a graph G and an associated opinion vector o (a list of users' opinions, where opinions are normalized between -1 and $+1$), the polarization measure, denoted $\delta_{G,o}$, is formulated as:

$$\delta_{G,o} = \sqrt{(o^+ - o^-)^\top L^+ (o^+ - o^-)}$$

In this equation, o^+ contains the positive opinions (and zeros elsewhere, i.e. $o_i^+ = \max(o_i, 0), \forall i$), o^- contains the absolute value of negative opinions, and L^+ is the Moore-Penrose pseudo-inverse of the graph's Laplacian matrix. The term L^+ embeds the network's topology, weighting the distance by the paths (or lack thereof) connecting individuals. Conceptually, the measure represents the network distance between the "centers of mass" of the opposing opinion groups. The authors demonstrate through synthetic and real-world data that this metric is uniquely sensitive to all three aspects of polarization, unlike measures that focus only on opinion distributions or local network assortativity. This measurement is further extended to multipolari polarization in [66].

Huang et al. [39] propose a partition-agnostic polarization measure based on the correlation between signed and unsigned random-walk dynamics. They define a signed random-walk transition matrix $\mathbf{M}(t)$, where matrix entries encode both topological distances and the signs of paths, and contrast this with its unsigned counterpart $|\mathbf{M}(t)|$. The polarization score for a node u is computed as the Pearson correlation between its signed and unsigned transition vectors: $\mathrm{Pol}(u;t) = \mathrm{corr}(|\mathbf{M}|_{:u}(t), \mathbf{M}_{:u}(t))$. Averaging these node-level scores gives the overall graph polarization. This methodology captures polarization at multiple scales (controlled by the time parameter t), and crucially, it does not require pre-defined community partitions.

3.4 Others

Amelkin et al. introduce **Social Network Distance (SND)** [89] to quantify the evolution of user opinions in social networks, taking into account both the network structure and the dynamics of competing polar views. SND frames opinion change as a transportation problem, allowing it to capture not just individual shifts, but also the collective patterns of opinion propagation across the network. This approach is effective for detecting events that trigger or intensify polarization within society. For example, when applied to real-world X (formerly Twitter) data, SND is able to identify spikes in opinion divergence around major political events, such as elections or controversial policy debates. These spikes signal moments when public sentiment becomes sharply divided, enabling researchers to pinpoint the timing and nature of polarization-triggering events.

Bozdag et al. [38] present a comprehensive framework for empirically measuring the manifestation of offline political segregation in online environments. They operationalize information diversity using Shannon entropy, calculating separate scores for a user's information input and output. Specifically, "source diversity" quantifies the political heterogeneity of tweets a user receives, while "output diversity" measures the diversity of political content a user disseminates. The disparity between these scores, together with an "input-output correlation" metric, serves as an indicator of the filter bubble effect. Importantly, their framework goes beyond conventional diversity metrics by capturing structural exclusion of minority voices. They introduce the concept of "openness" measured by two indicators: "minority reach" which reflects the network-wide penetration of minority viewpoints, and "minority exposure" representing the proportion of minority content in an individual's feed. This approach demonstrates that even when overall source diversity appears sufficient, significant segregation can persist through the marginalization of minority actors.

As discussed in the echo chamber detection section, Calderón et al. [4] introduce the **Echo Chamber Index (ECI)**, which quantifies echo chamber behavior on a given fan page by averaging the echoing scores of comments under a post. This provides a straightforward yet effective numerical measure of echo chamber dynamics at the post level.

In [90], the measurement of the echo chamber effect is formalized within the influence maximization framework through the **Influence Maximization with Echo Chamber(IMEC) problem**. The authors model echo chamber influence as an additional probabilistic mechanism: a user's likelihood of adopting information is increased if many members of their group have already adopted it. This group-level effect is mathematically represented using an Ising-model-inspired function, where the probability of group-based activation depends on the number of activated users and a group closeness parameter. By comparing information diffusion outcomes with and without this global group influence, the paper directly quantifies the impact of echo chambers on the overall spread.

Although not directly focused on social networks, Ge et al. [15] develop a quantitative framework to measure interest reinforcement in the context of e-commerce echo chambers. Their approach segments users' interaction histories into temporal blocks, representing each user's interests within a block as an aggregate embedding of the items they engaged with. By comparing a "Following Group"—users who frequently interact with recommendations—to a control "Ignoring Group", they examine changes in the structure of user interest clusters over time. The reinforcement effect is assessed using cluster validity indices: a smaller decrease in the Calinski-Harabasz (CHK) score indicates that clusters remain more compact, while a higher Adjusted Rand Index (ARI) suggests less user drift between clusters. Their results demonstrate that this method effectively quantifies the reinforcement phenomenon central to echo chambers.

Finally, there exist several works that design polarization or controversy metrics as optimization objectives for their mitigation strategies. These will be introduced in a subsequent section dedicated to mitigation approaches. It is important to note that there is no single universally accepted metric for measuring echo chambers; different metrics capture different aspects of echo chambers, such as topological structure, opinion distribution, or a combination of both and cater to different conditions. For example, if we can only obtain people's post content and no interaction information among users due to the privacy or terms of policies of the social media platform, we can embed people's generated content and infer their leanings with ML, then simply use the variance of leanings as an indicator of polarization or the echo chamber effect. On the other hand, if only people's interactions are available and we cannot extract their post content, topology-based metrics could be employed. When only internal and external edges are of concern or are available, the E-I index is one of the choices. If community size is also a concern (as opposing communities with similar sizes are generally considered more polarized compared to one very large and one very small community), BSI could take this into consideration.

Typically, these metrics do not quantify echo chambers directly; rather, they measure the effects associated with echo chambers. For example, they assess which network aligns more closely with the intuition or definition of an echo chamber, or better matches expected properties.

4 Echo Chamber Mitigation

Although this section focuses on echo chamber mitigation, we also include research on reducing related phenomena such as polarization, segregation, and controversy, as these efforts often aim to counter or mitigate echo chamber effects. For example, polarization is both a consequence of echo chambers and a factor that accelerates their formation [5]; thus, reducing polarization contributes to echo chamber mitigation. Since most studies do not explicitly distinguish between these terms, in this survey, we consider all such efforts to fall within the scope of echo chamber mitigation.

Existing social network-based approaches to mitigate echo chamber effects can be broadly categorized into three groups: (1) *Cross-group Promotion Approaches*, which involves adding edges between opposing groups or recommending posts from users with differing viewpoints to increase exposure to alternative perspectives; (2) *Opinion Dynamics-Based Approaches*, which simulate the evolution of user opinions through opinion dynamics models and seek to optimize a polarization metric that depends on the final opinions and/or the network topology structure, by modifying the topology, edge weights, or users' innate opinions (though the latter is less practical); and (3) *Agent Addition-Based Approaches*, which introduce artificial agents that strategically disseminate information with targeted leanings or ideologies to influence other users or affect recommender systems.

4.1 Cross-group Promotion

A common strategy for mitigating polarization is to bridge opposing groups and expose individuals to diverse perspectives. Many studies have explored methods such as adding cross-group connections or recommending contents with opposing viewpoints. By increasing the diversity of information people encounter, these approaches aim to prevent echo chambers and encourage more moderate ideologies.

Luo et al. [3] introduce a game-theoretic framework to mitigate segregation and echo chambers in social networks by incentivizing inter-community connections. They model user interactions as an edge formation game, where individuals trade off homophily (preference for same-group ties) against exogenous rewards for cross-community links. Their **Algorithmic Recommendation Mechanism (ARM)** leverages weak ties to encourage diverse connections, reshaping the Nash equilibrium from segregated networks to integrated ones. Simulations demonstrate ARM's efficacy in reducing segregation, particularly during polarizing events, offering a scalable mechanism design solution to counteract echo chambers.

Several studies have approached the problem by identifying and algorithmically bridging structural divides in social networks. Garimella et al. [42] propose an algorithmic approach to reduce polarization by strategically bridging opposing communities in an online discussion. They begin by modeling a controversial topic as a directed "endorsement graph" (e.g., a retweet network), which is then partitioned into two disjoint communities representing the opposing sides. The authors' objective is to reduce a specific quantitative metric, the Random-Walk Controversy (RWC) score, which measures the isolation of these communities by calculating the probability that a random walker remains within its starting community. To achieve this, their method recommends a small set of new edges ("bridges") to be added between the two sides. Acknowledging that not all recommendations are equally likely to be accepted in reality, they introduce a model for "acceptance probability" based on user polarity, shifting the goal to minimizing the expected controversy score. They propose an efficient heuristic algorithm (**ROV-AP**) that identifies the most effective bridges by primarily considering connections between high-degree users on opposite sides, thus offering a practical method for algorithmically mitigating online echo chambers.

Interian, Moreno, and Ribeiro [44] solve the problem of reducing network polarization by formulating it as an Integer Linear Program (ILP). Their method, which addresses the **Minimum-Cardinality Balanced Edge Addition Problem (MinCBEAP)**, is designed to find the smallest set of new edges, E', to add to a graph $G = (V, E)$. The ILP model uses binary variables to represent the selection of new edges and to track the shortest path distances. The core of the solution method is an optimization that minimizes the number of added edges while enforcing a key structural constraint: ensuring that every vertex v in a target polarized group A can reach a vertex outside of its group ($V \setminus A$) within a specified distance threshold D. The ILP formulation directly solves the following problem:

$$\begin{aligned} \min \quad & |E'| \\ \text{subject to} \quad & d_{G'}(v, V \setminus A) \le D, \quad \forall v \in A \\ \text{where} \quad & G' = (V, E \cup E') \end{aligned}$$

By solving this ILP model with standard optimization software, they compute the minimal set of edges required to structurally bridge the isolated groups in a network. Their subsequent work [45] further compares three ILP formulations and reports computational results on both simulated and real-world networks.

In [46], Haddadan et al. propose a method to reduce structural polarization in content networks by adding a budget of k links. They introduce the (Polarized) Bubble Radius (BR), which measures the expected random walk steps from a node v to a page with an opposing view. Nodes with a high BR are termed "parochial", and the network's overall structural bias is the sum of their BRs. The problem of minimizing the bias is framed as a submodular maximization problem, allowing for an efficient greedy algorithm. The **REPBUBLIK** algorithm iteratively adds links from source nodes that are selected based on a task-specific variant of Random-Walk Closeness Centrality (RWCC), thereby strategically creating shortcuts that have the broadest impact on reducing network-wide polarization.

Recent studies have also examined the application of the influence propagation paradigm within social networks to balance information exposure [91–94]. Garimella et al. [91] introduce an approach to mitigate filter bubbles and echo chambers in networks by focusing on balancing information exposure across opposing viewpoints. Their framework adopts a centralized perspective, seeking to maximize the number of users exposed to both sides of a controversial issue. The authors model the spread of information using the independent-cascade model and formalize the objective as maximizing the number of users who are either reached by both campaigns or by neither, thereby directly addressing the problem of information imbalance. They demonstrate that this balancing problem is NP-hard, and, crucially, that its objective function is neither monotone nor submodular, making it difficult to find efficient approximation solutions. To address this, they propose and analyze several greedy algorithms with approximation guarantees, and validate their effectiveness through experiments on real-world Twitter data spanning political and social controversies. With a similar idea, Tu et al. [93] formally define **co-exposure maximization (COEM)** as the task of selecting initial user sets for each campaign in order to maximize the expected number of users who receive information from both campaigns, considering the probabilistic nature of information spread in the network. They propose a greedy approximation algorithm that uses a submodular lower bound for the co-exposure objective, offering theoretical guarantees. Additionally, they introduce a scalable estimation method based on generalized random reverse-reachable sets, which enables efficient computation of expected co-exposure in large networks. Matakos et al. [92] also address filter bubble mitigation by maximizing the diversity of information exposure in social networks. However, their approach models both user and content leanings, and strategically recommends news articles to select users so that, as articles propagate, users are exposed to a wider range of viewpoints. The problem is formulated as a submodular optimization under matroid constraints, and solved efficiently using a novel sampling technique called reverse co-exposure sets. Considering the ignorance of the competition between opposing opinions propagating in previous studies, Banerjee et al. [94] address filter bubble mitigation by modeling the realistic competition between opposing viewpoints as they spread in a social network. They propose the RIC-FB model, which distinguishes between awareness and adoption of opinions, and incorporates a competition parameter that makes it harder for users to adopt a second, opposing viewpoint after adopting the first. Their approach rewards co-adoption (users adopting both viewpoints), thus directly targeting filter bubble reduction rather than just exposure balancing. They formulate the mitigation task as an optimization problem and prove its computational hardness, then introduce specialized algorithms—including a reverse-influence-sampling-based heuristic—to effectively select seeds for the counter-campaign. Experiments on real networks show their competition-aware methods outperform existing baselines, especially when competition is strong.

Beyond structural interventions, Orbach et al. [28] address echo chambers by introducing a novel **Natural Language Understanding (NLU)** task for detecting countering speeches. Rather than merely identifying content with opposing stances, their approach retrieves texts that directly refute the specific arguments presented in an input document. By automatically surfacing targeted rebuttals, their method aims to expose users to diverse viewpoints and foster a more balanced and informed perspective.

Although many studies have explored methods for promoting cross-group edges and have found that inter-group contact can foster compromise and mutual understanding in some contexts, "confirmation bias" [73] often impedes effective discourse and connection—especially especially when users behave strategically for profit or other motives. Bail et al. [72] also point out that attempts to expose people to a broad range of opposing political views on social media platforms like Twitter may not only be ineffective, but even counterproductive. Other studies indicate that exposure to opposing political views may trigger "backfire effects"[69], which can intensify political polarization[69–71]. It has also been discussed in [41] that the effectiveness of this approach may depend on the specific network context and the existing degree of polarization. Exposure to opposing views may reduce polarization during the initial or intermediate phases of polarization, but it is less effective once polarization is already strong. Bail et al. suggest that "future attempts to reduce political polarization on social media will most likely require learning which types of messages, tactics, or issue positions are most likely to create backfire effects." Depolarizing users by exposing them to viewpoints only slightly less radical than their own may be more effective [21, 75].

4.2 Opinion Dynamics-Based

There is a class of works focusing on simulating opinion dynamics with existing or designed models and optimizing polarization or similar metrics, which are computed using users' final opinion values after the opinion dynamics process. These studies achieve their goals by changing the network topology (adding or deleting edges), modifying edge weights, adjusting parameters in the opinion dynamics model (e.g., innate opinions or weights in the Friedkin-Johnsen model [74] (FJ model)), or by employing other strategies.

A foundational direction in this area is to optimize the structure or weights of the network to mitigate polarization. Musco et al. [50] formalize the problem of finding a network structure that minimizes both polarization and disagreement. They propose a **Polarization-Disagreement Index** based on the FJ model, which sums two terms: polarization (the variance of final opinions) and disagreement (the opinion differences across edges). Crucially, the authors prove this index is a convex function of the network's edge weights, which allows the optimal graph topology to be computed efficiently. Their analysis also shows that this optimal network can be well-approximated by a sparse graph with only $O(n/\epsilon^2)$ edges.

Building on the FJ model, several works propose new metrics or optimization methods to further enhance network robustness. Chen et al. [43] present a structural approach to conflict mitigation that focuses on minimizing the risk of conflict rather than the conflict itself for a single, known issue. Departing from methods that require specific opinion data, they propose network-level metrics that are independent of any particular opinion distribution. They define the **Average-Case Conflict Risk (ACR)** and **Worst-Case Conflict Risk (WCR)**, which quantify a network's inherent propensity for disagreement over all possible opinion configurations. They then introduce optimization algorithms (e.g., **coordinate descent**) to minimize these risk measures by making a small number of targeted edge additions or deletions. This strategy aims to create a more robust and resilient network topology that is less susceptible to polarization, regardless of the specific topic of controversy. Their empirical results show that minimizing the WCR is particularly effective, as it tends to reduce the average-case risk as well, leading to a more generally conflict-resistant network structure.

The dynamics of filter bubbles and content recommendation are also a focus of recent research. In their analysis of filter bubbles, [27] demonstrate how a "network administrator"—modeling a social media platform's content-filtering algorithm—can dramatically increase polarization by minimizing user disagreement. They also propose a "simple remedy" to mitigate this effect. Their solution involves modifying the administrator's objective function by adding an L^2 regularization term. This "**Regularized Dynamics**" approach discourages the algorithm from making large, concentrated changes to a few social connections. Instead, it incentivizes smaller, more distributed adjustments across many edges in the network. The authors show this method to be highly effective: in their experiments, the regularized model limited the increase in polarization to just 4%, compared to an over 4000% increase in the non-regularized model. Crucially, this was achieved while only minimally impacting user disagreement (an increase of at most 5%), suggesting that platforms could control the formation of polarizing echo chambers without significantly harming their engagement-driven business objectives.

Other research investigates network modifications under practical constraints. In their work, [64] propose methods for a centralized planner to reduce sociopolitical polarization by perturbing a social network's structure under a fixed budget. Using the FJ model, where expressed opinions $\mathbf{z}$ are determined by innate opinions $\mathbf{s}$ and the graph Laplacian $\mathbf{L}$ via the relation $\mathbf{z} = (\mathbf{I} + \mathbf{L})^{-1}\mathbf{s}$, they define polarization $P(\mathbf{z})$ as the variance of the expressed opinions. The authors first consider a setting where the planner has full knowledge of the population's opinions. They derive the exact change in polarization from adding an edge and propose two greedy heuristics: a **Coordinate Descent (CD)** strategy that iteratively adds the edge yielding the largest marginal decrease in polarization, and a simpler **Disagreement-**

Seeking (DS) strategy that adds an edge between the two individuals with the highest expressed disagreement, $(z_i - z_j)^2$. In a second, more robust setting, they analyze the problem where opinions are chosen adversarially to maximize polarization. They demonstrate this minimax problem is equivalent to maximizing the spectral gap (λ_2) of the graph's Laplacian. This motivates their third strategy, the **Fiedler Difference (FD)** heuristic, which adds edges between vertices on opposite sides of the network partition induced by the Fiedler vector—the eigenvector corresponding to λ_2. This approach aims to make the network structure more robust against divisive opinion configurations by bridging its most prominent communities.

Extending these approaches to more realistic settings, Cinus et al. [67] address the problem of mitigating polarization and disagreement in social networks by proposing a method to rebalance a user's social feed. Their approach is set within a directed graph context, where edges represent follower-followee relationships, and it is based on the FJ model. The core idea is to re-weight the influence of existing connections rather than creating new ones, thereby preserving each user's total engagement. The authors formulate an optimization problem that seeks to find a new row-stochastic adjacency matrix A^* that minimizes the sum of network polarization and disagreement at the equilibrium state of opinions. For a vector of innate opinions s, their objective function for a directed graph is given by:

$$f(A, s) = s^T(2I - A)^{-T}s + s^T(2I - A)^{-T}\frac{D_{in} - I}{2}(2I - A)^{-1}s$$

where D_{in} is the diagonal matrix of in-degrees. A key contribution is the analysis of this problem's properties, where they demonstrate that while the feasible set of matrices is convex (maintaining original sparsity and row-stochasticity), the objective function is not matrix-convex. To solve this challenging non-convex problem scalably, they develop an algorithm named **Laplacian-Constrained Gradient Descent (LcGD)**, which is based on projected gradient descent. The algorithm efficiently computes the gradient without explicit matrix inversion by solving linear systems and then projects the solution back onto the feasible set to maintain the constraints. Their work is notable for being one of the first to tackle this problem in the more realistic setting of directed graphs, providing a proper generalization of previous work on undirected networks.

Apart from modifying network structure or edge weights, some works explore interventions at the level of user opinions. Matakos et al. [63] propose and formalize two distinct, NP-hard problems for polarization reduction by convincing a small set of k individuals to adopt a neutral stance. The first method, **MODERATEINTERNAL**, models interventions that change an individual's core beliefs, such as through education. The goal is to select a set of k nodes, T_s, and set their internal opinions to zero. If s is the original vector of internal opinions and s' is the modified vector where the opinions for nodes in T_s are zero, the objective is to find the set T_s that minimizes the polarization of the resulting expressed opinions: $\min_{T_s, |T_s|=k} \|(\mathbf{L} + \mathbf{I})^{-1}s'\|^2$. The authors find that this strategy is most effective when targeting "fringe" nodes with the most extreme expressed opinions. On the other hand, convincing fringe nodes with extreme opinions to adopt a neutral stance is not realistic. In contrast, the second method, **MODERATEEXPRESSED**, models interventions that incentivize individuals to moderate their public statements. This approach involves selecting a set of k nodes, T_z, and fixing their expressed opinions to zero, which directly alters the opinion dynamics as these nodes now propagate neutrality. The objective is to choose the set T_z that minimizes the polarization index. This strategy is shown to be most effective when targeting central and influential nodes, as their moderated expression has the greatest cascading effect throughout the network. While this intervention assumes that individuals can be perfectly incentivized to propagate neutral public stances, in practice, the number of individuals who can be targeted (k) is typically constrained by practical considerations such as resource availability and intervention costs. As k increases, the potential to reduce polarization grows, but so does the cost and complexity of implementation. Thus, there is an inherent trade-off between the scale of the intervention and its feasibility. Nevertheless, the model offers valuable insight by providing an upper bound on the potential effectiveness of moderation-based interventions and can guide the design of more practical strategies.

Although opinion dynamics models allow researchers to simulate the evolution of opinions, design mitigation strategies, and evaluate their effectiveness based on final opinions, this class of methods has several limitations:

1. **Accuracy of Opinion Dynamics Models.** While many opinion dynamics models have been proposed to simulate opinion evolution (e.g., the FJ model, Sznajd model [76], and FJCB [77]—a Friedkin-Johnsen type model that incorporates confirmation bias to better capture echo chamber formation), their accuracy remains insufficiently analyzed and validated on real-world social networks. Moreover, parameter selection often lacks clear guidance or established rules, potentially limiting model performance. For instance, most methods rely on a simplified FJ model that assumes a uniform stubbornness value of 1 for all individuals. This assumption fails to reflect real-world scenarios, where individuals demonstrate varying degrees of stubbornness. Additionally, the model computes the final opinion as $z = (I+L)^{-1}s$, which implies that the sum or average of final opinions equals that of the innate opinions. This suggests that overall opinion remains unchanged after the opinion dynamics process, contradicting real-world observations where collective opinions can shift over time. Furthermore, the simplified FJ model assumes that interpersonal influence weights remain static throughout the process, whereas, in reality, these weights may evolve as individuals' opinions change due to factors such as exposure or personal leaning. Obtaining authentic innate opinions for users, which directly determine final opinions when the Laplacian matrix is given, is also nearly impossible. While the simplified model is frequently used for its analytical tractability and has been adopted in previous studies, these simplifications may significantly limit its ability to accurately capture the complexity of real-world opinion dynamics.

2. **Practicality of Proposed Strategies.** Although opinion dynamics can be simulated, the resulting mitigation strategies may be difficult to implement in practice. For example, in [63], the **MODERATEINTERNAL** intervention models changes to an individual's core beliefs through education or similar methods. However, attempting to educate users to alter their innate opinions is challenging and time-consuming, particularly in large-scale networks or when user identification is constrained by privacy concerns.

Despite these challenges, opinion dynamics-based methods offer a dynamic perspective for analyzing networks and user opinions. Some studies that balance polarization and disagreement (e.g., [50]) point to promising directions that consider both individual polarization and platform interests.

4.3 Agents Addition

Several studies have explored the strategy of introducing new agents into a network to disseminate ideas more effectively or to influence recommender systems, thereby promoting a more diverse flow of information among users.

Ghezelbash et al. [47] present an innovative analytical framework for strategically inducing polarization in social networks by selecting a minimal set of informed agents. Their approach models the social network as a linear dynamical system, where opinion formation is determined by an influence matrix A. The central insight is that steering the network toward a desired final opinion state, x_d, can be formulated as a control problem. By introducing the concept of "equilibratability", they show that the necessary set of informed agents corresponds exactly to the non-zero entries of the vector $(I-A)x_d$. Consequently, the challenge of selecting the fewest agents becomes a zero-norm minimization problem: identifying a target opinion state x_d that satisfies specific constraints (e.g., ensuring two subgroups attain distinct average opinions) while simultaneously minimizing the sparsity of $(I - A)x_d$. As this problem is NP-hard, the authors reformulate it as a computationally tractable Integer Linear Programming (ILP) problem. A notable finding from their analysis of Zachary's Karate Club network is that the optimal agents for inducing polarization are not necessarily the most highly connected "hubs", but rather "bridge" agents whose connections span opposing factions. This work offers a principled, optimization-based method for targeted intervention, providing a formal alternative to heuristic or simulation-based agent selection strategies. Their method represents a synthesis of opinion dynamics and agent addition techniques.

Although not directly focused on social networks, Rastegarpanah et al. [26] propose a novel approach to reducing polarization in recommender systems by introducing "antidote data". Their method involves injecting a small set of new, artificial users whose ratings are strategically optimized to minimize a chosen polarization metric, specifically the variance of predicted ratings for an item across the user population. This additional data acts as a constraint during the system's training phase, compelling the underlying model (e.g., matrix factorization) to learn item representations that yield less divergent predictions for the original users. Notably, this technique can effectively mitigate polarization without necessitating any changes to the core recommendation algorithm or the original dataset.

4.4 Others

Alatawi et al.[21] summarize several "human-focused prevention" strategies, which empower users to curate their own information feeds and thereby reduce bias. For a more detailed analysis, readers are encouraged to consult [21]. Additionally, there are numerous field-studies and survey-based mitigation strategies, as well as approaches focusing on recommender systems; readers are referred to [41] for further information.

5 Challenges and Opportunities

In this section, we outline the key challenges and potential future research directions in the field of echo chamber studies.

1. **Echo Chamber Detection and Measurement:**
   Users' opinions are central to detecting and measuring echo chambers. Some approaches rely on manually annotated labels in user posts to infer leanings, which requires a deep understanding of the topic and users' intentions. Furthermore, these approaches are not scalable. Others depend on existing scores for certain websites or public figures, which are not always generalizable to other datasets or topics. Automated methods—such as regression, classification models, and LLM-based approaches—have been applied to stance detection in text. Recent advances in LLMs have substantially boosted the accuracy of traditional methods, with F1 scores now exceeding 80%. Their application is still in an early, exploratory stage, and developing effective techniques to adapt and utilize these models for specific stance detection tasks remains an open research challenge. Additionally, the roles, background knowledge, and potential biases of LLMs in stance detection contexts require further investigation [95]. To enable more accurate and efficient opinion inference and support in-depth analysis, further improvements in stance detection—especially those tailored to social networks [96] and controversial topics—might be needed.

2. **Modeling Opinion Evolution:**
   Numerous opinion dynamics models have been proposed to study the evolution of users' opinions. Among these are models addressing polarization and echo chambers, which incorporate crucial

factors such as confirmation bias and social influence. However, as discussed previously, some models are overly simplistic and fail to capture the complex dynamics of real-world social networks. Moreover, there is a lack of analysis regarding the accuracy of these models on large-scale, long-term datasets where echo chambers may form and evolve. Developing opinion dynamics models that can accurately reflect and predict opinion evolution would greatly benefit echo chamber mitigation strategies and help minimize their negative side effects. In addition to improving theoretical models and evaluating them on large-scale datasets, conducting field studies to validate these models in real-world social networks is crucial. Such empirical validation can ensure that the models not only fit observed data but also reliably predict opinion dynamics and the formation of echo chambers. Additionally, field studies can also be conducted to identify factors contributing to the formation of echo chambers, which may help in designing more realistic opinion evolution models.

3. **Quantification Metrics:**
   There is currently no universal, widely accepted metric for quantifying echo chambers in research. Comparative studies of existing metrics are also lacking, making it difficult to assess whether these metrics reliably capture any specific characteristics of echo chambers and provide trustworthy scores. The GE method proposed in [65], which incorporates both network topology and opinion information, offers a promising direction. This method has been compared with other metrics like RWC on certain networks, demonstrating strong performance and sensitivity to both structural and opinion distribution factors in polarization quantification. Given the large number of metrics proposed for echo chamber and polarization measurement, deeper analysis and comparison based on large scale field studies are needed to uncover their strengths and limitations as effective measurement tools.

4. **Mitigation Strategies:**
   Practical mitigation strategies should balance the interests of platform companies, legal requirements, individual rights, ethical considerations, content diversity, and user engagement to foster healthier social network environments. As highlighted in [41], "no studies about removing (or adding) nodes for reducing the polarization were found in this review. However, this method is often used in practice for banning specific posts or accounts from social networks." While demoting or banning radical and influential users is often applied in practice (e.g., banning malicious accounts [41]), the effectiveness and consequences of such actions can be complex and not easily predictable. On one hand, user/post demotion may not necessarily trigger the backfire effects sometimes observed with cross-group promotion. On the other hand, high-profile bans, such as the removal of US President Trump's Facebook account, have resulted in significant controversy and political backlash, illustrating that such interventions can have side effects whose costs are not easily measurable. Beyond actions available to network operators, third parties—such as dedicated users, civil society organizations, or automated agents—may also help mitigate echo chambers and filter bubbles by disseminating diverse or corrective content, engaging in counter-speech, or promoting fact-checking initiatives. Furthermore, gamification strategies, such as incentivizing users to engage with diverse viewpoints or rewarding civil discourse, may also be promising approaches to encourage healthier interactions and reduce polarization.

5. **Distinguishing Echo Chambers from Polarization:**
   As noted in [14], "Another issue with the existing definition is the equating of polarization in the network with echo chambers, and subsequently, many of these studies attempt to propose an approach to address polarization". Consequently, much of the related work focuses on polarization mitigation. However, by definition, echo chambers should not be equated solely with polarization, as the "reinforcement" effect is a distinct characteristic, separate from polarity. Thus, an important

future research direction is to develop models and measurement techniques that accurately capture the reinforcement mechanisms unique to echo chambers. Detecting echo chambers by focusing specifically on these reinforcement dynamics—particularly during their formation and evolution—could provide deeper insights into their structure and impact. Furthermore, research into effective mitigation strategies should aim to disrupt or weaken the reinforcement processes that sustain echo chambers, rather than merely reducing overall polarization.

6. **Data Availability and Privacy:**
   There is a lack of large-scale, long-term datasets for researching echo chamber formation and evolution. Collecting such datasets, for example from platforms like Twitter, can be costly and challenging. Furthermore, due to privacy concerns, specific content from these platforms cannot be readily shared. Researchers using datasets from prior studies often need to re-collect content through tweet IDs or links. Therefore, better platform policies are needed—ones that both protect user privacy and facilitate social network research.

6 Conclusion

Echo chambers represent a critical challenge in the landscape of online social networks, with far-reaching consequences for public discourse, information diversity, and societal cohesion. In this survey, we have provided a comprehensive overview of echo chamber research, systematically reviewing approaches for detection, measurement, and mitigation. Our analysis highlights the diversity of methods available, spanning topological, semantic, and hybrid techniques, as well as the multitude of metrics employed in the literature to capture the nuanced effects of echo chambers.

Despite significant progress, key challenges remain. The field lacks universally accepted definitions and robust, generalizable metrics for quantifying echo chamber effects. Many mitigation strategies, while promising in simulation or small-scale studies, face practical and ethical constraints when deployed in real-world systems. Furthermore, the interplay between algorithmic design, user behavior, and societal context complicates the task of designing interventions that not only are effective, but also balance users' rights to diverse information, commercial interests of social media platforms, and practical feasibility.

Looking ahead, future research may address these challenges by developing more accurate models of opinion dynamics, improving the reliability of detection and measurement techniques, and designing mitigation strategies that balance competing interests. Interdisciplinary collaboration—spanning computer science, social science, psychology, and ethics—will be essential for advancing our understanding and management of echo chambers. By fostering a more nuanced and evidence-based approach to echo chambers, we can work towards healthier, more inclusive online environments that promote informed and diverse public discourse.

Acknowledgments

During the literature review and synthesis of related work, we used large language models to assist in generating initial draft summaries of surveyed papers, primarily using methodological descriptions and formulas from the original works. LLMs were also used to help polish and refine the written text throughout the manuscript. All LLM-generated content was subsequently reviewed, classified, and fact-checked by the authors, with connections and relationships between them established manually. The identification and analysis of drawbacks, challenges, and opportunities were carried out by the authors, with LLMs providing assistance for language clarity and editing where needed.

References

[1] Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G. & Quattrociocchi, W. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific Reports*. **6**, 37825 (2016)

[2] Chkhartishvili, A. & Kozitsin, I. Binary separation index for echo chamber effect measuring. *2018 Eleventh International Conference" Management Of Large-scale System Development"(MLSD*. pp. 1-4 (2018)

[3] Luo, R., Nettasinghe, B. & Krishnamurthy, V. Controlling segregation in social network dynamics as an edge formation game. *IEEE Transactions On Network Science And Engineering*. **9**, 2317-2329 (2022)

[4] Calderón, F., Cheng, L., Lin, M., Huang, Y. & Chen, Y. Content-based echo chamber detection on social media platforms. *Proceedings Of The 2019 IEEE/ACM International Conference On Advances In Social Networks Analysis And Mining*. pp. 597-600 (2019)

[5] Al Atiqi, M., Chang, S. & Deguchi, H. Agent-based approach to resolve the conflicting observations of online echo chamber. *2020 Joint 11th International Conference On Soft Computing And Intelligent Systems And 21st International Symposium On Advanced Intelligent Systems (SCIS-ISIS)*. pp. 1-6 (2020)

[6] Villa, G., Pasi, G. & Viviani, M. Echo chamber detection and analysis: A topology-and content-based approach in the COVID-19 scenario. *Social Network Analysis And Mining*. **11**, 78 (2021)

[7] Guerra, P., Meira Jr, W., Cardie, C. & Kleinberg, R. A measure of polarization on social media networks based on community boundaries. *Proceedings Of The International AAAI Conference On Web And Social Media*. **7**, 215-224 (2013)

[8] Minici, M., Cinus, F., Monti, C., Bonchi, F. & Manco, G. Cascade-based echo chamber detection. *Proceedings Of The 31st ACM International Conference On Information & Knowledge Management*. pp. 1511-1520 (2022)

[9] Cota, W., Ferreira, S., Pastor-Satorras, R. & Starnini, M. Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*. **8**, 1-13 (2019)

[10] Al Atiqi, M., Chang, S. & Hiroshi, D. Agent-based approach to echo chamber reduction strategy in social media. *2018 Joint 10th International Conference On Soft Computing And Intelligent Systems (SCIS) And 19th International Symposium On Advanced Intelligent Systems (ISIS)*. pp. 1301-1306 (2018)

[11] Botte, N., Ryckebusch, J. & Rocha, L. Clustering and stubbornness regulate the formation of echo chambers in personalised opinion dynamics. *Physica A: Statistical Mechanics And Its Applications*. **599** pp. 127423 (2022)

[12] Madsen, J., Bailey, R. & Pilditch, T. Large networks of rational agents form persistent echo chambers. *Scientific Reports*. **8**, 12391 (2018)

[13] Morini, V., Pollacci, L. & Rossetti, G. Toward a standard approach for echo chamber detection: Reddit case study. *Applied Sciences*. **11**, 5390 (2021)

[14] Mahmoudi, A., Jemielniak, D. & Ciechanowski, L. Echo chambers in online social networks: A systematic literature review. *IEEE Access*. **12** pp. 9594-9620 (2024)

[15] Ge, Y., Zhao, S., Zhou, H., Pei, C., Sun, F., Ou, W. & Zhang, Y. Understanding echo chambers in e-commerce recommender systems. *Proceedings Of The 43rd International ACM SIGIR Conference On Research And Development In Information Retrieval*. pp. 2261-2270 (2020)

[16] Garimella, K., Morales, G., Gionis, A. & Mathioudakis, M. Quantifying controversy on social media. *ACM Transactions On Social Computing*. **1**, 1-27 (2018)

[17] Newman, M. & Girvan, M. Finding and evaluating community structure in networks. *Physical Review E*. **69**, 026113 (2004)

[18] Fortunato, S. Community detection in graphs. *Physics Reports*. **486**, 75-174 (2010)

[19] Sikder, O., Smith, R., Vivo, P. & Livan, G. A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*. **10**, 5493 (2020)

[20] Prasetya, H. & Murata, T. A model of opinion and propagation structure polarization in social media. *Computational Social Networks*. **7** pp. 1-35 (2020)

[21] Alatawi, F., Cheng, L., Tahir, A., Karami, M., Jiang, B., Black, T. & Liu, H. A survey on echo chambers on social media: Description, detection and mitigation. *ArXiv Preprint ArXiv:2112.05084*. (2021)

[22] Cossard, A., Morales, G., Kalimeri, K., Mejova, Y., Paolotti, D. & Starnini, M. Falling into the echo chamber: the Italian vaccination debate on Twitter. *Proceedings Of The International AAAI Conference On Web And*

Social Media. **14** pp. 130-140 (2020)

[23] Nourbakhsh, A., Liu, X., Li, Q. & Shah, S. Mapping the Echo-chamber: Detecting and Characterizing Partisan Networks on Twitter. *Proceedings Of The 2017 International Conference On Social Computing, Behavioral-Cultural Modeling, & Prediction And Behavior Representation In Modeling And Simulation.* pp. 5-8 (2017)

[24] Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A. & Quattrociocchi, W. Mapping social dynamics on Facebook: The Brexit debate. *Social Networks.* **50** pp. 6-16 (2017)

[25] Du, S. & Gregory, S. The echo chamber effect in twitter: Does community polarization increase?. *Complex Networks & Their Applications V: Proceedings Of The 5th International Workshop On Complex Networks And Their Applications (COMPLEX NETWORKS 2016).* pp. 373-378 (2017)

[26] Rastegarpanah, B., Gummadi, K. & Crovella, M. Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. *Proceedings Of The Twelfth ACM International Conference On Web Search And Data Mining.* pp. 231-239 (2019)

[27] Chitra, U. & Musco, C. Analyzing the impact of filter bubbles on social network polarization. *Proceedings Of The 13th International Conference On Web Search And Data Mining.* pp. 115-123 (2020)

[28] Orbach, M., Bilu, Y., Toledo, A., Lahav, D., Jacovi, M., Aharonov, R. & Slonim, N. Out of the Echo Chamber: Detecting Countering Debate Speeches. arXiv 2020. *ArXiv Preprint ArXiv:2005.01157.*

[29] Interian, R. & Ribeiro, C. An empirical investigation of network polarization. *Applied Mathematics And Computation.* **339** pp. 651-662 (2018)

[30] Lelkes, Y. Mass polarization: Manifestations and measurements. *Public Opinion Quarterly.* **80**, 392-410 (2016)

[31] Markgraf, M. & Schoch, M. Quantification of echo chambers: a methodological framework considering multi-party systems. (2019)

[32] Ertan, G., Çarkoğlu, A. & Aytaç, S. Cognitive political networks: A structural approach to measure political polarization in multiparty systems. *Social Networks.* **68** pp. 118-126 (2022)

[33] Zhang, Y., Friend, A., Traud, A., Porter, M., Fowler, J. & Mucha, P. Community structure in Congressional cosponsorship networks. *Physica A: Statistical Mechanics And Its Applications.* **387**, 1705-1712 (2008)

[34] Dal Maso, C., Pompa, G., Puliga, M., Riotta, G. & Chessa, A. Voting behavior, coalitions and government strength through a complex network analysis. *PloS One.* **9**, e116046 (2014)

[35] Garcia, D., Abisheva, A., Schweighofer, S., Serdült, U. & Schweitzer, F. Ideological and temporal components of network polarization in online political participatory media. *Policy & Internet.* **7**, 46-79 (2015)

[36] Wolfowicz, M., Weisburd, D. & Hasisi, B. Examining the interactive effects of the filter bubble and the echo chamber on radicalization. *Journal Of Experimental Criminology.* **19**, 119-141 (2023)

[37] Emamgholizadeh, H., Nourizade, M., Tajbakhsh, M., Hashminezhad, M. & Esfahani, F. A framework for quantifying controversy of social network debates using attributed networks: biased random walk (BRW). *Social Network Analysis And Mining.* **10**, 90 (2020)

[38] Bozdag, E., Gao, Q., Houben, G. & Warnier, M. Does offline political segregation affect the filter bubble? An empirical analysis of information diversity for Dutch and Turkish Twitter users. *Computers In Human Behavior.* **41** pp. 405-415 (2014)

[39] Huang, Z., Silva, A. & Singh, A. Pole: Polarized embedding for signed networks. *Proceedings Of The Fifteenth ACM International Conference On Web Search And Data Mining.* pp. 390-400 (2022)

[40] Guyot, A., Gillet, A., Leclercq, É. & Cullot, N. ERIS: an approach based on community boundaries to assess polarization in online social networks. *International Conference On Research Challenges In Information Science.* pp. 88-104 (2022)

[41] Interian, R., G. Marzo, R., Mendoza, I. & Ribeiro, C. Network polarization, filter bubbles, and echo chambers: an annotated review of measures and reduction methods. *International Transactions In Operational Research.* **30**, 3122-3158 (2023)

[42] Garimella, K., De Francisci Morales, G., Gionis, A. & Mathioudakis, M. Reducing controversy by connecting opposing views. *Proceedings Of The Tenth ACM International Conference On Web Search And Data Mining.* pp. 81-90 (2017)

[43] Chen, X., Lijffijt, J. & De Bie, T. Quantifying and minimizing risk of conflict in social networks. *Proceedings Of The 24th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining.* pp. 1197-1205

(2018)

[44] Interian, R., Moreno, J. & Ribeiro, C. Reducing network polarization by edge additions. *Proceedings Of The 2020 4th International Conference On Intelligent Systems, Metaheuristics & Swarm Intelligence*. pp. 87-92 (2020)

[45] Interian, R., Moreno, J. & Ribeiro, C. Polarization reduction by minimum-cardinality edge additions: Complexity and integer programming approaches. *International Transactions In Operational Research*. **28**, 1242-1264 (2021)

[46] Haddadan, S., Menghini, C., Riondato, M. & Upfal, E. Repbublik: Reducing polarized bubble radius with link insertions. *Proceedings Of The 14th ACM International Conference On Web Search And Data Mining*. pp. 139-147 (2021)

[47] Ghezelbash, E., Yazdanpanah, M. & Asadpour, M. Polarization in cooperative networks through optimal placement of informed agents. *Physica A: Statistical Mechanics And Its Applications*. **536** pp. 120936 (2019)

[48] Shekatkar, S. Do zealots increase or decrease the polarization of social networks?. *Journal Of Complex Networks*. **8**, cnz036 (2020)

[49] Newman, M. & Girvan, M. Mixing patterns and community structure in networks. *Statistical Mechanics Of Complex Networks*. pp. 66-87 (2003)

[50] Musco, C., Musco, C. & Tsourakakis, C. Minimizing polarization and disagreement in social networks. *Proceedings Of The 2018 World Wide Web Conference*. pp. 369-378 (2018)

[51] Alatawi, F., Sheth, P. & Liu, H. Quantifying the echo chamber effect: an embedding distance-based approach. *Proceedings Of The International Conference On Advances In Social Networks Analysis And Mining*. pp. 38-45 (2023)

[52] Morales, A., Borondo, J., Losada, J. & Benito, R. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal Of Nonlinear Science*. **25** (2015)

[53] Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *Proceedings Of The National Academy Of Sciences*. **118**, e2023301118 (2021)

[54] Cinelli, M., Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. Echo chambers on social media: A comparative analysis. *ArXiv Preprint ArXiv:2004.09603*. (2020)

[55] Lai, M., Patti, V., Ruffo, G. & Rosso, P. Stance evolution and twitter interactions in an italian political debate. *Natural Language Processing And Information Systems: 23rd International Conference On Applications Of Natural Language To Information Systems, NLDB 2018, Paris, France, June 13-15, 2018, Proceedings 23*. pp. 15-27 (2018)

[56] Li, Z., Dong, Y., Gao, C., Zhao, Y., Li, D., Hao, J., Zhang, K., Li, Y. & Wang, Z. Breaking filter bubble: A reinforcement learning framework of controllable recommender system. *Proceedings Of The ACM Web Conference 2023*. pp. 4041-4049 (2023)

[57] Citraro, S. & Rossetti, G. Eva: Attribute-aware network segmentation. *International Conference On Complex Networks And Their Applications*. pp. 141-151 (2019)

[58] Zhang, P., Haq, E., Zhu, Y., Hui, P. & Tyson, G. Echo chambers within the russo-ukrainian war: The role of bipartisan users. *Proceedings Of The International Conference On Advances In Social Networks Analysis And Mining*. pp. 154-158 (2023)

[59] Karypis, G. & Kumar, V. METIS: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices. (1997)

[60] Amendola, M., Cavaliere, D., De Maio, C., Fenza, G. & Loia, V. Towards echo chamber assessment by employing aspect-based sentiment analysis and gdm consensus metrics. *Online Social Networks And Media*. **39** pp. 100276 (2024)

[61] Barberá, P., Jost, J., Nagler, J., Tucker, J. & Bonneau, R. Tweeting from left to right: Is online political communication more than an echo chamber?. *Psychological Science*. **26**, 1531-1542 (2015)

[62] Garimella, K., De Francisci Morales, G., Gionis, A. & Mathioudakis, M. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. *Proceedings Of The 2018 World Wide Web Conference*. pp. 913-922 (2018)

[63] Matakos, A., Terzi, E. & Tsaparas, P. Measuring and moderating opinion polarization in social networks. *Data Mining And Knowledge Discovery*. **31** pp. 1480-1505 (2017)

[64] Rácz, M. & Rigobon, D. Towards consensus: Reducing polarization by perturbing social networks. *IEEE*

Transactions On Network Science And Engineering. **10**, 3450-3464 (2023)
[65] Hohmann, M., Devriendt, K. & Coscia, M. Quantifying ideological polarization on a network using generalized Euclidean distance. *Science Advances*. **9**, eabq2044 (2023)
[66] Weidemann, C. Quantifying Multipolar Polarization. *ArXiv Preprint ArXiv:2405.16352*. (2024)
[67] Cinus, F., Gionis, A. & Bonchi, F. Rebalancing social feed to minimize polarization and disagreement. *Proceedings Of The 32nd ACM International Conference On Information And Knowledge Management*. pp. 369-378 (2023)
[68] Krackhardt, D. & Stern, R. Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*. pp. 123-140 (1988)
[69] Nyhan, B. & Reifler, J. When corrections fail: The persistence of political misperceptions. *Political Behavior*. **32**, 303-330 (2010)
[70] Lord, C., Ross, L. & Lepper, M. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.. *Journal Of Personality And Social Psychology*. **37**, 2098 (1979)
[71] Taber, C. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *American Journal Of Political Science*. **50**, 755-769 (2006)
[72] Bail, C., Argyle, L., Brown, T., Bumpus, J., Chen, H., Hunzaker, M., Lee, J., Mann, M., Merhout, F. & Volfovsky, A. Exposure to opposing views on social media can increase political polarization. *Proceedings Of The National Academy Of Sciences*. **115**, 9216-9221 (2018)
[73] Nickerson, R. Confirmation bias: A ubiquitous phenomenon in many guises. *Review Of General Psychology*. **2**, 175-220 (1998)
[74] Friedkin, N. & Johnsen, E. Social influence and opinions. *Journal Of Mathematical Sociology*. **15**, 193-206 (1990)
[75] Bail, C. Breaking the social media prism: How to make our platforms less polarizing. (Princeton University Press,2022)
[76] Sznajd-Weron, K., Sznajd, J. & Weron, T. A review on the Sznajd model—20 years after. *Physica A: Statistical Mechanics And Its Applications*. **565** pp. 125537 (2021)
[77] Chen, T., Wang, X. & Tsourakakis, C. Polarizing opinion dynamics with confirmation bias. *International Conference On Social Informatics*. pp. 144-158 (2022)
[78] Newman, M. Fast algorithm for detecting community structure in networks. *Physical Review E—Statistical, Nonlinear, And Soft Matter Physics*. **69**, 066133 (2004)
[79] Blondel, V., Guillaume, J., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal Of Statistical Mechanics: Theory And Experiment*. **2008**, P10008 (2008)
[80] Rosvall, M., Axelsson, D. & Bergstrom, C. The map equation. *The European Physical Journal Special Topics*. **178**, 13-23 (2009)
[81] Pons, P. & Latapy, M. Computing communities in large networks using random walks. *Computer And Information Sciences-ISCIS 2005: 20th International Symposium, Istanbul, Turkey, October 26-28, 2005. Proceedings 20*. pp. 284-293 (2005)
[82] Spohr, D. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*. **34**, 150-160 (2017)
[83] Törnberg, P. Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS One*. **13**, e0203958 (2018)
[84] Choi, D., Chun, S., Oh, H., Han, J. & Kwon, T. Rumor propagation is amplified by echo chambers in social media. *Scientific Reports*. **10**, 310 (2020)
[85] Metaxas, P. & Finn, S. The infamous# Pizzagate conspiracy theory: Insight from a TwitterTrails investigation. *Wellesley College Faculty Research And Scholarship*. **188** pp. 1-5 (2017)
[86] Asur, S., Huberman, B., Szabo, G. & Wang, C. Trends in social media: Persistence and decay. *Proceedings Of The International AAAI Conference On Web And Social Media*. **5**, 434-437 (2011)
[87] Colleoni, E., Rozza, A. & Arvidsson, A. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal Of Communication*. **64**, 317-332 (2014)
[88] Wollebæk, D., Karlsen, R., Steen-Johnsen, K. & Enjolras, B. Anger, fear, and echo chambers: The emotional basis for online behavior. *Social Media+ Society*. **5**, 2056305119829859 (2019)
[89] Amelkin, V., Bogdanov, P. & Singh, A. A distance measure for the analysis of polar opinion dynamics in

social networks. *ACM Transactions On Knowledge Discovery From Data (TKDD).* **13**, 1-34 (2019)
[90] Zhu, J., Ni, P., Tong, G., Wang, G. & Huang, J. Influence maximization problem with echo chamber effect in social network. *IEEE Transactions On Computational Social Systems.* **8**, 1163-1171 (2021)
[91] Garimella, K., Gionis, A., Parotsidis, N. & Tatti, N. Balancing information exposure in social networks. *Advances In Neural Information Processing Systems.* **30** (2017)
[92] Aslay, C., Matakos, A., Galbrun, E. & Gionis, A. Maximizing the diversity of exposure in a social network. *2018 IEEE International Conference On Data Mining (ICDM).* pp. 863-868 (2018)
[93] Tu, S., Aslay, C. & Gionis, A. Co-exposure maximization in online social networks. *Advances In Neural Information Processing Systems.* **33** pp. 3232-3243 (2020)
[94] Banerjee, P., Chen, W. & Lakshmanan, L. Mitigating filter bubbles under a competitive diffusion model. *Proceedings Of The ACM On Management Of Data.* **1**, 1-26 (2023)
[95] Zhang, B., Dai, G., Niu, F., Yin, N., Fan, X., Wang, S., Cao, X. & Huang, H. A survey of stance detection on social media: New directions and perspectives. *ArXiv Preprint ArXiv:2409.15690.* (2024)
[96] Zhang, C., Zhou, Z., Peng, X. & Xu, K. Doubleh: Twitter user stance detection via bipartite graph neural networks. *Proceedings Of The International AAAI Conference On Web And Social Media.* **18** pp. 1766-1778 (2024)
[97] Chen, J., Mang, Q., Zhou, H., Peng, R., Gao, Y. & Ma, C. Scalable Algorithm for Finding Balanced Subgraphs with Tolerance in Signed Networks. *Proceedings Of The 30th ACM SIGKDD Conference On Knowledge Discovery And Data Mining.* pp. 278-287 (2024)
[98] Bonchi, F., Galimberti, E., Gionis, A., Ordozgoiti, B. & Ruffo, G. Discovering polarized communities in signed networks. *Proceedings Of The 28th Acm International Conference On Information And Knowledge Management.* pp. 961-970 (2019)

A Survey on False Information Detection: From A Perspective of Propagation on Social Networks

Kun Xie
The Chinese University of Hong Kong
xiekun@se.cuhk.edu.hk

Sibo Wang
The Chinese University of Hong Kong
swang@se.cuhk.edu.hk

Abstract

The proliferation of false information in the digital age has become a pressing concern, necessitating the development of effective and robust detection methods. This paper offers a comprehensive review of existing false information detection techniques, approached from a novel perspective that emphasizes the propagation characteristics of misinformation. We introduce a new taxonomy that categorizes these methods into homogeneous and heterogeneous propagation-based approaches, providing a deeper understanding of the varying scopes and complexities involved in information dissemination. For each category, we present a formal problem formulation, review commonly used datasets, and summarize state-of-the-art methods. Additionally, we identify several promising directions for future research, including the creation of a unified benchmark suite, exploration of diverse information modalities, and development of innovative rumor debunking tasks. By systematically organizing the vast array of current techniques, this work offers a clear overview of the research landscape, aiding researchers and practitioners in navigating this complex field and inspiring further advancements.

1 Introduction

In the digital era, social media platforms facilitate both the rapid dissemination of information and the spread of false content, such as rumors and fake news. These can have severe consequences, including misleading the public [33], inciting panic [68], and influencing elections [1, 24]. Detecting and debunking false information early is crucial to mitigate its harmful impacts on society. The way information spreads through social networks can reveal its veracity [4, 45]. False information often exhibits distinct propagation patterns, such as rapid spread and specific dissemination strategies. By analyzing these propagation patterns and characteristics, researchers can identify false content more effectively and obtain deeper insights into how misinformation spreads.

This study seeks to offer a comprehensive understanding of propagation-based false information detection, addressing a gap left by existing surveys. Previous works [16, 21, 57, 63] primarily focus on text content-based rumor detection using general deep learning techniques. Other studies [35, 48] emphasize the fusion of multiple modalities, integrating text and images to enhance detection capabilities. Additionally, some research targets specific aspects of false information detection. For example, [31] concentrates on early detection, while [2] examines detection within particular populations. Despite these contributions, the propagation characteristics of false information on social networks remain underexplored. This perspective offers valuable insights that extend beyond traditional content analysis, highlighting the dynamic nature of misinformation spread and its implications for detection strategies. In recent years, several surveys [14, 32, 46, 56] have summarized methods for false information detection utilizing Graph Neural Networks (GNNs) [17, 29]. However, these surveys predominantly concentrate on GNN techniques, lacking a comprehensive review and analysis from the perspective of propagation information. While propagation information can indeed be modeled using graph topology, it is important to note that GNNs are not the sole technology employed in this domain. Architectures such as Transformers [3, 11, 66] and RNNs [8, 45] have also been extensively studied and applied. Consequently, focusing solely on GNN-based methods does not provide a complete overview of propagation-based approaches. Our work seeks to address this gap by offering a more holistic understanding of false information detection through the lens of propagation dynamics.

Specifically, we begin by reviewing the definition of false information and introducing the concept of information propagation on social networks. We categorize propagation based on its homogeneous or heterogeneous nature. Homogeneous propagation involves the dissemination of a source post through user interactions such as comments and retweets, forming a single-type propagation graph where all nodes represent posts. This type of propagation provides additional perspectives and debates that aid in assessing the veracity of the original news. In contrast, heterogeneous propagation encompasses a broader social context, including user metadata, semantic information, and cross-platform discussions. This rich context forms a heterogeneous social context graph with multiple node types, offering a comprehensive view of news propagation and enhancing veracity understanding. These two categories reflect different scopes and complexities of information propagation, each presenting unique insights and challenges for rumor detection.

Based on this categorization, we have reviewed false information detection methods that utilize both homogeneous and heterogeneous propagation. For each category, we first provide a formal problem formulation to help readers better understand the task. We then review the commonly used datasets and summarize existing methods. For methods based on homogeneous propagation, we categorize them from the perspective of propagation into three subcategories: 1) methods that model the dynamics of propagation, 2) methods that enhance robustness in the face of complex and noisy real-world propagation, and 3) methods that utilize Large Language Models (LLMs) to aid in understanding propagation. For methods based on heterogeneous propagation, we categorize them based on the types of social context they incorporate into two subcategories: 1) user-related context and 2) other context. Finally, we propose several feasible future research directions, including the development of an off-the-shelf unified benchmark suite and exploring alternative approaches to mitigate the impact of false information, to promote the advancement of this field. Figure 1 illustrates the taxonomy proposed in this paper.

The contributions of this work can be summarized as follows:

- *New Perspective*. This study reviews and synthesizes existing false information detection methods from the perspective of propagation, categorizing them into homogeneous and heterogeneous types. This approach provides insights into the varying scopes and complexities involved in information dissemination.

- *New Taxonomy*. We propose a novel taxonomy for propagation-based false information detection methods, initially dividing them based on their use of homogeneous or heterogeneous propagation. This taxonomy aids in systematically organizing the extensive array of existing methods.

- *Comprehensive Review and Summary*. This work offers a thorough and detailed review of existing methods for false information detection. For each propagation category, we present a formal problem formulation, review commonly used datasets, and summarize state-of-the-art methods. This comprehensive overview provides readers with a holistic understanding of the current research landscape in this field.

This paper is organized as follows: Section 2 introduces the concept of false information and commonly used terms in the field. Section 3 explores the characteristics and distinctions between homogeneous and heterogeneous propagation. Sections 4 and 5 provide comprehensive summaries of existing false

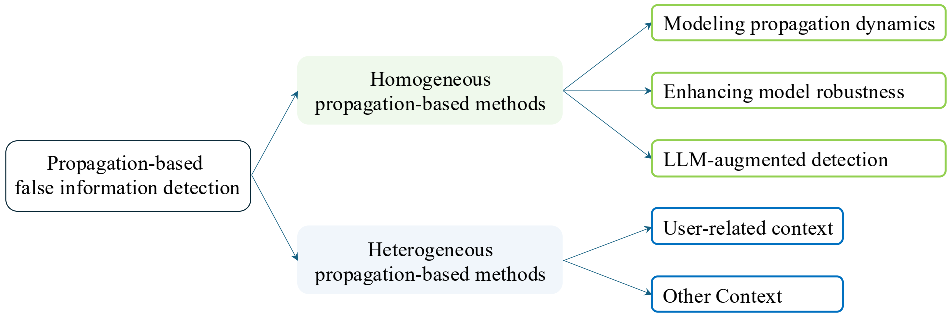


Figure 1: Taxonomy of propagation-based false information detection methods.

information detection methods based on homogeneous and heterogeneous propagation, respectively. Each section includes a formal problem formulation, a review of commonly used datasets, and an in-depth analysis of state-of-the-art methods. Finally, Section 6 discusses potential future research directions, highlighting opportunities for further advancement in this area.

2 False Information on Social Media

To comprehend false information on the web and social media, Kumar and Shah [30] systematically introduces this concept by categorizing it according to intent and knowledge content. Based on intent, false information can be classified as *misinformation*, which is created without the intention to mislead, and *disinformation*, which is deliberately produced to deceive and mislead readers [13]. From the perspective of knowledge content, false information can be divided into *opinion-based*, where a unique ground truth does not exist (as in the case of product reviews on e-commerce platforms), and *fact-based*, which involves falsehoods about entities that have a unique, verifiable ground truth [64].

In the field of false information detection, two terms are frequently encountered. The first term is *fake news* [57], which refers to deliberately false or verifiably inaccurate news content, falling within the category of disinformation under the aforementioned taxonomy of false information. The second term is *rumor*, defined as unverified and instrumentally relevant information statements in circulation [16, 43]. Unlike fake news, rumors may ultimately be proven true or false [93]. Despite their conceptual differences, these two terms are often used interchangeably in the context of social media false information detection, with their subtle distinctions being largely overlooked in previous studies [14]. Following this established practice in the literature, this survey reviews previous false information detection studies targeting both fake news and rumors.

False information can be disseminated through various types of media on social networks. In existing datasets, the common carriers include *micro-posts* on platforms such as Twitter[1] and sina Weibo[2], as well as published fake news *articles*. These carriers exhibit distinct characteristics: micro-posts are typically brief and colloquial, while news articles are longer and more structured, with separate components such as titles and content. While these structural differences are noteworthy, they are not the focus of our discussion. Instead, this paper emphasizes how false information propagates through social networks, regardless of its carrier type. In the next section, we present two primary approaches that existing works have employed to capture these propagation patterns.

[1] https://x.com/
[2] https://m.weibo.cn/

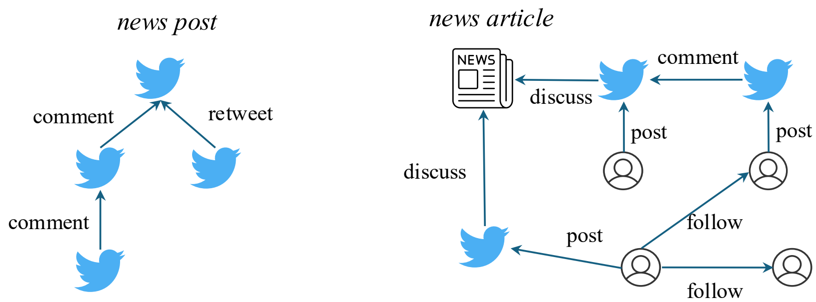


Figure 2: Illustration of homogeneous propagation (left) and heterogeneous propagation (right).

3 Information Propagation on Social Networks

The propagation of false information on social networks can be categorized into two distinct types based on the nature of propagation:

- **Homogeneous propagation** refers to the dissemination of a source post on social media, where other users interact with it through comments or retweets. These interactions, which may include expressing opinions on the source news or engaging in discussions with other comments, provide additional perspectives, supplementary information, and debates regarding the original news. The textual content of these retweets and comments offers further evidence for classifying the veracity of the original news. Moreover, the reply relationships between these discussions form a graph topology, linking the relevant texts together and further facilitating the prediction of veracity. We term this homogeneous propagation because it involves only a single type of propagation, forming a homogeneous propagation graph where all nodes represent posts on social media.

- **Heterogeneous propagation**, in contrast, encompasses the spread of information across broader social networks. Specifically, in addition to direct interactions with the original news, user metadata such as profiles and friendships, semantic information of the original news like keywords and domains, and related discussions of the news on other social platforms are all captured as social context. This rich social context provides a comprehensive view of the news propagation and discussion, enhancing the understanding of the news's veracity. We refer to this as heterogeneous propagation because it involves multiple types of propagation forms. This heterogeneous propagation can be modeled with a heterogeneous social context graph, where nodes represent not just social media posts but also news articles, cross-platform discussions, and user accounts, while the edges connect different types of nodes, reflecting explicit and implicit information flow on social networks.

These two categories reflect the different scopes and complexities of information propagation on social networks, each offering unique insights and challenges for rumor detection. Figure 2 illustrates these two types of propagation. The left sub-figure displays the homogeneous propagation of a post on social media, with commenting and retweeting interaction posts forming the homogeneous propagation tree. Conversely, the right sub-figure demonstrates the propagation of a news article across complex social networks, where the news article is connected to related discussion posts on other social media platforms, and the user profiles and user friendship network are also collected, together forming a heterogeneous social context graph.

Based on these two distinct propagation mechanisms, we proceed to provide a comprehensive analysis of each approach. In Section 4, we present a formal problem formulation for homogeneous propagation-based rumor detection, followed by an overview of commonly used datasets and a systematic review of existing methodologies. Similarly, Section 5 examines heterogeneous propagation-based false information detection, detailing its mathematical formulation, popular datasets, and a thorough survey of current research efforts.

4 Methods Based on Homogeneous Propagation

4.1 Problem Formulation

The task under this category is to classify a given post as either a rumor or a non-rumor, based on interactions occurring on the same social media platform, include commenting and retweeting the source post.

Formally, the propagation topology of the given source post can be modeled by a homogeneous graph, $G_p = (V_p, E_p, X_p)$, where the subscript p denotes the index of the source post. Specifically, the graph G_p is acyclic and tree-structured, as each comment or retweet is directed towards a single parent post. The node set V_p comprises all nodes in the graph G_p, with v_0 as the root representing the initial source post, and each v_i for $i \in \{1, \cdots, |V_p| - 1\}$ representing a comment or retweet post. The edge set E_p includes all edges in the graph G_p, where each edge (v_i, v_j) signifies the comment or retweet relationship between two posts, i.e., post v_j comments on or retweets post v_i. The set X_p contains the text content of all posts, where i-th row of X_p denotes the text attributes of the corresponding node v_i. The source post, along with its direct propagation graph, is referred to as an event [4], and the dynamics of commenting and retweeting are sometimes called conversation threads [45].

The objective is to predict the veracity of the given post using both the text content and the direct propagation information G_p. The definition of rumor veracity varies across datasets. For instance, in the Twitter15 [44] and Twitter16 [44] datasets, posts are categorized into *1)* non-rumors, *2)* false rumors, *3)* true rumors, and *4)* unverified rumors. Conversely, in other datasets like PHEME5 [90, 92] and PHEME9 [91], posts are classified as *1)* rumors and *2)* non-rumors.

In a typical supervised learning framework, a function $f : \mathcal{G} \to \mathcal{Y}$ is learned to predict veracity labels for posts with propagation graphs, where $\mathcal{G}$ represents the space of events, and $\mathcal{Y}$ denotes the space of veracity labels. The function f is trained using a labeled training set $\mathcal{G}_{train}$, where the ground-truth veracity label y_p is available for each $G_p \in \mathcal{G}_{train}$. The learned function f is then applied to classify unlabeled events in the test set $\mathcal{G}_{test}$. Most existing propagation-based rumor detection studies follow this supervised learning approach, although some employ self-supervised learning or transfer learning paradigms, as will be discussed later.

Figure 3 illustrates an example event from the Weibo22 dataset [87], showcasing how false information about a flood spread on the Sina Weibo platform. The post claims that Yichang City in China experienced severe flooding, allegedly due to heavy rainfall and increased water discharge from the Three Gorges Dam. As this information circulated on the platform, users engaged through reposts and comments, reflecting the propagation threads and patterns of the source post. Notably, the comment posts contributed valuable, diverse perspectives and discussions, playing a crucial role in collectively verifying the authenticity of the source information. For instance, local users reported only minor waterlogging in subsequent comments and called for thorough fact-checking. By abstracting both the source post and its associated comments/reposts as nodes, the entire propagation process can be represented as a homogeneous graph, as depicted on the right side of the figure.

4.2 Datasets

We have compiled a collection of datasets commonly utilized in existing studies on homogeneous propagation-based detection of false information. Detailed descriptions of these datasets are provided

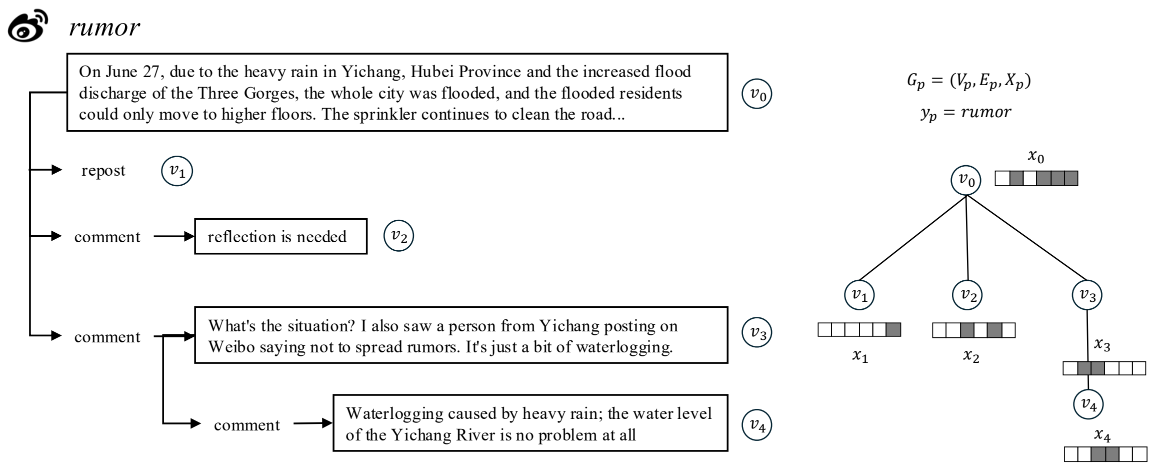


Figure 3: An example event from Weibo22 [87] containing false information related to a flood disaster, and the corresponding graph G_p modeling the propagation threads.

Table 1: Statistics of commonly used datasets, PHEME5 [90, 92], PHEME9 [91], Weibo-COVID19 [36], and Twitter-COVID19 [36], for direct propagation-based false information detection.

	PHEME5	PHEME9	Weibo-COVID19	Twitter-COVID19	Weibo	Weibo21	Weibo22
# Source posts	5,802	6,425	399	400	4,664	7,567	4,174
# Comment/retweet posts	103,212	105,354	26,687	406,185	3,805,656	184,775	961,962
Avg. # comment per event	18	16	67	1015	816	24	230
# Non-rumor	3,830	4,023	146	148	2,351	3,605	2,087
# Rumor	1,972	2,402	253	252	2,313	3,962	2,087

below:

- The **PHEME5** [90, 92] dataset was constructed by harvesting tweets from Twitter that pertain to newsworthy events with the potential to incite and propagate rumors. This dataset collected tweets based on five such events: *the Ferguson unrest, the Ottawa shooting, the Sydney crisis, the Charlie Hebdo shooting*, and *the Germanwings flight crash*. The collected tweets were then identified as rumors and non-rumors by journalists based on the source tweet contents and their associated comment threads.

- The **PHEME9** [91] dataset includes posts and propagation graphs derived from comments and reposts on Twitter. It encompasses tweets related to nine events, including breaking news likely to generate multiple rumors and known rumor events. Specifically, the dataset covers five sudden news events: *the Ferguson unrest, the Ottawa shooting, the Sydney crisis, the Charlie Hebdo shooting*, and *the Germanwings flight crash*. Additionally, it includes four specific rumors: *the Prince concert in Toronto, the Gurlitt collection, Putin's disappearance*, and *Michael Essien contracted Ebola*. Similar to PHEME5, journalists annotated the source posts for the collected tweets as either rumors or non-rumors, along with their propagation threads.

- The **Weibo** [43] dataset was constructed based on Sina Weibo, a large Chinese microblog platform. The dataset comprises a set of known rumors sourced from the Sina community management center[3], documenting various instances of false information. An equivalent number of non-rumor events were gathered by randomly crawling posts not reported as rumors.
- The **Weibo21** [50] dataset is a multi-domain fake news dataset curated based on the Sina Weibo platform. It includes rumor events collected from reports by the Weibo Community Management Center, while non-rumor events were collected randomly and further verified by NewsVerify[4]. All news items were annotated across nine domains: *Science, Military, Education, Disasters, Politics, Health, Finance, Entertainment,* and *Society* using crowd-sourcing.
- The **Weibo22** [87] dataset was designed to address the issue of incomplete and untraceable posts in the previous Weibo and Weibo21 datasets. Specifically, Weibo22 provides original comment text and comprehensive author profile details. Events in the dataset are categorized as rumors or non-rumors based on information from the Weibo Community Management Center and the China Internet Joint Rumor Debunking Platform[5].Over half of the posts in Weibo22 relate to the COVID-19 pandemic.
- The **Weibo-COVID19** and **Twitter-COVID19** [36] datasets were developed focusing on the COVID-19 pandemic. The Weibo-COVID19 dataset comprises rumor posts on Sina Weibo related to COVID-19, annotated by the Sina community management center, and non-rumor posts collected randomly. The Twitter-COVID19 dataset extends text-only claims from a COVID-19 rumor dataset [26] by incorporating propagation threads from Twitter. These datasets include micro posts in Chinese and English, respectively.
- The **Twitter15** and **Twitter16** [44] datasets extend previous text-only datasets by incorporating propagation trees. Twitter15 was constructed based on the framework proposed in [39], and Twitter16 was an extension of the dataset in [43]. Furthermore, the original binary labels (rumors and non-rumors) were refined to include four categories: non-rumors, false rumors, true rumors, and unverified rumors, in accordance with the veracity tags on articles from rumor debunking websites[6].

Tables 1 and 2 present common statistics for nine popular datasets. Specifically, Table 1 provides statistics for seven datasets where news posts are classified into rumors and non-rumors, while Table 2 displays statistics for two datasets with news posts classified into four categories: non-rumors, false rumors, true rumors, and unverified rumors. It is important to note that the statistics of these datasets may vary over time due to the dynamic nature of social networks, where users frequently delete posts or deactivate their accounts. Consequently, this can lead to inconsistencies in dataset statistics across different studies.

4.3 Method Review

In this section, we review related works that detect false information on social media based on direct propagation threads. We categorize existing methods into three categories: 1) methods that focus on modeling the dynamics of propagation, 2) methods that focus on improving the robustness of propagation-based detection, and 3) methods that utilize large language models (LLMs) to augment propagation. We elaborate on the related methods from these aspects in the following sections.

[3] `https://service.account.weibo.com/`
[4] `https://www.newsverify.com/`
[5] `https://www.piyao.org.cn/`
[6] `snopes.com` and `Emergent.info`

Table 2: Statistics of commonly used datasets, Twitter15 [44] and Twitter16 [44], for direct propagation-based false information detection.

	Twitter15	Twitter16
# Source posts	1,490	818
# Comment/retweet posts	331,612	204,820
Avg. # comments per event	223	251
# Non-rumor	374	205
# False rumor	370	205
# True rumor	372	205
# Unverified rumor	374	203

4.3.1 Dynamics of Propagation

RvNN [45] and BiGCN [4] model the dynamics of propagation threads through two distinct directions: the bottom-up direction, which models rumor dispersion, and the top-down direction, which models rumor propagation. In the bottom-up tree, responsive nodes always point to the nodes they are responding to, resembling a citation network where a response acts like a reference. Conversely, the top-down structure aligns with the natural flow of information, where a link from node v_i to v_j indicates that v_j reads the information from v_i and responds to it. This setup simulates how information cascades from the source tweet to all its receivers. Based on these two directed tree structures, RvNN and BiGCN utilize recursive neural models [8] and graph neural networks [29] for rumor representation and classification, respectively. PLAN [27] and RDLT [83] further extend these directed relationships to pairwise interactions using learnable attention mechanisms. Specifically, PLAN employs the attention mechanism in a transformer network to capture long-range relationships among comments, while RDLT uses it to enhance focus on long-tail comments. DDGCN [60] models the propagation dynamics with a graph convolution network that captures structural information at different time stages, and then combines the representations from different stages with a temporal fusion unit. CT-RvNN [55] enhances the propagation model by incorporating comment time intervals as edge weights and constructs a coding tree [34] over the weighted propagation graph to extract essential structural information.

4.3.2 Model Robustness

In real-world scenarios, social media posts are often accompanied by unreliable responses, such as malicious comments designed to enhance the credibility of false information, and irrelevant noise comments unrelated to the source post. These factors significantly interfere with the accurate detection of rumors [42, 78, 91]. Therefore, enhancing the robustness of rumor detectors against noise, perturbations, and adversarial attacks, and developing models that can maintain high predictive performance in complex and noisy social media environments, is a crucial research direction. EBGCN [71] models comment uncertainty by adaptively adjusting edge weights in propagation graphs from a probabilistic perspective, which are then utilized in the aggregation of GCNs. FGCN [72] introduces a neuro-fuzzy method based on fuzzy theory [52] to adapt edge weights according to predefined membership. DCE-RD [86] designs a subgraph generation strategy to produce diverse counterfactual evidence, employing a Determinantal Point Process (DPP) based loss to enhance evidence diversity. The generated evidence is aggregated to provide robust veracity prediction. GARD [62] enhances detection robustness by capturing semantic evolution information through self-supervised learning, conducting feature reconstruction from both local and global perspectives. EIN [22] addresses sensitivity to data quality by integrating epidemiological knowledge [6, 70], modeling propagation dynamics and user stance with susceptible states, and utilizing LLMs to annotate comment stances. Similarly, JSDRV [77] employs LLMs for post-level stance annotation and develops a reinforcement learning-based model for joint stance detection and rumor veracity prediction.

Contrastive learning, a self-supervised approach, ensures graphs maintain similar embeddings under different perturbations by emphasizing invariant features, thereby reducing sensitivity to noise and enhancing model robustness and generalization [7, 18]. This approach is also utilized to improve the robustness of rumor detectors. RDEA [19] pre-trains a GIN [75] encoder by maximizing mutual information between views under different perturbations, then fine-tunes the encoder for rumor veracity prediction. Node masking, edge dropping, and subgraph sampling operations are utilized for graph perturbation. RDCL [40] further proposes six perturbation methods under node-based and topology-based categories, maximizing consistency between two perturbed graphs of the same original graph, and minimizing the distance between perturbed and original graphs from the same class. A hard positive strategy also enhances the performance of the contrastive learning framework. GACL [61] utilizes contrastive learning to perceive differences between propagation graphs of the same and different classes, proposing an Adversarial Feature Transformation module to generate hard negative samples. CRFB [41] introduces a two-component (true-false) beta mixture model to distinguish true and false negative examples in contrastive learning, employing a CNN-based model to capture consistent and complementary information between two augmented propagation structures. RAGCL [9] observes the wide and shallow characteristics of propagation trees and proposes augmenting original graphs following three principles: exempting root nodes, retaining deep reply nodes, and preserving lower-level nodes in deep sections, based on which node centrality-based importance scores are incorporated to generate augmented views. FADE [85] improves representation quality with contrastive learning among augmented graphs under an adaptive strategy, mitigating event bias by subtracting event-only predictions.

Recent studies have demonstrated the effectiveness of adversarial training in improving robustness against noise and attacks [88]. AARD [59] considers both text content features and topology context when generating adversarial responses, utilizing the adversarial response generator to improve rumor detector robustness. GACL [61] also employs adversarial training to produce conflicting samples and hard negative samples, enhancing the effectiveness and robustness of the rumor detector.

Generation-based methods are also employed to enhance rumor detector robustness under certain conditions. GenFEND [51] focuses on alleviating exposure bias and enhancing comment diversity, adopting large language models (LLMs) as user simulators and comment generators. KPG [87] comprehensively considers both noisy propagation threads and spreading-limited rumors at early stages, proposing two interdependent modules to select key propagation graphs from generation-enhanced candidate graphs, utilizing a reinforcement learning framework to alternately update these modules. D^2 [74] aims to enhance early rumor detection by predicting possible diffusion paths based on limited propagation at early stages and users' social relationships.

Furthermore, some methods target improving transferability among rumors in different languages or topics. ACLR [36] overcomes domain and language restrictions via language alignment and a novel supervised contrastive training paradigm, developing an adversarial augmentation mechanism to enhance the robustness of low-resource rumor representation. To assess transferability under different test topics and languages, the Weibo-COVID19 dataset is utilized as the low-resource test set for evaluating proposed methods, while the Twitter15 and Twitter16 datasets serve as the well-resourced training set. Similarly, the Twitter-COVID19 dataset is used as the low-resource test set, and the Weibo dataset is employed as the well-resourced training set. T^3RD [84] introduces test-time self-supervised learning to enhance rumor detection performance on low-resource datasets, performing graph- and node-level contrastive learning as auxiliary tasks, and introducing feature alignment constraints to balance knowledge derived from the training set and test samples. FNDCD [15] employs a reweighting strategy based on classification confidence and propagation structure regularization to reduce domain-specific biases, enhancing the detection of unseen fake news with new topics and domains.

4.3.3 LLM-augmented Detection

In recent years, there has been a growing interest in leveraging Large Language Models (LLMs) for rumor detection. Despite their remarkable capabilities in text understanding and generation, LLMs face significant challenges when dealing with complex structured information, particularly in effectively processing intricate topologies. Specifically, LeRuD [38] designs specific prompts to guide the model's focus on crucial clues such as writing styles, commonsense errors, and rebuttals or conflicts in comments. It also divides the entire propagation information into a Chain-of-Propagation to alleviate the cognitive burden on LLMs. ARG [20] prompts the LLM to generate multi-perspective rationales for rumors and utilizes smaller language models, such as BERT [11], to integrate and infer rumor veracity from these rationales. SePro [82] employs Graph Attention Networks [37] and community detection methods [53] to identify and extract key subgraphs from the complex information network, thereby simplifying the structure that the LLM needs to process.

5 Methods Based on Heterogeneous propagation

5.1 Problem Formulation

As introduced in Section 4.1, the task of rumor detection based on homogeneous propagation information involves predicting the veracity of the source post based on the commenting/retweeting interactions that form a propagation tree, making it a graph-level classification problem. In contrast, rumor detection with heterogeneous propagation incorporates a broader spectrum of information collected from social networks. This approach can be viewed as a node classification problem on heterogeneous graphs, where different types of nodes and edges represent various elements of the social context. Specifically, the social context typically includes: 1) users who published the source, comment, or retweet posts; 2) keywords extracted from the text content of the source news; and 3) semantically related posts discussing the same news event on other social platforms. To integrate all this information, a heterogeneous graph is constructed, with different types of nodes and edges representing the complex relationships within the social context.

Unlike rumor detection based on direct propagation information, which primarily focuses on the structure and dynamics of information spread, approaches with social context-based propagation introduce a higher level of complexity. Different methods in this category often adopt varied problem definitions, focusing on distinct aspects of the social context. To facilitate a comprehensive understanding, we abstract a foundational problem formulation that serves as a common ground for different variants of the problem.

Specifically, let S denote the set of news posts, T denote the set of comment/retweet posts interacting with the news posts in S, and U denote the set of users on the social network. The set $\mathcal{U}$ includes users who directly interacted with the news posts (i.e., by publishing the news posts or replying to them with comments/retweets), as well as users indirectly associated with the news posts (i.e., having a friend relationship with direct users). These sets collectively form the different types of nodes in a heterogeneous graph, $V = S \cup T \cup U$. In this heterogeneous graph, the edges encompass various types of relationships, denoted by the edge set E. Specifically, edges between nodes in S and T indicate commenting or retweeting interactions. Edges between nodes in U and S represent users publishing news posts $s_i \in S$, or commenting on/retweeting news posts $s_i \in S$, while edges between $u_i \in U$ and $t_i \in T$ represent u_i posting the comment/retweet post t_i. Additionally, edges between nodes in U capture social relationships, such as interactions or friendships. Thus, the heterogeneous graph can be represented as $G = (V, E)$. The goal of this task is to learn a prediction function $f : \mathcal{S} \times \mathcal{G} \to \mathcal{Y}$, which classifies the news posts $s_i \in S$ into rumors (fake news) or non-rumors (real news) along with the heterogeneous social context propagation graph G, where $\mathcal{Y}$ is the space of veracity labels. The learning process is based on a set of labeled news posts $\mathcal{S}_{train}$, and the learned function f is expected to accurately predict the veracity label of unlabeled news posts in $\mathcal{S}_{test}$: $\hat{y}_i = f(s_i, G)$. In some existing methods, only a subset of the node/edge types in the heterogeneous graph G are utilized to enhance the effectiveness of rumor detection, while other approaches employ additional information to refine the types and relationships within G.

5.2 Datasets

In this section, we introduce the common datasets used by false information detection methods that leverage heterogeneous propagation. In addition to the Twitter15, Twitter16, and Weibo datasets introduced in Section 4.2, the following datasets are also utilized:

- **Fakeddit** [49] is a large-scale benchmark dataset from Reddit[7] that includes both text and image content. This dataset supports news classification at different granularities, including 2-class, 3-class, and 6-class classifications. Specifically, the 2-way classification determines whether a sample is fake or true. The 3-way classification distinguishes between completely true samples, direct quotes from propaganda posters, and fake samples. The 6-way classification includes the following labels: *True, Satire/Parody, Misleading Content, Imposter Content, False Connection*, and *Manipulated Content*.

- **GossipCop** [58] collects news from GossipCop[8], a website that provides rating scores for entertainment stories on a scale from 0 to 10, reflecting the degree from fake to real. The dataset includes stories with scores under 5 as fake news, and collects related discussion posts on Twitter, along with their commenting/retweeting interactions, engaged user profiles, user friendship information, and spatial locations provided in user profiles.

- **PolitiFact** [58] gathers news content based on annotations from PolitiFact[9], a fact-checking website where political news is evaluated as fake or real by journalists and domain experts. Similar to the GossipCop dataset, the PolitiFact dataset also includes semantically related tweet posts on Twitter, along with rich propagation information on social contexts, including both comment/retweet interactions and comprehensive user information associated with the posts.

- **FakeHealth** [10] is collected from the healthcare information review website, Health News Review[10], and contains over 2000 news articles, 500k posts, and 27k user profiles, along with user networks.

- **Mc-Fake** [47] comprises labeled fake news from existing datasets and real news from reliable sources, covering five topics: *Politics, Entertainment, Health, Covid-19*, and *Syria War*. Related tweets, retweets, and replies, along with the corresponding users on the Twitter platform, are retrieved, forming a dataset with rich social contexts.

[7] https://www.reddit.com/
[8] https://www.gossipcop.com/
[9] https://www.politifact.com/
[10] https://www.healthnewsreview.org/

5.3 Method Review

In this section, we compile and review existing methods for false information detection that leverage heterogeneous propagation information on social networks. Unlike the methods discussed in Section 4, which focus on homogeneous propagation patterns, the methods reviewed here incorporate a broader range of contextual information available on social networks. Specifically, we classify these methods into two main categories based on the types of additional context they utilize: *1)* user-related context and *2)* other context.

5.3.1 User-Related Context

In this category, methods collect information about engaged users as social context, providing complementary evidence to assist in detecting false information. The user context not only provides connections across different news events, implying implicit information dissemination, but also reflects user preferences through user profiles and friendship relationships, offering useful clues for veracity classification. Specifically, GLAN [80] integrates user information to model local and global propagation on social networks using a heterogeneous graph, involving news posts, commenting posts, and the users publishing these posts. An attention mechanism is employed to enhance the representation of news posts by fusing the semantic information in the constructed graph. SMAN [81] further categorizes users into two types: 1) publishers, the authors of news posts, and 2) comment users who post comments on news posts. Additionally, SMAN annotates the credibility of a subset of users based on their historical publishing and reposting behaviors as supervised information, and jointly predicts rumor veracity and user credibility. FANG [54] follows this user type refinement and proposes an inductive graph learning framework [17] that produces representations effectively and efficiently. UPFD [12] is inspired by the confirmation bias theory, which suggests that the probability of users spreading fake news is related to their preferences, which can be extracted from users' historical posts. Specifically, UPFD models the target news and user historical post contents, extracts user-user interaction information, and fuses these endogenous preferences and exogenous context together. DUCK [65] extracts information from the user-user interaction network, the post-post commenting network, and conversation threads using Graph Attention Networks (GAT) [67], BERT [11] combined with GAT, and transformers [3]. PSIN [47] applies a divide-and-conquer strategy to model the heterogeneous relations and employs three variants of Graph Attention Networks to model post-post, user-user, and post-user subgraphs. An additional adversarial topic discriminator is adopted to learn topic-agnostic features for enhanced veracity prediction. DECOR [73] computes the number of common engaged users between news articles to form a news engagement graph. DECOR further reveals that the degree distributions of fake and real news in this engagement graph show a clear difference, and proposes a Degree-Corrected Stochastic Block model to enhance fake news detection. PSGT [89] focuses on user-user interaction relationships, proposing a noise-reduction self-attention mechanism based on the information bottleneck principle and a novel relational propagation graph as a position encoding for the graph Transformer.

Special learning settings, such as adversarial attack and transfer learning, are also explored. For instance, QSA-AC [69] considers attacking the user-news engagement network by injecting bots into the network to obfuscate GNN-based fake news detectors. QSA-AC follows a surrogate-based approach, leveraging the query process to enhance the effectiveness of the attack. Meanwhile, MMHT [79] utilizes user engagement and historical preference information for cross-domain fake news detection. Specifically, MMHT disentangles news content and user engagement from a macro perspective and disentangles veracity-relevant and veracity-irrelevant features from a micro perspective. These features are then utilized to facilitate more effective knowledge transfer.

5.3.2 Other Context

In addition to leveraging user information as social context, methods in this category broaden the scope of social context by incorporating various types of information. For instance, HDGCN [25] collects the temporal and domain information of news articles and proposes a neighbor sampling strategy along with a novel hierarchical attention mechanism to enhance news embedding and classification. DiHAN [5] constructs a heterogeneous graph that includes the core subjects of the news and their temporal information. It extracts meta-paths considering both node/edge types and the chronological interactions between news items, employing a hierarchical attention mechanism to integrate different types of meta-path-based temporal information. Temporal information is also utilized in DWAN [28] for train-test splitting, introducing a new approach where the model is trained on earlier news and tested on subsequent news. DWAN improves detection performance under this temporality-aware setting by using a Graph Structure Learning framework to reassign weights to existing edges based on a learned edge weight estimator. FinerFact [23] mines social context information through semantic relations to enable fine-grained reasoning. Specifically, FinerFact extracts keywords from news articles and posts on other social media platforms, using them as connectors between news, discussion posts, and the users who publish these posts. This forms a heterogeneous graph rich in social contextual information. The credibility of users is combined with the graph topology to reason about the veracity of rumors. Following this graph construction approach, SureFact [76] proposes a reinforced subgraph generation method and develops a Hierarchical Path-aware Kernel Graph Attention Network to reason about rumor veracity based on the generated subgraphs. This subgraph reasoning paradigm not only provides clear interpretability of the prediction results but also enhances generalization.

6 Future Work

6.1 Development of a Comprehensive Benchmark for False Information Detection

The field of false information detection currently lacks a practical, off-the-shelf benchmark that can standardize research and facilitate direct comparisons. This deficiency is particularly evident in three critical aspects: unified train-validation-test splitting, standardized evaluation process, and unified data platform.

Firstly, in existing studies, the train-test split is often conducted independently by different researchers, leading to non-comparable outcomes across studies and inconsistency with real-world scenarios. To address these challenges, we propose the development of a benchmark that includes commonly used datasets and provides a unified train-validation-test split. This split should be based on meaningful dimensions, such as topic or time, to better simulate real-world conditions and assess the robustness and generalizability of detection methods. For instance, splitting the dataset by topic or time stamp, with the training and test sets containing entirely different topics, can effectively evaluate a method's ability to generalize across diverse subjects. This approach also allows researchers to analyze how generalization performance varies with different topics.

Secondly, another critical gap in current research is the lack of a unified evaluation process. Different methods often employ distinct evaluation codes and adopt various metrics, making it difficult to compare their performance directly. A standardized benchmark should include a consistent evaluation protocol, with clearly defined metrics and evaluation scripts. This will ensure that all methods are assessed under the same conditions, facilitating fair comparisons and promoting the development of more effective detection techniques.

Thirdly, the dispersion of datasets across different studies also poses challenges for researchers. A comprehensive, off-the-shelf benchmark platform that integrates multiple datasets would significantly facilitate research in this area. Such a platform would provide a one-stop solution for researchers to access and work with standardized datasets, reducing the time and effort required to prepare data and allowing them to focus more on developing and testing new methods. This centralized approach would also promote collaboration and reproducibility, driving the field forward more efficiently.

6.2 Diversifying Information Carriers for False Information Detection

Current propagation-based false information detection primarily focuses on news or microblogs, where text is the dominant information carrier, occasionally supplemented by images. However, with the rapid growth of short video platforms, false information disseminated in video format has become a significant concern. The detection of false information in video content and the construction of relevant datasets should be prioritized in future research.

Short videos offer a unique set of challenges and opportunities for false information detection. Unlike text, which can be analyzed through natural language processing techniques, video content involves multiple modalities, including visual, auditory, and textual elements. Developing methods that can effectively integrate and analyze these multimodal features will be crucial for accurate detection. Additionally, creating datasets that capture the nuances of false information spread through videos will provide a valuable resource for researchers to develop and validate new detection algorithms. This diversification of information carriers will not only address the evolving landscape of information dissemination but also enhance the robustness and applicability of false information detection techniques.

6.3 Mitigating the Impact of False Information

Detecting and removing false information from social platforms (e.g., by deleting posts) is a straightforward approach to mitigate the impact of false information. However, this method has several drawbacks. First, it may mistakenly remove content that is not false information, thereby affecting user experience. Second, it does not address the root causes of why people believe and propagate false information. Finally, this method heavily relies on platform operations and may not comprehensively solve the problem of false information. Therefore, how to effectively reduce the impact of potentially false information on social networks is a critical issue that requires in-depth research.

One potential solution is to utilize bots on social platforms to disseminate corrective information. This approach can automatically push accurate information to users once false information is detected, thereby mitigating the spread of misinformation. However, several key aspects need to be considered for this method to be effective. The content of corrective information needs to be based on both the original information and the propagated information. This is because the propagated information reflects the community's understanding of the original message, and generating corrective posts from this understanding can increase the likelihood of acceptance. Furthermore, the dissemination strategy for corrective information should be precisely tailored by integrating the propagation information of the original post and relevant social context information.

Despite its potential, this approach currently faces many challenges. The problem definition and evaluation of such a correction mechanism are still unclear. For example, how to measure the effectiveness of corrective information and determine its role in reducing the spread of false information remains an open question. Moreover, there is a lack of high-quality datasets for researching this issue. Existing datasets mostly focus on the detection of false information, while data related to the generation, dissemination, and user feedback of corrective information are relatively scarce. This makes it difficult for researchers to systematically study and validate relevant methods.

7 Conclusion

This study systematically reviews existing false information detection methods through the lens of information propagation on social networks. We begin by categorizing information dissemination into homogeneous and heterogeneous propagation. Based on this categorization, we introduce a comprehensive taxonomy that encompasses existing detection methods. We then provide a detailed overview and summary of related methods within these two categories, covering formal problem formulations, popular datasets, and detailed approaches, with an emphasis on the utilization of propagation dynamics and social context. Furthermore, we identify several promising avenues for future research, including the development of standardized evaluation frameworks, innovative strategies to mitigate the impact of false information, and the exploration of diverse information carriers.

Acknowledgements

This work is supported by the RGC GRF grant (No. 14217322) and the Tencent WeChat Rhino-Bird Focused Research Program.

References

[1] Jonathan Bakdash, Char Sample, Monica Rankin, Murat Kantarcioglu, Jennifer Holmes, Sue Kase, Erin Zaroukian, and Boleslaw Szymanski. The future of deception: Machine-generated and manipulated images, video, and audio? In *2018 International Workshop on Social Sensing (SocialSens)*, pages 2–2. IEEE, 2018.

[2] Fabiana Battista, Tiziana Lanciano, and Antonietta Curci. A survey on the criteria used to judge (fake) news in italian population. *Brain and Behavior*, 15(2):e70315, 2025.

[3] Iz Beltagy, Matthew E Peters, and Arman Cohan. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*, 2020.

[4] Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. Rumor detection on social media with bi-directional graph convolutional networks. In *AAAI*, pages 549–556, 2020.

[5] Ya-Ting Chang, Zhibo Hu, Xiaoyu Li, Shuiqiao Yang, Jiaojiao Jiang, and Nan Sun. Dihan: A novel dynamic hierarchical graph attention network for fake news detection. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 197–206, 2024.

[6] You Chang and Mart CM de Jong. A novel method to jointly estimate transmission rate and decay rate parameters in environmental transmission models. *Epidemics*, 42:100672, 2023.

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

[8] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[9] Chaoqun Cui and Caiyan Jia. Propagation tree is not deep: Adaptive graph contrastive learning approach for rumor detection. In *AAAI*, pages 73–81, 2024.

[10] Enyan Dai, Yiwei Sun, and Suhang Wang. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 853–862, 2020.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

[12] Yingtong Dou, Kai Shu, Congying Xia, Philip S Yu, and Lichao Sun. User preference-aware fake news detection. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 2051–2055, 2021.

[13] Don Fallis. A functional analysis of disinformation. *IConference 2014 Proceedings*, 2014.

[14] Shuzhi Gong, Richard O Sinnott, Jianzhong Qi, and Cecile Paris. Fake news detection through graph-based neural networks: A survey. *arXiv preprint arXiv:2307.12639*, 2023.

[15] Shuzhi Gong, Richard Sinnott, Jianzhong Qi, and Cecile Paris. Unseen fake news detection through casual debiasing. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 981–985, 2025.

[16] Bin Guo, Yasan Ding, Lina Yao, Yunji Liang, and Zhiwen Yu. The future of false information detection on social media: New perspectives and trends. *ACM Computing Surveys (CSUR)*, 53(4):1–36, 2020.

[17] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

[18] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[19] Zhenyu He, Ce Li, Fan Zhou, and Yi Yang. Rumor detection on social media with event augmentations. In *SIGIR*, pages 2020–2024, 2021.

[20] Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 22105–22113, 2024.

[21] Fiza Gulzar Hussain, Muhammad Wasim, Seemab Hameed, Abdur Rehman, Muhammad Nabeel Asim, and Andreas Dengel. Fake news detection landscape: Datasets, data modalities, ai approaches, their challenges, and future perspectives. *IEEE Access*, 2025.

[22] Wei Jiang, Tong Chen, Xinyi Gao, Wentao Zhang, Lizhen Cui, and Hongzhi Yin. Epidemiology-informed network for robust rumor detection. In *Proceedings of the ACM on Web Conference 2025*, pages 3618–3627, 2025.

[23] Yiqiao Jin, Xiting Wang, Ruichao Yang, Yizhou Sun, Wei Wang, Hao Liao, and Xing Xie. Towards fine-grained reasoning for fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5746–5754, 2022.

[24] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, Yu Wang, and Jiebo Luo. Detection and analysis of 2016 us presidential election related rumors on twitter. In *Social, Cultural, and Behavioral Modeling: 10th International Conference, SBP-BRiMS 2017, Washington, DC, USA, July 5-8, 2017, Proceedings 10*, pages 14–24. Springer, 2017.

[25] Zhezhou Kang, Yanan Cao, Yanmin Shang, Tao Liang, Hengzhu Tang, and Lingling Tong. Fake news detection with heterogenous deep graph convolutional network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 408–420. Springer, 2021.

[26] Debanjana Kar, Mohit Bhardwaj, Suranjana Samanta, and Amar Prakash Azad. No rumours please! a multi-indic-lingual approach for covid fake-tweet detection. In *2021 grace hopper celebration India (GHCI)*, pages 1–5. IEEE, 2021.

[27] Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8783–8790, 2020.

[28] Junghoon Kim, Junmo Lee, Yeonjun In, Kanghoon Yoon, and Chanyoung Park. Revisiting fake news detection: Towards temporality-aware evaluation by leveraging engagement earliness. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pages 559–567, 2025.

[29] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[30] Srijan Kumar and Neil Shah. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559*, 2018.

[31] Lazarus Kwao, Yang Yang, Jie Zou, and Jing Ma. A survey of approaches to early rumor detection on microblogging platforms: Computational and socio-psychological insights. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 15(1):e70001, 2025.

[32] Batool Lakzaei, Mostafa Haghir Chehreghani, and Alireza Bagheri. Disinformation detection using graph neural networks: a survey. *Artificial Intelligence Review*, 57(3):52, 2024.

[33] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer,

Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[34] Angsheng Li and Yicheng Pan. Structural information and dynamical complexity of networks. *IEEE Transactions on Information Theory*, 62(6):3290–3339, 2016.

[35] Xianghua Li, Jiao Qiao, Shu Yin, Lianwei Wu, Chao Gao, Zhen Wang, and Xuelong Li. A survey of multimodal fake news detection: A cross-modal interaction perspective. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2025.

[36] Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Guang Chen. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. *arXiv preprint arXiv:2204.08143*, 2022.

[37] Qiang Liu, Feng Yu, Shu Wu, and Liang Wang. Mining significant microblogs for misinformation identification: an attention-based approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(5):1–20, 2018.

[38] Qiang Liu, Xiang Tao, Junfei Wu, Shu Wu, and Liang Wang. Can large language models detect rumors on social media? *arXiv preprint arXiv:2402.03916*, 2024.

[39] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1867–1870, 2015.

[40] Guanghui Ma, Chunming Hu, Ling Ge, Junfan Chen, Hong Zhang, and Richong Zhang. Towards robust false information detection on social networks with contrastive learning. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pages 1441–1450, 2022.

[41] Jiachen Ma, Jing Dai, Yong Liu, Meng Han, and Chunyu Ai. Contrastive learning for rumor detection via fitting beta mixture model. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 4160–4164, 2023.

[42] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1751–1754, 2015.

[43] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. 2016.

[44] Jing Ma, Wei Gao, and Kam-Fai Wong. Detect rumors in microblog posts using propagation structure via kernel learning. Association for Computational Linguistics, 2017.

[45] Jing Ma, Wei Gao, and Kam-Fai Wong. Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics, 2018.

[46] Alaa Safaa Mahdi and Narjis Mezaal Shati. A survey on fake news detection in social media using graph neural networks. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, 16(2):23–41, 2024.

[47] Erxue Min, Yu Rong, Yatao Bian, Tingyang Xu, Peilin Zhao, Junzhou Huang, and Sophia Ananiadou. Divide-and-conquer: Post-user interaction network for fake news detection on social media. In *Proceedings of the ACM web conference 2022*, pages 1148–1158, 2022.

[48] Mohamed Mostafa, Ahmad S Almogren, Muhammad Al-Qurishi, and Majed Alrubaian. Modality deep-learning frameworks for fake news detection on social networks: a systematic literature review. *ACM Computing Surveys*, 57(3):1–50, 2024.

[49] Kai Nakamura, Sharon Levy, and William Yang Wang. r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. *arXiv preprint arXiv:1911.03854*, 2019.

[50] Qiong Nan, Juan Cao, Yongchun Zhu, Yanyan Wang, and Jintao Li. Mdfend: Multi-domain fake news detection. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 3343–3347, 2021.

[51] Qiong Nan, Qiang Sheng, Juan Cao, Beizhe Hu, Danding Wang, and Jintao Li. Let silence speak: Enhancing fake news detection with generated comments from large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 1732–1742, 2024.

[52] CV Negoita and DA Ralescu. On fuzzy optimization. *Kybernetes*, 6(3):193–195, 1977.

[53] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.

[54] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. Fang: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 1165–1174, 2020.

[55] Xingyu Peng, Junran Wu, Ruomei Liu, and Ke Xu. Rumor detection on social media with temporal propagation structure optimization. *arXiv preprint arXiv:2412.08316*, 2024.

[56] Huyen Trang Phan, Ngoc Thanh Nguyen, and Dosam Hwang. Fake news detection: A survey of graph neural network methods. *Applied Soft Computing*, 139:110235, 2023.

[57] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36, 2017.

[58] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188, 2020.

[59] Yun-Zhu Song, Yi-Syuan Chen, Yi-Ting Chang, Shao-Yu Weng, and Hong-Han Shuai. Adversary-aware rumor detection. In *ACL/IJCNLP*, pages 1371–1382, 2021.

[60] Mengzhu Sun, Xi Zhang, Jiaqi Zheng, and Guixiang Ma. Ddgcn: Dual dynamic graph convolutional networks for rumor detection on social media. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 4611–4619, 2022.

[61] Tiening Sun, Zhong Qian, Sujun Dong, Peifeng Li, and Qiaoming Zhu. Rumor detection on social media with graph adversarial contrastive learning. In *WWW*, pages 2789–2797, 2022.

[62] Xiang Tao, Liang Wang, Qiang Liu, Shu Wu, and Liang Wang. Semantic evolvement enhanced graph autoencoder for rumor detection. In *WWW*, pages 4150–4159, 2024.

[63] Gabrielle Taylor, Wenting Jiang, Xiao Qin, Ashish Gupta, et al. Misinformation detection: A survey of ai techniques and research opportunities. *Foundations and Trends® in Information Systems*, 8(2):66–147, 2024.

[64] Jeffrey E Thomas. Statements of fact, statements of opinion, and the first amendment. *Calif. L. Rev.*, 74: 1001, 1986.

[65] Lin Tian, Xiuzhen Jenny Zhang, and Jey Han Lau. Duck: Rumour detection on social media by modelling user and comment propagation networks. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4939–4949, 2022.

[66] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[67] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1079–1088, 2010.

[68] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *science*, 359(6380): 1146–1151, 2018.

[69] Lanjun Wang, Zehao Wang, Le Wu, and An-An Liu. Bots shield fake news: Adversarial attack on user engagement based fake news detection. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 2369–2378, 2024.

[70] Rong-Hua Wang, Zhen Jin, Quan-Xing Liu, Johan van de Koppel, and David Alonso. A simple stochastic model with environmental transmission explains multi-year periodicity in outbreaks of avian flu. *PloS one*, 7 (2):e28873, 2012.

[71] Lingwei Wei, Dou Hu, Wei Zhou, Zhaojuan Yue, and Songlin Hu. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection. In *ACL/IJCNLP (1)*, pages 3845–3854, 2021.

[72] Lingwei Wei, Dou Hu, Wei Zhou, Xin Wang, and Songlin Hu. Modeling the uncertainty of information propagation for rumor detection: A neuro-fuzzy approach. *IEEE Trans. Neural Networks Learn. Syst.*, 35 (2):2522–2533, 2024.

[73] Jiaying Wu and Bryan Hooi. Decor: Degree-corrected social graph refinement for fake news detection. In *SIGKDD*, pages 2582–2593, 2023.

[74] Haowei Xu, Chao Gao, Xianghua Li, and Zhen Wang. D2: Customizing two-stage graph neural networks

for early rumor detection through cascade diffusion prediction. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pages 568–576, 2025.

[75] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

[76] Ruichao Yang, Xiting Wang, Yiqiao Jin, Chaozhuo Li, Jianxun Lian, and Xing Xie. Reinforcement subgraph reasoning for fake news detection. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2253–2262, 2022.

[77] Ruichao Yang, Wei Gao, Jing Ma, Hongzhan Lin, and Bo Wang. Reinforcement tuning for detecting stances and debunking rumors jointly with large language models. *arXiv preprint arXiv:2406.02143*, 2024.

[78] Xiaoyu Yang, Yuefei Lyu, Tian Tian, Yifei Liu, Yudong Liu, and Xi Zhang. Rumor detection on social media with graph structured adversarial learning. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 1417–1423, 2021.

[79] Xuankai Yang, Yan Wang, Xiuzhen Zhang, Shoujin Wang, Huaxiong Wang, and Kwok Yan Lam. A macro- and micro-hierarchical transfer learning framework for cross-domain fake news detection. In *Proceedings of the ACM on Web Conference 2025*, pages 5297–5307, 2025.

[80] Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, and Songlin Hu. Jointly embedding the local and global relations of heterogeneous graph for rumor detection. In *2019 IEEE international conference on data mining (ICDM)*, pages 796–805. IEEE, 2019.

[81] Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, and Songlin Hu. Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning. *arXiv preprint arXiv:2012.04233*, 2020.

[82] Yirong Zeng, Xiao Ding, Bibo Cai, Ting Liu, and Bing Qin. Exploring large language models for effective rumor detection on social media. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2537–2552, 2025.

[83] Guixian Zhang, Rongjiao Liang, Zhongyi Yu, and Shichao Zhang. Rumour detection on social media with long-tail strategy. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2022.

[84] Huaiwen Zhang, Xinxin Liu, Qing Yang, Yang Yang, Fan Qi, Shengsheng Qian, and Changsheng Xu. T3rd: Test-time training for rumor detection on social media. In *Proceedings of the ACM Web Conference 2024*, pages 2407–2416, 2024.

[85] Jiajun Zhang, Zhixun Li, Qiang Liu, Shu Wu, Zilei Wang, and Liang Wang. Evolving to the future: Unseen event adaptive fake news detection on social media. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 4273–4277, 2024.

[86] Kaiwei Zhang, Junchi Yu, Haichao Shi, Jian Liang, and Xiaoyu Zhang. Rumor detection with diverse counterfactual evidence. In *SIGKDD*, pages 3321–3331, 2023.

[87] Yusong Zhang, Kun Xie, Xingyi Zhang, Xiangyu Dong, and Sibo Wang. Rumor detection on social media with reinforcement learning-based key propagation graph generator. In *Proceedings of the ACM on Web Conference 2025*, pages 2742–2753, 2025.

[88] Mengnan Zhao, Lihe Zhang, Jingwen Ye, Huchuan Lu, Baocai Yin, and Xinchao Wang. Adversarial training: A survey. *arXiv preprint arXiv:2410.15042*, 2024.

[89] Junyou Zhu, Chao Gao, Ze Yin, Xianghua Li, and Jürgen Kurths. Propagation structure-aware graph transformer for robust and interpretable fake news detection. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4652–4663, 2024.

[90] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. Learning reporting dynamics during breaking news for rumour detection in social media. arxiv 2016. *arXiv preprint arXiv:1610.07363*, 2016.

[91] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3): e0150989, 2016.

[92] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. Exploiting context for rumour detection in social media. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I 9*, pages 109–123. Springer, 2017.

[93] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. Detection and resolution of rumours in social media: A survey. *Acm Computing Surveys (Csur)*, 51(2):1–36, 2018.

Lightweight Influence-Aware News Recommendation in Social Media

Yuting Feng[1], Bogdan Cautis[2]

[1]Department of Electrical and Computer Engineering (ECE), National University of Singapore
[2]Infocomm Technology Cluster, Singapore Institute of Technology

Abstract

News recommendation systems are generally based on the semantic content of news items and user profiles, whereas the underlying recommendation scenario is ignored. With the increasing popularity of social media as pathways to news, making personalized news recommendations in specific social scenarios deserves attention, where the information diffusion patterns and influence mechanisms therein are exploited. We consider in this paper a diffusion and influence-aware perspective on the news recommendation problem, and we propose a *lightweight deep learning* approach for it, called LISM. This approach targets news recommendation in micro-blogging platforms, such as Twitter or Weibo, whose extreme *data velocity* demands a satisfactory trade-off between the model's complexity and its effectiveness. LISM is a content-based deep recommendation model that seeks to represent news and users, while also taking into consideration the observed news diffusions (cascades). We use graph embeddings – node representations that are indicative of news diffusion patterns – leading to valuable social-related information for recommendations. To merge the semantics and social-related representations of news, a specially designed convolutional neural network for joint feature representation (SCNN) is used as the news encoder, while an attention model automatically aggregates the different interests of users. We conduct experiments on two real-world datasets, showing that LISM outperforms the state-of-the-art news recommendation approaches, while exhibiting a good complexity vs. effectiveness tradeoff.

1 Introduction

The tremendous growth of the Web and e-commerce has led to the development of recommender systems, benefiting from a rich research literature in recent years [40]. Many techniques have been proposed in various application scenarios, from traditional collaborative filtering to deep learning-based approaches. In essence, they reason about the relationships between users and items (such as music, books, movies, and news), to identify the items that users might be most interested in and to generate personalized recommendations based on them. Recommender systems are nowadays ubiquitous on the Web, as one of the main engines for its success. One particular scenario that has become highly popular is the one of *online news recommendation*, which raises major specific challenges, pertaining to the highly dynamic, short-lived nature of news.

At the same time, with the advent of social networks, a lot of information – including news – spreads online, generating interest in the analysis of influence and information diffusion, under the formal scope of *influence estimation* [14, 22, 42] and *influence maximization* [25, 30]. While the main application of information diffusion studies in social networks is viral marketing, there are other and quite diverse potential applications, among which recommender systems [27].

In the social media context, it is natural to revisit the "matching" problem by taking into consideration social links and interactions, observing and exploiting how the information that is up for recommendation may also spread in an underlying network through *word-of-mouth* mechanisms. On one hand, such mechanisms – e.g., likes, shares, reposts / retweets, notifications – enable information to propagate rapidly and touch upon large audiences. On the other hand, social media brings *credibility* to the messages that are being conveyed, as recent studies show that we are more inclined to pay attention to a referral from a *friend* or from a known *influencer* [3].

Naturally, this renders influence and information diffusion important dimensions for any recommendation problems in social media, and even more so in the case of news. Indeed, social media is an increasingly popular platform for news dissemination and consumption, accounting now for about half of the market share [23] (reaching even 65% in some countries). Importantly, both the style of news and the patterns of news consumption have been reshaped by social media.

In the *micro-blogging* context, illustrated by hugely popular platforms such as Twitter and Sina Weibo, one may adopt news items not only based on content and preferences, but also based on the perceived influence from others with respect to news adoption. Such influence may be *local*, i.e., directly exerted by friends or followed users, or *global*, i.e., indirectly exerted by some influential users (a.k.a. "social whales" in the social marketing jargon).

The state-of-the-art ML-based news recommendation approaches (e.g., [13, 46, 51, 63]), which can be labelled as influence or social-agnostic, are generally exploiting the semantic content of news items and the user profiles, with some recent studies adopting a sequential perspective on the users' adoption patterns. Similar to the recent work of [11], we adopt in this paper an influence-aware perspective on news recommendation, with the main goal of *finding a sweet spot between the model's complexity and its effectiveness*. It is well-known that about half a billion tweets are published daily on Twitter [10], leading to billions of daily engagements. Therefore, we believe that only by achieving such a practically relevant complexity vs. effectiveness trade-off we can handle such unprecedented rates at which news posts and engagements thereof occur in micro-blogging applications.

For performance reasons and in order to achieve such an acceptable trade-off, we do not model the problem as a sequential recommendation task, shedding light on the fact that news adoptions in micro-blogging do exhibit temporal diversity, a finding that is similar to the one of the recent study of [54], on online news recommendation.

Starting from a joint history of observed news adoptions and news propagations (cascades) in a micro-blogging network, we propose a deep learning approach which follows a content-based architecture, while also taking into consideration the propagations thereof. Our main contributions are the following:

- We propose a lightweight deep learning-based model, called LISM, designed for news recommendation in social scenarios, where the susceptibility of a target user to news depends on both its semantics and the social impact it carries.

- We use information diffusion patterns to estimate the impact a piece of news may have on the target user, by leveraging the participation of users in news dissemination (the news cascades) and by representing users through graph embeddings.

- We endow news items with diffusion-related information (users involved in diffusion) besides semantics, by a specially designed CNN-based approach (SCNN) to align and fuse multi-source embeddings for joint feature extraction.

- We perform extensive experiments on two real-world datasets (including a publicly available one), confirming that (i) news diffusion patterns can significantly improve the effectiveness of news recommendation and, (ii) the proposed model strikes a balance between complexity and effectiveness, being therefore applicable in the highly dynamic contexts that motivate our work.

2 Related Work

2.1 Deep Learning in Recommender Systems

Early research on recommender systems is mainly about collaborative filtering (CF) [44] or content-based models [35]. The CF models make recommendations based on known user-item pairs, but they usually face cold-start problems, while the content-based mdoels leverage attribute information about users and items. Hybrid models [26] combine the benefits of the two directions.

With the increase in data volumes that recommender systems have to process, traditional models face problems such as computation cost, sparsity, and scalability. To tackle them, deep learning-based models were used recently, leading to many state-of-the-art ideas (e.g., the pioneering work on Restricted Boltzmann Machines [43]), with many studies aiming to capture complex, high-order dependencies between users and items via deep learning models [4, 19].

Among them, some CNN-based approaches are used as feature extractors from unstructured data – text [8], video [24], music [5], etc – while Recurrent Neural Networks (RNNs) are used in sequential modeling to observe the dynamic evolution of user preferences [7] or in session-based recommendation tasks [20].

2.2 Graphs in Recommender Systems

Graph analysis has been systematically explored by researchers to get insights on the relations and interactions between entities. Some of the most common tasks for graph analysis are (i) link prediction / recommendation [31], (ii) node classification [2], (iii) clustering [41], or (iv) visualization [1].

Generally, graphs are either described by an adjacency matrix or are projected (embedded) into a low-dimensional space [16]. The interest of the latter is that it gives an elegant way to transpose discrete data into a differentiable space, where machine learning techniques can apply for downstream analysis, among which recommendation task. With the increasing attention received by graph neural networks (GNNs) in recent research, many works [48, 57, 60] seek to exploit the effectiveness of GNNs in recommender systems. Such models incorporate both the graph structure and node attributes to represent users and items, by fusing neighboring node embeddings.

2.3 Social-Aware Recommender Systems

The underlying idea of recommendation algorithms in social media is to exploit social concepts such as homophily (attraction to similarity) and influence (actions and behavior evolution due to social connectivity). We discuss next such works, among which a few are focusing on news. Some of the pioneering social-aware models incorporate the social links, as indicators of similarity in collaborative filtering [18, 32]. Deep neural networks (and in particular GNNs) have also been used in the social networking scenario, to aggregate social-related information. E.g., the work of [9] builds user-item and user-user graphs, merging user-item interactions and social neighbourhoods for recommendation. In the DICER approach of [12], the authors exploit a dual-side deep context, which leverages social connectivity for modeling both user interests and item attractiveness. Beyond the first-order social neighbourhood interest of a user-user graph, DiffNet++ [56] models also the recursive diffusion process in the graph, under social influence.

Few works have taken steps in the direction of influence-aware news recommendation. The early work of [38] analyzed news in social media for recommendation, but their *Friends-Rank* strategy is to select and directly serve tweets from the user's friends, therefore limiting its scope drastically. More recently, [59] uses Monte Carlo Tree Search to mine high-order connectivity among users, and proposes a multi-armed bandit algorithm for selecting social friends for user representations.

Diffusion patterns or propagation cascades of news are not considered for user modelling in these works. In comparison, going beyond social similarity, we aim to exploit how information items may be diffused and adopted by socially interconnected users. We use a deep *behavior-driven* network to represent users as influencers (posters, i.e., news cascade initiators) or influencees (reposters, i.e., news cascade participants).

2.4 Personalized News Recommendation

News recommendation is a rather particular task in recommender systems because of the timeliness, diversity, and heterogeneity of news [28]. In addition, the explosion of new media content creators, besides the traditional media ones (newspapers, agencies, etc), makes personalized recommendation even more challenging. Following the success of deep learning in NLP, some of the models derived from information retrieval [33], for comparing query and document similarity, have been transferred to news recommendation, focusing on the highly condensed language in news. In state-of-the-art personalized news recommendation models, researchers generally focus on user representations based on user-item interactions (clicked news), using attention mechanisms to aggregate diverse interests, based on users' history (adopted news titles [49]) and news content [50]. Additional information such as news categories (by either a multi-level attention mechanism to fuse information in [51], or by a multi-head self attention mechanism in [52]), dwell time [50], shares or likes [55] are also used to represent news. In addition, with the development of general-purpose knowledge graphs and knowledge embeddings, discovering news connections based on external knowledge connections has been a recent trend in news recommender systems [46, 63].

In [53], the authors explore the benefits of *pre-trained language models* (PLMs), such as BERT [6], to mine the deep semantic information of news. PLMs are shown to have stronger text modeling ability than shallow models that are built directly from the news corpus. In a similar vein, the work of [62] learns news representations by pre-training with BERT, proving the generalization ability of this approach in cold-start scenarios.

Following the success of GNNs, recent studies also adopt a graph perspective on users and news, building for instance a bipartite graph where users and news are both represented as nodes, and neighboring nodes are respectively aggregated in the representation of news or users [13]. Similarly, [21] builds a user-news-topic graph to represent long-term interests of users.

All these deep learning news recommendation approaches traditionally focus on the news and adoptions thereof, ignoring the underlying social dimension to news.

The very recent work of [11] showed that, when it comes to news recommendation in a social networking scenario, relevant relations among users are revealed not only by common clicked news, but also by implicit influence. They propose a deep learning approach that is tailored for news recommendation in micro-blogging, modeled as a *sequential recommendation task*. Building on the assumption that users tend to adopt similar news successively – a prevalent assumption in the news recommendation literature – their sequential model incorporates temporal aspects, such as attractiveness and timeliness.

We adopt in this paper an influence-aware perspective on news recommendation, similar to [11]. Accordingly, LISM exploits jointly a history of observed news adoption and news propagation (diffusion cascades), bridging the content-based modeling with representations learned from the cascades, capturing local and global influence patterns on news. However, our model follows different architectural choices, which lead to a solution that is arguably on par with the one of [11] in terms of effectiveness, while being much more efficient. We believe that reduced training complexity is paramount in the micro-blogging context, which is one the most dynamic mediums for user-generated online content. By design, our method is not only less computationally intensive, but also less demanding in terms of input data dimensions for training, and therefore more versatile. Indeed, the temporal / sequential details of news adoptions may not always be (readily) available for the predictive model, therefore requiring approaches that do not depend on them. Finally, our work also provides similar insights and conclusions on online news adoption as [54], which shows that a sequential modeling of the news recommendation problem may be suboptimal.

3 Social Graph Representations

To model the influence exerted among users in social graphs, we build upon prior work on analyzing social interactions and representing users, in the context of a recommendation task. We review here the embedding methods used in our work to extract node representations in a social graph. A graphical illustration of the node representations discussed in this section is provided in Figure 1.

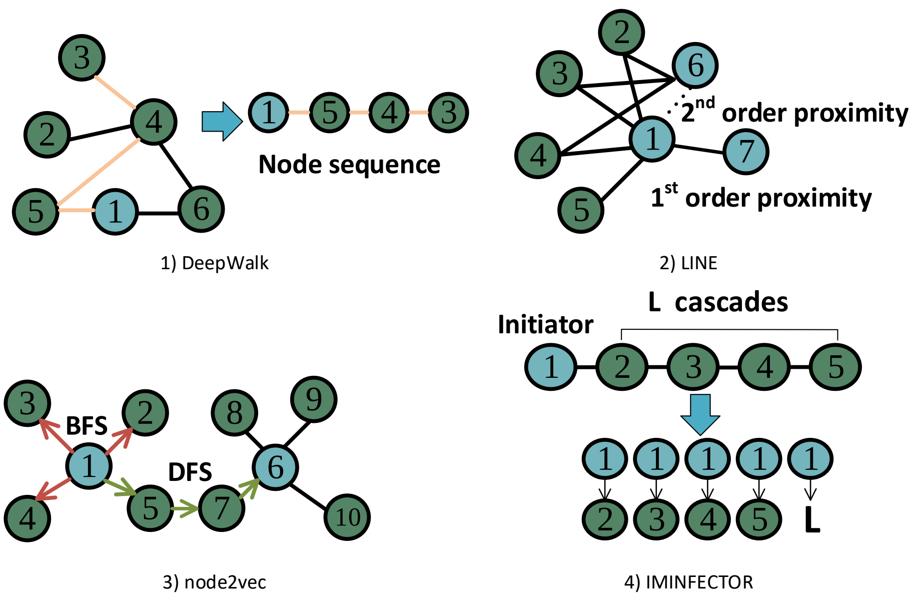


Figure 1: Topology or influence based node embeddings.

For a weighted social graph $G = (V, E, \omega)$, where V are the nodes (users), E are the edges, and ω weights each edge, a graph embedding corresponds to a mapping $\phi : V \rightarrow \mathbb{R}^g$, representing each node by a g-dimensional vector that preserves certain graph relationships and similarities. We review the following methods:

- *DeepWalk* [37] generates node embeddings by maximizing the probability of observing the last κ nodes and the next κ nodes in a random walk of length $2\kappa + 1$ centering at node v_i. With the node sequence $S_{v_i} = \{v_{i-\kappa}, ..., v_{i-1}, v_{i+1}, ..., v_{i+\kappa}\}$, the objective is

$$\min_{\phi}[-\log\Pr(S_{v_i}|\phi(v_i))] \tag{1}$$

  with $\phi(v_i) = \boldsymbol{v}_i$ the g-dimensional embedding of v_i.

- *LINE* [45] uses the adjacency matrix and embeddings to represent jointly probability distributions for each node pair, by the minimized KL-divergence of distributions. For 1st order proximity, the two distributions and objective functions are

$$\hat{p_1}(v_i, v_j) = \omega_{ij}/\Omega \tag{2}$$

$$p_1(v_i, v_j) = 1/[1 + \exp(- < \phi(v_i), \phi(v_j) >] \tag{3}$$

$$O_1 = - \sum_{(i,j)\in E} \omega_{ij} \log p_1(v_i, v_j), \tag{4}$$

where $\Omega = \sum_{(i,j)\in E} \omega_{ij}$ is the sum of all weights in the adjacency matrix $\boldsymbol{\Omega} \in \mathbb{R}^{|V|\times|V|}$ of G, and $\phi(v_i) = \boldsymbol{v}_i$, $\phi(v_j) = \boldsymbol{v}_j$ are the node embeddings of v_i and v_j. The distributions and objective function for 2nd order proximity are

$$\hat{p_2}(v_j|v_i) = \omega_{ij}/\delta_i \tag{5}$$

$$p_2(v_j|v_i) = \frac{\exp(\phi(v_i), \phi(v_j))}{\sum_{k=1}^{|V|} \exp(\phi(v_k), \phi(v_i))} \tag{6}$$

$$O_2 = - \sum_{(i,j)\in E} \omega_{ij} \log p_2(v_j|v_i), \tag{7}$$

where δ_i is the out-degree of node v_i.

- *node2vec* [17] extends *DeepWalk* to weighted, directed graphs, introducing biased random walks, based on Depth-First Search (DFS) and Breadth-First Search (BFS) strategies. Hyper-parameters μ_1, μ_2 are used to set the search bias β:

$$\beta_{\mu_1\mu_2}(v_j, v_i) = \begin{cases} \frac{1}{\mu_1}, & \text{if} \quad d_{ji} = 0 \\ 1, & \text{if} \quad d_{ji} = 1 \\ \frac{1}{\mu_2}, & \text{if} \quad d_{ji} = 2 \end{cases} \tag{8}$$

where d_{ji} is the shortest path distance between nodes v_j and v_i. Combined with edge weights $w_{(i-1)i}$, the unnormalized transition probability is

$$\pi_{v_{i-1}v_i} = \beta_{\mu_1\mu_2}(v_j, v_i) \cdot \omega_{(i-1)i} \tag{9}$$

The ith node in the sequence is generated by

$$P(v_i|v_{i-1}) = \begin{cases} \frac{\pi_{v_{i-1}v_i}}{\Gamma}, & \text{if} \quad (v_i, v_{i-1}) \in E \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

with Γ being a normalizing constant.

Besides the aforementioned *topology-based* graph representation learning, some novel deep learning-based approaches addressing graph-related tasks in an end-to-end way (*GNN models*) have attracted increasing attention recently [58]. As our focus goes beyond structural relationships in the social graph, to the influence a user – albeit distant - may exert on another one and its potential impact on news adoptions, we also consider the *cascade-based* IMINFECTOR approach of [34], which proposes a deep learning model for node representations of *influencer* and *influencee* (susceptible) users.

For the graph embedding task, IMINFECTOR takes a set of cascades – time-order sequence of timestamped adoptions in the form of $[(v_1, t_1), ..., (v_j, t_j), ..., (v_{m_x}, t_{m_x})]$ – where v_1 is the cascade initiator, all other v_j nodes are the influenced users in the propagation process, and m_x denotes the length of the cascade for item x. A time-sensitive sampling is used to select influenced nodes from cascades as training data. The hidden layer output is set as $\boldsymbol{z}_{t,i} = \boldsymbol{O}_i \boldsymbol{T}^T + \boldsymbol{b}_t$, for influencer v_i and target user v_t. $\boldsymbol{O} \in \mathbb{R}^{I\times g}$

provides the influencer (cascade initiator) embeddings, $\boldsymbol{T} \in \mathbb{R}^{|V| \times g}$ provides the target (susceptible) user embeddings, and $\boldsymbol{b}_t$ is the bias. I is the number of cascade initiators in the training data. The output is

$$\hat{\boldsymbol{y}}_t = \frac{e^{\boldsymbol{z}_{t,i}}}{\sum_{v_{i'} \in V} e^{\boldsymbol{z}_{t,i'}}} \tag{11}$$

with $\boldsymbol{y}_t$ being the one-hot representation of target node v_t to minimize the cross-entropy loss function

$$L_t = \boldsymbol{y}_t \log(\hat{\boldsymbol{y}}_t). \tag{12}$$

We will use the trained target node embeddings matrix $\boldsymbol{T}$ to represent users in the news diffusion network.

4 Problem Formulation

In an online, micro-blogging context, news adoptions are no longer observed by clicks, but instead by posting (tweeting) or reposting (retweeting) actions. The input for our news recommendation approach comes in the form of a social network and a cascade history, which also gives us the adoptions.

A news item will in general be first posted by a source, the cascade initiator, and then it will be adopted (posted) by other users involved in that diffusion process, the cascade participants. When a cascade reaches a user, she may become aware of who posted initially and who adopted that news item before, increasing her awareness about it and thus swaying the adoption decision.

With a social graph representation where each user is endowed with an embedding vector, news items can be described as records $x = \{W_x, V_x\}$, where W_x is the semantic information of the news content, and V_x represents the social-related information. Semantic information W_x is composed of a list of words in news titles $W_x = [w_1, w_2, ..., w_{n_x}]$, with n_x the number of words in the news title. (For simplicity and computation efficiency we only take titles into consideration here, but this can be easily generalized to include the news abstract / content.) Let V_x denote the list of involved users $V_x = [v_1, v_2, ..., v_{m_x}]$, obtained from the diffusion cascade of news item x, $L_x = [(v_1, t_1), \ldots, (v_{m_x}, t_{m_x})]$, which is a time-ordered sequence of timestamped adoptions by nodes (users), with $\{v_1, \ldots, v_{m_x}\} \subseteq V$, m_x being the number of users involved in the propagation of x, and v_1 being the cascade initiator.

With the representation for each news in the social scenario, a user i's adoption history in LISM will be represented as a set $\{x_1, x_2, ..., x_{s_i}\}$, where each x_j represents a news item adopted by user i, and s_i the total number of news items adopted by user i. For a user i with a known adoption history, our recommendation task is to predict the probability that i will click on some candidate news x_k.

5 LISM for News Recommendation

We present in this section the overall framework of LISM, followed by details on how we compute node embeddings while preserving the structural relationships and accounting for the observed diffusion patterns, then the design of SCNN for multi-source embeddings of words and nodes, and finally the application of an attention mechanism to aggregate the users' history.

5.1 LISM Framework

LISM takes a candidate news and a target user's history as input, and outputs the corresponding adoption probability. The structure of LISM is shown in Figure 2, and each component is detailed next, passing bottom-up over this framework.

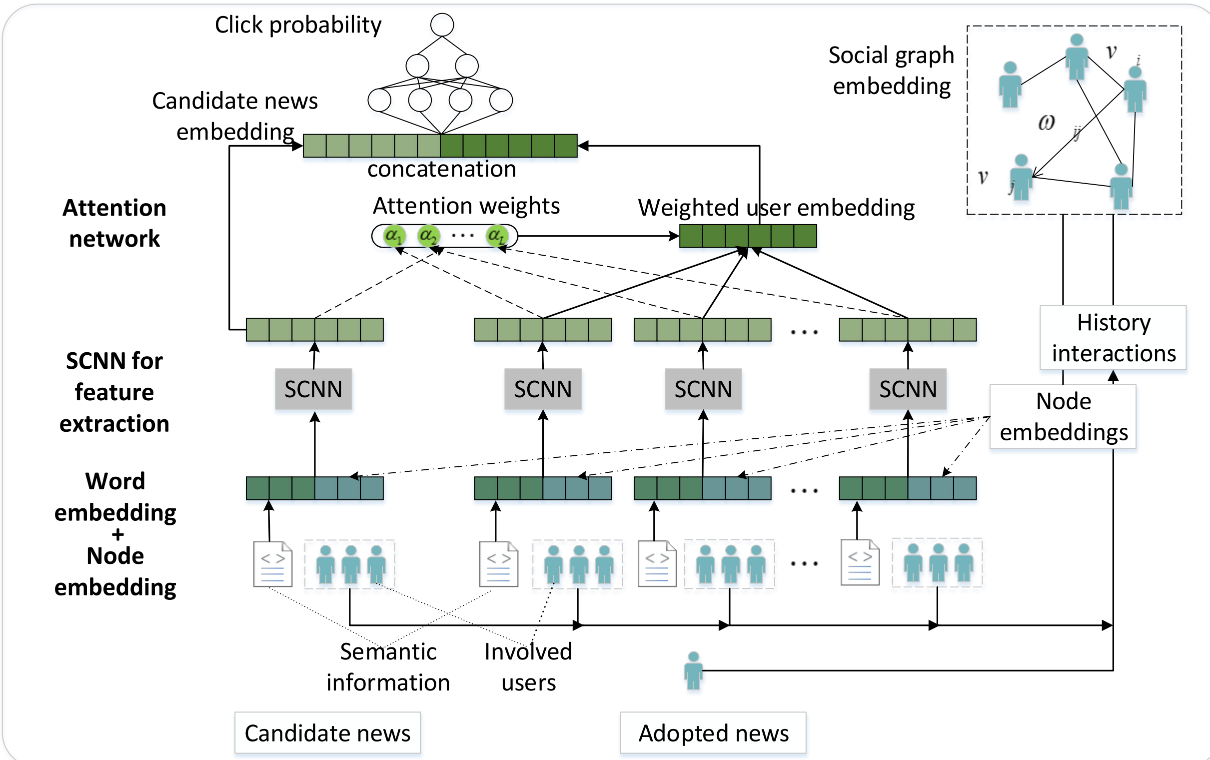


Figure 2: LISM framework.

In LISM a news item is an embedding matrix, composed of the news title represented as a word sequence $\boldsymbol{W}_x = [\boldsymbol{w}_1, \boldsymbol{w}_2, ...\boldsymbol{w}_{n_x}]^T \in \mathbb{R}^{n_x \times f}$, and the users in the cascade sequence represented as $\boldsymbol{V}_x = [\boldsymbol{v}_1, \boldsymbol{v}_2, ...\boldsymbol{v}_{m_x}]^T \in \mathbb{R}^{m_x \times g}$. $\boldsymbol{w}$ is the word embedding vector of w, and $\boldsymbol{v}$ is the node embedding vector of v, f and g being the embedding dimensions for word and node vectors respectively. The word vectors are pre-learned from a large corpus, while the node vectors are the output of a pre-trained node embedding model.

The process of user node embeddings in the social graph was discussed previously. The weights of edges in social graph and the feature extraction with SCNN will be detailed in the following two sections respectively. The representations of the candidate news and of the user's adopted news provided by SCNN are fed into an attention-based network, to allocate weights to each piece of history w.r.t. the candidate news. This weighted user history embedding and candidate news embedding are then concatenated as the input to a Deep Neural Network (DNN), to get the click probability.

5.2 Influence-Aware Social Graph

To endow news items with diffusion-related information (users involved in diffusion), besides semantics, we can either use a cascade-based diffusion network to project V into a vector space, as $\phi : V \rightarrow \mathbb{R}^g$, by topology-based embedding techniques designed for weighted, directed graphs, such as node2vec [17], or alternatively, we can directly use the cascades to obtain embeddings that capture the influencer / susceptible representation of nodes, as in [34].

For the former alternative, as in [15], we start by building a *diffusion graph* (V, E, ω) from the initial social graph $G = (V, E)$ and the cascades. For $(v_i, v_j) \in E$, we look at how often tweets from v_i are retweeted by v_j. More precisely,

$$\omega_{ij} = \frac{R_{i2j}}{\sum_{v_k} R_{k2j}} \tag{13}$$

where ω_{ij} is the influence v_i has on v_j, R_{i2j} is the number of times v_j reposts from v_i, and v_k iterates over all the nodes that v_j has reposted from. In this way, the structural relationship and historical interactions among nodes are both present in (V, E, ω); for simplicity, hereafter G will denote this diffusion graph.

Alternatively, the work of [34] focuses on influence maximization based on cascades solely, using cascade-based node embeddings. Since the focus of our paper is not on influence maximization, but on leveraging the diffusion patterns to capture a user's adoption likelihood, we can pre-train such an embedding model on the news cascades, leading to node representations $\boldsymbol{v}_i, i \in V$.

A comparison over the impact on the prediction task of the topology-based node representations and the cascade-based ones will be detailed in the experiments section. We stress here that only the latter are time-dependent, but these can be pre-trained on a relevant – yet not necessarily the latest – history of cascades. In fact, the frequency by which the node embedding layer is re-trained can be seen as knob on the efficiency vs. effectiveness tradeoff.

5.3 SCNN for Social-Related Multi-Source Feature Extraction

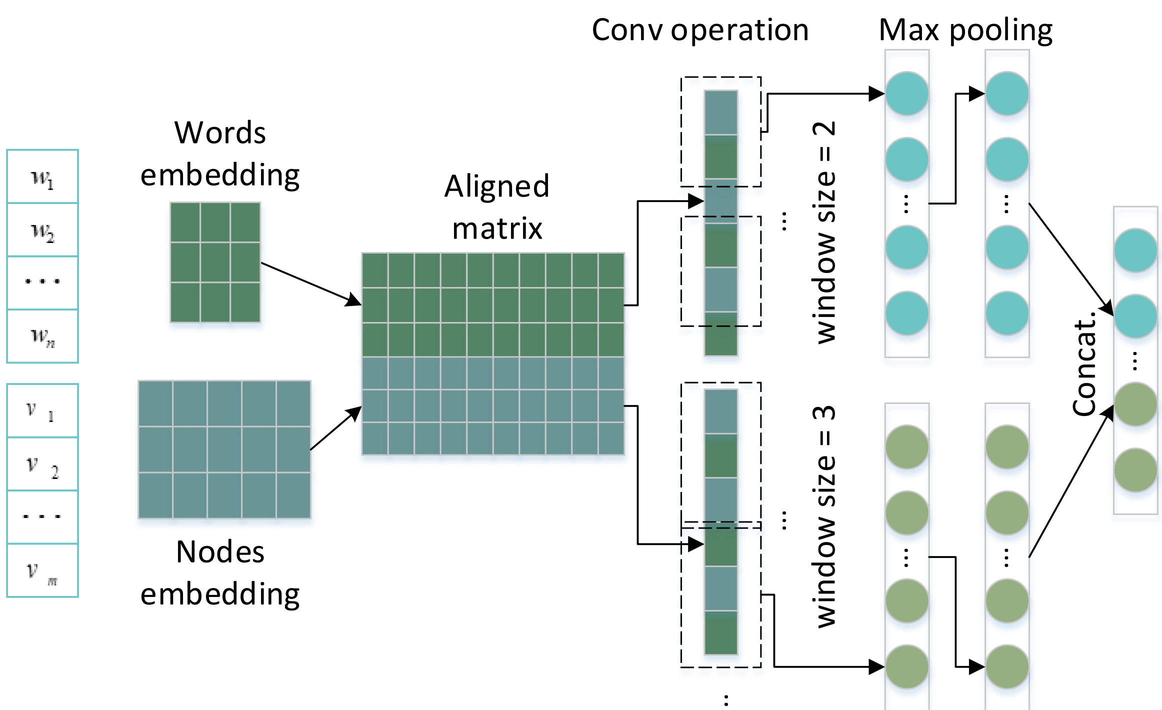


Figure 3: SCNN for word and node feature extraction.

Inspired by the similarity between node embeddings and word embeddings [37], we design a special CNN for news encoding in order to (i) align node embeddings with word embeddings, eliminating the heterogeneity of vector spaces, and (ii) fuse and extract features for joint representation of multi-source embeddings (Figure 3).

Following the notations from Section 4, with pre-trained node embeddings $\boldsymbol{v}_i$ obtained as in Section 5.2, and word embedding pre-learned from a large corpus, the representation embedding of news x for both semantic and diffusion-related information is

$$\boldsymbol{X} = \{\boldsymbol{W}_x, \boldsymbol{V}_x\} = \{\boldsymbol{w}_1\boldsymbol{w}_2...\boldsymbol{w}_{n_x}\boldsymbol{v}_1\boldsymbol{v}_2...\boldsymbol{v}_{m_x}\} \tag{14}$$

with $\boldsymbol{w} \in \mathbb{R}^f$ and $\boldsymbol{n} \in \mathbb{R}^g$. For a generic x, we simplify notation hereafter writing n and m instead of n_x and m_x. To align the word and node vectors, a transformation function ρ is applied on node vectors to map the node embeddings into the word space, while preserving the original node relationships, leading to

$$\boldsymbol{E} = [\boldsymbol{w}_1\boldsymbol{w}_2...\boldsymbol{w}_n\rho(\boldsymbol{v}_1)\rho(\boldsymbol{v}_2)...\rho(\boldsymbol{v}_m)] \tag{15}$$

where $\boldsymbol{E} \in \mathbb{R}^{(m+n)\times f}$ and the transformation function being either the linear one $\rho(\boldsymbol{v}) = \boldsymbol{M}\boldsymbol{v}$ or the non-linear one $\rho(\boldsymbol{v}) = \sigma(\boldsymbol{M}\boldsymbol{v}) + \boldsymbol{b}$, with $\boldsymbol{M} \in \mathbb{R}^{f\times g}$ the trainable transformation matrix, $\boldsymbol{b} \in \mathbb{R}^f$ the trainable bias, and σ either the sigmoid function or the tanh one.

SCNN with multiple filters extracts features from the multi-source embedding matrix $\boldsymbol{E}$. Each $\boldsymbol{h} \in \mathbb{R}^{f\times l}$ filter, with varying window size l, is applied to a sub-matrix $\boldsymbol{E}_{i+l-1}$ and a new feature is generated as

$$c_i^h = f(\boldsymbol{h} \cdot \boldsymbol{E}_{i:i+l-1} + \boldsymbol{b}) \tag{16}$$

with $i = 1, 2, \ldots, n+m-l+1$ and the bias $\boldsymbol{b} \in \mathbb{R}^f$. A feature map $[c_1^h, c_2^h, \cdots, c_{n+m-l+1}^h]$ is obtained going through the embedding matrix $\boldsymbol{E}$. After a max-over-time pooling operation on the feature map, the largest feature is chosen as

$$c_{max}^h = \max\{c_1^h, c_2^h, \cdots, c_{n+m-l+1}^h\} \tag{17}$$

For multiple filters, the final representation of news, comprising both word and node / diffusion information, is the concatenation of the largest feature from each filter $\boldsymbol{h}^i$, for $i = 1, \ldots, \gamma$, as follows:

$$\boldsymbol{e} = [c_{max}^{h_1}, c_{max}^{h_2}, \cdots, c_{max}^{h_\gamma}] \tag{18}$$

5.4 Attention Mechanism for User Interests

Assuming a user i with history of adopted news $\{x_1^i, x_2^i, \cdots, x_{s_i}^i\}$, after SCNN for joint feature extraction, the final representation is written as $\{\boldsymbol{e}_1^i, \boldsymbol{e}_2^i, \cdots, \boldsymbol{e}_{s_i}^i\}$. The same SCNN is applied on the candidate news k, leading to representation $\boldsymbol{e}_k$.

Considering that a user may have a wide range of interests, and that previously adopted news pertain to various topics of interest, instead of systematically taking an average of all embeddings $\boldsymbol{e}^i = \frac{1}{s_i}\sum_{j=1}^{s_i} \boldsymbol{e}_j^i$ of adopted news, an attention network is used to characterize the user's preferences w.r.t. different candidate news $\boldsymbol{e}_k$. To illustrate the difference of impact that each adopted news $\boldsymbol{e}_j^i$ has on the candidate one $\boldsymbol{e}_k$, the attention network automatically allocates weights to adopted news to aggregate a user's preferences w.r.t. $\boldsymbol{e}_k$. For a given candidate k represented as $\boldsymbol{e}_k$, the impact weight that clicked news $\boldsymbol{e}_j^i$ in user i's history has on candidate news $\boldsymbol{e}_k$ is

$$\alpha_{k,j}^i = \text{softmax}(\mathcal{N}(\boldsymbol{e}_k, \boldsymbol{e}_j^i)) = \frac{\exp(\mathcal{N}(\boldsymbol{e}_k, \boldsymbol{e}_j^i))}{\sum_{j=1}^{s_i}\exp(\mathcal{N}(\boldsymbol{e}_k, \boldsymbol{e}_j^i))} \tag{19}$$

where $\mathcal{N}$ is a DNN serving as the attention network, which takes the candidate news $\boldsymbol{e}_k$ and one piece of adopted news $\boldsymbol{e}_j^i$ as input and outputs the corresponding impact weight between the two. The softmax function is used for normalization. With impact weights, the entire history can then be summed up as

$$\boldsymbol{e}^i = \sum_{j=1}^{s_i} \alpha_{k,j}^i \boldsymbol{e}_j^i \tag{20}$$

Table 1: Dataset description.

	Twitter	*Weibo*
# users	248,195	692,833
# retweet records	4,999,535	31,211,347
# original tweets	4,566,942	232,978
# news	441,632	13,153
avg. # words per title	6.94	7.26
median length of diffusion chain	2.74	23
max length of diffusion chain	124	31009

Finally, the concatenation of the candidate news embedding $\boldsymbol{e}_k$ and the user's integrated history embedding $\boldsymbol{e}^i$ is fed into another DNN $\mathcal{D}$ to predict the probability that i adopts the candidate news $\boldsymbol{e}_k$, by

$$P_{k,i} = \mathcal{D}(\boldsymbol{e}_k, \boldsymbol{e}^i). \tag{21}$$

6 Experimental Evaluation

Datasets For our experiments, we use the same datasets as [11], originating from the two most popular micro-blogging platforms in China (*S*ina Weibo) and worldwide (*Twitter*). The news articles have been crawled, starting from the links appearing in tweets.

While the *Twitter* dataset is collected through the Twitter API, the *Weibo* one is a publicly available dataset [61], especially designed for information diffusion studies. Both datasets are diffusion-oriented, consisting of tweets pertaining to news and the follower / followee network of the posting users, where for each tweet of a piece of news (adoption), the information on whether it is a retweet or an original tweet, and the retweet / original user are recorded. The main statistics for these datasets are in Table 1.

6.1 Baselines

The following state-of-the-art methods are used as baselines in our experiments.

- **LibFM** [39] is a classical factorization model in feature engineering to estimate interactions. It has been widely used in recommendation tasks for click-through rate (CTR) prediction based on users' and items' features. Here, the news title embedding and the node embedding of target users are used as input features.
- **DeepWide** [4] is a deep learning-based model, containing a non-linear part (deep) and linear part (wide) to learn feature interactions. News title embeddings and node embeddings of users involved in the news diffusion process are concatenated as input.
- **DeepFM** [19] is an end-to-end factorization machine-based deep neural model, having a shared input with "wide" and "deep" parts compared to DeepWide; it uses the same input as DeepWide.
- **DCN** [47] introduces a cross network that is more efficient in learning bounded-degree feature interactions, while keeping the benefit of DNN in high dimensional nonlinear feature learning. In the experiments, we use the same input as DeepWide.

- **DKN** [46] recommends news by exploiting an external knowledge graph to capture relationships between news, and links the candidate news with the target user's history of adopted news both at knowledge-level and at semantic-level. To compare with DKN, experiments are carried out by also incorporating the knowledge entity embeddings into LISM (denoted as LISM(+)).
- **DAN** [63] enhances the framework of DKN with knowledge-entity type information, and uses an LSTM mechanism to model the sequential evolution of users' interests. Accordingly, for this baseline method, we use knowledge entities, entity types, as well as an LSTM to aggregate the users' history.
- **GERL** [13] builds a bipartite graph where users and news are both represented as nodes, and neighboring nodes are respectively aggregated in the representation of news or users, along with the semantic information for the recommendation. For this baseline method, we build the same kind of graph.
- **NAML** [51] exploits semantics and news categories, by a multi-level attention mechanism encoding news and users. For this baseline method, we use the topic distribution vector for categories and the news' titles for the semantics dimension.
- **IGNiteR** [11] is the extremely recent and most related approach, studying the same problem we consider in this paper, influence-aware news recommendation in micro-blogging scenarios, modeled as a sequential recommendation task. They follow the assumption that users tend to adopt similar news successively, and their sequential model incorporates temporal aspects, such as attractiveness and timeliness.
- **DICER** [12] is a social-aware recommendation model, which leverages the social connections among users to build a user-user graph, complementing the user-item and item-item graphs constructed from the click history. To compensate for the semantic dimension, we replace the ID embedding matrix of news items with the averaged news title embedding.

Note: we did not compare with the NRMS model [52], as it bears significant similarities with NAML on the attention mechanism. Also, we leave as a future extension the use of PLMs in both our model and in NAML (basically corresponding to [53]'s NAML-BERT), but also in the UNBERT approach of [62]. Finally, other news recommendation models – such as CPRS [50] or FedRec [55] – were not included in our comparison, as they pertain to aspects that are orthogonal to our framework and at the same time available only in proprietary data from commercial platforms like MSN News. Finally, we stress that for completeness purposes we include in our comparison DAN and IGNiteR, even though they exploit as additional information the adoption timestamps.

6.2 Experimental Setup

Our algorithm was implemented and evaluated using Tensorflow. As the two datasets record users' behaviour over a span of 3 years, for each user in our predictive framework an observation window of 3 months of clicked history is used; in this way, a user from the application may be split into several ones in our framework. We randomly sampled as negative samples from the news released during the given period and not clicked by the target user. We split the train / test set by the timeline, so that around 85% of data is used for training and 15% is used for testing; we randomly sample 10% from the training set for validation. To remove outliers, we also used other filters as follows: the maximum number of clicked news per user is set to 20, the maximum title length is set to 10 for *Twitter* and to 15 for *Weibo*, and the maximum number of users involved in a diffusion is set to 30. In the training phase, the ratio between negative samples and positive samples is 4, and in the test phase the maximum number of negative samples is set to 10. One year worth of diffusion cascades has been used in the training phase, to obtain the node embedding of users. We used node embeddings with dimensionality $g = \{50, 100, 200, 300\}$. As to semantic embeddings, we chose the Chinese word vectors pre-trained on Weibo [29] corpus with word2vec of the Chinese words for *Weibo* dataset, and Glove [36] vectors pre-trained on Twitter corpus for the *Twitter* dataset. The word embedding dimensionality is $f = \{50, 100, 200, 300\}$. In the SCNN part, the window size is set from the combination of $l = \{1, 2, 3, 4\}$, and in each window size, the filter number varies in $\{32, 64, 128, 256\}$. Further discussion on the variants and hyper-parameters will be given in Section 6.3.

The best set of parameters is determined by the performance in the validation phase, with the average AUC, F1, MRR, NDCG@5, NDCG@10 scores as the evaluation metrics. We did our best to tune the baseline methods. In DKN, DAN, GERL, DICER and NAML, the datasets are pre-processed in the same way (sampling ratio, user history, title length), while the knowledge-entity length is set to 10 for DKN and DAN, and the number of neighboring user is set to 20 in GERL. IGNiteR uses the same one year worth of diffusion cascades to obtain the node embeddings. In DICER, the number user / item links is set to 10 per item, and the number of friends in DICER is set to 30. The attention hidden layer size is set to 200 for NAML and GERL, 256 for DKN and DAN, and [128, 64] for DICER. The hidden layer structure of DeepWide model is set as [512, 512], and for DeepFM and DCN as [256, 256]. For LibFM, the number of factors is set as 50. The dimensions of word embeddings and node embeddings are set as $f = 100$ and $g = 100$. Each experiment is repeated for 10 times and the average performance is reported in the results.

6.3 Results Analysis

In this section, we present the comparison of LISM with the baselines, the comparison among variants of LISM, as well as the impact of hyper-parameters on its performance.

6.3.1 Comparison with Baselines

Table 2: Comparison with baseline methods on $Weibo$.

Models	AUC	F1	MRR	NDCG@5	NDCG@10
LibFM	65.70 ± 0.45	63.70 ± 1.50	15.84 ± 1.09	55.37 ± 0.59	57.95 ± 0.21
DeepFM	68.16 ± 0.46	65.29 ± 0.98	16.13 ± 1.80	59.47 ± 1.23	61.15 ± 1.01
DeepWide	68.30 ± 2.03	66.96 ± 1.48	16.26 ± 2.12	61.36 ± 0.78	62.08 ± 0.61
DCN	69.57 ± 2.021	67.65 ± 0.50	16.90 ± 0.91	61.87 ± 0.56	63.14 ± 0.49
DKN	71.82 ± 1.38	69.97 ± 2.26	18.56 ± 1.09	64.38 ± 1.23	66.15 ± 1.58
DAN	72.69 ± 0.95	71.87 ± 0.53	19.79 ± 0.63	65.13 ± 0.77	67.49 ± 1.22
DICER	73.53 ± 0.65	71.28 ± 0.93	20.06 ± 0.28	66.03 ± 1.27	68.59 ± 1.53
GERL	73.71 ± 1.28	71.54 ± 0.68	20.59 ± 0.84	66.32 ± 0.59	68.98 ± 0.52
NAML	74.85 ± 0.78	72.15 ± 1.02	21.63 ± 1.06	67.44 ± 1.22	69.38 ± 1.28
IGNiteR	76.92 ± 0.58	73.85 ± 1.01	23.63 ± 0.36	69.59 ± 0.91	72.01 ± 0.78
LISM	75.94 ± 1.34	72.56 ± 1.05	23.44 ± 0.67	68.35 ± 1.78	71.08 ± 0.51
LISM(+)	$\mathbf{76.37 \pm 1.86}$	$\mathbf{72.89 \pm 1.12}$	$\mathbf{23.75 \pm 0.26}$	$\mathbf{68.78 \pm 1.2}$	$\mathbf{71.59 \pm 1.48}$

Table 3: Comparison with baseline methods on *Twitter*.

Models	AUC	F1	MRR	NDCG@5	NDCG@10
LibFM	60.72 ± 1.17	59.80 ± 1.32	13.30 ± 0.59	57.76 ± 0.67	59.59 ± 0.41
DeepFM	65.23 ± 1.12	64.30 ± 1.24	13.89 ± 1.31	61.56 ± 0.96	63.21 ± 0.97
DeepWide	65.78 ± 1.25	64.96 ± 0.92	14.07 ± 0.52	61.96 ± 0.62	63.88 ± 0.84
DCN	66.40 ± 1.13	65.35 ± 1.06	14.58 ± 0.31	62.47 ± 0.76	64.02 ± 2.01
DKN	68.72 ± 0.96	67.02 ± 0.86	15.22 ± 0.28	64.17 ± 1.32	66.35 ± 1.45
DAN	69.83 ± 0.34	69.28 ± 0.59	15.62 ± 0.54	64.83 ± 1.34	67.02 ± 1.33
DICER	70.38 ± 1.72	69.08 ± 1.95	15.79 ± 0.21	65.31 ± 1.72	67.70 ± 1.56
GERL	70.39 ± 1.38	69.48 ± 1.68	16.35 ± 0.32	65.54 ± 2.59	68.44 ± 2.52
NAML	71.05 ± 2.18	70.15 ± 1.92	16.54 ± 0.55	67.19 ± 0.72	70.58 ± 1.23
IGNiteR	74.82 ± 0.58	72.25 ± 0.53	17.97 ± 0.63	69.57 ± 1.22	72.75 ± 0.46
LISM	72.87 ± 1.05	71.68 ± 1.80	17.47 ± 0.18	69.01 ± 0.95	72.08 ± 1.25
LISM(+)	$\mathbf{73.41 \pm 1.83}$	$\mathbf{72.53 \pm 1.70}$	$\mathbf{17.93 \pm 0.41}$	$\mathbf{69.43 \pm 1.36}$	$\mathbf{72.77 \pm 1.24}$

In the comparison with baselines, in LISM the dimensions of node and word embeddings are both 100 and the number of filters is 200. As to the graph embedding method, the best result of LISM on both *Weibo* and *Twitter* are obtained with the node representations generated by the IMINFECTOR cascade-based graph embedding method, trained on the diffusion cascades. Both the Chinese and English knowledge entity embeddings of *Weibo* and *Twitter* respectively are generated by the TransE model for knowledge graph embedding. We can see from the results of Table 2 and Table 3 that LISM outperforms all the baselines on news data from social media, with the exception of our direct competitor (IGNiteR), which validates once more the motivation to exploit diffusion patterns in news recommendation.

In terms of effectiveness, we can observe that LISM is on par with IGNiteR. The latter performs slightly better both in the Twitter case (on four out of the five metrics) and in the Weibo case (on three out of the five metrics). This comes as a pleasant surprise, since, as discussed in the following section, LISM is significantly lighter computationally, and does not require knowledge of the time dimension (adoption timestamps) and the sequential user adoption history this dimension enables.

Within the class of generic DL-based models, LibFM performs worse than other models, which indicates that the factorization model fails to capture nonlinear interactions in news and the social graph. DeepFM makes up for this with its deep part for learning high-dimensional interaction. DCN and DeepWide outperform DeepFM in both datasets. Considering the performance of LibFM, it is possible that the factorization model is less effective in learning graph information.

We can note that, with the exception of DICER, the performance of the neural news recommendation models overpasses that of the generic deep recommendation models. Although DICER is not designed for news recommendation, with the social-aware information integrated in the user-user graph, it performs quite well under in our social scenario. NAML performs the best among the baseline methods, showcasing the efficiency of the multi-view attention mechanism that aggregates news features. GERL and DICER perform almost comparably well on both datasets, with GERL slightly outperforming the other. Our intuition is that these models share common ideas in the building of graph structures encompassing users and news, with DICER leveraging the social relations and a well-designed GNN to construct such graphs, while GERL balances the results with a dedicated news transformer for semantic information.

We can also see that although knowledge-level information brings improvements in both datasets, the ones in Twitter are larger. A possible explanation may be that knowledge-level information weighs more for the recommendation in Twitter than in Weibo.

6.3.2 Comparison among LISM variants

To give insights into LISM, experiments are carried out to compare the variants of the model. From Table 5, we can see that with node embeddings, LISM performs better on all metrics, which confirms that the influence-aware user representations improve news recommendation on social platforms. Among the graph embedding methods, the best results on *Weibo* and *Twitter* are obtained with the influence-base embedding, which generates node representations based on cascades, while node2vec performs best among the topology-based methods.

Table 4: Comparison among LISM variants in *Weibo*.

Variants	Weibo				
	AUC	F1	MRR	NDCG@5	NDCG@10
Node embedding	$\mathbf{75.94 \pm 1.34}$	$\mathbf{72.56 \pm 1.05}$	$\mathbf{23.44 \pm 0.67}$	$\mathbf{68.35 \pm 1.78}$	$\mathbf{71.08 \pm 0.51}$
Without node embedding	72.08 ± 1.78	70.13 ± 2.15	17.87 ± 0.78	64.65 ± 1.57	67.21 ± 1.67
IMINFECTOR	$\mathbf{75.94 \pm 1.34}$	$\mathbf{72.56 \pm 1.05}$	$\mathbf{23.44 \pm 0.67}$	$\mathbf{68.35 \pm 1.78}$	$\mathbf{71.08 \pm 0.51}$
Node2vec	73.55 ± 0.89	71.39 ± 1.21	19.43 ± 0.35	66.07 ± 1.02	68.63 ± 1.17
LINE	73.03 ± 1.74	70.97 ± 0.81	18.63 ± 0.22	65.67 ± 1.32	68.14 ± 1.01
DeepWalk	72.68 ± 1.57	70.78 ± 1.31	18.03 ± 0.45	65.12 ± 1.24	67.77 ± 1.06
Attention	$\mathbf{75.94 \pm 1.34}$	$\mathbf{72.56 \pm 1.05}$	$\mathbf{23.44 \pm 0.67}$	$\mathbf{68.35 \pm 1.78}$	$\mathbf{71.08 \pm 0.51}$
Without attention	70.18 ± 1.05	68.54 ± 2.37	16.77 ± 0.86	62.78 ± 1.02	64.19 ± 1.97
Knowledge	$\mathbf{76.37 \pm 1.86}$	$\mathbf{72.89 \pm 1.12}$	$\mathbf{23.75 \pm 0.26}$	$\mathbf{68.78 \pm 1.2}$	$\mathbf{71.59 \pm 1.48}$
Without knowledge	75.94 ± 1.34	72.56 ± 1.05	23.44 ± 0.67	68.35 ± 1.78	71.08 ± 0.51
Linear transformation	75.60 ± 1.32	72.53 ± 1.03	23.42 ± 0.85	$\mathbf{68.37 \pm 1.73}$	71.04 ± 0.30
Non-linear transformation	$\mathbf{75.94 \pm 1.34}$	72.56 ± 1.05	$\mathbf{23.44 \pm 0.67}$	68.35 ± 1.78	$\mathbf{71.08 \pm 0.51}$
Zero-padding	75.61 ± 1.04	$\mathbf{72.58 \pm 1.15}$	23.36 ± 0.81	68.12 ± 1.88	70.97 ± 0.98

In the two datasets, the use of an attention mechanism greatly improves performance. The main reason is that the attention network can automatically aggregate a user's diverse interests with respect to the candidate news. Performance is improved on both datasets when knowledge entities are incorporated in CTR prediction, which corroborates with the DKN [46] conclusions about the importance of knowledge-level connections in news recommendation. In terms of transformation function to align node and word vectors, there is no option that is significantly better than the other.

6.3.3 Training Complexity

Our training complexity analysis is related to the various hyper-parameters of the methods we consider. Since each model has its own set of parameters to be tuned, we cannot guarantee here absolute fairness in the comparison. When using computation time as the metric for efficiency / data complexity, a common phenomenon in deep learning is that with the complexity of the neural network going up (number of layers, number of cells, etc.), the predictive performance may be improved by a small margin, while the computation time increases by a large margin. However, we strive to avoid biases as much as possible by running the various models based on the same parameter setting for their common stages or components. Note that the parameter setting may now be different from the one used in the previous effectiveness comparison, where we used the best tuned hyper-parameters for each model.

Table 5: Comparison among LISM variants in *Twitter*.

Variants	Twitter				
	AUC	F1	MRR	NDCG@5	NDCG@10
Node embedding	$\mathbf{72.87 \pm 1.05}$	$\mathbf{71.68 \pm 1.80}$	$\mathbf{17.77 \pm 0.18}$	$\mathbf{69.01 \pm 0.95}$	$\mathbf{72.08 \pm 1.25}$
Without node embedding	68.74 ± 1.07	65.78 ± 1.53	14.49 ± 0.70	63.85 ± 0.92	65.95 ± 1.73
IMINFECTOR	$\mathbf{72.87 \pm 1.05}$	$\mathbf{71.68 \pm 1.80}$	$\mathbf{17.47 \pm 0.18}$	$\mathbf{69.01 \pm 0.95}$	$\mathbf{72.08 \pm 1.25}$
Node2vec	70.16 ± 0.85	69.68 ± 1.56	15.70 ± 0.73	65.03 ± 1.25	67.38 ± 0.97
LINE	69.56 ± 1.27	66.28 ± 1.09	15.13 ± 0.49	64.73 ± 1.68	66.87 ± 0.85
DeepWalk	69.12 ± 2.27	65.98 ± 2.23	14.79 ± 0.66	64.21 ± 1.79	66.37 ± 1.90
Attention	$\mathbf{72.87 \pm 1.05}$	$\mathbf{71.68 \pm 1.80}$	$\mathbf{17.77 \pm 0.18}$	$\mathbf{69.01 \pm 0.95}$	$\mathbf{72.08 \pm 1.25}$
Without attention	67.94 ± 2.57	65.18 ± 2.93	13.98 ± 1.69	63.35 ± 1.74	65.75 ± 0.89
Knowledge	$\mathbf{73.41 \pm 1.83}$	72.53 ± 1.70	$\mathbf{17.93 \pm 0.21}$	$\mathbf{69.43 \pm 1.36}$	$\mathbf{72.77 \pm 1.24}$
Without knowledge	72.87 ± 1.05	71.68 ± 1.80	17.77 ± 0.18	69.01 ± 0.95	72.08 ± 1.25
Linear transformation	72.80 ± 1.15	71.58 ± 1.69	$\mathbf{17.82 \pm 0.15}$	$\mathbf{69.02 \pm 1.55}$	72.04 ± 2.25
Non-linear transformation	$\mathbf{72.87 \pm 1.05}$	$\mathbf{71.68 \pm 1.80}$	17.77 ± 0.18	69.01 ± 0.95	$\mathbf{72.08 \pm 1.25}$
Zero-padding	72.83 ± 1.95	71.58 ± 1.93	17.69 ± 0.25	68.97 ± 1.59	72.03 ± 1.57

For all models, we sampled from 700 users in *Weibo*, and we generated (i) around 20000 samples for LibFM, DeepWide, DeepFM, DCN, DKN, DAN and NAML, (ii) around 11000 samples for GERL and DICER, and (iii) 7500 samples for LISM and IGNiteR; for GERL, DICER (resp. LISM), we made sure that enough graph neighbours (resp. cascade nodes) are available for training. From the diffusion cascades, we sampled the diffusion information from a 6-months period to generate the cascade-based node embeddings, for both LISM and IGNiteR[1]. For the CNN-based models, the number of filters is set to 200, and the kernel size is set to 3. For the models using an attention mechanism, the number of attention layer units is 200. The other settings are as mentioned in the previous section. The batch size is 200, and the number of epochs is 10. For the performance vs. complexity tradeoff, we use the computation time (order of minutes) for each epoch as the complexity indicator, while NDCG@5 is used as the performance indicator.

We can see from Figure 4 that although LibFM, DeepWide, DeepFM and DCN are relatively fast, they perform significantly worse than the neural news recommendation models.

In contrast to Sec. 6.3.1, where all the models are fine-tuned for optimal performance, IGNiteR shows a clearer dependence on the number of training cascades and on the news adoption timespan. With the diffusion cascades used for training spanning a period of 6 months, although IGNiteR outperforms LISM, the sequential analysis of users' behavior and of news adoption lifecycles increases its complexity. LISM achieves by far the best complexity vs. effectiveness tradeoff, being runner-up and close to IGNiteR's performance, while working with significantly fewer samples and thus having a reduced training time, compared to all the other neural news recommendation models. Within this class, DKN seems to be the most costly method.

Regarding data scalability, we found in our experiments that the training time of LISM evolves linearly with the size of the training data (further details are omitted here).

[1] The training process for the graph neural network takes around 4 minutes for each epoch, with 10 epochs in total. The number of involved nodes was 1.17M.

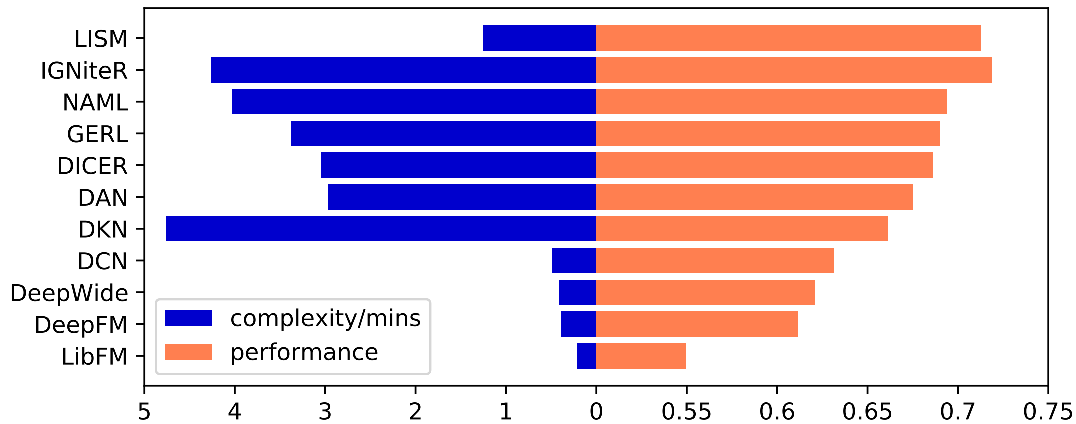


Figure 4: Performance vs. complexity tradeoff.

6.3.4 Analysis on Hyper-Parameters

Parameter setting is important for the performance of LISM. In Figure 5, we analyze the sensitivity of the AUC and F1 metrics to the embedding dimension of words and nodes, and to the CNN's hyper-parameters.

1. *Word embedding dimension and node embedding dimension.* We select from all the combinations of dimensions f and g from $\{50, 100, 200, 300\}$. For a given node embedding dimension g, the performance of LISM improves as the word embedding dimension f increases, but it decreases when the f continues going up, since a large dimension of word embedding may introduce noise and lead to the inability to retain effective information. However, there is an exception for node embedding dimension $g = 100$, which maintains the performance even with large word embedding dimension.

2. *CNN hyper-parameters.* The best performance is obtained with a filter at 128, in all combinations between window size and filter size. LISM's performance drops when the window size gets complicated, perhaps due to over-fitting. Compared to the AUC metric, the results on F1 seem to be more sensitive to the variation of window size, while for AUC the changes are smoother.

7 Conclusion

We propose in this paper LISM, a lightweight content-based deep learning model for news recommendation in micro-blogging scenarios. To incorporate the awareness about news due to social influence, we represent users by embeddings obtained through methods leveraging the diffusion history (cascades), in such a way that news is endowed with social-related information. A special designed CNN model is applied to deal with the joint representation of news and an attention mechanism allows us to aggregate users' diverse interests with respect to candidate news. Extensive experiments are conducted on two real-world datasets, including a publicly available one. They show that news diffusion patterns can significantly improve the effectiveness of news recommendation, compared to influence and social-agnostic models, even when the problem is not (or cannot be) modeled as a sequential recommendation one. Furthermore, they show that the proposed model strikes a balance between complexity and effectiveness, being

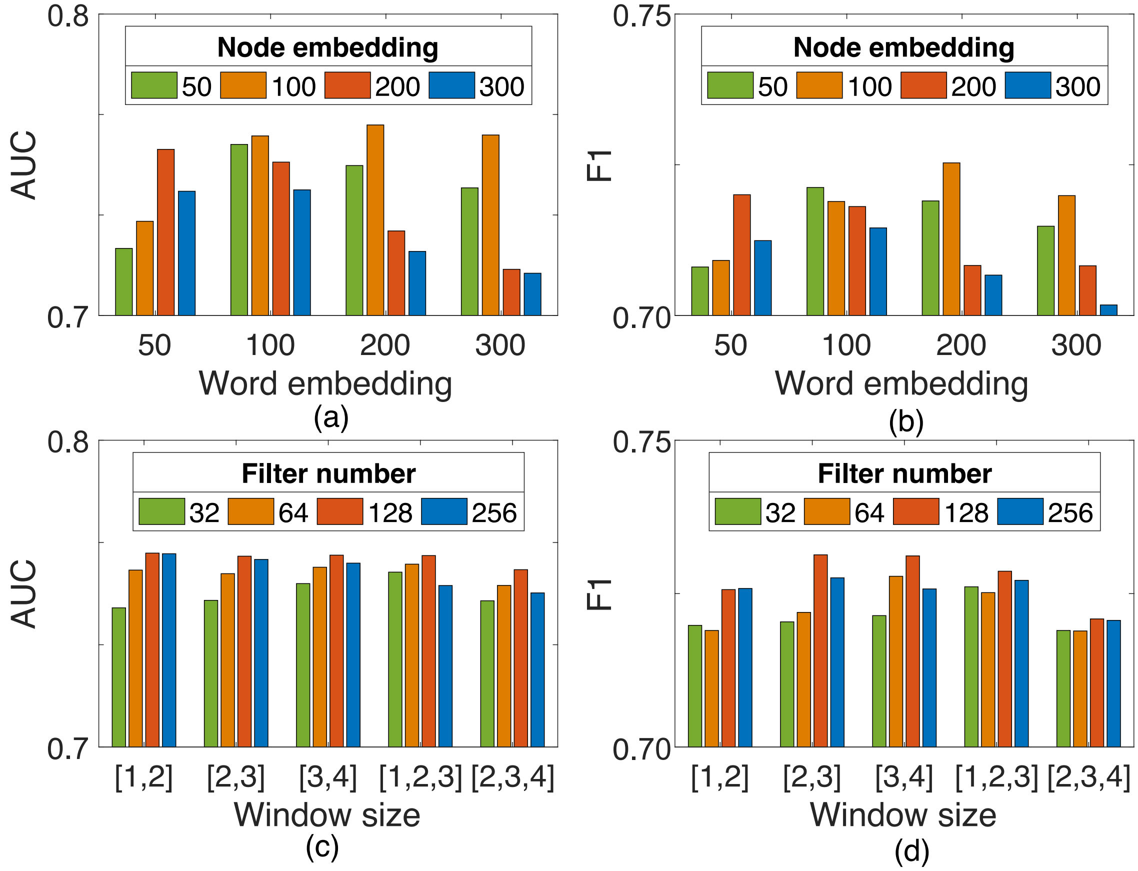


Figure 5: Parameter analysis in LISM.

therefore applicable in the highly dynamic (high data velocity) context of micro-blogging platforms. We believe that our model is generic enough to be applied in any social media setting, when adoptions and diffusion traces are available, including those where adoption timestamps may not be known. A future extension to our work is the use of PLMs for a deeper semantic understanding of news.

References

[1] Batagelj, V., Mrvar, A.: Pajek—analysis and visualization of large networks. In: Graph drawing software (2004)

[2] Bhagat, S., Cormode, G., Muthukrishnan, S.: Node classification in social networks. In: Social network data analytics (2011)

[3] Bughin, J.: Getting a sharper picture of social media's influence. McKinsey Quarterly (2015)

[4] Cheng, H.T., et al.: Wide & deep learning for recommender systems. In: Workshop on deep learning for recommender systems (2016)

[5] Choi, K., Fazekas, G., Sandler, M., Cho, K.: Convolutional recurrent neural networks for music classification. In: ICASSP (2017)

[6] Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2019)

[7] Donkers, T., Loepp, B., Ziegler, J.: Sequential user-based recurrent neural network recommendations. In: RecSys (2017)

[8] Dos Santos, C., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: COLING (2014)
[9] Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., Yin, D.: Graph neural networks for social recommendation. In: The World Wide Web Conference. pp. 417–426 (2019)
[10] Fedoryszak, M., Frederick, B., Rajaram, V., Zhong, C.: Real-time event detection on social data streams. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019 (2019)
[11] Feng, Y., Cautis, B.: Igniter: News recommendation in microblogging applications (extended version). arXiv preprint arXiv:2210.01942 (2022)
[12] Fu, B., Zhang, W., Hu, G., Dai, X., Huang, S., Chen, J.: Dual side deep context-aware modulation for social recommendation (2021)
[13] Ge, S., Wu, C., Wu, F., Qi, T., Huang, Y.: Graph enhanced representation learning for news recommendation. In: Proceedings of The Web Conference 2020. pp. 2863–2869 (2020)
[14] Gomez-Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. TKDD (2012)
[15] Goyal, A., Bonchi, F., Lakshmanan, L.V.S.: A data-based approach to social influence maximization. Proc. VLDB Endowement (2011)
[16] Goyal, P., Ferrara, E.: Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems (2018)
[17] Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: SIGKDD (2016)
[18] Guo, G., Zhang, J., Yorke-Smith, N.: Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In: Proceedings of the AAAI conference on artificial intelligence. vol. 29 (2015)
[19] Guo, H., Tang, R., Ye, Y., Li, Z., He, X.: DeepFM: a factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247 (2017)
[20] Hidasi, B., Quadrana, M., Karatzoglou, A., Tikk, D.: Parallel recurrent neural network architectures for feature-rich session-based recommendations. In: RecSys (2016)
[21] Hu, L., Li, C., Shi, C., Yang, C., Shao, C.: Graph neural news recommendation with long-term and short-term interest modeling. Information Processing & Management **57**(2), 102142 (2020)
[22] Huang, H., Yan, Q., Chen, L., Gao, Y., Jensen, C.S.: Statistical inference of diffusion networks. IEEE TKDE (2019)
[23] Institute, R.: Digital News Report 2022 (2020), `https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022`
[24] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: CVPR (2014)
[25] Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: SIGKDD (2003)
[26] Kouki, P., Fakhraei, S., Foulds, J., Eirinaki, M., Getoor, L.: HyPER: A flexible and extensible probabilistic framework for hybrid recommender systems. In: RecSys (2015)
[27] Leskovec, J., Singh, A., Kleinberg, J.: Patterns of influence in a recommendation network. In: PAKDD (2006)
[28] Li, M., Wang, L.: A survey on personalized news recommendation technology. IEEE Access (2019)
[29] Li, S., Zhao, Z., Hu, R., Li, W., Liu, T., Du, X.: Analogical reasoning on chinese morphological and semantic relations. arXiv:1805.06504 (2018)
[30] Li, Y., Fan, J., Wang, Y., Tan, K.: Influence maximization on social graphs. TKDE (2018)
[31] Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. Journal of the American society for information science and technology (2007)
[32] Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: WSDM (2011)
[33] Neculoiu, P., Versteegh, M., Rotaru, M.: Learning text similarity with siamese recurrent networks. In: Workshop on Representation Learning for NLP (2016)
[34] Panagopoulos, G., Malliaros, F., Vazirgiannis, M.: Multi-task learning for influence estimation and maximization. IEEE TKDE (2020)

[35] Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Adaptive web (2007)
[36] Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: EMNLP (2014)
[37] Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: KDD (2014)
[38] Phelan, O., McCarthy, K., Smyth, B.: Using twitter to recommend real-time topical news. In: RecSyS (2009)
[39] Rendle, S.: Factorization machines with LibFM. ACM TIST (2012)
[40] Ricci, F.: Recommender systems: Models and techniques. (2014)
[41] Saerens, M., Fouss, F., Yen, L., Dupont, P.: The principal components analysis of a graph, and its relationships to spectral clustering. In: ECML (2004)
[42] Saito, K., Nakano, R., Kimura, M.: Prediction of information diffusion probabilities for independent cascade model. In: KES (2008)
[43] Salakhutdinov, R., Mnih, A., Hinton, G.: Restricted boltzmann machines for collaborative filtering. In: ICML (2007)
[44] Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Adaptive web (2007)
[45] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: LINE: Large-scale information network embedding. In: WWW (2015)
[46] Wang, H., Zhang, F., Xie, X., Guo, M.: DKN: Deep knowledge-aware network for news recommendation. In: WWW (2018)
[47] Wang, R., Shivanna, R., Cheng, D., Jain, S., Lin, D., Hong, L., Chi, E.: DCN-M: Improved deep & cross network for feature cross learning in web-scale learning to rank systems. arXiv:2008.13535 (2020)
[48] Wang, X., He, X., Wang, M., Feng, F., Chua, T.S.: Neural graph collaborative filtering. In: Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. pp. 165–174 (2019)
[49] Wu, C., Wu, F., An, M., Huang, J., Huang, Y., Xie, X.: NPA: neural news recommendation with personalized attention. In: SIGKDD (2019)
[50] Wu, C., Wu, F., Qi, T., Huang, Y.: User modeling with click preference and reading satisfaction for news recommendation. In: IJCAI (2020)
[51] Wu, C., Wu, F., An, M., Huang, J., Huang, Y., Xie, X.: Neural news recommendation with attentive multi-view learning. arXiv preprint arXiv:1907.05576 (2019)
[52] Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., Xie, X.: Neural news recommendation with multi-head self-attention. In: EMNLP-IJCNLP'19 (2019)
[53] Wu, C., Wu, F., Qi, T., Huang, Y.: Empowering news recommendation with pre-trained language models. In: SIGIR (2021)
[54] Wu, C., Wu, F., Qi, T., Li, C., Huang, Y.: Is news recommendation a sequential recommendation task? In: SIGIR '22: The 45th International ACM SIGIR (2022)
[55] Wu, C., Wu, F., Qi, T., Liu, Q., Tian, X., Li, J., He, W., Huang, Y., Xie, X.: Feedrec: News feed recommendation with various user feedbacks. In: WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022 (2022)
[56] Wu, L., Li, J., Sun, P., Hong, R., Ge, Y., Wang, M.: Diffnet++: A neural influence and interest diffusion network for social recommendation. IEEE Transactions on Knowledge and Data Engineering (2020)
[57] Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., Tan, T.: Session-based recommendation with graph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 346–353 (2019)
[58] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S.Y.: A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems (2020)
[59] Xiao, W., Zhao, H., Pan, H., Song, Y., Zheng, V.W., Yang, Q.: Beyond personalization: Social content recommendation for creator equality and consumer satisfaction. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 235–245 (2019)
[60] Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W.L., Leskovec, J.: Graph convolutional neural networks for web-scale recommender systems. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 974–983 (2018)
[61] Zhang, J., Liu, B., Tang, J., Chen, T., Li, J.: Social influence locality for modeling retweeting behaviors. In:

IJCAI (2013)

[62] Zhang, Q., Li, J., Jia, Q., Wang, C., Zhu, J., Wang, Z., He, X.: Unbert: User-news matching bert for news recommendation. In: IJCAI (2021)

[63] Zhu, Q., Zhou, X., Song, Z., Tan, J., Guo, L.: DAN: Deep attention neural network for news recommendation. In: AAAI (2019)

Kandinsky Meets Social Conversation: Towards Abstract Art-Inspired Visualization Abstraction for Mobile Devices

Castillo Clarence Fitzerald Gumtang, Chun Leong Tham, Sourav S. Bhowmick
Nanyang Technological University (NTU), Singapore
clarence001@e.ntu.edu.sg, ctham005@e.ntu.edu.sg, assourav@ntu.edu.sg

Abstract

Social media platforms enable users and organizations to share content–such as text, images, and videos–collectively referred to as anchor posts. These posts often attract hundreds of comments, forming rich social conversations. Traditionally, these comment threads are displayed in a hierarchical list format, which users must manually browse to understand the discussion dynamics. However, this process can be tedious and inefficient, discouraging users from conducting multi-faceted exploration of the conversation landscape. In this paper, we introduce a novel visualization abstraction called *Several Concentric Circles (SCC)*, which transforms textual social discussions to a digital "artwork" to provide an interactive, mobile-friendly, and visually appealing alternative to conventional list-based abstraction. SCC addresses the limitations of traditional formats by integrating principles from two seemingly distinct domains–abstract art and computer science–to foster both aesthetic appeal and user engagement. To validate our approach, we developed a prototype system named Kandinsky Mobile, which brings our visualization framework to life and supports effective exploration of social discussions.

1 Introduction

Social media platforms (*e.g., Facebook*, *YouTube*, X) enable individuals and organizations to upload content such as posts, images, or videos–collectively referred to as *anchor posts*. Other users can engage with these posts by reading and commenting on them, resulting in social conversations (*i.e.,* discussions). In essence, anchor posts are essentially social media posts with accompanied comments, facilitating discussion and interaction. Posts can feature various elements, including visuals, text, and interactive components, to capture user interest and build community engagement. Comments typically appear in hierarchical list-based format and serve as textual or multimedia responses that allow users to share opinions, pose questions, and take part in the ongoing conversation. With the rapid rise of mobile technology, these interactions increasingly occur via mobile devices. For instance, mobile phones are often the primary means by which users access X[1].

A large number of online users read comments associated with an anchor post for various reasons such as to learn about opinions of others, to be entertained or amused, to get information, to compare one's view on a topic with others, etc [7]. Unfortunately, this can be cognitively challenging when one has to peruse a large number of comments. To mitigate this cognitive burden, most social media platforms automatically display comments deemed *"top"*, *"most relevant"*, or *"most recent"*, as determined by proprietary machine learning (ML) algorithms. These algorithms are typically optimized to maximize user engagement and platform profitability [35]. It is worth noting that in recent times, such ML-driven

[1] https://blog.X.com/en_us/a/2013/new-compete-study-primary-mobile-users-on-twitter

exposition models in social media sites (*e.g., Facebook*) have led to *"proliferation of extremism, bullying, hate speech, disinformation, conspiracy theory, and rhetorical violence"* [35].

1.1 Motivation

Despite the prevalence of the list-based visualization abstraction of social media discussions with the goal to optimize engagement and profit, it suffers from several key limitations as follows.

- First, reading beyond the top or most relevant/recent comments can be tedious and time-consuming, particularly when an anchor post has a large volume of responses. In many ways, this process resembles navigating a lengthy hierarchical text document. This is especially aggravated in a mobile device due to its small screen size and unique interaction style [10, 26, 29]. In particular, none of the social media sites allow a user to get a bird's-eye view of the conversation landscape, highlighting discussion threads that have garnered considerable attention and interactions between various comments and users. Such view can enable users to choose *efficiently any* conversation that they *wish* to peruse instead of steered by ML-based algorithms.

- Second, users are generally unable to perform interactive search within the discussion threads, making it difficult to locate specific topics, keywords, or user contributions across potentially vast comment sections. Additionally, there is limited support for temporal navigation, what we refer to as "time traveling" through the discussion space, where users could explore how conversations evolved over time or revisit specific moments in the discussion's history. This lack of temporal and semantic interactivity significantly hinders deeper exploration, analysis, and understanding of social discourse.

- Third, while aesthetics is a well-established criterion for usability [13] and a key component of effective human-data interaction frameworks [50], the conventional list-based presentation of social media comments tends to lack aesthetic appeal as it prioritizes individual comment readability over aesthetics. This linear, text-heavy format can appear cluttered, monotonous, and visually unengaging, which may detract from the overall user experience and hinder meaningful engagement with the content.

- Last but not the least, ML-based exposition of *relevant* comments, while designed to surface the most engaging or relevant content, can inadvertently lead users into "filter bubbles" [35]. These algorithms tend to prioritize comments that align with a user's prior interactions, preferences, or inferred interests, thereby limiting exposure to diverse viewpoints within the broader discussion landscape. As a result, users may be systematically steered away from diverse opinions or novel perspectives, undermining the potential for open dialogue, critical engagement, and the healthy exchange of ideas.

To address the aforementioned limitations, at a high-level we aim to *transform* the hierarchical textual comments linked to an anchor post into an aesthetically pleasing, interactive "artwork". This concept has gained momentum recently with the emergence of text-to-image generation systems [4, 6], which are capable of producing digital images and artworks from natural language prompts. To this end, we introduce a novel *abstract art-inspired visualization abstraction* for representing social discussions centered around an anchor post. It enables us to achieve the following set of key visualization goals while laying the groundwork for addressing the aforementioned limitations.

1. *Canvas Construction:* Aesthetically pleasing, complete, and mobile-friendly bird's eye view of the social discussions landscape.

2. *On-demand Conversation Visualization:* On-demand, mobile-friendly visualization and exploration of comments and discussion threads.

3. *Canvas Search:* Interactive keyword and comment similarity search on the discussions and visualization of search results.

4. *Time Travel:* Time travel of the discussions landscape.

Specifically, Goals 1 and 2 address the first and third limitations, while Goals 3 and 4 target the second. Together, all four are intended to establish a foundation for tackling the final limitation.

1.2 Overview

Devising visualization abstraction that realize the aforementioned goals is challenging. While challenges such as on-demand exploration, interactive search, mobile-friendly visualization, and temporal navigation fall squarely within the domain of computer science, the dimension of aesthetics is traditionally rooted in the arts. Consequently, we seek inspirations from the abstract arts of the famous Russian painter Wassily Kandinsky (1866-1944) to build a novel visualization abstraction called *Several Concentric Circles* (SCC) that lays the foundation for transforming the textual comment space of an anchor post to a digital, interactive artwork. The data model of SCC consists of a set of colorful *circles* and *shape buckets* that contain these circles. Each circle represents a comment in a social discussion. The social conversation landscape with respect to an anchor post is then *transformed* to an aesthetically pleasing, interactive collection of colorful circles and concentric circles on a mobile display canvas inspired by the artworks of Kandinsky using a set of *operations*. In particular, more users are involved in a discussion, the larger and more colorful the concentric circles are on the canvas. Specifically, these operations facilitate realization of all the aforementioned visualization goals. We also introduce a prototype implementation on *YouTube*, named Kandinsky-Mobile[2] , which realizes the SCC abstraction to demonstrate the proposed visualization goals. It is worth noting that while text-to-image generation systems have gained increasing popularity with the rise of generative AI, Kandinsky-Mobile [22] and its desktop variant [38] predate these systems. In summary, this paper makes the following contributions.

- We propose a novel visualization abstraction, *Several Concentric Circles (SCC)*, inspired by abstract arts of Wassily Kandinsky, to "paint" the landscape of social discussions surrounding an anchor post. To the best of our knowledge, this is the first work to bridge the domains of abstract art and data visualization in the context of visualizing social media conversations.

- To systematically realize the aforementioned visualization goals, we introduce a set of novel *SCC operations*–a collection of core interaction and rendering operations grounded in the design principles of our SCC abstraction. These operations serve as the foundation for achieving the aforementioned goals. Building on these operations, we develop Kandinsky Mobile, a prototype system that demonstrates the practical viability of our approach on mobile devices.

- Our initial experimental evaluation and user study highlight the effectiveness and superiority of SCC compared to traditional list-based visualization abstraction in achieving the aforementioned goals.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents a brief background on Kandinsky's artworks and inspiration behind our SCC abstraction and Kandinsky

[2] Kandinsky-Mobile and its desktop variant are demonstrated in [22] and [38], respectively.

Mobile. Sections 4 and 5 describe the architecture of Kandinsky Mobile and the SCC abstraction that underpins it. We describe how the aforementioned visualization goals are realized using SCC in Section 6. We report our preliminary performance study in Section 7. The final section concludes the paper with reflections on the proposed visualization paradigm and outlines open challenges and future directions in this interdisciplinary area.

2 Related Work

Conversation visualization. Most germane to our research are efforts on visualizing social comments and conversations. Research on this arena dates back to the 1990s and primarily focuses on identifying the temporal or structural relationship between conversations.

Donath *et al.* [14] proposed *Chat Circles*, a graphical interface for visualizing synchronous conversations. The interface represents each participation by a colored circle on the screen in which his or her posts appear. The interface *Conversation Landscape* was also developed to visualize the conversational archive of *Chat Circles*. It provides a 2-dimensional visualization of a chat group where the x-axis represents participants identified by color and the y-axis represents timeline of postings by participants. Each post is shown as a horizontal line, with the length reflecting the message's length. This design provides a bird's-eye view of the activeness of participants in the discussion over time and the interaction patterns of the conversation, making dominant users and surge of discussions recognizable. The limitation of this design is that, for discussion among a large number of participants but with low contribution rate, it may render a long horizontal axis but sparse lines.

Conversation Thumbnails [51] visualizes the thread structure of conversations using tree-visualization techniques. Engdahl *et al.* [17] used *Squarified Treemap Layout* to visualize threaded discussion forums on devices with small screen size. The technique uses squarified treemaps to display the contents of discussion forums. Each thread from discussion forums is rendered as a colored rectangle in a treemap. The size of the rectangle is either proportional to the number of articles in the thread, or reflects the relevance of the query for that thread when searching is performed. The user is first presented an overview of the top level threads for each forum. As the stylus is moved across the screen, the details of the selected thread is displayed in the popup. This approach makes use of 100% of the limited screen space and provides an intuitive overview of the discussion landscape, allowing the user to easily compare popularity among threads. The most (resp. less) discussed thread is placed in the upper left (resp. bottom right) corner of the map. However, the ability to retrieve detailed information and analyze temporal trends is missing.

PeopleGarden [52] creates a garden of flowers to visualize users' interactions over time. Each flower represents a user and each petal is a post by him/her. Each user is portrayed as a flower, whose petals represent the user's posts, and height reflecting the amount of time the user has been in the discussion. A flower "blossoms" as a user becomes more engaged in a discussion. Together, users' interactions in the social environment form a garden. This design achieves compact information representation about participants involved in online discussions, but fails to provide details related to the discussion. Furthermore, flowers might overlap and become difficult to analyze once the more participants are involved.

ToPIN [48] is a system for analyzing and visualizing comments on online educational videos. It provides a qualitative perspective by examining the specifics of individual comments, including their type, timestamp, and how they cluster and evolve over time. Additionally, it incorporates *ThemeRiver* [23] to offer a complementary quantitative view, highlighting trends and shifts within the collection of time-stamped comments.

ConVis [27] visualizes blog conversations through a thread overview that reveals their structure and sentiment, alongside a faceted overview that highlights key topics and contributors. More recently, [49] used sentiment analysis on *Yahoo! News* article comments about Covid19 testing, classifying them as negative, positive, or neutral, and leverage a bubble chart for visualizing public sentiment and its monthly changes based on response volume and approval rates. *EmoVis* [28] is a framework for visualizing sentiment shifts in social network comment sections as a way to track public discourse over time. Using a pre-trained model, it predicts emotional scores from user comments and maps them to various sentiment categories. However, this framework does not propose any novel visualization abstraction. These efforts are complimentary to our work as we do not focus on visualizing sentiments.

None of the aforementioned visualization schemes for conversation draw inspiration from abstract art. Moreover, they lack a systematic implementation of visualization abstractions based on a defined data model and a corresponding set of operations. In contrast, Kandinsky Mobile utilizes the proposed SCC abstraction to design and build an interactive visualization framework that represents the landscape of social discussions on mobile devices.

Text-to-artwork generation. With the emergence of *Generative Artificial Intelligence (GenAI)*, several large-scale text-to-art models and systems have been introduced [4, 6]. A shared characteristic among these models is their capability to generate digital artwork based on user-written prompts in natural language. In a similar vein, Kandinsky Mobile seeks to convert textual user comments, rather than explicit prompts, into visually appealing artwork, while avoiding reliance on large generative models. Additionally, unlike Kandinsky Mobile, these systems are not tailored to a specific artistic style, such as that of Kandinsky.

Recently, Zhou *et al.* [55] developed a text-to-image dataset centered on the works of Wassily Kandinsky and introduced a generative method that allows users to automatically produce personalized artwork in Kandinsky's style, based on their preferences–removing the need for explicit prompts. In contrast to SCC, their approach does not rely on any data model for creating artwork in Kandinsky's style. The objectives are also different. While they focus on generating personalized digital artworks in the style of Kandinsky, our aim is to create Kandinsky-inspired visual abstractions for social discussions.

Complex circle-based network data visualization. The use of complex circular geometries is well-established within the visualization community, particularly for illustrating various properties and structures of networks [9, 42, 54]. For instance, *hybrid layout* techniques, as described in [42], combine multiple layout strategies to effectively capture and display the topological features of networks. In a related approach, the work presented in [9] introduces composite circular encodings, where *Treemap visualizations*, typically used for hierarchical data, are integrated within circular frameworks to represent different network layers within a unified visual metaphor. Similarly, the *Graph Thumbnails* technique [54] leverages circle packing to provide an overview visualization of hierarchical network structures. In the domain of property graph visualization, the concept of *composite* circles (*i.e.,* primary circles surrounded by smaller satellite circles) has been recently utilized to visually encode both property graph queries and schema structures [11, 40], offering an intuitive and spatially compact representation of complex property graph semantics. These efforts are primarily driven by structural, semantic, or functional considerations and are not informed by any particular tradition of abstract artistic expression. The SCC abstraction draws deliberate inspiration from abstract art. Specifically, it employs vibrant, colorful concentric circles as a central visual motif to aesthetically represent the dynamic conversations among social media users in relation to a given anchor post.

User interface design for mobile devices. The need to find effective user interfaces for mobile devices is critical. In order to save space on small screens, Noirhomme-Fraiture *et al.* [43] suggested to use concise and precise information. Giller *et al.* [20] observed that users performed significantly better when they could scroll, instead of tapping on widget elements tabs for page-to-page navigation, and that

Figure 1: **[Best viewed in color]** Artworks of Wassily Kandinsky: (Left) *Squares with Concentric Circles* (1913); (Middle) *Several Circles* (1926) [2]; (Right) *Heavy Circles* (1927) [3].

scrolling vertically rather than horizontally led to better user experience. Vertical scrolling should be used when displaying the long list of comments, such that the number of taps can be minimized [46]. To overcome the limited interaction with a finger touch (*i.e.,* the thumb), the application must be fat-finger friendly [5, 10, 45]. Due to the limited motion available to the thumb, Hoober [24, 25] observed based on collected data that users prefer to view and touch the center of the screen. People prefer reading content at the center of the screen and are better at tapping at the center of the screen. Kandinsky Mobile embodies these key insights through the implementation of the SCC operations.

3 Inspirations Behind *Kandinsky Mobile*

In this section, we present relevant abstract artworks of Wassily Kandinsky that serve as inspiration to our SCC abstraction. In particular, we articulate how his artistic philosophy in these artworks informed the design of Kandinsky Mobile. In the subsequent sections, we shall elaborate on the realization of this design philosophy.

3.1 Wassily Kandinsky: Father of Abstract Art

As the "father" of abstract art, Wassily Kandinsky (1866-1944) is known for his lyrical style and innovative theories on non-representational art[3]. As opposed to traditional realism that focused on physical matter and materialism, he exploited the evocative interrelation between *color* and *form* to create an aesthetic experience. A recent study reported that Kandinsky's art has impacted aesthetic experience emphasizing that modern art is not meaningless and has value because of its artistic expression [12].

Color. Kandinsky's color theory is built around three primary colors, *Red*, *Yellow*, and *Blue*, and three primary tones, *Black*, *White*, and *Grey* [32]. By combining the primary colors, he derives three secondary colors, *Orange*, *Green*, and *Violet*. He classifies these six chromatic colors based on perceived temperature: red, yellow, and orange are categorized as *warm*, while green, blue, and violet are considered *cold*. Tone in Kandinsky's framework refers to the lightness or darkness of a composition, determined by the ratio of black to white. Each color is associated with a distinct spiritual emotion and exerts an influence on surrounding colors [32], similar to how elements in a visual or perceptual system interact and modify one another. This anticipates modern theories in computational aesthetics and HCI, where color is used not just for visibility but also for emotional engagement and user experience design.

[3]Representational art is designed to represent real life, and non-representational art is the opposite.

Form. Geometrical elements took on increasing importance in Kandinsky's painting – particularly the circle. He proposed a structured theory of visual tension, defining graphical elements in terms of their shape, spatial position, and orientation. He begins with the primitive unit, the *Point*, and derives two higher-order constructs: the *Line* and the *Plane* [33]. Lines are categorized based on the type of tension they exhibit, resulting in three forms: *Straight*, *Curved*, and *Angular* [33]. Similarly, planes are subject to tension and can be represented as *Triangles*, *Squares*, or *Circles* [33]. These seven geometric primitives serve as the core components in Kandinsky's visual compositions, akin to how basic shapes form the building blocks in computational visual systems.

Kandinsky asserts that the overall composition is shaped by the dynamic interrelationships among all visual elements. It is defined not only by the placement of individual elements but also by the connections they form with one another and the structure of grouped elements within the visual space.

3.2 Circles in Kandinsky's Artworks

The aforementioned artistic philosophy of Kandinsky can be experienced in his works such as *Circles in a Circle* (1923) [1], *Several Circles* (1926) [2], and *Heavy Circles* (1927) [3] (Figure 1). We focus on these as they serve as inspiration to the design of SCC and Kandinsky Mobile.

Squares with Concentric Circles (1913). *Squares with Concentric Circles* (Figure 1 (left)) is a key work by Kandinsky that exemplifies his use of abstraction, color theory, and spiritual symbolism. The painting features concentric circles within squares, where the square acts as a static container, providing contrast to the dynamic, fluid nature of the circles. This creates a tension between the infinite (circle) and the finite (square), a recurring theme in Kandinsky's work that explores the balance between opposing forces, such as the spiritual vs. the material and freedom vs. constraint.

The concentric circles are layered with different colors, which not only create visual movement but also evoke emotional depth. While the circles vary in color and size, they maintain harmony within the square, reflecting Kandinsky's belief in balance and the significance of geometric forms in expressing emotional and spiritual resonance.

Several Circles (1926). Kandinsky regarded the circle as the most "spiritual" of all geometric forms. In *Several Circles* (Figure 1 (middle)), he isolates this shape to investigate its expressive possibilities. Unlike his earlier works, which often included a variety of shapes and intersecting lines, this work is dedicated solely to only one form, the circle. This could focus instead on colour, mass and the relative positions of the circles, minimizing visual noise and emphasizing their symbolic meaning. The circles differ in size, color, transparency, and placement, generating a visually intricate yet harmonious composition. This approach reflects Kandinsky's belief that effective composition arises not from symmetry, but from the dynamic interplay between forms. Each circle interacts with the others through its color and placement, illustrating his idea that elements in a composition influence one another.

Kandinsky also applies his color theory to these forms, using contrasts of temperature (warm vs. cool) and tone (light vs. dark) to evoke emotional intensity. Color affects not only the individual circle's feel but also its relationship to the rest of the composition. The largest circle, surrounded by a glowing halo, commands attention and becomes a focal point. The careful positioning of each circle reflects Kandinsky's theory of pictorial tension, where shapes exert visual "forces" on one another based on their size, orientation, and location. This interaction generates a sense of dynamic equilibrium–one that feels alive despite the static nature of the shapes.

Heavy Circles (1927). In *Heavy Circles* (Figure 1 (right)), Kandinsky builds on his exploration of geometric abstraction, emphasizing visual weight, density, and contrast. Similar to *Several Circles*, the composition is centered around two large circles, with a series of smaller circles placed in the corners of the artwork. The painting introduces variations in line, color, and complexity, differentiating it from his

Table 1: Relationship between Kandinsky's artworks and the features of Kandinsky Mobile.

ID	*Artwork*	*Feature in Kandinsky's artworks*	*Feature in Kandinsky Mobile*	*Semantics*	*SCC Operations*
K1	Several Circles, Heavy Circles	Circles are the only geometric forms to minimize visual noise.	Circles are the only geometric shape in SCC.	Each circle represents a comment.	paint
K2	Squares with Concentric Circles	A collection of colorful concentric circles layered with different colors to create visual movement and emotional depth.	Nucleus concentric circles layered with different colors.	Represents replies to a comment.	paint
K3	Several Circles, Heavy Circles	Circles of different size and color may overlap on the circumference.	Peripheral concentric circles.	Represents replies to a reply (depth greater than or equal to 2 in a conversation network).	paint
K4	Squares with Concentric Circles, Several Circles, Heavy Circles	Usage of different colors, contrast of temperature and tone.	Color coding of circles using different temperature and tone.	Different users in a conversation.	paint
K5	Several Circles, Heavy Circles	Two larger circles are the focal points.	Larger concentric circles are focal points.	Represents conversations that have garnered a lot of interactions.	paint, balance
K6	Several Circles, Heavy Circles	Circles are strategically positioned to minimize visual clutter and highlight interplay between forms.	Circles are positioned strategically to reduce cluttering and showing interactions in a conversation.	Represents the interactions in the conversation space.	balance, update
K7	Squares with Concentric Circles	Squares encapsulate the concentric circles.	Squares are used to encapsulate circles that satisfy search results.	Represents results of interactive search.	update
K8	Several Circles, Heavy Circles	Dark background of the canvas.	Dark background of *Kandinsky Canvas*.	Choice of background color for Kandinsky Mobile.	paint
K9	Heavy Circles	Color intensity is used to show the interplay of heavier and lighter circles.	Color intensity is lowered in interactive search and time travel.	Lower intensity circles represent conversations that are not part of search results or not within a specific time window.	paint, update

earlier work *Several Circles*. The circles in *Heavy Circles* have a "heavier" appearance compared to the lighter, airier forms in *Several Circles*. Kandinsky manipulates color saturation, size, and placement to create a sense of mass, making the shapes feel anchored and impactful, both visually and psychologically. His theory of visual tension is evident in the balance between heavier and lighter circles. The "heaviness" of some areas is counteracted by lighter, more dispersed forms, creating a dynamic equilibrium that guides the viewer's eye. The interplay of color, size, and opacity also adds emotional depth, with darker tones evoking seriousness and brighter accents adding intensity.

3.3 Kandinsky's Circles to Kandinsky Mobile

Our system draws inspiration from these colorful, concentric, and overlapping circles in the artworks of Kandinsky. Table 1 gives an overview of the "mapping" between the three aforementioned artworks of Kandinsky and corresponding features in SCC and Kandinsky Mobile. The key features of these artworks that inspire the design of Kandinsky Mobile are as follows:

- Both *Several Circles* and *Heavy Circles* center around a single geometric element–the circle–to minimize visual noise. Likewise, the SCC abstraction exploits only circles to visualize the social conversation landscape (**K1**, **K3**).

- Variations in size, color, placement, and transparency of circles create a dynamic yet cohesive visual interaction in the artworks. The circles may be concentric or overlapping. Similarly, Kandinsky Mobile utilizes these same visual attributes–size, color, and placement–to depict the structure and semantics of social conversations (**K2-K4**, **K6**).

- In *Several Circles* and *Heavy Circles*, two prominent circles dominate the composition, serving as visual focal points. In parallel, Kandinsky Mobile positions large circles to highlight conversations that have attracted substantial engagement (**K5**).

- Kandinsky's approach to color resonates with contemporary digital color models. His use of warm versus cool tones and varying saturation not only generates visual rhythm but also communicates psychological states. Similarly, Kandinsky Mobile uses dark "canvas" and applies warm and cool colors along with saturation levels to encode information about the conversational landscape. For instance, color saturation is used for conveying time travel (**K8**, **K9**).

- The structured arrangement of concentric circles within squares in *Squares with Concentric Circles* informs our design choice to highlight results of interactive search in static square containers (**K7**).

4 The Several Concentric Circles (SCC) Abstraction

In this section, we present an *abstract art-inspired abstraction* of the social discussion landscape called *Several Concentric Circles (SCC)* whose design is inspired by the aforementioned artworks of Kandinsky. We begin with the architecture of Kandinsky Mobile system that realizes the SCC abstraction.

4.1 Architecture

Kandinsky Mobile is intended for users with normal vision who are experienced with mobile devices and social media conversations. Figure 2 depicts the architecture. A comment associated with a social post (*i.e., anchor post*) comprises the content, the author information (identifier and name), the date/time of comment, number of likes it has garnered (if any), and the *reply id* in which the comment is replied to. A set of such comments among individuals engenders social discussions. The goal of Kandinsky Mobile is to facilitate the four visualization goals introduced in Section 1 w.r.t the discussions associated with an anchor post.

Given a user-selected anchor post, all comments associated with it and their attributes are retrieved by the *Retrieval Service* via the social media data API. These comments and aforementioned attributes are stored in a key-value store by the *Persistence Layer* and indexes are created for efficient access to them. Furthermore, for each comment, topics are identified using Latent Dirichlet Allocation (LDA). The *Operator Layer* is responsible for the implementation of the *abstract art-inspired visualization abstraction* of Kandinsky Mobile. Specifically, it realizes the *map*, *paint*, *balance*, *retrieve*, *lookup*, and *update* operations which we shall describe in the sequel. The *Service Layer* exploits these operations to support the four goals of visualization, namely, *canvas construction*, *on-demand conversation visualization*, *canvas search*, and *time travel* (Recall from Section 1.1). A user interacts with these services through the *Presentation Layer*.

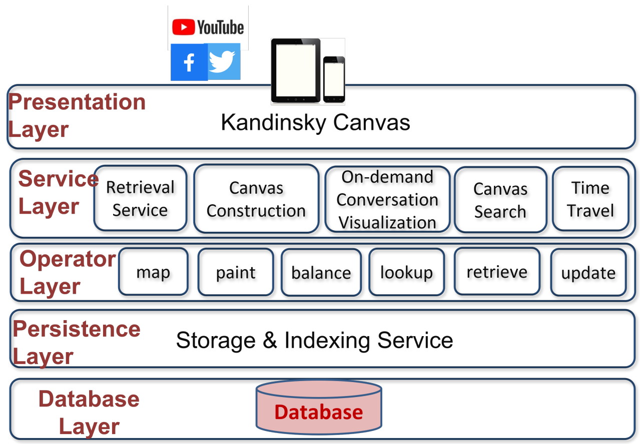


Figure 2: Architecture of Kandinsky Mobile.

4.2 The Data Model

We now present the SCC abstraction on which Kandinsky Mobile is built. Specifically, it aims to address the limitations of list-based abstractions highlighted earlier. At first glance, it may seem that we can address these limitations by switching to a directed graph-based data model where nodes represent comments and edges represent interactions between them (*e.g.,* replies, mentions, retweets). Although such purely computing-driven abstraction is reasonable to visualize social conversation, it becomes cognitively challenging to an end user when the graph size becomes large especially in small-screen mobile devices. In particular, large and dense graphs overload the human perception and cognitive systems, resulting in poor performance of relatively complex tasks such as identifying relationship in graphs [30, 53]. The study by [53] revealed that a graph with more than 100 edges can induce significant cognitive load on the end users. Unfortunately, many online discussion threads of anchor posts involve hundreds to thousands of comments.

The common limitation of list- or graph-based visualization abstraction is their ineffectiveness in exposing aesthetically pleasing bird's eye view of the discussion landscape that may galvanize end users to explore *any* part of the *entire* space on-demand. Hence, we depart from such traditional abstraction schemes and *explore the possibility of drawing inspirations from art to design a data model and operations on social discussions.* The rationale for pursuing this direction is aesthetics and engagement are closely associated with music and art. Integrating them with computational paradigms pave the way to address these limitations effectively.

At first glance, it may seem that the marriage of two disciplines that are poles apart, abstract art and computing, is quixotic. Certainly, as remarked in Section 2, this issue is hardly explored in the context of social discussion visualization. As an initial proof that such endeavour is indeed within the realm of possibility, we draw inspirations from the aesthetic art forms of Kandinsky (Section 3) for designing an abstraction that potentially leads to aesthetically pleasing and engaging visualization of a social discussion landscape.

Several Concentric Circles Data Model (SCC). Every abstraction in computer science consists of a data model and a way of manipulating the data. Social discussions typically adopt a hierarchical structure with the *original comment* (*i.e.,* comment which directly responds to an anchor post) as the root. We refer to it as *conversation network*. Figure 3(a) is an example of a conversation network rooted at Comment 1 where ids to comments are assigned according to their creation time. The SCC data model for visualizing the conversation networks of an anchor post (*i.e.,* discussion landscape) consists of the following:

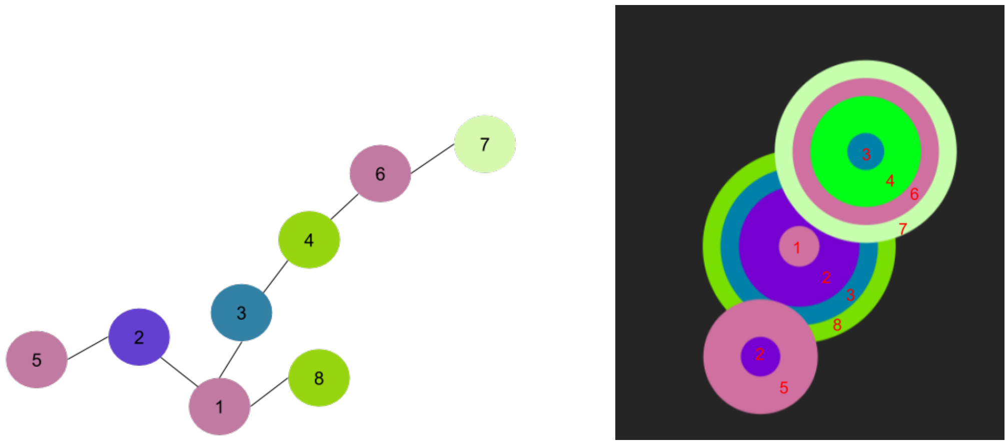


Figure 3: **[Best viewed in color]** A conversation network and its SCC abstraction.

1. A collection of colored shapes C. Each shape $c \in C$ represents a comment in a conversation network.
2. A set of *shape buckets* S (bucket for brevity) where $|S| \leq |C|$ in a 2D display space with dimension D.
3. A *map* function map from C to S. Each shape bucket s contains a subset of those shapes c of C such that $map(c) = s$.

The set of operations are:

1. Compute $map(c)$ where c is a member of C.
2. $paint(\cdot)$ generates a visualization of a shape bucket s or a comment c.
3. $balance(\cdot, D)$ places the shape buckets S or a set of comments C' in the 2D display with dimension D. It returns the (x, y) coordinates of these objects.
4. $retrieve(\cdot)$ retrieves the content of a shape bucket s or a circle/comment c.
5. $lookup(s.c, p)$ returns `true` if c in s satisfies a user-defined *predicate* p.
6. $update(\cdot, param\_list)$ updates $s \in S$ or $s.c$ as specified in the $param\_list$ and returns the updated s or c.

In this work, we consider circle as the shape for representing a comment. We refer to the abstraction as *Several Concentric Circles* (SCC). As we shall see next, the implementation of the *map-paint-balance* abstraction is inspired by the paintings of Kandinsky. Similar to the aforementioned artworks, circles are the *motif*[4] of the SCC abstraction (**K1** in Table 1). We refer to the 2D display space as *Kandinsky Canvas* (canvas for brevity).

[4] Motif is an element or combination of elements repeated often enough in a composition to become a dominant feature in a painting.

Algorithm 1 The map algorithm

```
Require: circles - array of circles representing the comments of the post to visualize
Ensure: the root circle of each shape bucket
function MAP(circles)
    roots ← []
    circleMap ← new Map⟨commentId, circle⟩
    for each circle in circles do
        circleMap.SET(circle.comment.id, circle)
        if circle has parent then
            parentCircle ← circleMap.GET(circle.comment.parentId)
            parentCircle.children.PUSH(circle)
        else
            roots.PUSH(circle)
        end if
    end for
    return roots
end function
```

5 The SCC Operations

In this section, we describe the implementation of the operations in the SCC abstraction. Intuitively, any implementation of the operations aim to facilitate automatic *transformation* of each conversation network to an aesthetically pleasing, color-coded shape bucket and then judicious *placement* of these buckets on the *Kandinsky Canvas* to galvanize search and exploration.

5.1 Map Operation

The map operation $map(\cdot)$ maps each comment $c \in C$ to a shape bucket $s \in S$. Specifically, all comments in a conversation network are mapped to the same shape bucket. For example, all the comments in Figure 3(a) are mapped to the same shape bucket, which is visualized in Figure 3(b) using the paint operation. Hence, if there are k conversation networks associated with an anchor post, then k shape buckets are created (*i.e.,* $|S| = k$) and the comments are distributed to these buckets.

Algorithm 1 outlines the map operation. It accepts the array of circles to display on the canvas as input and returns the root circle of every shape bucket. The return value is chosen so that the buckets can then be passed to the paint operation, which is often invoked after the map operation. The time complexity of the operation is $O(|C|)$ where $|C|$ is the number of circles.

5.2 Paint Operation

Intuitively, the paint operation $paint(\cdot)$ is responsible for "painting" on the *Kandinsky Canvas*. Since the SCC data model involves two types of data, circles (*i.e.,* comments) and shape buckets, which leads to two primary goals: generating visual representations for a shape bucket and the comments it contains. We discuss them in turn.

Visual representation of a shape bucket. "Painting" a shape bucket on the canvas is inspired by Kandinsky's artworks. Each comment in a shape bucket s is represented by a colored circle where distinct colors (or color hues) are used as identifiers of the commentators within a conversation network N_s represented by the bucket (**K1** in Table 1). Color-coding these circles is a natural way to gain attention [16, 44] and visually convey different contributors and comments involved in a conversation. Replies to a comment are visualized as *concentric circles* (**K2** in Table 1). The radius of a circle is a function of the number of likes and its position in the sequence of concentricity represents its

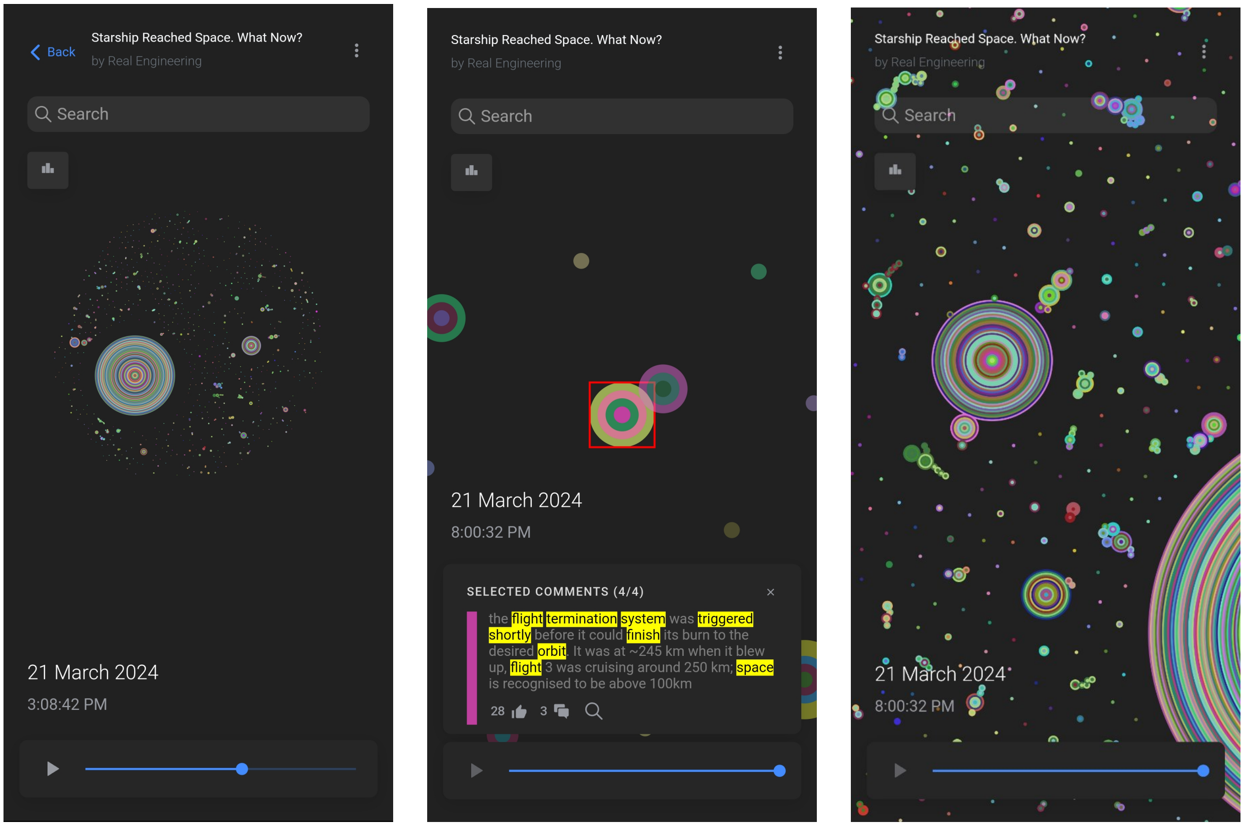


Figure 4: **[Best viewed in color]** Screenshots of Kandinsky Mobile: (left) Bird's eye view of social conversation associated with an anchor post; (middle) Comments associated with a discussion thread; (right) Examples of concentric circles and peripherals.

chronological order (ties are broken arbitrarily) in N_s. The innermost circle of a concentric circle is called a *pivot*. Intuitively, there are two types of concentric circles, *nucleus* and *peripheral*. Concentric circle resulted from an original comment and its replies is referred to as *nucleus*. In other words, in a conversation network, the root and its children form the nucleus with the root being the pivot. On the other hand, replies at depth equal to or greater than two in a conversation network (*e.g.,* replies to a reply) form separate concentric circles, referred to as *peripherals*. Peripherals are positioned on the periphery of the nucleus (**K3**).

The paint operation takes the comments associated with a shape bucket s as input and constructs the conversation network N_s. Then it generates the concentric circles by traversing N_s to visually represent the shape bucket. First, the pivot and nucleus are created and then the peripherals, if any. The size of a shape bucket is set to at least 9.2 mm by default based on the recommendation from Parhi *et al.* [45] that target sizes should be at least this value for discrete single-target tasks in *fat thumb-friendly* devices without degrading performance and preference. Note that most mobile users rely on single-touch and most use their thumbs as the only input method [5, 10]. The color codes of commentators are assigned based on the number of distinct commentators in N_s as well as the selected colors of the adjacent commentators in order to maintain visual contrast[5]. To this end, nodes in N_s are annotated with color codes that are chosen based on the commentators and colors of their parents and siblings (**K4**). These codes are then exploited during the generation of the concentric circles.

[5]Contrast in arts is a large difference between two things; for example, warm and cool colors, light and shadow. Contrasting colors and textures add excitement, emphasis and interest to a work of art.

Algorithm 2 The paint algorithm.

```
Require: buckets - the array of root circles of each shape bucket
Ensure: the nucleus concentric circle of each shape bucket
function PAINT(buckets)
    shapeBuckets ← []
    Generate radius scale from like count of all comments
    for each bucket in buckets do
        PAINTCIRCLE(bucket)
    end for
    for each bucket in buckets do
        ConcentricCircle ← PAINTCONCENTRICCIRCLE(bucket)
        shapeBuckets.PUSH(concentricCircle)
    end for
    return shapeBuckets
end function
```

For example, consider Figure 3(b), which "paints" the shape bucket associated with the conversation network in Figure 3(a). The color codes of the circles follow that of the comments. The paint operation creates the nucleus using Comments 1 (pivot), 2, 3, and 8. Note that 1 is the original post whereas 2, 3, and 8 are replies to it. Since the depths of Comments 4, 5, 6, and 7 in the network are greater or equal to 2, two peripherals are generated by the paint operation with Comments 2 and 3 being the pivots, respectively. Note that paint may recursively create peripheral-of-peripheral if the hierarchy of comments is deep (*e.g.,* Figure 4(right)). Observe that conversations with more participants result in painting of a larger and more colorful circles (**K5** in Table 1). On the other hand, the shape bucket of an original comment that has not garnered any comments degenerates to a single-color circle. Figure 4 (left) depicts results of the paint operation on the shape buckets associated with an anchor post.

Algorithm 2 outlines the paint operation. It invokes two procedures, *PaintCircle* and *PaintConcentricCircle*, to paint individual circles and concentric circles, respectively. The *PaintCircle* is a recursive function such that given a pivot circle, it paints the circle by modifying its visual properties and is recursively called on the next circle in the sequence of the bucket. The *PaintConcentricCircle* method, on the other hand, is another recursive function that paints concentric circles as a whole. Given a pivot circle, it then creates the concentric circle object by grouping this pivot with its replies. It is then recursively called to generate the peripheral concentric circles of replies that are of depth 2 with respect to this pivot circle (*i.e.,* reply to a reply). The time complexity of the operation is $O(|C|)$ where $|C|$ is the number of circles.

Visual representation of a comment. Given a comment associated with a shape bucket, topics in a textual content are highlighted in *yellow* (a primary color in Kandinsky's artworks) by paint. Colored bars are drawn on each comment item to represent the contributor (same color code) and the comment's *normalized* like count (using thickness of the bar). Figure 4 (middle) depicts an example of a comment.

Connection to visualization theories. Observe that the visual representation of shape buckets is grounded on the *expressiveness* and *effectiveness criteria* of visualization [39]. The former determines whether a graphical language can express the desired information whereas the latter identify which of these graphical language is the most effective for a given situation at exploiting the capabilities of the output medium and the human vision system. Specifically, Mackinlay [39] identified and ranked perceptual tasks for encoding quantitative, ordinal and nominal data. The paint operation is cognizant of these visualization theories and principles. Specifically, it only encodes comments that are associated with a conversation network – nothing more, nothing less. It utilizes perceptual tasks such as position, shape, color hue and area for effective visualization. Since the set of comments is nominal data, position and color hue are the most effective perceptual tasks for representing them. Furthermore, it uses the area of a (concentric) circle to capture the number of comments garnered by a conversation network. Since this is a quantitative data, area is considered as moderately effective in the ranking of perceptual tasks, which is exploited by Kandinsky Mobile. Note that higher-ranked tasks for quantitative data (*e.g.,* length, angle, slope) are not conducive for visually representing shape buckets.

5.3 Balance Operation

Given the visual representations of shape buckets and comments, the goal of the balance[6] operation *balance*$(\cdot)$ is to arrange them on the *Kandinsky Canvas* with dimension D *effectively*. It operates at two levels, the set of shape buckets and comments in a shape bucket. We discuss them in turn.

Arrangement of shape buckets. Given the set of shape buckets representing k conversation networks, we need to *effectively* place them on a mobile device interface with dimension D. This is a challenging problem since D is small and there can be many conversation networks that are large. First, the placement of the shape buckets has to be *fat thumb-friendly* [8, 10]. That is, the size of the objects and distance between them need to be palatable to fat thumb interaction. Second, it is paramount to maintain the visual clarity of the bird's-eye view of the entire discussion landscape on the canvas. These can be conflicting goals when k grows. Due to fixed D, larger k restricts the distance between the shape buckets and their sizes, adversely impacting effective data visualization. Hence, it is essential to judicially arrange the shape buckets on the canvas such that "significant" conversations are visualized effectively to engage end users and then use pinch-and-spread gesture to explore various parts of the canvas.

Given a set of visual representation of the shape buckets, the balance operation extends force-directed simulation layout and exploits fat thumb-friendly heuristics to implement it. The minimum space between elements in a mobile device is recommended to 2mm [8]. Although this is easy to achieve for icons (*e.g.,* search), there may be too many shape buckets on the canvas to impose the recommended size. Furthermore, when the rest fingers are grasping the device, the range of motion available to the thumb is more limited, thus limiting the area of the screen a user's thumb can reach. Hence, the balance operation utilizes the following three strategies:

1. Shape buckets that correspond to conversations that have garnered significant attention are placed at around center of the screen for fat thumb-friendly access [24, 25] (**K5**). In particular, touch targets in the center can be smaller - as small as 7mm, while corner target sizes must be about 12mm [25]. For example, conversations that have garnered significant attention[7] in Figure 4 (left) appear in the center of the screen.

2. A padded collision force of 30px (7.94mm) is added to the force-directed simulation layout to make the concentric circles push each other, avoiding overlaps and providing adequate space for fat-thumb support (**K6**).

3. Zooming is enabled so that one can zoom into a specific area of the display to view shape buckets in that area.

Based on the above strategies the balance operation assigns a position (*i.e.,* (x, y) co-ordinates of the pivot) to each shape bucket and returns the set of shape buckets for display on the canvas. Algorithm 3 outlines the procedure for arranging the shape buckets. The worst-case time complexity is $O(klogk + k^2)$

[6] Balance in arts is a principle of design. The arrangement of elements in a work of art (including size and number of objects) that achieves a sense of equality.

[7] In arts, this is called *focal point* – the most important part or area in a work of art.

Algorithm 3 The balance algorithm.

Require: Shape buckets S

Ensure: *none*

```
function BALANCE(concentricCircle S)
    S.EnableZoom() /* Strategy #3 */
    S.InitializeGravitionalForce() /* Strategy #1 */
    S.InitialiseRepellentForce() /* Strategy #2 */
    S.ForceSimulation()
end function
```

where k is the total number of concentric circles. It may invoke the update operation (discussed later) to modify the sizes of the concentric circles in a shape bucket, if necessary, to realize the placement of the buckets in *incremental mode*. Observe that the visual representation and arrangement of the shape buckets are aligned with the similarity, continuation, closure, proximity, and symmetry of Gestalt principles for describing visual perception [31, 47].

Arrangement of comments. Due to limited space on mobile devices, balance restricts the arrangement of a comment to fixed height adjusted as a portion of the device's length. To enhance user navigation, it is positioned at the bottom of the device to facilitate fat-thumb interaction and vertical scrolling (*e.g.,* Figure 4 (middle)).

5.4 Retrieve Operation

The retrieve operation $retrieve(\cdot)$ is designed to efficiently retrieve comments associated with a shape bucket. Given a selected shape bucket s (*e.g.,* through touch), it retrieves the content of the conversation associated with s (*e.g.,* Figure 4 (middle)) by leveraging the indexes. The current implementation supports two types of selection: (a) If one selects anywhere inside a shape bucket then the *entire* conversation network is retrieved. That is, we do not allow fine-grained selection of an individual concentric circle since it can be physically challenging in a mobile device. (b) If a peripheral is selected, then the conversation associated with it is retrieved only.

5.5 Lookup Operation

The lookup operation $lookup(\cdot)$ on a circle c (comment) in a shape bucket returns a Boolean value depending on whether the c satisfies a user-specified predicate p. Note that p can be a set of keywords (*e.g.,* "Florida") or a comment. In the case of the former, it returns true if c contains the keywords. In the latter case, it returns true if c contains comment that is *similar* to the specified comment.

5.6 Update Operation

The update operation $update(\cdot)$ updates a shape bucket on the *Kandinsky Canvas* by modifying various properties on demand. Given a shape bucket s, it may add new circles (*i.e.,* new comments in the conversation network) to s or modify properties such as color, size, and positions of existing circles in s or *highlight* it. Typically, it is invoked by the aforementioned operations. For instance, in paint the number of peripherals influences the size of the circles since we do not want the peripherals to obscure the nucleus or one another. Hence, it is necessary to judiciously set the size and positioning of them to optimize perceptibility. As more circles are added to a shape bucket, the update operation maintains these properties accordingly. Similarly, it can resize a shape bucket in response to the balance operation to optimize arrangement of shape buckets on the canvas. The lookup operation also invokes

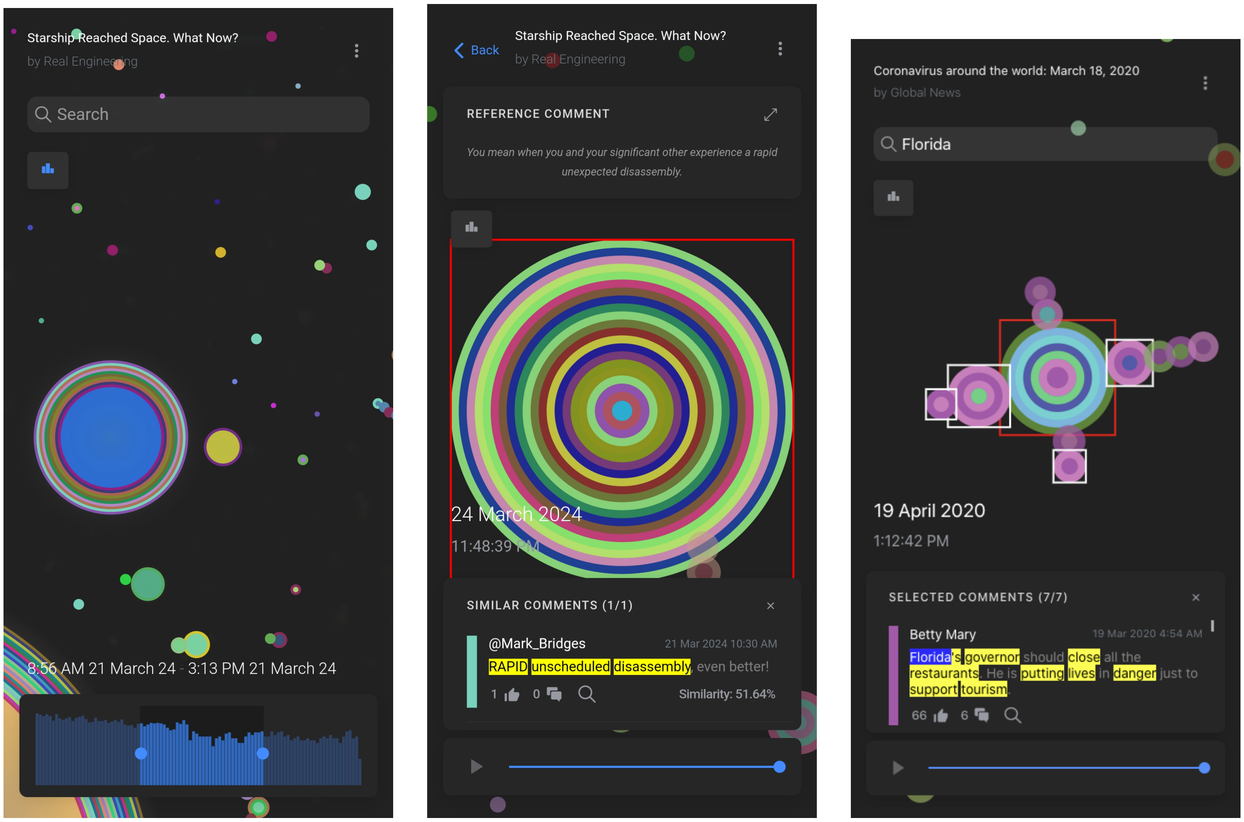


Figure 5: **[Best viewed in color]** Screenshots of Kandinsky Mobile: (left) Time travel; (middle) Similar comment search; (right) Keyword search.

it to highlight a shape bucket that satisfies the predicate by drawing a white boundary box around a circle/bucket (**K7**). An example is given in Figure 5 (right).

6 Implementing The Visualization Goals

In this section, we first briefly describe how the proposed SCC abstraction enables us to achieve the visualization goals mentioned in Section 1 with ease. We invite the reader to view the video of Kandinsky Mobile at `https://youtu.be/ay_0LpRRBQE` for an example. Next, we highlight the superiority of SCC abstraction to the popular list-based abstraction widely used in social media w.r.t. our goals.

6.1 SCC-based Visualization

Goal 1: Bird's eye view of discussion landscape. It is intuitive to visualize the discussions landscape associated with an anchor post using the SCC abstraction. Specifically, we provide two ways to generate a bird's eye view, *spectral* and *incremental*. In the *spectral mode*, the entire set of conversation networks is displayed on the canvas in one shot. This can be achieved by the following sequence of operations: [map, paint, balance]. In the *incremental mode*, the canvas is incrementally updated according to the chronological order of the comments. That is, the shape buckets are visualized incrementally over time. Note that this can be achieved by invoking the update operation with the aforementioned operations. Figure 4 (left) depicts a bird's eye view of an anchor post. The canvas has a dark background similar to *Several Circles* and *Heavy Circles* (**K8**). Observe that a user gets a fair

Table 2: Anchor posts for experimental study.

PostID	Category	Num. of comments	PostID	Category	Num. of comments
FJ1wak5zOhQ	News & Politics	63	G0WtbB-G6fc	Self-help	2564
tTBsGArSCuE	Self-help	127	_bcfxty39Cw	Science & Technology	2657
PfPelzWeSvY	Self-help	240	mY8RqxCt_Fk	Entertainment	2759
gn4S7lQ111M	Education	278	HHPX_QdI804	Education	2817
pvo6hb8rs40	Entertainment	417	bWr-DA5Wjfw	News & Politics	2891
5LpaVYxTedE	Education	465	qOOnBTaHG_Q	Education	3003
-Nm9l5qRWQU	News & Politics	602	HffWFd_6bJ0	Self-help	3069
m6qFCdHvYuI	Science & Technology	747	TlSOMfDX-yY	Science & Technology	3228
NFLjeyd2M0k	Science & Technology	819	2SkLBP0QtSQ	Entertainment	3307
g1pb2aK2we4	Education	961	eaFyTIsowqY	Science & Technology	3444
RThSD70hiTg	Self-help	1059	Ra4W_aztfHA	News & Politics	3545
oRt5NPY7yYA	Science & Technology	1163	P6FORpg0KVo	Self-help	3612
DAKRcS_XeEI	News & Politics	1211	B9SptdjpJBQ	Self-help	3696
EFRUL7vKdU8	Science & Technology	1332	qu6GMpQfyUU	Education	3799
pA5hNeJe1_I	Entertainment	1418	sm18t0m0hRc	Entertainment	3938
VsBSCj2oqwU	Entertainment	1458	MA7TQNExbRg	Education	4079
HFYv-rk4v9Y	News & Politics	1615	tPuOeg4zGQQ	News & Politics	4217
Y2M2Dgmm958	Self-help	1701	xX96xng7sAE	Education	4306
BaptBhVD1r0	News & Politics	1775	YMZcp0EQO2s	Self-help	4356
wl_s54MBsmY	Entertainment	1910	3kwDVw0u4Kw	Science & Technology	4535
c-jALYCMOz0	Entertainment	1998	-CcQ4jKn8aE	News & Politics	4616
rD5goS69LT4	Self-help	2095	sXJg9J81acY	Science & Technology	4685
8AaPHNlrwFU	Entertainment	2159	0r2x7G0hwCw	Education	4768
M2pzKxMNKOM	Education	2290	fkbZ60JXXB8	Entertainment	4921
1CHt6Yo6sVE	News & Politics	2430	kRzgCylePjk	Science & Technology	4984

Table 3: Distribution of comments in anchor posts used in experiments.

Number of Comments	Number of Samples
1-1000	10
1001-2000	10
2001-3000	10
3001-4000	10
4001-5000	10

opportunity to *quickly* select *any* of the shape buckets to peruse the corresponding conversation instead of getting steered by ML-based algorithms.

Goal 2: On-demand visualization of conversation. Given the constructed *Kandinsky Canvas*, one can explore the canvas space and select any shape bucket or peripheral to view the corresponding comments. This is achieved by the following sequence of operations on the selected bucket or peripheral: [retrieve, paint, balance].

Goal 3: Canvas search. The *canvas search problem* takes the user-specified keywords (resp. comment) as input and returns an updated canvas highlighting shape buckets containing the keywords (resp. similar comments). The sequence of operations to realize this is [lookup, update]. For each circle in a shape bucket, the lookup operation returns true if it contains the user-specified keyword or similar comment. Then the update operation surrounds the corresponding shape bucket or peripheral with a white bounding box (**K7** in Table 1). Figure 5 (right) depicts an example of keyword search. For similar comment search, it also lowers the color intensity of the shape buckets (*i.e.,* modifies the color properties of the circles) that do not contain similar comments to make them less visible (**K9**). Selecting one of the highlighted shape buckets retrieves and displays its comments using [retrieve, paint, update, balance] where the update operation highlights the matching keywords in yellow. Figure 5 (middle) depicts an example of similar comment search. The reference comment is shown at the top and similar comment is displayed below along with the shape bucket.

Goal 4: Time travel. Lastly, the SCC operations enable time travel to visualize the discussion landscape (*i.e., Kandinsky Canvas*) at a particular time period (referred to as *Spectrum Filter* in

Table 4: Experimental environment.

Machine	Xiaomi Poco X3 Pro
Operating System / Version	Android OS / Version 12
Processor	Qualcomm Snapdragon 860
Memory	8GB
Internet Connection / Speed	Wi-Fi Connection / 76.8Mbps

Kandinsky Mobile). Given a user-specified time duration $[t_1, t_2]$ and the canvas, comments that are not published within the specified time frame are blurred out (**K9**). This can be achieved by invoking update to "repaint". That is, change the color property of the circles in each shape bucket that are not published between t_1 and t_2. Figure 5 (left) depicts an example.

6.2 List-based vs SCC Abstractions

All major commercial social media sites visualize social discussions in form of scrollable and expandable list. In particular, comments that are automatically exposed to a user are selected by secret-sauce machine learning-based algorithms that aim to optimize engagement and profit. Such abstraction is suboptimal in providing a bird's eye view of the discussion landscape. In contrast, the SCC abstraction enables end users to visualize and explore the entire landscape with ease. Importantly, it gives control back to them to decide which conversation network they would like to engage with instead of feeding them conversations selected by opaque ML algorithms. Furthermore, SCC has a richer set of operations that support aesthetically pleasing visualization and on-demand search and exploration of the social discussions. Lastly, SCC can be easily used to create effective, aesthetically pleasing visualization for solutions to important problems. For example, one can use any state-of-the-art technique to detect sentiments [34] of the discussions and then invoke the update operation to modify the properties of shape buckets to visualize different sentiments in an conversation. Similarly, it can emphasize the comments of opinion leaders in an anchor post by dimming the colors of circles associated with non-leaders. Such visualization is not easily realizable in existing list-based abstraction.

7 Performance Study

Kandinsky Mobile is built using *Ionic 4*, a framework for building hybrid mobile applications. It currently supports anchor posts in *YouTube*. In this section, we report preliminary performance results of Kandinsky Mobile. We also report an initial small-scale user study to report the potential usefulness of the abstract art-inspired SCC abstraction. The codebase is available at `https://github.com/chun-leong/kandinsky-mobile`.

7.1 Efficiency and Scalability

Experimental Setup. We selected 50 *YouTube* anchor posts under five different categories (news & politics, self-help, education, entertainment, and science & technology) containing up to around 5000 comments. Each category contains 10 anchor posts. Table 2 reports the *postid*, *category*, and the *number of comments* of these anchor posts. Table 3 shows the distribution of the comments.

Table 4 reports the system environment for our experimental study. We ensure that no other application processes are running on the machine. Also, we ensure that any Kandinsky Mobile application data is deleted prior to running each experiment.

An experiment on each anchor post is executed 3 times to obtain the average runtime performance. In order to measure end-to-end performance of Kandinsky Mobile, instead of measuring the performance

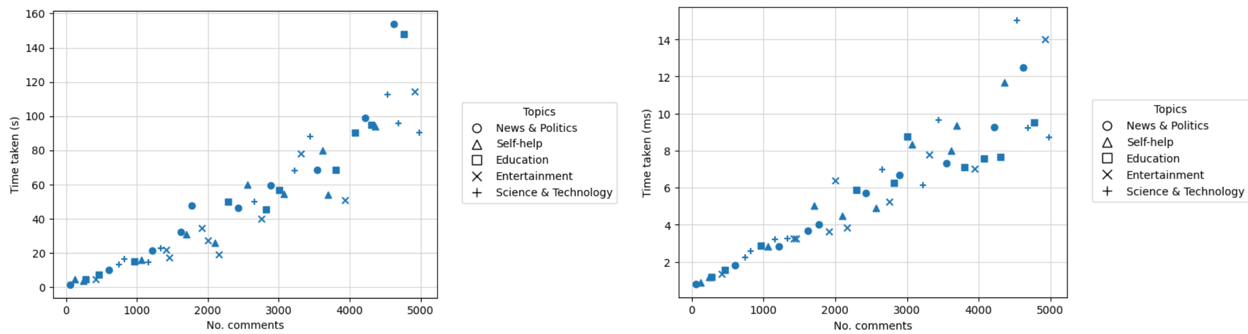


Figure 6: Performance of the extraction stage: (left) Response time; (right) Storage time for post metadata.

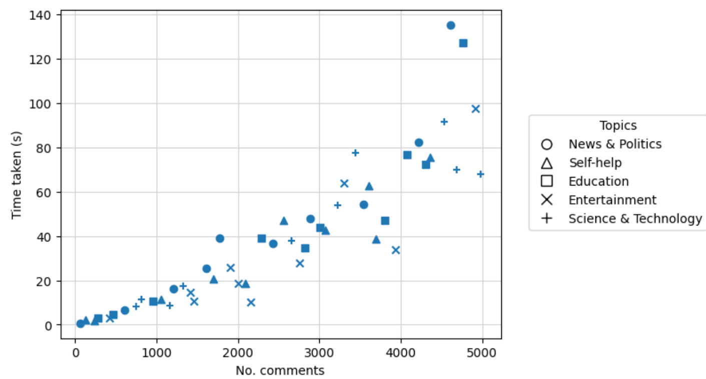


Figure 7: Performance of the representation stage.

of individual SCC operations, we measure the performances of the three stages associated with any anchor post: *extraction*, *representation*, and *canvas construction*. We elaborate on them in the turn.

Extraction. The *extraction* stage in Kandinsky Mobile is responsible for retrieving the post and comment data as well as converting them into a form usable by the other services in the Kandinsky Mobile application. The time taken for the extraction stage includes fetching the data through the social media platform's data API, parsing the information, and saving the data to the persistent storage. Note that these are the processes needed to be completed before the anchor post data can be utilized by Kandinsky Mobile.

Figure 6 (left) plots the response time of the extraction stage. We can make the following observations. First, it grows linearly with the number of comments. Second, given that a large majority of anchor posts typically draw lesser than 3000 comments, the extraction stage can be completed within a minute for these posts.

Note that majority of the extraction time is dominated by the data retrieval step through the *YouTube* API. Figure 6 (right) reports the time required to save each post's metadata to the persistent storage. Observe that it grows linearly with the number of comments and can be completed efficiently.

Representation. The *representation* stage in Kandinsky Mobile performs the analysis of comment data. Specifically, in this stage topic modeling is performed, which has a time complexity of $O(m)$ where where m is the total number of words in all comments. The results obtained from performing the representation stage are used in the exploration of the conversation landscape. Figure 7 plots the results. Observe that for majority of the anchor posts, the representation stage can be performed within a minute.

Canvas Construction. This stage is responsible for representing and visualizing comments on the

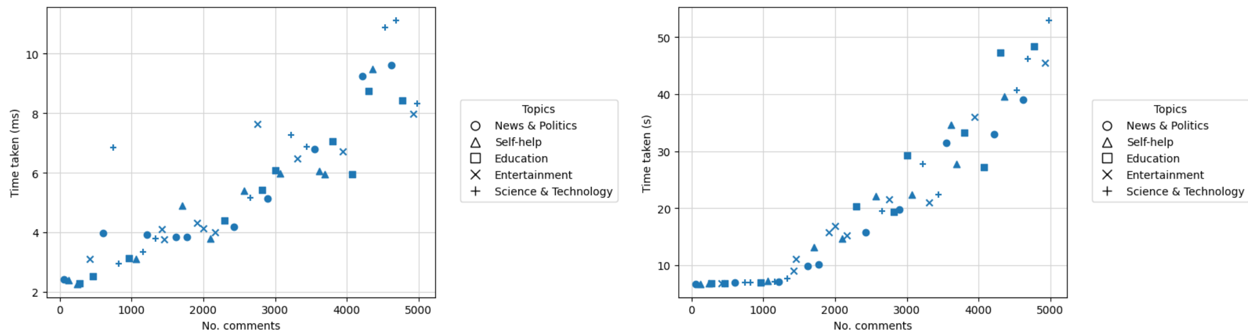


Figure 8: Performance of the canvas construction stage: (left) Canvas preparation; (right) Canvas drawing.

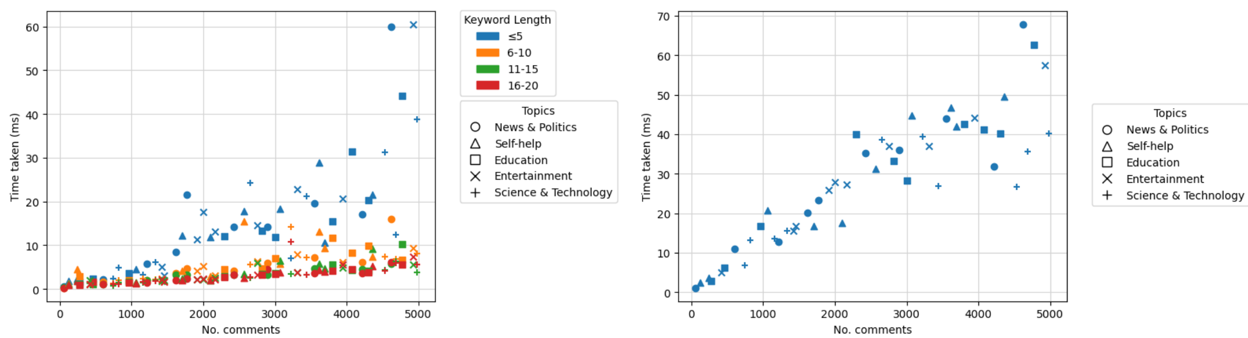


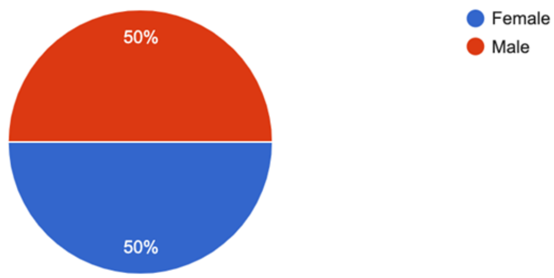
Figure 9: **[Best viewed in color]** Performance of the visualization goals: (left) Keyword search; (right) Similar comment search.

canvas. In other words, it performs the *Canvas Construction Service*, which invokes the map, paint, and balance operators. This stage is the final stage of visualizing the conversation landscape of a post. After which, users may carry out their exploration of the landscape as they desire.

The canvas construction stage can be decomposed into two sub-stages, *canvas preparation* and *canvas drawing*. *Canvas preparation* includes generating the circle entities from the comments. Hence, the map operation is performed in this sub-stage. *Canvas drawing* includes performing the paint and balance operations on the canvas entities. Figure 8 (left) reports the performance of the canvas preparation stage. Observe that the running time increases linearly to the number of comments. It is very efficient and can finish processing the largest number of comments within 12 msec.

The performance of canvas drawing is reported in Figure 8 (right). Observe that we can generate the canvas within a minute for posts with around 5000 comments. For majority of posts with less than 3000 comments, it can generate the canvas within 20 seconds.

Performance of our visualization goals. Lastly, we report the performance of some of the visualization goals described in Section 6. The bird's eye view generation time (Goal 1) is similar to the canvas construction time reported above. The running time for canvas search (Goal 3) and time travel (Goal 4) is less than 70 msec and hence highly interactive. Figure 9 plots the performance of keyword search for different keyword set size and similar comment search. Observe that they grow linearly with the number of comments.



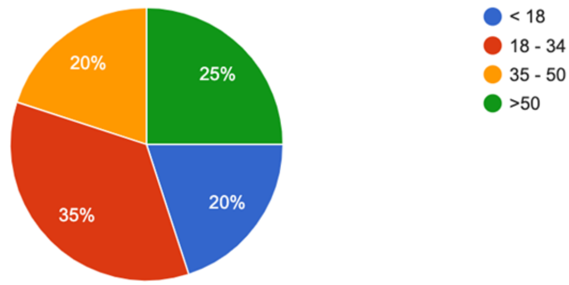


Figure 10: **[Best viewed in color]** Profile of participants in the user study.

7.2 User Study

Next, we undertake a small-scale user study to investigate the usefulness and usability of Kandinsky Mobile w.r.t. the four visualization goals. Specifically, we investigate the following research questions:

RQ1: How is the experience, engagement, and usability of Kandinsky Mobile as an end-to-end mobile app?

RQ2: How effective is Kandinsky Mobile w.r.t. the four visualization goals?

RQ3: How does the SCC-based abstraction in Kandinsky Mobile compare with the popular list-based abstraction?

Note that we do not focus on how closely the design of Kandinsky Mobile is aligned to the artworks of Kandinsky, as it would be unrealistic to expect our participants to possess expertise in his body of work. We first present the setup for our user study and then highlight the key results.

Participants profile. 20 unpaid volunteers initially participated in the study in accordance to HCI research that recommends at least 10 participants [19, 36]. The gender and age distributions of the participants are plotted in Figure 10. As Kandinsky Mobile is intended for users with normal vision and familiarity with social interactions on social media platforms, we collected feedback from them regarding their experiences with social media posts, usage of mobile devices, and experience in using digital technology. All subjects reported to have normal vision. Figure 11 depicts the survey form and the results on a Likert scale of 1-10 with higher the score the greater is the agreement with the statement. Observe that majority of the users use their mobile phones a lot, regularly interact with social media posts, and are savvy with digital technology. Specifically, three participants (two female, one male) indicated limited exposure to social media posts. Consequently, we replaced them with three individuals of the same gender and demographic background who had adequate experience with social conversations, to better isolate the impact of Kandinsky Mobile from confounding factors such as limited exposure to social media discourse.

Tasks. Next, we presented a scripted tutorial on how to use Kandinsky Mobile. All participants were then requested to use their own mobile phones to use it and explore social discussions in *YouTube*. The specific anchor posts were chosen by them and they were instructed to perform the following tasks using *both* list-based and SCC-based visualizations for each post in random order:

(a) find conversations that have attracted significant discussions (related to Goal 1);

(b) select conversations of interest and read them (related to Goal 2);

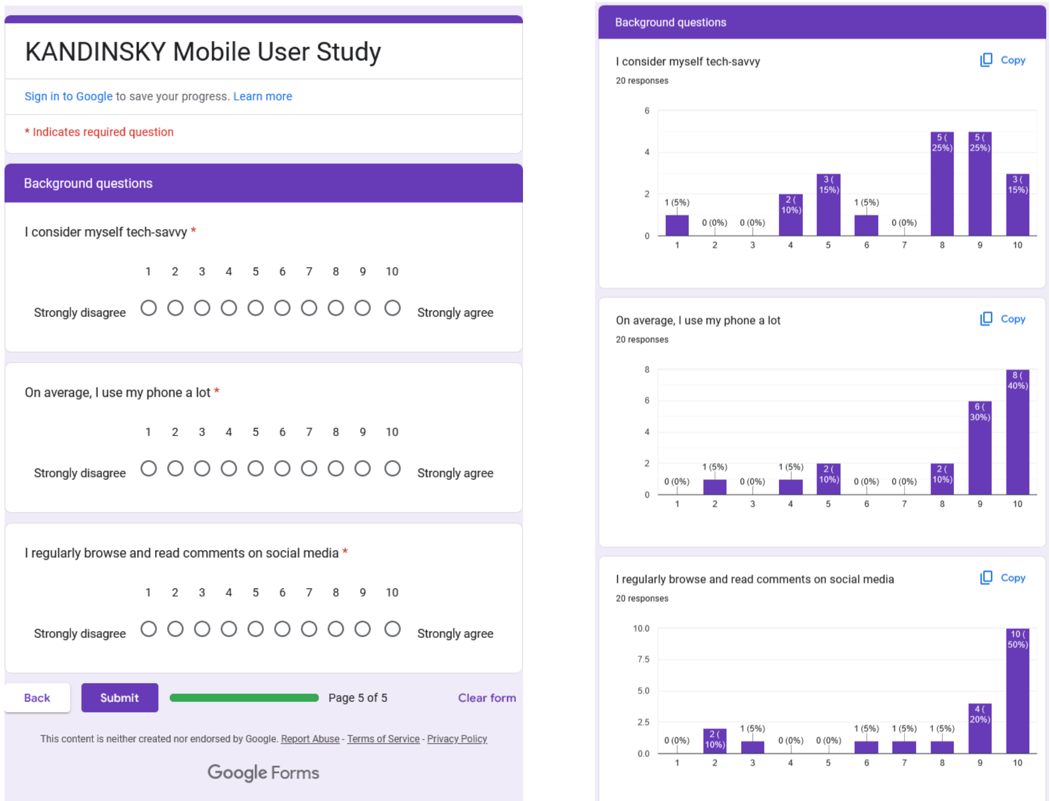


Figure 11: Experience of participants of the user study.

(c) find discussions containing comments with some keywords-of-interest (related to Goal 3);

(d) find similar comments to a chosen comment (related to Goal 3); and

(e) find all discussions that have occurred up to a particular timepoint in the history (related to Goal 4).

Finally, they were asked to answer a survey. Each subject gave a rating in the Likert scale of 1-10 for each question in the survey with higher the score the greater is the agreement with it. On average each participant performed the above tasks on 7.2 anchor posts.

Quantitative Results. We first seek feedback on the overall experience, engagement, and usability of Kandinsky Mobile (**RQ1**). Figure 12 reports the results. Most of them gave positive ratings. Next, we drill down to gather feedback on how effectively Kandinsky Mobile supports the previously stated goals (**RQ2**). The results are illustrated in Figure 13. Notably, participants generally rated its usefulness positively across all four goals, with the lowest rating being 7. Lastly, we requested the subjects to compare their experience of abstract art-inspired visualization with the prevailing list-based visualization of social conversations (**RQ3**). Figure 14 reports the results. While all subjects found the visualization superior to list-based visualization, they found Kandinsky Mobile laggy. This is due to the time taken to construct the *Kandinsky Canvas* as reported earlier.

Subjective Feedback. We also asked the participants to provide unstructured feedback on what they liked (resp. disliked) about Kandinsky Mobile. Table 5 reports some representative feedback. Observe that the subjects likes the usability and aesthetics of the interface and found it innovative. Specifically, they find the canvas "beautiful", highlighting the aesthetic appeal of Kandinsky Mobile. On the other

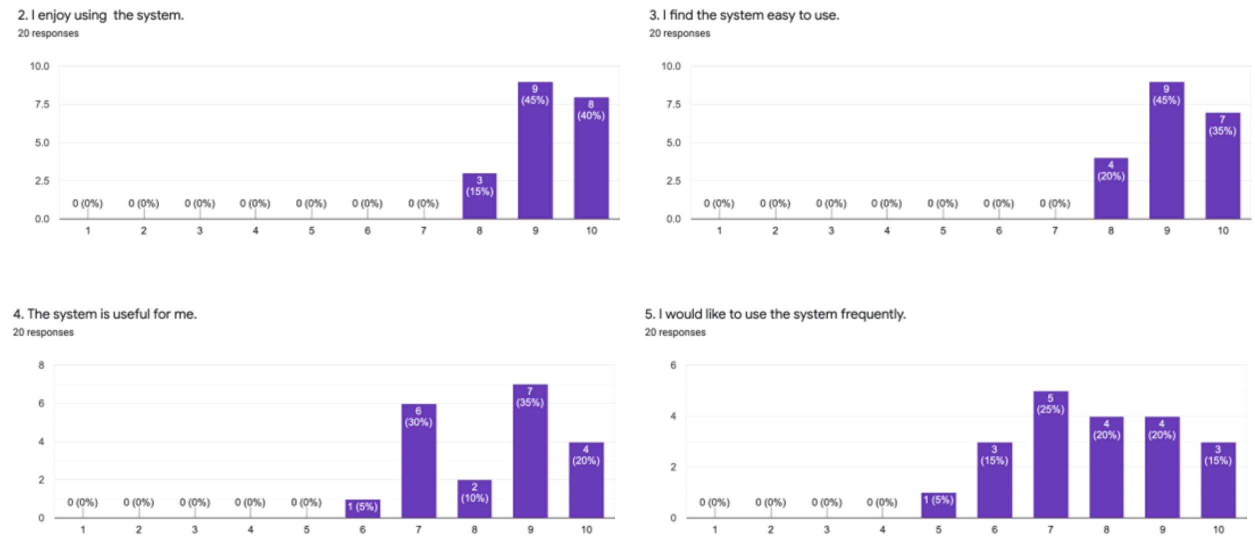


Figure 12: Overall experience with Kandinsky Mobile (**RQ1**).

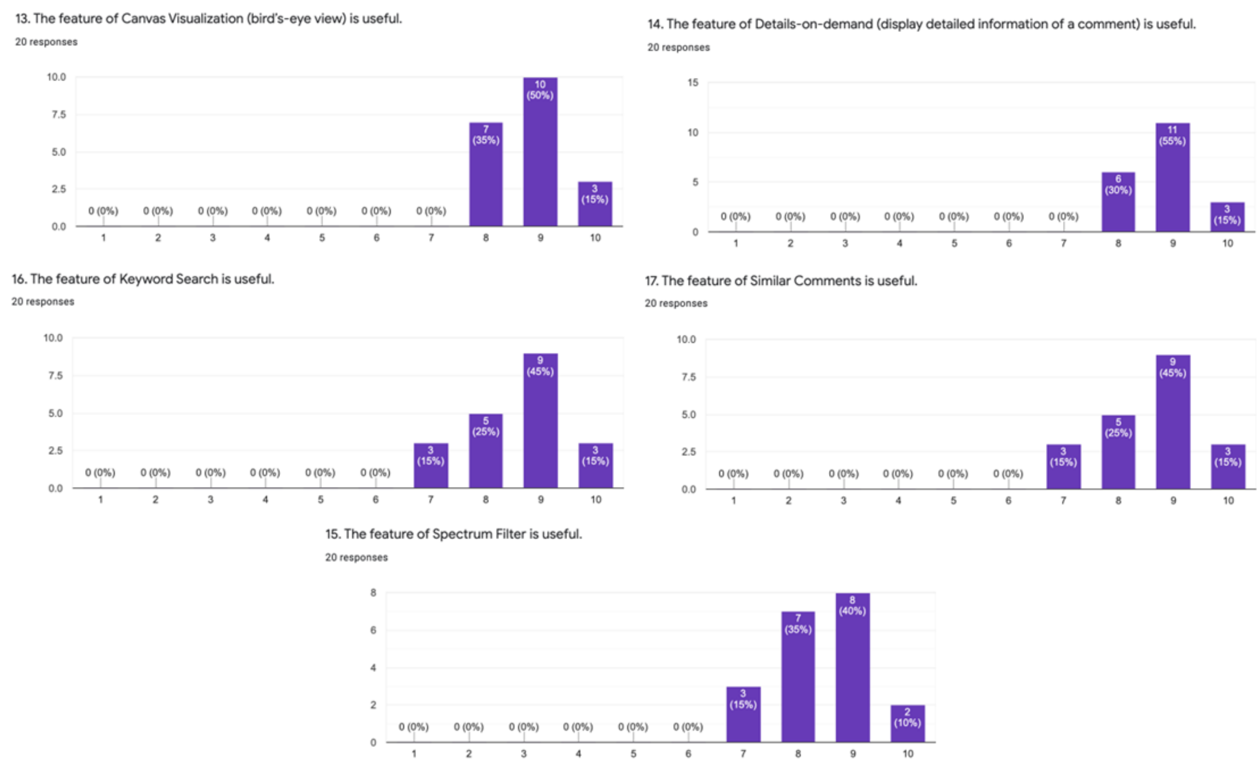


Figure 13: Usefulness of Kandinsky Mobile for supporting the four goals (**RQ2**): (Top) Bird's eye view of the conversation landscape and on-demand visualization of conversation. (Middle) Interactive search; (Bottom) Time travel.

hand, they commented on the delay in canvas construction and pointed out that the current version is limited to the *YouTube* platform only. The participants would like the current framework to be extended

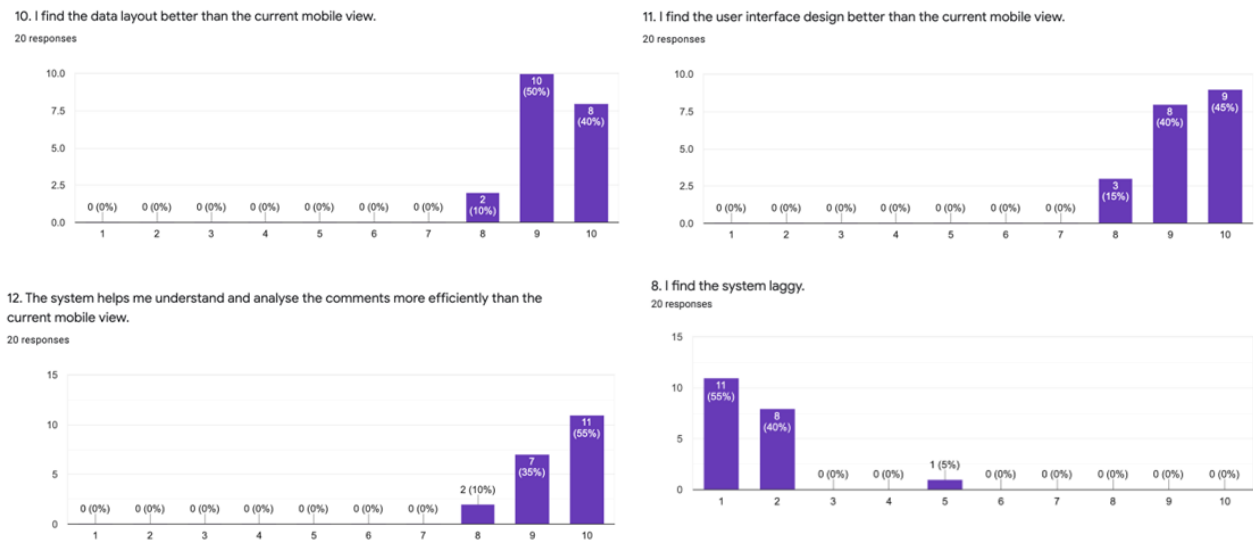


Figure 14: Comparison of abstract art-inspired visualization and list-based visualization (**RQ3**).

Table 5: Subjective feedback.

Question	Response
What do you like most about Kandinsky Mobile?	“easy to use”, “the features and functions are very useful”, “the UI looks nice”, “Nice design”, “Innovative visualization of displaying comments”, “Cool UI”, “Beautiful!”
What do you like least about Kandinsky Mobile?	“Cannot export information”, “Only supports Youtube”, “Time taken to create the canvas can be long for conversations with many comments”
What other features do you think the system should provide?	“Export data”, “Support for more platforms such as fb, X, insta”, “Sentiment analysis”, “Comment summary”, “Support other languages such as Chinese”, “Add to Google Play Store”

to other social media platforms and languages as well as support vertical applications such as sentiment analysis and comment summarization.

8 Reflections & Future Work

This paper introduces a novel visualization paradigm called *Several Concentric Circles* (SCC), which draws inspiration from the abstract art of Wassily Kandinsky and integrates it with digital technology to enable intuitive, visually engaging, and effective exploration of social discussions on mobile platforms. Unlike conventional list-based interfaces that dominate social media applications, SCC emphasizes aesthetic composition and spatial metaphors to represent conversational dynamics in a more immersive and meaningful way. At the heart of this paradigm is a working prototype we developed, named Kandinsky Mobile, which leverages the expressive potential of abstract forms and spatial layout to overcome the cognitive and navigational limitations posed by traditional linear-scrolling interfaces. To the best of our knowledge, this fusion of abstract artistic principles with mobile interaction design for the purpose of social discussion visualization represents a previously unexplored direction, first addressed in our recent demonstration of Kandinsky Mobile [22].

Enhancing Kandinsky Mobile. Our preliminary performance study demonstrates the superior potential of the SCC abstraction for visualizing social conversations. Nevertheless, there are several opportunities for improvement. First, there are opportunities to enhance the efficiency and scalability of

Kandinsky Mobile as highlighted by the participants in our user study. Although the interactive search and time travel can be undertaken efficiently, the performance of canvas construction can be further improved. In this context, while data retrieval time is the primary bottleneck during the extraction stage, both the representation and canvas drawing stages can take up to a minute for very large conversations. Therefore, there is potential for improvement by leveraging parallelism, as the k shape buckets can be potentially "painted" concurrently. Additionally, although the results of our initial user study are encouraging, we plan to conduct a more extensive study with a larger group of participants to gain deeper insights into the practical impact of Kandinsky Mobile.

Second, although the subjects in our user study generally find the SCC-based visualization aesthetically pleasing (Table 5), it is important to undertake a rigorous and systematic analysis of aesthetics of Kandinsky Mobile. In this context, we aim to investigate how color theories can inform the visual representation of a large user base. While Kandinsky's artworks typically feature only a few colored circles arranged on the canvas, Kandinsky Mobile employs a much broader color palette to accommodate the number of users.

Third, in the current framework we assume the target audience to be users with normal vision. It is interesting to explore how we can extend SCC to support individuals suffering from color blindness. Furthermore, as mentioned by some participants, it is important to integrate Kandinsky Mobile with additional social media platforms.

Lastly, anchor posts that garner thousands of comments may demand a 3D visualization of the canvas for more effectiveness. The intuition is to show "significant" conversations first and enable a user to "travel" through the depth of the canvas to explore less significant conversations. Solution to this problem demands marrying ideas from 3D mobile interactions [18, 21] with visual art-based visualization schemes.

Beyond Social Conversations. Although we focus on visualizing social media discussions, the underlying principles of SCC are broadly applicable and can be readily extended to other domains characterized by hierarchical or threaded interactions. One such example is online learning environments, such as discussion boards used in MOOCs or university learning management systems, where student-instructor and peer-to-peer exchanges often form complex, multi-level conversation trees. By adapting SCC to these contexts, it is possible to enhance user engagement, improve navigability, and provide a more aesthetically enriching and cognitively supportive interface for exploring educational dialogues. Beyond discussion boards, a recent study [37] took inspirations from SCC to visualize student feedback.

An intriguing direction for future exploration is the extension of the SCC abstraction to support the aesthetic and effective visualization of large-scale networks. Dunne and Shneiderman [15] proposed a compelling technique in this space, where common network motifs (*fans*, *connectors*, *cliques*) are identified and replaced with compact, visually distinct glyphs. This approach can be viewed as an instantiation of the SCC paradigm, where shapes are circles and lines representing nodes and edges, respectively, and are grouped into higher-level shape buckets that correspond to glyphs. Consequently, the map-paint-balance operations "paint" the glyph-based network on a display space by carefully laying out the shapes and shape buckets (*i.e.,* nodes, edges, and glyphs).

Beyond network visualization, SCC also presents opportunities as a flexible substrate for a variety of vertical applications related to social discussion analysis. For instance, sentiment analysis results could be visually encoded by leveraging color, texture, or opacity of circles–enabling users to intuitively perceive emotional dynamics within complex social conversations. Similarly, it can be the substrate for effective visualization and exploration of online hostility [41].

Acknowledgments. We would like to acknowledge Wen Shan Yap, a professional artist in Singapore for sharing his insights on abstract art and works of Wassily Kandinsky.

References

[1] Circles in a Circle (1923). https://www.philamuseum.org/collection/object/51019.

[2] Several Circles (1926). https://www.guggenheim.org/artwork/1992.

[3] Heavy Circles (1927). https://www.nortonsimon.org/art/detail/P.1953.216/.

[4] Mid Journey. https://www.midjourney.com/home.

[5] Mobile Usability Testing: Fat Finger Syndrome. https://ux247.com/mobile-usability-testing-fat-finger-syndrome/.

[6] Stable Diffusion Model. https://github.com/AUTOMATIC1111/stable-diffusion-webui.

[7] Survey of Commenters and Comment Readers. https://mediaengagement.org/research/survey-of-commenters-and-comment-readers/#:~:text=People%20Read%20Comments%20to%20Learn,or%20amused%20by%20others'%20comments.

[8] UX for Mobile: The Rise of Fat-Finger Design. https://www.outsystems.com/blog/posts/ux-for-mobile-fat-finger-design/.

[9] D. Archambault, T. Munzner, D. Auber. GrouseFlocks: Steerable Exploration of Graph Hierarchy Space. *IEEE Trans. Vis. Comput. Graph.*, 14(4): 900-913, 2008.

[10] S. Boring, D. Ledo, et al. The Fat Thumb. *In MobileHCI*, 2012.

[11] K. Chanthatrojwong, S. S. Bhowmick, B. Choi. LICS: Towards Theory-Informed Effective Visual Abstraction of Property Graph Schemas. *In ACM SIGMOD*, 2025.

[12] K. N. Diaz. Kandinsky's Nonrepresentational Art: Finding Meaning Through Aesthetic Experience. *ETD Collection for University of Texas, El Paso*. AAI1494343, https://scholarworks.utep.edu/dissertations/AAI1494343, 2011.

[13] A. Dix, J. Finlay, G. Abowd, R. Beale. Human-computer Interaction, 2nd edn, *Pearson Education Ltd*, Harlow, 1998.

[14] J. Donath, K. Karahalios, F. B. Viégas. Visualizing Conversation. *In HICSS*, 1999.

[15] C. Dunne, B. Shneiderman. Motif Simplification: Improving Network Visualization Readability with Fan, Connector, and Clique Glyphs. *In CHI*, 2013.

[16] M. A. Dzulkifli, M. F. Mustafar. The Influence of Colour on Memory Performance: A Review. *Malays J Med Sci.*, 20(2): 3-9, 2013.

[17] B. Engdahl, M. Köksal, G. Marsden. Using Treemaps to Visualize Threaded Discussion Forums on PDAs. *In CHI*, 2005.

[18] D. Fiorella, A. Sanna, F. Lamberti. Multi-touch User Interface Evaluation for 3D Object Manipulation on Mobile Devices. *J. Multimodal User Interfaces*, 4(1): 3-10, 2010.

[19] L. Faulkner. Beyond the Five-user Assumption: Benefits of Increased Sample Sizes in Usability Testing. *Behavior Research Methods, Instruments, & Computers*, 35(3), 2003.

[20] V. Giller, R. Melcher, J. Schrammel, R. Sefelin, M. Tscheligi. Usability Evaluations for Multi-device Application Development Three Example Studies. *In the Proceedings of the International Conference on Mobile Human-Computer Interaction*, Springer, Berlin, Heidelberg, 2003.

[21] E. S. Goh, M. S. Sunar, A. W. Ismail. 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review. *IEEE Access*, 7: 40581-40601, 2019.

[22] C. C. F. Gumtang, S. S. Bhowmick. Kandinsky Mobile: Abstract Art-Inspired Interactive Visualization of Social Discussions on Mobile Devices. *In MM*, Interactive Arts Track, 2021.

[23] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. ThemeRiver: Visualizing Thematic Changes in Large Document Collections. *IEEE Transactions on Visualization and Computer Graphics*, 8, 1, 2002.

[24] S. Hoober. Design for Fingers, Touch, and People, Part 1. Retrieved from `https://www.uxmatters.com/mt/archives/2017/03/design-for-fingers-touch-and-people-part-1.php`, March 2017.

[25] S. Hoober. Design for Fingers, Touch and People, Part 3. Retrieved from `https://www.uxmatters.com/mt/archives/2017/07/design-for-fingers-touch-and-people-part-3.php`, July 2017.

[26] S. Hoober, E. Berkman. Designing Mobile Interfaces. *OReilly*, Beijing, 2012.

[27] E. Hoque, G. Carenini. ConVis: A Visual Text Analytic System for Exploring Blog Conversations. *Computer Graphics Forum*, 33, 3, 2014.

[28] I. Hossain, S. Puppala, Md. J. Alam, S. Talukder, Z. Talukder. A Visual Approach to Tracking Emotional Sentiment Dynamics in Social Network Commentaries. *In ICWSM*, 2024.

[29] K. Y. Huang. Challenges in Human-computer Interaction Design for Mobile Devices. *In Proc. of the World Congress on Engineering and Computer Science*, 1(6), 2009.

[30] W. Huang, P. Eades, S.H. Hong. Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective. *Inf. Vis.*, 8(3), 2009.

[31] J. Johnson. Designing With The Mind In Mind: Simple Guide to Understanding User Interface Design Guidelines. *Morgan Kaufmann*, 2020.

[32] W. Kandinsky. Concerning the Spiritual in Art. *Dover Publications, Inc,* 1977.

[33] W. Kandinsky. Point and Line to Plane. *Courier Corporation,* 1979.

[34] W. M. Kouadri, M. Ouziri, S. Benbernou, K. Echihabi, T. Palpanas, I. B. Amor. Quality of Sentiment Analysis Tools: The Reasons of Inconsistency. *PVLDB*, 14(4), 2020.

[35] D. Lauer. Facebook's Ethical Failures are not Accidental; They are Part of the Business Model. *AI and Ethics 1*, 395-403, 2021.

[36] J. Lazar, J.H. Feng, H. Hochheiser. Research Methods in Human-Computer Interaction. John Wiley & Sons, 2010.

[37] F. Lin, C. Li, W. Ying R. Lim, Y.-H. Lee. Empower Instructors with Actionable Insights: Mine and Visualize Student Written Feedback for Instructors' Reflection. *Comput. Educ. Artif. Intell.*, 8, 2025.

[38] C. Lui, S. S. Bhowmick, A. Jatowt. KANDINSKY: Abstract Art-Inspired Visualization of Social Discussions. *In SIGIR*, 2019.

[39] J. Mackinlay. Automating the Design of Graphical Presentations of Relational Information. *ACM Trans. on Graphics*, 5(2), 1986.

[40] J. Ma, S. S. Bhowmick, B. Choi, L. Tay. Theories and Principles Matter: Towards Visually Appealing and Effective Abstraction of Property Graph Queries. *In SIGMOD*, 2023.

[41] D. Murthy, S. Sharma. Visualizing YouTube's Comment Space: Online Hostility as a Networked Phenomena. *New Media Soc.* 21(1), 2019.

[42] C. Nobre, M. D. Meyer, M. Streit, A. Lex. The State of the Art in Visualizing Multivariate Networks. *Comput. Graph. Forum*, 38(3): 807-832, 2019.

[43] M. Noirhomme-Fraiture, F. Randolet, L. Chittaro, G. Custinne. Data visualizations on Small and Very Small Screens. *In Proceedings of the 11th International Symposium on Applied Stochastic Models and Data Analysis*, 2005.

[44] Y. Pan. Attentional Capture by Working Memory Contents. *Can J Exp Psychol.* 64(2), 2012.

[45] P. Parhi, A. K. Karlson, B. B. Bederson. Target Size Study for One-handed Thumb use on Small Touchscreen Devices. *In MobileHCI*, 2006.

[46] D. Raneburger, D. Alonso-Ríos, R. Popp, H. Kaindl, J. Falb. A User Study with GUIs Tailored for Smartphones. *In IFIP Conference on Human-Computer Interaction*, Springer, Berlin, Heidelberg, 2013.

[47] A. S. Reber. Gestalt Psychology. *The Penguin Dictionary of Psychology*, Viking, 1985.

[48] C.-Y. Sung, X.-Y. Huang, Y. Shen, F.-Y. Cherng, W.-C. Lin, H.-C. Wang. ToPIN: A Visual Analysis Tool for Time-anchored Comments in Online Educational Videos. *In CHI Extended Abstracts*, 2016.

[49] T. Tanaka, M. Chosokabe, K. Tanimoto, S. Tsuchiya. Visualization of the Social Atmosphere Using Comments on News Sites. GDN 2022: 105-114

[50] E. Z. Victorelli, et al. Understanding Human-data Interaction: Literature Review and Recommendations for Design. *Int. J. Hum. Comput. Stud.*, 134: 13-32, 2020.

[51] M. Wattenberg, D. Millen. Conversation Thumbnails for Large-scale Discussions. *In CHI Extended Abstracts*, 2003.

[52] R. Xiong, J. Donath. PeopleGarden: Creating Data Portraits for Users. *In UIST*, 1999.

[53] V. Yoghourdjian, et al. Exploring the Limits of Complexity: A Survey of Empirical Studies on Graph Visualization. *Visual Informatics* 2, 2018.

[54] V. Yoghourdjian, T. Dwyer, K. Klein, K. Marriott, M. Wybrow. Graph Thumbnails: Identifying and Comparing Multiple Graphs at a Glance. *IEEE Trans. Vis. Comput. Graph.*, 24(12): 3081-3095, 2018.

[55] A. L. Zhou, Y.-A. Wang, W. Wu, K. Zhang. Human Aesthetic Preference-Based Large Text-to-Image Model Personalization: Kandinsky Generation as an Example. *In CVPR*, 2024.

Non-profit Org.
U.S. Postage
PAID
Los Alamitos, CA
Permit 1398

IEEE Computer Society
10662 Los Vaqueros Circle
Los Alamitos, CA 90720-1314