

How Cohesive Are Community Search Results on Online Social Networks?: An Experimental Evaluation

Yining Zhao
yining002@e.ntu.edu.sg
Nanyang Technological University
Singapore

Nastassja L. Fischer
nastassja.fischer@nie.edu.sg
National Institute of Education, Singapore
Singapore

Sourav S Bhowmick
assourav@ntu.edu.sg
Nanyang Technological University
Singapore

SH Annabel Chen
annabelchen@ntu.edu.sg
Nanyang Technological University
Singapore

Abstract

Recently, numerous *community search* methods for large graphs have been proposed, at the core of which is defining and measuring *cohesion*. This paper experimentally evaluates the effectiveness of these community search algorithms *w.r.t.* cohesiveness in the context of online social networks. Social communities are formed and developed under the influence of group cohesion theory, which has been extensively studied in social psychology. However, current *generic* methods typically measure cohesiveness using *structural* or *attribute-based* approaches and overlook domain-specific concepts such as group cohesion. We introduce five novel *psychology-informed cohesiveness measures*, based on the concept of *group cohesion* from social psychology, and propose a novel framework called CHASE for evaluating eight representative community search algorithms *w.r.t.* these measures on online social networks. Our analysis reveals that there is no clear correlation between structural and psychological cohesiveness, and no algorithm effectively identifies psychologically cohesive communities in online social networks. This study provides new insights that could guide the development of future community search methods.

CCS Concepts

• **Information systems** → **Retrieval effectiveness**; **Data mining**.

Keywords

Community search; Cohesive; Social networks; Empirical analysis

ACM Reference Format:

Yining Zhao, Sourav S Bhowmick, Nastassja L. Fischer, and SH Annabel Chen. 2025. 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)*, July 13–18, 2025, Padua, Italy. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3726302.3729997>



This work is licensed under a Creative Commons Attribution 4.0 International License. *SIGIR '25, Padua, Italy*

© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1592-1/2025/07
<https://doi.org/10.1145/3726302.3729997>

1 Introduction

Real-world networks frequently exhibit community structures [41], making *community search (CS)* a key area of focus in graph search [34]. It seeks to retrieve *cohesive* subgraphs from an input graph based on query nodes. Communities extracted by CS algorithms have potential applications in event organization, recommendations, protein complex identification, and targeted advertising [34, 39]. In particular, human beings thrive in communities, forming connections through shared experiences and values, which promote cooperation, resilience, and a sense of belonging [8, 98]. Thus, community search is particularly important for applications involving human interaction, such as those in online social networks.

Extensive research on the community search problem has primarily focused on enhancing the efficiency, scalability, and quality of results [34]. Central to this issue is the concept of *cohesion* within communities. Most CS algorithms define communities through *subgraph cohesiveness metrics* derived from structural network information, such as *k-core*, *k-truss*, and *k-ECC*, along with attribute data [34]. These measures are *generic*, meaning they are not tied to any specific application domain. Recently, several studies have investigated *learning-based* methods for identifying cohesive subgraphs, which integrate both structural information and vertex attributes into neural network models, thereby removing the need for predefined rules for communities [31, 39, 43, 60, 94, 95]. *In this paper, we evaluate the cohesiveness of communities produced by these existing techniques within online social networks.*

In online social networks, communities are fundamentally characterized as groups of individuals connected through “invisible bonds”, a phenomenon that has been widely studied in social psychology as *cohesion* or *group cohesion* since the 1950s [35, 110]. Most psychology studies recognize cohesion as *multi-dimensional* [18, 70, 80] and have demonstrated its validity in various contexts, such as sports [17, 75], work [66, 73], and technology-mediated learning environments [106], as well as in both offline and online settings [16, 38, 68]. Consequently, algorithmically identified communities can be hypothesized to manifest psychologically grounded cohesion properties. Notably, structural cohesiveness metrics used in the CS problem have their origins in social psychology [37, 96].

Existing CS methods evaluate their solutions in three main ways [29, 59, 63, 71, 94, 95, 107]: (a) by comparing retrieved communities to known ground-truth communities using various performance metrics, though the interpretation and definition of ground

truth vary [55, 57, 69, 101, 104]; (b) by assessing communities against those produced by other algorithms through “effectiveness metrics” or “goodness metrics” that measure community characteristics [10, 63], especially when the ground truth is unavailable; and (c) by conducting case studies to demonstrate the interpretability and significance of the communities identified by CS algorithms.

None of the existing evaluation approaches systematically assesses the effectiveness of cohesiveness measures in networks involving interacting individuals. Consequently, they fail to address key questions specific to online social networks, such as: (1) Do the generic cohesiveness measures from different CS techniques accurately capture group cohesion in online social networks? (2) How cohesive are the communities identified by these techniques when evaluated using measures based on group cohesion in social psychology?

In this paper, we address these key questions about the cohesiveness of communities identified by community search (CS) algorithms in online social networks. We begin by discussing how research on group cohesion can inform the assessment of cohesion in online social networks. We critique the limitations of existing generic cohesiveness measures used by various CS techniques in capturing the multi-dimensional nature of group cohesion and introduce a set of *psychology-informed cohesiveness measures* grounded in social psychology’s definition of cohesion to address them. Additionally, we present a framework called *CHASE* to conduct extensive experiments comparing the cohesiveness of results from eight representative CS techniques using real-world social network data.

Our experimental evaluation reveals three key findings. First, existing algorithms, when using the same query node, often identify communities that differ significantly in terms of structural and psychology-based cohesiveness, with varying query hit rates. Second, recent learning-based algorithms tend to produce communities with low structural cohesiveness or fail to identify valid communities. Third, there is no clear relationship between structural cohesiveness and psychology-based cohesion. More importantly, no current algorithm identifies cohesive communities in online social networks that align with the concept of cohesion in social psychology.

In summary, this work makes the following novel contributions. (a) It is the first study to assess the cohesiveness of social network communities generated by various community search methods, analyzed through the lens of cohesion theories in social psychology. (b) It introduces a set of psychology-informed cohesiveness measures for evaluation, which is unavailable in existing community search evaluation toolkits. (c) The experimental findings presented in this work are novel and have not been reported in previous studies.

2 Related Work

Fang et al. [34] evaluated CS results from traditional algorithms on simple and attributed graphs. For simple graphs, it experimentally assesses the average running time and reports community quality using four generic structure-based measures: diameter, degree, density, and clustering coefficient. For attributed graphs, the evaluation focuses on keyword-based CS solutions for efficiency and F1 scores. In social psychology, cohesion measures are developed based on a specific definition and are tailored to a target population.

Their quality is assessed through *reliability* and *validity* [12], focusing on how well these measures align with theoretical constructs [13, 84]. Recently, Bhowmick et al. [79] introduced the paradigm of *psychology-informed design* and reviewed existing social influence computation techniques that embrace it. None of these efforts undertakes a systematic evaluation of the cohesiveness of CS solutions on online social networks through the lens of social psychology.

3 The Community Search Problem

In this section, we review the community search problem and describe a taxonomy of existing CS algorithms. We begin by introducing some terminology to facilitate the exposition of the paper.

Terminology. An *online social network* can be modeled as a directed multigraph $\mathcal{G} = (\mathcal{U}, \mathcal{E})$, where \mathcal{U} indicates the set of users (vertices) and \mathcal{E} includes edges representing individual and interaction activities. *Individual activity* occurs when a user u_i posts information at time t without targeting anyone, represented as a self-loop $e_{ii}^t \in \mathcal{E}$. *Interaction activity* refers to an interaction between user u_i and u_j at time t , forming a directed edge $e_{ij}^t \in \mathcal{E}$. An *induced subgraph* $H = (\mathcal{U}_H, \mathcal{E}_H)$ of \mathcal{G} consists of a vertex subset from \mathcal{G} and all edges between them.

Problem Definition. Based on [34, 86], the community search problem can be formalized as follows:

PROBLEM 3.1. [Community Search (CS) Problem] *Given a graph $G = (V, E)$ and a set of query nodes $Q \subseteq V$, the community search problem seeks to find a community that contains Q and satisfies the following properties:*

- (i) *Connectivity: vertices in the community are connected;*
- (ii) *Cohesiveness: vertices in the community are intensively linked to each other based on a particular goodness metric.*

Note that \mathcal{G} can be reduced to a simple graph $G = (V, E)$ where $V = \mathcal{U}$ and each set of parallel edges in \mathcal{E} between vertices can be reduced to undirected or directed connections.

Taxonomy of CS Algorithms. Following [95], we classify existing CS works into *traditional* and *learning-based* categories.

Traditional CS algorithms. Generally, traditional CS algorithms formulate a community by specifying its subgraph properties [34]. The commonly adopted cohesiveness properties are structural-based, including *k-core* [6, 28, 86, 103], *k-truss* [2, 47, 65, 107], *k-clique* [105], and *k-ECC* [22, 45, 46]. These metrics are often referred to as *structural cohesiveness metrics* (*structural metrics* for brevity). Additionally, studies incorporate attribute information to formalize cohesiveness by refining existing structural metrics or proposing new *attribute cohesiveness metrics* alongside the structural ones. Examples of the former approach include *attributed pkd-truss community* [99], *(k, θ)-community* [71], *(k, p)-core* [67], and *β -temporal core* [63]. Examples of the latter include *keyword cohesiveness* [33], *spatial cohesiveness* [25, 32], and *query cohesiveness* [29].

Learning-based CS algorithms. Due to the structural inflexibility of traditional CS algorithms, recently learning-based methods have been proposed to retrieve communities by learning their hidden patterns without relying on predefined structures. Depending on their reliance on ground-truth communities for training, they can be further categorized as *supervised* [24, 31, 43, 50, 95], *semi-supervised* [60], and *unsupervised* [94]. However, most learning-based solutions weakly correlate with community cohesiveness, struggle with

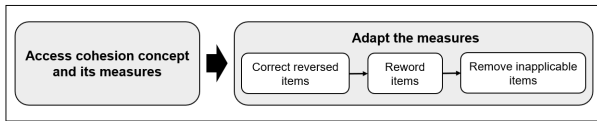


Figure 1: Workflow for adapting cohesiveness measures.

large graphs due to neural network constraints, and risk of model invalidation when the graph topology changes [95].

4 Cohesiveness Framework for Social Networks

This section outlines the methodology for measuring cohesion based on social psychology and highlights the limitations of current generic cohesiveness measures used in community search techniques. The goal is to determine whether these generic measures align with social psychology’s concept of cohesion, which is crucial for effective community retrieval in online social networks.

4.1 Methodology

In social psychology, measurements are developed from proposed concepts or adapted from existing ones to fit different study subjects and contexts [13, 83]. With well-studied cohesion concepts and measures available, in Figure 1, we summarize a workflow to adapt existing cohesiveness measures in social psychology to the context of online social networks. Here, “measures” particularly refer to questionnaires, which are widely used in psychological studies to assess cohesion [36, 81, 106].

We start by selecting a cohesion concept from psychological studies and accessing its *measures*. To ensure consistent positive scoring of cohesiveness, *reversed-word items*, introduced alongside *positively-worded items* to reduce response style bias [89], are corrected. Besides, two common adaptation methods used in group cohesion studies—rewording and removing inapplicable items [15, 87]—are applied to fit the remaining items into our study context.

Cohesion Concept and Measures. In this study, we adopt a widely accepted definition proposed by Carron [17], which defines cohesion 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”. It recognizes cohesion as *multidimensional* and *dynamic* that changes its extent and forms throughout group formation, development, maintenance, and dissolution [18].

To measure this concept, a conceptual model has been proposed that identifies two key perceptions binding members to their group: a member’s perception of the group as a whole (labeled as *group integration (GI)*) and a member’s personal attraction to the group (labeled as *individual attractions to group (ATG)*). Each perception is further divided into task and social aspects, resulting in four correlated constructs that influence group cohesion: *group integration-task (GI-T)*, *group integration-social (GI-S)*, *individual attractions to group-task (ATG-T)*, and *individual attractions to group-social (ATG-S)* [20]. The *Group Environment Questionnaire (GEQ)* derived from this model consists of 18 items and has demonstrated its reliability and validity in various contexts [5, 14, 20]. The cohesion score for a group is calculated by averaging individual GEQ responses [19].

Measure Adaptation. Regarding our study context, we focus on the social scales of the GEQ, which emphasize social-oriented motivations driving user interactions [3, 30, 85, 90]. Following our

Table 1: Adapted GEQ for online social networks.

Scale	Item
ATG-S	1. I enjoy being a part of the social activities of this group.
	5. Some of my best friends are in this group.
	7. I enjoy interactions in my group more than in other groups.
	9. For me, this group is one of the most important social groups to which I belong.
GI-S	11. Members of our group would rather interact with each other than post information individually.
	13. Our group members frequently interact with each other.

workflow, the GEQ is adapted into Table 1, which includes four items for the ATG-S scale and two items for the GI-S scale.

4.2 Comparison

Table 2 summarizes whether existing measures (introduced in Section 3) capture the cohesion features outlined in the adapted GEQ (Table 1). Observe that both ATG-S and GI-S items incorporate underlying structural information, yet in distinct ways. In particular, ATG-S places greater emphasis on personal opinions, carrying more psychological insights that influence a user’s membership. To reflect this distinction, items are labeled “S” for structural information and “P” for psychological information. “√”, “△”, and “×” denote that an item is fully captured, partially captured, or not captured by existing measures, respectively.

Intuitively, all existing structural cohesiveness measures, which focus on the subgraph interconnections, successfully capture Item 13. While strong topological connections may suggest users’ enjoyment, friendship formation, and group engagement (Items 1, 5, 7, and 9), these psychological elements are not explicitly characterized. Similarly, it remains unclear to what extent *keywords cohesiveness*, which captures users’ attribute similarity, and *spatial cohesiveness*, which reflects spatial relationships, accurately represent ATG-S items. While *query cohesiveness* incorporates network information to characterize the user’s temporal activeness, its representation of psychology-based cohesiveness is still implicit and insufficient.

Notably, none of the existing metrics capture users’ willingness to interact within a group (Item 11). Furthermore, they cannot be combined to capture all items in Table 1.

5 Psychology-informed Cohesiveness Measures

The preceding section highlighted the limitations of the generic cohesiveness measures in capturing psychological cohesion in social networks. To further assess and compare their effectiveness, in this section, we propose a set of *psychology-informed measures* that mathematically quantifies group cohesion indicators described in items within Table 1.

Intuitively, for ATG-S items, our formulation focuses on the relationship between individual users and the community, yielding individual scores, which are then averaged into a community score. For GI-S items, the community is treated as a single unit to report a group-level score. All measures are positively scored.

5.1 Measures for ATG-S Items

Enjoyment Index (EI). Item 1 emphasizes the user’s enjoyment level during social interactions. Enjoyment is often measured concerning positive emotions or needs fulfillment during interactions [52, 62, 90]. Accordingly, we identify the *enjoyment index* as a cohesiveness indicator, quantified by the cumulative sentiments elicited during interactions.

Table 2: Comparison of existing CS cohesiveness metrics and social psychology-based cohesion.

Item (Info)	k -core	k -truss	k -clique	k -ECC	Attributed pkd -truss community	(k, θ) -community	(k, p) -core	β -temporal core	Keyword cohesiveness	Spatial cohesiveness	Query cohesiveness
1 (S & P)	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
5 (S & P)	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
7 (S & P)	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
9 (S & P)	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
11 (S)	\times	\times	\times	\times	\times	\times	\times	\times	\times	\times	\times
13 (S)	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\times	\times	\times

Empirical findings indicate that an individual's sentiment is shaped by the sentiments and frequency of prior interactions [7, 52]. Individuals prioritize emotion-congruent stimuli and respond more quickly to emotional stimuli than neutral ones [40, 54, 72, 93, 102]. Therefore, we focus on the sentiment polarity of each interaction (positive, neutral, or negative, represented as 1, 0, or -1). We model the impact of sentiments as a *sentiment-aware excitation function*, which quantifies the *excitation degree* of previous interactions for the sentiment of the current interaction. Then the *elicited sentiment* can be modeled as the sentiment conveyed in the interaction scaled by the *excitation degree*.

The sentiment-aware excitation function is constructed based on a basic excitation function, represented as a linear combination of the basic decay functions $\Phi(\cdot)$ [77] while introducing a sign function to determine whether past activities $\mathcal{E}^{\prec t}$ enhance or diminish the sentiment of the interaction e^t at time t , as shown below.

$$\lambda(t, e^t, \mathcal{E}^{\prec t}) = \lambda_0 + \sum_{e^{t'} \in \mathcal{E}^{\prec t}} \text{sgn}(\text{Senti}(e^{t'}), \text{Senti}(e^t)) \cdot \Phi(t - t')$$

Wherein, λ_0 represents the basic sentiment perception level elicited by a single interaction. We set $\lambda_0 = 1$ to ensure that the elicited sentiment matches the original sentiment in the absence of prior activities. The function $\text{Senti}(\cdot) \in [-1, 1]$ extracts the sentiment from activities, and the sign function is defined as $\text{sgn}(\text{Senti}(e^{t'}), \text{Senti}(e^t)) = \text{Senti}(e^{t'}) \cdot \text{Senti}(e^t)$. The sentiment extraction methods and decay function selection will be introduced in Section 6. To ensure non-negative excitation, we set $\lambda(t, e^t, \mathcal{E}^{\prec t}) = \max(\lambda(t, e^t, \mathcal{E}^{\prec t}), 0)$. Formally, the *elicited sentiment* is defined as below.

Definition 5.1. [Elicited Sentiment] Given an interaction activity e^t at time t , the elicited sentiment is defined as the interaction sentiment scaled by the sentiment excitation degree as follows:

$$ESenti(t, e^t, \mathcal{E}^{\prec t}) = \text{Senti}(e^t) \cdot \lambda(t, e^t, \mathcal{E}^{\prec t}).$$

Let $\mathcal{E}_{H,i}^{t_{cur}} = \{e_{xy}^t \mid e_{xy}^t \in \mathcal{E}_H, \{x, y\} = \{i, j\}, j \neq i, t \leq t_{cur}\}$ be activities where user u_i interacts with other group members up to time t_{cur} . Assuming that more recent activities better capture user interests [82], we weight them by a decay function $\Phi(\cdot)$ and define the *enjoyment index* as follows.

Definition 5.2. [Enjoyment Index (EI)] Given a graph \mathcal{G} , community H , current time t_{cur} , and a user $u_i \in H$, u_i 's EI is defined as the cumulative elicited sentiments generated during interactions between u_i and others in H by t_{cur} :

$$EI_i(\mathcal{G}, H, t_{cur}) = \sum_{e^t \in \mathcal{E}_{H,i}^{t_{cur}}} ESenti(t, e^t, \mathcal{E}_{H,i}^{\prec t}) \cdot \Phi(t_{cur} - t),$$

where $\mathcal{E}_{H,i}^{\prec t} = \{e_{xy}^{t'} \mid e_{xy}^{t'} \in \mathcal{E}_H, \{x, y\} = \{i, j\}, j \neq i, t' < t\}$.

Sentimental Interaction Tendency (SIT). Now consider Item 5 in Table 1. Having best friends in the group indicates cohesiveness, prompting the need to measure friendships among group members. Friendship generally develops through the exchange of information [21] and is characterized by reciprocity, closeness, and emotional attachment [4, 74]. Compared to ordinary friends, best friends share greater intimacy, reciprocity, and stability [11, 78, 97]. For simplicity, we propose *sentimental interaction tendency* to characterize friendships developed through cumulative information exchange, with higher values denoting closer ties.

Let $\mathcal{M}_{H,i}^{t_{cur}} = \{u_j \mid u_j \in \mathcal{U}_H, j \neq i, \exists t_1, t_2 \leq t_{cur} : e_{ij}^{t_1}, e_{ji}^{t_2} \in \mathcal{E}_H\}$ be the set of subgroup members that mutually interact with u_i . Then, for each $u_j \in \mathcal{M}_{H,i}^{t_{cur}}$, define the interactions between u_i and u_j as $\mathcal{E}_{H,\{i,j\}}^{t_{cur}} = \{e_{xy}^t \mid e_{xy}^t \in \mathcal{E}_H, \{x, y\} = \{i, j\}, t \leq t_{cur}\}$. The cumulative, time-decaying sentiments in these interactions are computed as follows.

$$Accu_{ij}(\mathcal{G}, H, t_{cur}) = \sum_{e^t \in \mathcal{E}_{H,\{i,j\}}^{t_{cur}}} ESenti(t, e^t, \mathcal{E}_{H,\{i,j\}}^{\prec t}) \cdot \Phi(t_{cur} - t)$$

Note that $Accu_{ij}(\cdot)$ is symmetric, i.e., $Accu_{ij}(\mathcal{G}, H, t_{cur}) = Accu_{ji}(\mathcal{G}, H, t_{cur})$, reflecting the reciprocity of friendship.

Definition 5.3. [Sentimental Interaction Tendency (SIT)] Given a graph \mathcal{G} , community H , current time t_{cur} , and a user $u_i \in H$, u_i 's SIT is defined as the accumulated elicited sentiments between u_i and each user with whom they have mutual interactions up to t_{cur} :

$$SIT_i(\mathcal{G}, H, t_{cur}) = \sum_{u_j \in \mathcal{M}_{H,i}^{t_{cur}}} Accu_{ij}(\mathcal{G}, H, t_{cur}).$$

Comparative Enjoyment Degree (CED). Users often engage in multiple groups on online social networks [91]. They perceive greater cohesiveness in a specific group when it positively contributes to their enjoyment, as indicated by Items 7 and 9. This leads us to propose the *comparative enjoyment degree* to measure how much more enjoyment a user gets from interactions within a community than from interactions outside that community.

Definition 5.4. [Comparative Enjoyment Degree (CED)] Given a graph \mathcal{G} , community H , current time t_{cur} , and a user $u_i \in H$, the CED of u_i at time t_{cur} is defined as the difference between the enjoyment derived from interactions within and outside H by t_{cur} :

$$CED_i(\mathcal{G}, H, t_{cur}) = \sum_{e^t \in \mathcal{E}_{H,i}^{t_{cur}}} ESenti(t, e^t, \mathcal{E}_{H,i}^{\prec t}) \cdot \Phi(t_{cur} - t) - \sum_{e^t \in \mathcal{E}_{\mathcal{G} \setminus H,i}^{t_{cur}}} ESenti(t, e^t, \mathcal{E}_{\mathcal{G} \setminus H,i}^{\prec t}) \cdot \Phi(t_{cur} - t),$$

where $\mathcal{E}_{\mathcal{G} \setminus H,i}^{t_{cur}} = \{e_{xy}^t \mid e_{xy}^t \in \mathcal{E}_{\mathcal{G}}, \{x, y\} = \{i, j\}, u_j \notin \mathcal{U}_H, t \leq t_{cur}\}$ represents u_i 's interactions with users outside H .

Table 3: Dataset statistics.

Dataset	$ \mathcal{U} $	$ \mathcal{E} $	$ \mathcal{T} $	Density	Deg_{avg}
BTW17	7,721	12,839	11,890	0.0002	3.33
Chicago_COVID	4,917	246,633	246,001	0.0102	100.32
Crawled_Dataset144	10,388	119,836	116,639	0.0011	23.07
Crawled_Dataset26	21,263	108,538	104,575	0.0002	10.21

5.2 Measures for GI-S Items

Group Interaction Preference (GIP). Item 11 captures the group’s preference for group interaction over individual posting, which can be formulated as *group interaction preference (GIP)*. Let $\mathcal{E}_{H,inter}^{t_{cur}} = \{e_{ij}^t \mid e_{ij}^t \in \mathcal{E}_H, j \neq i, t \leq t_{cur}\}$ and $\mathcal{E}_H^{t_{cur}} = \{e_{ij}^t \mid e_{ij}^t \in \mathcal{E}_H, t \leq t_{cur}\}$ denote the set of interaction activities and all activities in H up to t_{cur} , respectively, where $\mathcal{E}_{H,inter}^{t_{cur}} \subseteq \mathcal{E}_H^{t_{cur}}$. The GIP ranges from 0 to 1 and is defined as follows.

Definition 5.5. [Group Interaction Preference (GIP)] Given a graph \mathcal{G} , community H , and current time t_{cur} , GIP is defined as the ratio of group members’ interactions to their overall activities by t_{cur} :

$$GIP(\mathcal{G}, H, t_{cur}) = |\mathcal{E}_{H,inter}^{t_{cur}}| / |\mathcal{E}_H^{t_{cur}}|.$$

Group Interaction Density (GID). Lastly, Item 13 emphasizes the interplay between member relationships and interaction frequency. We quantify it by extending the *weighted density* [64] with a temporal normalization. Let the observation window be $W_{t_{cur}} = (t_{cur} - t_0) / \tau_t$, where t_0 is the start of observation and τ_t is the time-unit (e.g., day, week, month). We call the resulting metric the *group interaction density (GID)*.

Definition 5.6. [Group Interaction Density (GID)] Given a graph \mathcal{G} , community H , and current time t_{cur} , GID measures the interaction frequency per possible user pair per unit time within H by t_{cur} :

$$GID(\mathcal{G}, H, t_{cur}) = |\mathcal{E}_{H,inter}^{t_{cur}}| / (|\mathcal{U}_H| \cdot (|\mathcal{U}_H| - 1) \cdot W_{t_{cur}}).$$

Remark. Note that these five measures for assessing cohesion can be merged into a single *cohesiveness score* through an aggregated linear combination of weights. However, our work does not aim to calculate such a score. Instead, the evaluation shall focus on analyzing the quality of results produced by existing CS techniques w.r.t. each of these measures, enabling us to get a fine-grained understanding of how they perform against *each* item in Table 1.

6 Experimental Settings

In this section, we introduce the datasets, representative community search algorithms, evaluation framework, and parameter settings used to set up the experiments. The codebase is available at [108].

6.1 Dataset

Due to the lack of sentiment data in most publicly available datasets (e.g., SNAP [58]), we constructed four real-world datasets from X (formerly *Twitter*) to facilitate evaluation. Table 3 summarizes their topological characteristics, where $|\mathcal{U}|$, $|\mathcal{E}|$, and $|\mathcal{T}|$ denote the numbers of vertices, edges, and edge timestamps, respectively. Deg_{avg} represents the average node degree, with self-loops and parallel edges counted.

Data Collection. The BTW17 dataset contains tweets related to the 19th German election campaigns, with over 1.2 million tweets

As a social scientist, your task is to analyze the sentiment of a series of user tweets extracted from Twitter. Please assign a sentiment label from ["Positive", "Neutral", "Negative"] for each tweet. In situations where the sentiment is difficult to definitively classify, please provide your best estimation of the label. Here is the text:

Figure 2: Prompt for sentiment analysis.

from active users [56]. The Chicago_COVID dataset contains tweets related to the COVID-19 pandemic and geolocated near Chicago [42]. Due to its large volume, we sampled 125k tweets for this study. Besides, we collected Crawled_Dataset144 and Crawled_Dataset26 datasets in December 2022 using the official API [76]. Starting with single users randomly selected from real-time tweets, we employed snowball sampling to gather the users’ original tweets and associated replies (conversation utterances). Newly identified users were recorded for future data collection. This process resulted in over 200k original tweets for each dataset. For all four datasets, we extracted metadata for each user, including *user_id*. The tweet metadata includes *conversation_id*, *tweet_id*, *user_id*, *in_reply_to_user_id*, *tweet_text*, and *timestamp*. The number of users and tweet data elements varied across datasets. We retained only English tweets, removed retweets, quoted tweets, duplicates, and empty entries, and cleaned the tweet text by removing newlines and URLs.

Sentiment Analysis. The preprocessed tweets were then subjected to a sentiment analysis task to determine their sentiment scores. Following common practice, we treated the task as a three-class classification problem [23]. To ensure superior analysis results, we adopted *Llama3-8B*, an open-source Large Language Model (LLM) from Meta [1], and used a prompt-based approach. The prompt was adapted from [53] to assign sentiment labels (see Figure 2), which were mapped to 1, 0, and -1, respectively. Both *top_p* and *temperature* were set to 0.01 for more deterministic responses. Only the tweets that were successfully assigned labels were retained.

Network Construction. Networks were constructed based on the *reply* relationships (i.e., who replies to whom), including interaction and self-posting tweets (i.e., initial posts not directed at specific users). Isolated nodes were removed. Since the constructed networks consist of components with their size distribution highly skewed, only the largest component in each network was considered, resulting in the networks reported in Table 3.

6.2 Representative CS Algorithms

We selected representative CS algorithms based on the following criteria: (a) their codebases are publicly available; (b) they cover the two categories introduced in Section 3; and (c) they consider various types of information to characterize cohesiveness, potentially capturing different items in the adapted GEQ (Table 2). As a result, we chose three k -core-based algorithms (ALS, WCF-CRC, CSD), three k -truss-based algorithms (ST-Exa, Repeeling+, I2ACSM), and two unsupervised learning algorithms (TransZero-LS, TransZero-GS), as summarized in Table 4. Although [29] could be a useful technique for our study (based on Table 2), unfortunately, its codebase is not available. The last column in Table 4 indicates whether the algorithms use ground truth for evaluation, with most of them not using it. Note that the ground truth adopted in the three approaches is not based on group cohesion theory and lacks data elements (e.g., temporality and sentiments) to calculate our proposed measures. Therefore, we cannot exploit them for our evaluation. We now briefly describe these algorithms.

Table 4: Representative community search algorithms.

Category	Algorithm	Graph Type	Ground Truth
Traditional	ALS [63]	Undirected Temporal	No
	WCF-CRC [92]	Dynamic	No
	CSD [26]	Directed Multigraph ¹	No
	ST-Exa [107]	Undirected	No
	Repeeling+ [61]	Streaming Directed	No
	I2ACSM [99]	Attributed and Weighted	Yes
Learning-based	TransZero-LS [94]	Undirected	Yes
	TransZero-GS [94]	Undirected	Yes

ALS identifies the β -temporal proximity core with the largest β in a temporal graph [63], integrating temporal proximity with structural cohesiveness seamlessly. It requires specifying the tele-transportation probability α to compute temporal proximity. WCF-CRC searches on dynamic networks, represented as a sequence of time-variant weighted graph instances ($\mathbb{G} = G_{t_1}, G_{t_2}, \dots, G_{t_T}$) [92]. It aims to find a *Time Interval based* (θ, k) -core *Reliable Community (CRC)* with the highest *reliability score*, where a (θ, k) -core in a G_t is a k -core where all edge weights are at least θ . CSD finds a (k, l) -core in the directed graph, where k and l denote the minimum in-degree and out-degree of subgraph nodes [26].

ST-Exa identifies a subgraph with the largest *min-support* based on the query node, while satisfying the size constraint $[l, h]$. It is parameter-free, eliminating the need to specify the cohesiveness constraint. Repeeling+ is a peeling-based algorithm for finding a maximal D -truss containing the query node in the snapshot digraph G^t [61]. G^t is induced from a streaming directed graph using a sliding window length τ and a *stride* β . The algorithm’s parameters need to be tuned according to the dataset characteristics to ensure valid results. I2ACSM aims to search an attributed pkd -truss community that includes the query node in an Attributed and Weighted Graph (AWG) [99], ensuring that any vertex in the k -truss subgraph is within distance d from the query node and the maximum path influence from the query node to any subgraph node exceeds p [99].

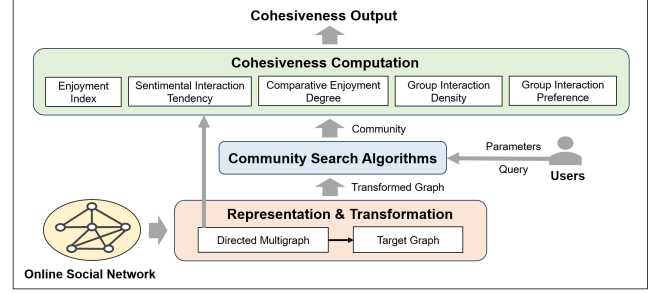
TransZero-LS and TransZero-GS are proposed under *TransZero*, a two-phase pre-trained Graph Transformer-based community search framework that aims to find communities without ground-truth data [94]. In the *offline* phase, *CSGphormer* is pre-trained to learn node representations that encode both community information and graph topology. In the *online* phase, the community score is calculated by measuring the representation similarity between the query and the remaining graph nodes, and an *expected score gain function* is introduced to search communities with maximum expected score gain (*ESG*). TransZero-LS and TransZero-GS are two heuristic algorithms based on *ESG* for community identification.

6.3 The CHASE Framework

Figure 3 presents our framework coined CHASE (CoHesiveness evAluation for Social nEtworks) for evaluating CS algorithms, which includes query generation, parameter selection, graph transformation, community mapping, and cohesiveness evaluation.

We first generate Q queries ($Q = 100$ in our study) per dataset, each with a single query node. Queries are randomly selected from nodes in the top 50% by degree. Next, for each algorithm, datasets are transformed into algorithm-specific graph types to support its community retrieval. The varying parameter values are chosen

¹The original paper referred to the graphs as directed, but upon closer examination, we found they are more accurately characterized as directed multigraphs.

**Figure 3: The CHASE framework.**

from the original paper. For dataset-sensitive parameters, we manually test values to ensure meaningful results. For algorithms with multiple parameters, all value combinations are considered.

Using the query nodes and parameters, we obtain Q communities from the transformed graph by applying the algorithm. To ensure fairness, each community is mapped back to the directed multigraphs for evaluation (*i.e.*, cohesiveness computation). In other words, a mapped community is extracted according to nodes in the initial community and includes all edges among them. For algorithms that provide multiple solutions for a single query under various parameter combinations, the cohesiveness of these communities is averaged before being compared with other techniques.

6.4 Parameter Settings

The parameter settings for the ALS, I2ACSM, TransZero-LS, and TransZero-GS follow the original papers, with influence values and attributes for I2ACSM generated following the methods mentioned in [99]. Dataset-dependent parameters for the remaining algorithms are adjusted for meaningful results. Specifically, for WCF-CRC, the query time intervals for BTW and the rest of the datasets are set to $t \in \{2, 3\}$ and $t \in \{4, 6, 8, 10\}$, respectively, based on the features of transformed datasets. In CSD, k and l are varied from 1 to 10. The size constraint $[l, h]$ in ST-Exa ranges from $[1, 10]$ to $[71, 80]$. The window size, stride size, k_c , and k_f in Repeeling+ vary across datasets; these parameters for each dataset are manually tested and reported in Table 5.

We use an exponential decay function, commonly applied in fields such as biology, physics, and network analysis [44, 51, 88], to calculate cohesiveness measures. The function is defined as $\Phi(t_{cur} - t) = e^{-\lambda(t_{cur} - t)}$, where t_{cur} is the latest timestamp in each dataset. The decay rate λ is selected from $\{0.0001, 0.0005, 0.001, 0.005, 0.01\}$, with 0.0001 as the default. When calculating GID values, we note that the four datasets span different observation lengths and that no standard choice of τ_t exists. Since $W_{t_{cur}}$ uniformly scales all GID scores of communities within a dataset without affecting their relative ranking, we omit $W_{t_{cur}}$ for clarity and simplicity. As users in the four datasets lack consistent information except their ids, we use only *user_id* as the node feature during *CSGphormer* training.

7 Experimental Evaluation

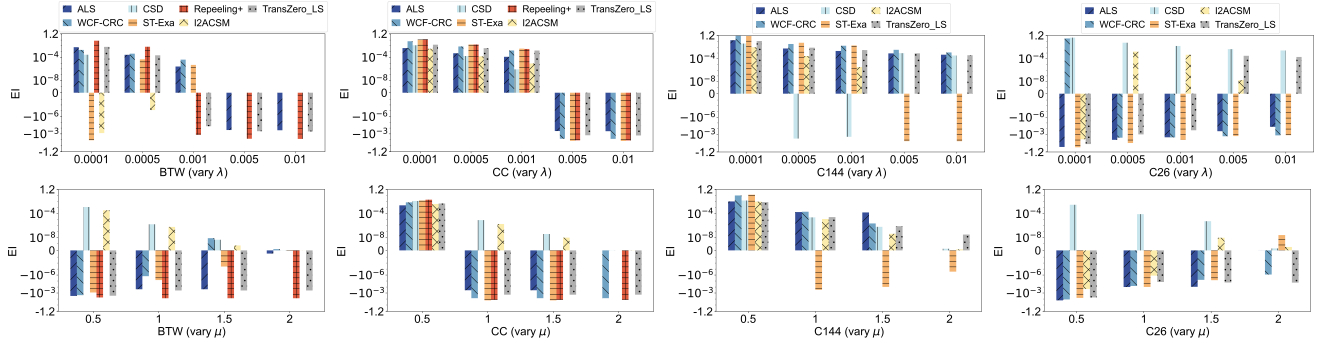
In this section, we conduct an in-depth evaluation of the representative CS algorithms and report key results. The reader may refer to [109] for detailed results. We first assess the structural cohesiveness of their identified communities, then evaluate their quality

Table 5: Parameter settings in Repeeling+ algorithm.

Dataset	[Window]	[Stride]/[Window](%)	k_c	k_f
BTW17	{10k, 50k, 100k, 500k, 1M, 5M, 10M, 50M}	{0.01, 0.05, 0.1, 0.5, 1, 5, 10}	{0, 1}	{0, 1}
Chicago_COVID	{10k, 50k, 100k, 500k, 1M, 5M, 10M, 50M, 100M}	{0.01, 0.05, 0.1, 0.5, 1, 5, 10}	{0, 1, 2, 3}	{0, 1, 2, 3}
Crawled_Dataset144	{100k, 500k, 1M, 1.5M, 2M, 2.5M}	{0.001, 0.005, 0.01, 0.05, 0.1}	{0, 1, 2, 3}	{0, 1, 2, 3}
Crawled_Dataset26	{100k, 500k, 1M, 1.5M, 2M, 2.5M}	{0.001, 0.005, 0.01, 0.05, 0.1}	{0, 1, 2, 3}	{0, 1, 2, 3}

Table 6: Evaluation of structural cohesiveness of CS solutions across different datasets.

Algorithm	BTW17				Chicago_COVID				Crawled144				Crawled26			
	d	Size	Deg_{min}	Q_{hit}	d	Size	Deg_{min}	Q_{hit}	d	Size	Deg_{min}	Q_{hit}	d	Size	Deg_{min}	Q_{hit}
ALS	9.24	5029.18	1.00	100	16.21	3998.33	1.18	100	15.36	9385.31	1.00	100	14.87	17043.60	1.15	100
WCF-CRC	6.88	242.80	2.19	45	10.71	280.91	8.77	52	5.64	451.59	11.12	93	9.86	287.89	3.93	39
CSD	5.67	24.00	0.67	3	4.16	19.70	59.52	10	7.70	204.22	11.14	34	6.15	125.29	8.84	20
ST-Exa	4.49	44.96	2.20	100	5.60	44.88	13.09	100	3.01	44.58	26.62	100	3.60	44.74	5.59	100
Repeeling+	8.46	3228.53	1.49	100	6.80	998.94	18.20	100	INF	INF	INF	INF	INF	INF	INF	INF
I2ACSM	0.55	1.61	1.14	100	1.22	3.48	101.28	100	0.97	2.41	5.95	100	0.42	1.42	4.37	100
TransZero-LS	10.00	3031.42	1.00	100	10.07	458.28	12.22	100	11.72	4536.41	1.11	100	9.10	2367.10	1.36	100
TransZero-GS	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0

**Figure 4: EI of communities (top row: community cohesiveness; bottom row: impact of time-decay functions).**

using five psychology-informed cohesiveness measures. Learning-based algorithms were run on a Linux server with an Intel Xeon Platinum 8255C (2.50 GHz) and an RTX 3080 GPU (10 GB VRAM). Other experiments were conducted on a Linux server with an Intel Xeon Platinum 8352V (2.10 GHz) and 120 GB RAM. Recall that we do not focus on efficiency and scalability analysis here, as our goal is to understand the effectiveness of CS algorithms. The reader may refer to [34] for studies related to the former.

7.1 Structural Cohesiveness

We use three commonly used evaluation metrics to assess the structural cohesiveness of CS algorithms [34, 67, 100]: diameter (d), size, and minimum degree (Deg_{min}). Here, the Deg_{min} calculation includes self-loops and parallel edges. Generally, higher community quality is indicated by lower diameter and size, and higher minimum degree [34]. We also report the *query hit rate* Q_{hit} (%), which is the ratio of queries that successfully return communities across all parameter combinations. Algorithms that fail to complete within 5 days are marked as *INF* and excluded from the evaluation.

Table 6 reports the results. We can make the following observations: (1) although ALS, Repeeling+, and TransZero-LS have high Q_{hit} , they tend to generate larger yet sparser communities characterized by greater degrees and sizes, and lower Deg_{min} values; (2) I2ACSM identifies very small communities across all datasets; and (3) compared to aforementioned algorithms, WCF-CRC, CSD, and ST-Exa are better at identifying structurally cohesive communities. However, WCF-CRC and CSD generally achieve lower hit rates,

indicating difficulty in identifying target communities. Among all algorithms, ST-Exa successfully returns communities for all queries and demonstrates superior performance across all datasets, partially due to its size-constrained nature. Note that the communities (or node lists) returned by TransZero-GS are disconnected, likely because its candidate selection ignores edge connections. Hence, its results are left blank.

The poor performance of ALS may stem from the smaller time scales used in our datasets, which hinder its ability to find nodes with greater temporal proximity. However, the original paper does not discuss the time scale and its impact on CS results. Additionally, identifying extremely large or small communities may arise from graph sparsity after multigraph transformations. The unpromising performance of the two learning-based methods may be attributable to insufficient node features in our datasets, limiting CSGphormer’s ability to learn adequate information for calculating community scores. On the other hand, this highlights these methods’ stringent data requirements and their limitations when applied to datasets not specifically designed for model training.

7.2 Psychology-informed Cohesiveness

Next, we report evaluation results using our psychology-informed cohesiveness measures.

Enjoyment Index (EI). The top row of Figure 4 illustrates that community cohesiveness, as indicated by EI values, varies across algorithms and is affected by the importance assigned to recent activities. For example, communities identified by Repeeling+ on

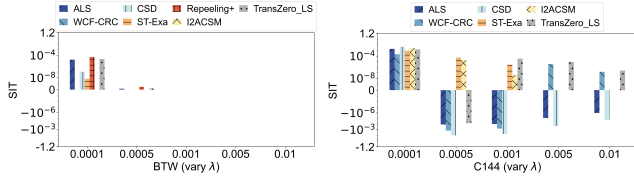


Figure 5: SIT of communities identified by CS algorithms.

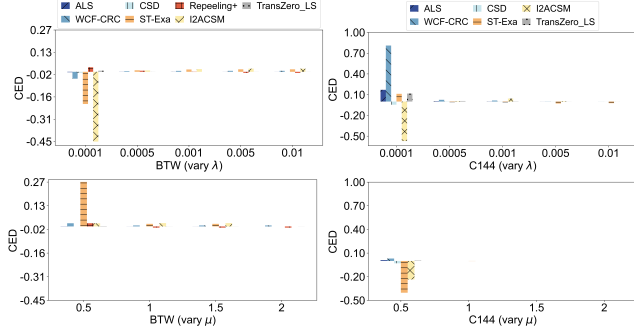


Figure 6: CED of communities identified by CS algorithms.

the BTW show higher enjoyment when $\lambda \in \{0.0001, 0.0005, 0.001\}$. However, increasing the weight of recent activities reduces enjoyment due to recent negative interactions. Notably, communities identified by CSD mostly exhibit positive EI values across all datasets and are less affected by the λ .

We also examine the impact of a different time-decay function on EI, as shown in the bottom row of Figure 4. Here, we use the polynomial decay function, $\Phi(t - t_{cur}) = (t_{cur} - t + 1)^{-\mu}$ ($\mu > 0$), which models a slower decline in the importance of past activities [27]. We vary μ from 0.5 to 2 in steps of 0.5. With a stronger and more persistent impact of recent interactions, EI values become more consistent across decay rates, and communities identified by CSD still largely maintain positive EI scores. Overall, none of the algorithms retrieve communities with consistently high EI scores.

Sentimental Interaction Tendency (SIT). SIT scores are sensitive to decay rates and the choice of decay functions. Nevertheless, the SIT results are distinguishably different from the EI results. Figure 5 shows results on BTW and C144. On BTW, as λ increases, no algorithms effectively capture mutual relationships among users. Furthermore, results on C144 reveal that communities with predominantly positive EI scores can still include negative mutual relationships (negative SIT values), and vice versa. Although CSD and Repeeling+ consider directed edges, and ALS and WCF-CRC incorporate temporal information, their communities are still inadequate to achieve higher SIT scores.

Comparative Enjoyment Degree (CED). Figure 6 plots the CED results. When $\lambda = 0.0001$, some communities exhibit strong internal enjoyment, while others prefer external interactions, as indicated by the polarity of the CED scores. However, as the decay rate increases, the CED values for all communities converge toward zero, meaning the algorithms cannot find communities that increasingly prefer internal interactions over time. Similar to EI and SIT, CED is highly sensitive to the choice of decay function, reflecting the complexity of user interactions, as CED is calculated based on sentiment, timing, and where interactions occur. No algorithms consistently capture these complex interaction dynamics.

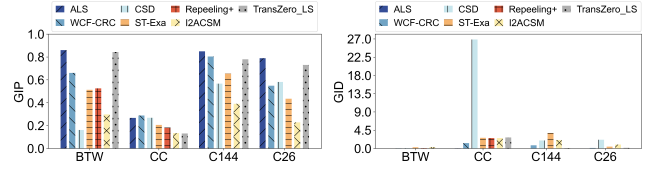


Figure 7: GIP and GID of CS communities.

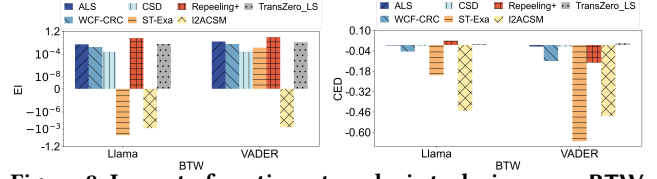


Figure 8: Impact of sentiment analysis techniques on BTW.

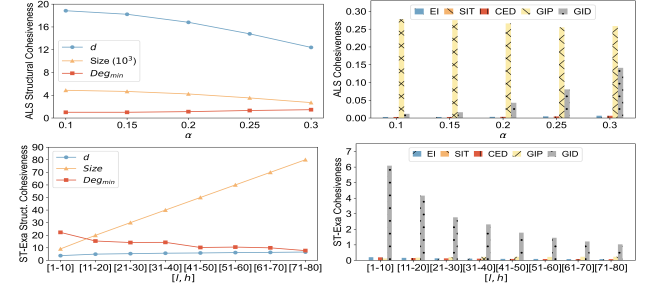


Figure 9: Impact of parameter selection.

Group Interaction Preference (GIP). The left plot in Figure 7 reports the GIP values. ALS, WCF-CRC, and TransZero-LS tend to identify highly interactive communities. However, as noted in Table 6, these high GIP values may result from the large community sizes returned by these algorithms, which include most interactions. Similar patterns are observed with higher GIP values for CSD communities on the C144 and C26 datasets and lower scores for I2ACSM communities across all datasets.

Group Interaction Density (GID). The right plot in Figure 7 reports that users within communities from CSD, ST-Exa, and I2ACSM generally engage more with each other. Note that these high-GID communities also have smaller sizes and higher Deg_{min} .

7.3 Impact of Sentiment Analysis Techniques

To examine the impact of sentiment analysis techniques on evaluation and to demonstrate the broad applicability of our cohesiveness metrics, we adopt a popular lexicon-based approach, VADER [48], to relabel the datasets. We recompute sentiment-related cohesiveness measures (*i.e.*, EI, SIT, and CED), and compare these results with those from Llama3-labeled datasets. Figure 8 presents two results on BTW datasets. In most cases, the cohesiveness scores are largely unaffected by sentiment analysis techniques. However, when the labeling technique changes to VADER, the ST-Exa and I2ACSM communities respond more strongly, with their values either switching polarity (*e.g.*, ST-Exa's EI value on BTW) or showing obvious changes (*e.g.*, ST-Exa's and I2ACSM's CED values on C144).

7.4 Impact of Parameter Selection

Next, we discuss how parameter selection affects structural and psychology-informed cohesiveness using two single-parameter algorithms, ALS and ST-Exa, on the CC dataset.

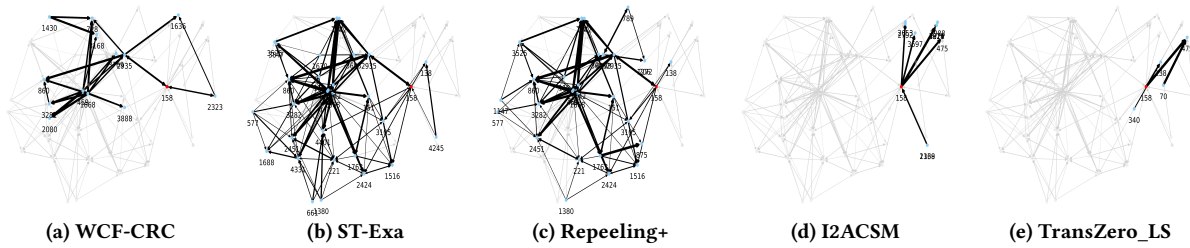


Figure 10: Case study on the CC dataset with the query node marked in red.

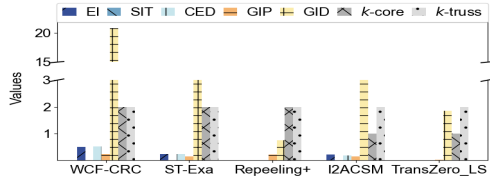


Figure 11: Comparison of two types of cohesiveness.

For ALS, increasing the *teleportation probability* α results in smaller communities with higher structural cohesiveness and increasing psychological cohesiveness, except GIP. Notably, GID improves significantly with larger α , while other scores show minor increases (Figure 9, top row). For ST-Exa, members of smaller, structurally dense communities consistently achieve higher EI, CED, and GID. In contrast, members of larger communities tend to engage more, as observed in slightly higher GIP (Figure 9, bottom row).

7.5 Case Study

Finally, we present a case study using the CC dataset to compare CS results from various algorithms, evaluating cohesiveness with our proposed measures and two additional structural metrics: k -core and k -truss. The structural metrics are computed after converting the CC into an undirected graph. Here, we present the results for the query node “158”, as most algorithms return its communities. To simplify visualization, we only display algorithms that return communities of size 5–30. If algorithms produce multiple communities under different parameters, the largest one is selected. Besides, self-loops within the networks are not displayed, and edge widths reflect the number of interactions between users (See Figure 10).

We observe that communities returned by different algorithms vary significantly. The first three algorithms identify larger communities with more internal connections. While the communities from I2ACSM and TransZero-LS exhibit frequent interactions between users, their overall connectivity is sparse. For structural cohesiveness, the first three communities are 2-core and 2-truss. However, despite being truss-based, Repeeling+ and I2ACSM fail to find more prominent k -truss communities. I2ACSM and TransZero-LS communities only meet the 1-core requirement. *In other words, existing CS techniques fail to return consistent results for the same query node.*

Interestingly, our comparison of two types of cohesiveness across five communities reveals no explicit correlation between them (Figure 11). The 2-core and 2-truss communities identified by the first three algorithms lack consistent higher psychological cohesiveness. Furthermore, all methods struggle to capture psychological cohesiveness. *That is, existing techniques return communities whose features are misaligned with the notion of cohesion in social psychology.*

8 Reflections & Future Work

In this paper, we conduct the first study to extensively evaluate the cohesiveness of communities in online social networks produced by state-of-the-art CS algorithms. Our study reveals that communities identified by existing methods for the same query node lack consensus, and their structural and psychological cohesiveness vary greatly, with no clear correlation between them. Despite the notion of cohesion being strongly rooted in social psychology theories, none of these generic techniques performs well on online social networks *w.r.t.* the five psychological cohesiveness measures that are inspired by research in group cohesion. This is primarily because the existing one-size-fits-all cohesiveness measures of CS techniques do not systematically incorporate theories from social psychology into their design. This issue is further compounded by the fact that current CS techniques are either not assessed using “ground-truth” communities or rely on communities that are built without considering theories of group cohesion.

Despite these significant results, our evaluation has some limitations. Firstly, the design of our measures is constrained by sentiment analysis tools, and the simple three-class classification in our sentiment analysis may overlook nuanced sentiments inherent in specific comments. Secondly, our study is based on a single social network platform (*i.e.*, X). However, given the fundamental limitations in the design of existing cohesiveness measures for human-interaction datasets (Section 4.2), we believe that the key insights reported in this paper will also hold in other large online social networks. Thirdly, our study cannot use ground-truth communities for benchmarking because no public ground-truth datasets have been created based on cohesion theories.

We believe that the findings of our study can inspire a reevaluation of current community search algorithms by incorporating insights from social psychology, leading to the development of more effective solutions. Our proposed measures may be incorporated into search techniques to better identify cohesive communities in online social networks. In particular, it is paramount to create ground-truth community datasets for community search that are grounded in social psychology theories. This will not only enable the validation of our proposed psychology-informed cohesiveness measures but also assist existing CS techniques in developing more effective cohesiveness measures and algorithms for practical usage. Lastly, while this work concentrates on the community search problem, it is clear that our findings could also inspire a rethinking of various *community-centric* social search and analytics problems involving human interactions, such as community detection [9, 49].

References

- [1] AI@Meta. 2024. Llama 3 Model Card. (2024). https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md
- [2] Esra Akbas and Peixiang Zhao. 2017. Truss-based community search: a truss-equivalence based indexing approach. *Proc. VLDB Endow.* 10, 11 (Aug. 2017), 1298–1309. doi:10.14778/3137628.3137640
- [3] Fatima Alsalem. 2019. Why do they post? Motivations and uses of snapchat, Instagram and twitter among Kuwait college students. *Media Watch* 10, 3 (2019), 550–567.
- [4] Yair Amichai-Hamburger, Mila Kingsbury, and Barry H Schneider. 2013. Friendship: An old concept with a new meaning? *Computers in Human Behavior* 29, 1 (2013), 33–39.
- [5] Arden J Anderson and Stacy Warner. 2017. Social network analysis as a complimentary tool to measuring team cohesion. *Journal of Sport Behavior* 40, 1 (2017), 3.
- [6] Nicola Barbieri, Francesco Bonchi, Edoardo Galimberti, and Francesco Gullo. 2015. Efficient and effective community search. *Data mining and knowledge discovery* 29 (2015), 1406–1433.
- [7] Liad Bareket-Bojmel and Golan Shahaar. 2011. Emotional and interpersonal consequences of self-disclosure in a lived, online interaction. *Journal of Social and Clinical Psychology* 30, 7 (2011), 732–759.
- [8] Roy F Baumeister and Mark R Leary. 2017. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Interpersonal development* (2017), 57–89.
- [9] Punam Bedi and Chhavi Sharma. 2016. Community detection in social networks. *Wiley interdisciplinary reviews: Data mining and knowledge discovery* 6, 3 (2016), 115–135.
- [10] Ali Behrouz, Farnoosh Hashemi, and Laks V. S. Lakshmanan. 2022. FirmTruss Community Search in Multilayer Networks. *Proc. VLDB Endow.* 16, 3 (Nov. 2022), 505–518. doi:10.14778/3570690.3570700
- [11] Thomas J Berndt. 1982. Fairness and friendship. In *Peer relationships and social skills in childhood*. Springer, 253–278.
- [12] Anol Bhattacherjee. 2012. *Social science research: Principles, methods, and practices*. USA.
- [13] Amy Blackstone. 2018. *Principles of sociological inquiry: Qualitative and quantitative methods*. Saylor Academy Open Textbooks.
- [14] Chris Blanchard, Pauline Poon, Wendy Rodgers, and Bruce Pinel. 2000. Group environment questionnaire and its applicability in an exercise setting. *Small group research* 31, 2 (2000), 210–224.
- [15] Kenneth A Bollen and Rick H Hoyle. 1990. Perceived cohesion: A conceptual and empirical examination. *Social forces* 69, 2 (1990), 479–504.
- [16] Yuejun Lawrence Cai. 2023. Strengthening perceptions of virtual team cohesiveness and effectiveness in new normal: A hyperpersonal communication theory perspective. *Asian Business & Management* (2023), 1.
- [17] Albert V Carron. 1982. Cohesiveness in sport groups: Interpretations and considerations. *Journal of Sport psychology* 4, 2 (1982), 123–138.
- [18] Albert V Carron and Lawrence R Brawley. 2000. Cohesion: Conceptual and measurement issues. *Small group research* 31, 1 (2000), 89–106.
- [19] Albert V Carron, Lawrence R Brawley, Mark A Eys, Steven Bray, Kim Dorsch, Paul Estabrooks, Craig R Hall, James Hardy, Heather Hausenblas, Ralph Madison, et al. 2003. Do individual perceptions of group cohesion reflect shared beliefs? An empirical analysis. *Small group research* 34, 4 (2003), 468–496.
- [20] Albert V Carron, W Neil Widmeyer, and Lawrence R Brawley. 1985. The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of Sport and Exercise psychology* 7, 3 (1985), 244–266.
- [21] Gloria Hong-Yee Chan and T Wing Lo. 2014. Do friendship and intimacy in virtual communications exist? An investigation of online friendship and intimacy in the context of hidden youth in Hong Kong. *Revista de Cercetare si Interventie Sociala* 47 (2014), 117.
- [22] Lijun Chang, Xuemin Lin, Lu Qin, Jeffrey Xu Yu, and Wenjie Zhang. 2015. Index-based optimal algorithms for computing steiner components with maximum connectivity. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. 459–474.
- [23] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A Survey on Evaluation of Large Language Models. 15, 3, Article 39 (March 2024), 45 pages. doi:10.1145/3641289
- [24] Jiazun Chen, Yikuan Xia, and Jun Gao. 2023. CommunityAF: An Example-Based Community Search Method via Autoregressive Flow. *Proc. VLDB Endow.* 16, 10 (June 2023), 2565–2577. doi:10.14778/3603581.3603595
- [25] Lu Chen, Chengfei Liu, Rui Zhou, Jianxin Li, Xiaochun Yang, and Bin Wang. 2018. Maximum co-located community search in large scale social networks. *Proc. VLDB Endow.* 11, 10 (June 2018), 1233–1246. doi:10.14778/3231751.3231755
- [26] Yankai Chen, Jie Zhang, Yixiang Fang, Xin Cao, and Irwin King. 2021. Efficient community search over large directed graphs: An augmented index-based approach. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 3544–3550.
- [27] Graham Cormode, Vladislav Shkapenyuk, Divesh Srivastava, and Bojian Xu. 2009. Forward decay: A practical time decay model for streaming systems. In *2009 IEEE 25th international conference on data engineering*. IEEE, 138–149.
- [28] Wanyun Cui, Yanghua Xiao, Haixun Wang, and Wei Wang. 2014. Local search of communities in large graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (Snowbird, Utah, USA) (SIGMOD '14)*. Association for Computing Machinery, New York, NY, USA, 991–1002. doi:10.1145/2588555.2612179
- [29] Badhan Chandra Das, Md Musfique Anwar, Md Al-Amin Bhuiyan, Iqbal H Sarker, Salem A Alyami, and Mohammad Ali Moni. 2021. Attribute driven temporal active online community search. *IEEE Access* 9 (2021), 93976–93989.
- [30] Muhterem Dindar and Nihal Dulkadir Yaman. 2018. #IUseTwitterBecause: content analytic study of a trending topic in Twitter. *Information Technology & People* 31, 1 (2018), 256–277.
- [31] Shuheng Fang, Kangfei Zhao, Yu Rong, Zhixun Li, and Jeffrey Xu Yu. 2024. Inductive Attributed Community Search: To Learn Communities Across Graphs. *Proc. VLDB Endow.* 17, 10 (Aug. 2024), 2576–2589. doi:10.14778/3675034.3675048
- [32] Yixiang Fang, Reynold Cheng, Yankai Chen, Siquang Luo, and Jiafeng Hu. 2017. Effective and efficient attributed community search. *The VLDB Journal* 26, 6 (Dec. 2017), 803–828. doi:10.1007/s00778-017-0482-5
- [33] Yixiang Fang, Reynold Cheng, Siquang Luo, and Jiafeng Hu. 2016. Effective community search for large attributed graphs. *Proc. VLDB Endow.* 9, 12 (8 2016), 1233–1244. doi:10.14778/2994509.2994538
- [34] Yixiang Fang, Xin Huang, Lu Qin, Ying Zhang, Wenjie Zhang, Reynold Cheng, and Xuemin Lin. 2020. A survey of community search over big graphs. *The VLDB Journal* 29 (2020), 353–392.
- [35] Leon Festinger. 1950. Informal social communication. *Psychological review* 57, 5 (1950), 271.
- [36] Donelson R Forsyth. 2021. Recent advances in the study of group cohesion. *Group Dynamics: Theory, Research, and Practice* 25, 3 (2021), 213.
- [37] Santo Fortunato. 2010. Community detection in graphs. *Physics reports* 486, 3-5 (2010), 75–174.
- [38] Charles E Galyon, Eleanore CT Heaton, Tiffany L Best, and Robert L Williams. 2016. Comparison of group cohesion, class participation, and exam performance in live and online classes. *Social Psychology of Education* 19 (2016), 61–76.
- [39] Jun Gao, Jiazun Chen, Zhao Li, and Ji Zhang. 2021. ICS-GNN: lightweight interactive community search via graph neural network. *Proc. VLDB Endow.* 14, 6 (Feb. 2021), 1006–1018. doi:10.14778/3447689.3447704
- [40] David Garcia, Arvid Kappas, Dennis Küster, and Frank Schweitzer. 2016. The dynamics of emotions in online interaction. *Royal Society open science* 3, 8 (2016), 160059.
- [41] Michelle Girvan and Mark EJ Newman. 2002. Community structure in social and biological networks. *Proceedings of the national academy of sciences* 99, 12 (2002), 7821–7826.
- [42] Rachel Gordon. 2023. Chicago COVID-19 Twitter Data. doi:10.7910/DVN/TPHAQM
- [43] Farnoosh Hashemi, Ali Behrouz, and Milad Rezaei Hajidehi. 2023. CS-TGN: Community Search via Temporal Graph Neural Networks. In *Companion Proceedings of the ACM Web Conference 2023 (Austin, TX, USA) (WWW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 1196–1203. doi:10.1145/3543873.3587654
- [44] Russell K Hobbie, Bradley J Roth, Russell K Hobbie, and Bradley J Roth. 2007. Exponential growth and decay. *Intermediate physics for medicine and biology* (2007), 31–47.
- [45] Jiafeng Hu, Xiaowei Wu, Reynold Cheng, Siquang Luo, and Yixiang Fang. 2016. Querying Minimal Steiner Maximum-Connected Subgraphs in Large Graphs. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management (Indianapolis, Indiana, USA) (CIKM '16)*. Association for Computing Machinery, New York, NY, USA, 1241–1250. doi:10.1145/2983323.2983748
- [46] Jiafeng Hu, Xiaowei Wu, Reynold Cheng, Siquang Luo, and Yixiang Fang. 2017. On minimal steiner maximum-connected subgraph queries. *IEEE Transactions on Knowledge and Data Engineering* 29, 11 (2017), 2455–2469.
- [47] Xin Huang, Hong Cheng, Lu Qin, Wentao Tian, and Jeffrey Xu Yu. 2014. Querying k-truss community in large and dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (Snowbird, Utah, USA) (SIGMOD '14)*. Association for Computing Machinery, New York, NY, USA, 1311–1322. doi:10.1145/2588555.2610495
- [48] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.
- [49] Muhammad Aqib Javed, Muhammad Shahzad Younis, Siddique Latif, Junaid Qadir, and Adeel Baig. 2018. Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications* 108 (2018), 87–111.
- [50] Yuli Jiang, Yu Rong, Hong Cheng, Xin Huang, Kangfei Zhao, and Junzhou Huang. 2022. Query driven-graph neural networks for community search: from non-attributed, attributed, to interactive attributed. *Proc. VLDB Endow.* 15, 6

- (Feb. 2022), 1243–1255. doi:10.14778/3514061.3514070
- [51] Thomas Karagiannis, Jean-Yves Le Boudec, and Milan Vojnović. 2007. Power law and exponential decay of inter contact times between mobile devices. In *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking* (Montréal, Québec, Canada) (*MobiCom '07*). Association for Computing Machinery, New York, NY, USA, 183–194. doi:10.1145/1287853.1287875
- [52] Hiroaki Kawamichi, Sho K Sugawara, Yuki H Hamano, Kai Makita, Takatori Kochiyama, and Norihiro Sadato. 2016. Increased frequency of social interaction is associated with enjoyment enhancement and reward system activation. *Scientific reports* 6, 1 (2016), 24561.
- [53] Kiana Kheiri and Hamid Karimi. 2023. Sentimentgpt: Exploiting gpt for advanced sentiment analysis and its departure from current machine learning. *arXiv preprint arXiv:2307.10234* (2023).
- [54] Johanna Kissler, Cornelia Herbert, Irene Winkler, and Markus Junghofer. 2009. Emotion and attention in visual word processing—An ERP study. *Biological psychology* 80, 1 (2009), 75–83.
- [55] Christine Klymko, David Gleich, and Tamara G Kolda. 2014. Using triangles to improve community detection in directed networks. *arXiv preprint arXiv:1404.5874* (2014).
- [56] Nane Kratzke. 2017. *The #BTW17 Twitter Dataset - Recorded Tweets of the Federal Election Campaigns of 2017 for the 19th German Bundestag*. doi:10.5281/zenodo.835735
- [57] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2007. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data* 1, 1 (March 2007), 2–es. doi:10.1145/1217299.1217301
- [58] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>.
- [59] Jianxin Li, Xinjue Wang, Ke Deng, Xiaochun Yang, Timos Sellis, and Jeffrey Xu Yu. 2017. Most influential community search over large social networks. In *2017 IEEE 33rd international conference on data engineering (ICDE)*. IEEE, 871–882.
- [60] Ling Li, Siqiang Luo, Yuhai Zhao, Caihua Shan, Zhengkui Wang, and Lu Qin. 2023. COCLEP: Contrastive Learning-based Semi-Supervised Community Search. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 2483–2495.
- [61] Xuankun Liao, Qing Liu, Xin Huang, and Jianliang Xu. 2024. Truss-Based Community Search over Streaming Directed Graphs. *Proc. VLDB Endow.* 17, 8 (May 2024), 1816–1829. doi:10.14778/3659437.3659440
- [62] Aleck Lin, Shirley Gregor, and Michael Ewing. 2009. Understanding the nature of online emotional experiences: a study of enjoyment as a web experience. In *Proceedings of the 11th International Conference on Electronic Commerce*. 259–268.
- [63] Longlong Lin, Pingpeng Yuan, Hong-Hua Li, Chunxue Zhu, Hongchao Qin, Hai Jin, and Tao Jia. 2024. QTCS: Efficient Query-Centered Temporal Community Search. *Proc. VLDB Endow.* 17, 6 (May 2024), 1187–1199. doi:10.14778/3648160.3648163
- [64] Guimei Liu, Limsoon Wong, and Hon Nian Chua. 2009. Complex discovery from weighted PPI networks. *Bioinformatics* 25, 15 (2009), 1891–1897.
- [65] Qing Liu, Minjun Zhao, Xin Huang, Jianliang Xu, and Yunjun Gao. 2020. Truss-based Community Search over Large Directed Graphs. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data* (Portland, OR, USA) (*SIGMOD '20*). Association for Computing Machinery, New York, NY, USA, 2183–2197. doi:10.1145/3318464.3380587
- [66] Luo Lu. 2015. Building trust and cohesion in virtual teams: the developmental approach. *Journal of organizational effectiveness: People and performance* 2, 1 (2015), 55–72.
- [67] Zhao Lu, Yuanyuan Zhu, Ming Zhong, and Jeffrey Xu Yu. 2022. On time-optimal (k, p)-core community search in dynamic graphs. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 1396–1407.
- [68] Annmarie A Lyles, Colleen Loomis, Scherezade K Mama, Sameer Siddiqi, and Rebecca E Lee. 2018. Longitudinal analysis of virtual community perceptions of cohesion: The role of cooperation, communication, and competition. *Journal of Health Psychology* 23, 13 (2018), 1677–1688.
- [69] Omid Madani, Sai Ananth Averinini, and Shashidhar Gandham. 2022. A Dataset of Networks of Computing Hosts. In *Proceedings of the 2022 ACM on International Workshop on Security and Privacy Analytics*. 100–104.
- [70] Janet McLeod and Kathryn Von Treuer. 2013. Towards a cohesive theory of cohesion. (2013).
- [71] Xiaoye Miao, Yue Liu, Lu Chen, Yunjun Gao, and Jianwei Yin. 2022. Reliable community search on uncertain graphs. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 1166–1179.
- [72] Paula M Niedenthal and Marc B Setterlund. 1994. Emotion congruence in perception. *Personality and Social Psychology Bulletin* 20, 4 (1994), 401–411.
- [73] Ravi Paul, John R Drake, and Huigang Liang. 2016. Global virtual team performance: The effect of coordination effectiveness, trust, and team cohesion. *IEEE transactions on professional communication* 59, 3 (2016), 186–202.
- [74] Pamela Paxton and James Moody. 2003. Structure and sentiment: Explaining emotional attachment to group. *Social Psychology Quarterly* (2003), 34–47.
- [75] Anthony T Pescosolido and Richard Saavedra. 2012. Cohesion and sports teams: A review. *Small Group Research* 43, 6 (2012), 744–758.
- [76] X Developer Platform. 2022. Twitter Academic Research API. <https://docs.x.com/x-api/introduction>. Accessed: 2022-12.
- [77] Mohammad Masoud Rahimi, Elham Naghizade, Mark Stevenson, and Stephan Winter. 2023. SentiHawkes: a sentiment-aware Hawkes point process to model service quality of public transport using Twitter data. *Public Transport* 15, 2 (2023), 343–376.
- [78] Karen A Roberto and Priscilla J Kimboko. 1989. Friendships in later life: Definitions and maintenance patterns. *The International Journal of Aging and Human Development* 28, 1 (1989), 9–19.
- [79] Sourav S Bhowmick, Hui Li, S. H. Annabel Chen, and Yining Zhao. 2024. Social Psychology Meets Social Computing: State of the Art and Future Directions. In *Companion Proceedings of the ACM Web Conference 2024* (Singapore, Singapore) (*WWW '24*). Association for Computing Machinery, New York, NY, USA, 1250–1253. doi:10.1145/3589335.3641242
- [80] Eduardo Salas, Rebecca Grossman, Ashley M Hughes, and Chris W Coultas. 2015. Measuring team cohesion: Observations from the science. *Human factors* 57, 3 (2015), 365–374.
- [81] Jessica M Santoro, Aurora J Dixon, Chu-Hsiang Chang, and Steve WJ Kozlowski. 2015. Measuring and monitoring the dynamics of team cohesion: Methods, emerging tools, and advanced technologies. In *Team cohesion: Advances in psychological theory, methods and practice*. Emerald Group Publishing Limited, Leeds, 115–145.
- [82] Iqbal H Sarker, Alan Colman, and Jun Han. 2019. Recencyminer: mining recency-based personalized behavior from contextual smartphone data. *Journal of Big Data* 6, 1 (2019), 1–21.
- [83] John J Shaughnessy, Eugene B Zechmeister, and Jeanne S Zechmeister. 2000. *Research methods in psychology*. McGraw-Hill.
- [84] Kelly G Shaver. 2015. *Principles of social psychology*. Psychology Press.
- [85] Dong-Hee Shin. 2010. Analysis of online social networks: A cross-national study. *Online Information Review* 34, 3 (2010), 473–495.
- [86] Mauro Sozio and Aristides Gionis. 2010. 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* (Washington, DC, USA) (*KDD '10*). Association for Computing Machinery, New York, NY, USA, 939–948. doi:10.1145/1835804.1835923
- [87] Kevin S Spink and Albert V Carron. 1994. Group cohesion effects in exercise classes. *Small Group Research* 25, 1 (1994), 26–42.
- [88] Juliette Stehlé, Alain Barrat, and Ginestra Bianconi. 2010. Dynamical and bursty interactions in social networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 81, 3 (2010), 035101.
- [89] Javier Suárez Álvarez, Ignacio Pedrosa, Luis M Lozano, Eduardo García Cueto, Marcelino Cuesta Izquierdo, José Muñiz Fernández, et al. 2018. Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30 (2018).
- [90] Sue Yeon Syn and Sanghee Oh. 2015. Why do social network site users share information on Facebook and Twitter? *Journal of Information Science* 41, 5 (2015), 553–569.
- [91] Chenhao Tan and Lillian Lee. 2015. All who wander: On the prevalence and characteristics of multi-community engagement. In *Proceedings of the 24th International Conference on World Wide Web*. 1056–1066.
- [92] Yifu Tang, Jianxin Li, Nur Al Hasan Haldar, Ziyu Guan, Jiajie Xu, and Chengfei Liu. 2022. Reliable community search in dynamic networks. *Proc. VLDB Endow.* 15, 11 (July 2022), 2826–2838. doi:10.14778/3551793.3551834
- [93] David Vinson, Marta Ponari, and Gabriella Vigliocco. 2014. How does emotional content affect lexical processing? *Cognition & emotion* 28, 4 (2014), 737–746.
- [94] Jianwei Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, and Ying Zhang. 2024. Efficient Unsupervised Community Search with Pre-Trained Graph Transformer. *Proc. VLDB Endow.* 17, 9 (Aug. 2024), 2227–2240. doi:10.14778/3665844.3665853
- [95] Yuxiang Wang, Xiaoxuan Gou, Xiaoliang Xu, Yuxia Geng, Xiangyu Ke, Tianxing Wu, Zhiyuan Yu, Runhui Chen, and Xiangying Wu. 2024. 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* (Washington DC, USA) (*SIGIR '24*). Association for Computing Machinery, New York, NY, USA, 1680–1690. doi:10.1145/3626772.3657771
- [96] Stanley Wasserman and Katherine Faust. 1994. Social network analysis: Methods and applications. (1994).
- [97] Carolyn Weisz and Lisa F Wood. 2005. Social identity support and friendship outcomes: A longitudinal study predicting who will be friends and best friends 4 years later. *Journal of Social and Personal Relationships* 22, 3 (2005), 416–432.
- [98] Barry Wellman. 2001. Physical place and cyberspace: The rise of personalized networking. *International journal of urban and regional research* 25, 2 (2001), 227–252.
- [99] Xiaoqin Xie, Mingjie Song, Chiming Liu, Jiaming Zhang, and Jiahui Li. 2021. Effective influential community search on attributed graph. *Neurocomputing* 444 (2021), 111–125.
- [100] Tianyang Xu, Zhao Lu, and Yuanyuan Zhu. 2022. Efficient triangle-connected truss community search in dynamic graphs. *Proceedings of the VLDB Endowment* 16, 3 (2022), 519–531.

- [101] Jaewon Yang and Jure Leskovec. 2012. Defining and evaluating network communities based on ground-truth. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics* (Beijing, China) (MDS '12). Association for Computing Machinery, New York, NY, USA, Article 3, 8 pages. doi:10.1145/2350190.2350193
- [102] Suyong Yang, Wenbo Luo, Xiangru Zhu, Lucas S Broster, Taolin Chen, Jinzhen Li, and Yuejia Luo. 2014. Emotional content modulates response inhibition and perceptual processing. *Psychophysiology* 51, 11 (2014), 1139–1146.
- [103] Kai Yao and Lijun Chang. 2021. Efficient size-bounded community search over large networks. *Proc. VLDB Endow.* 14, 8 (April 2021), 1441–1453. doi:10.14778/3457390.3457407
- [104] Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. 2017. Local Higher-Order Graph Clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada) (KDD '17). Association for Computing Machinery, New York, NY, USA, 555–564. doi:10.1145/3097983.3098069
- [105] Long Yuan, Lu Qin, Wenjie Zhang, Lijun Chang, and Jianye Yang. 2017. Index-based densest clique percolation community search in networks. *IEEE Transactions on Knowledge and Data Engineering* 30, 5 (2017), 922–935.
- [106] Andrew Zamecnik, Cristina Villa-Torrano, Vitomir Kovanović, Georg Grossmann, Srećko Joksimović, Yannis Dimitriadis, and Abelardo Pardo. 2022. The cohesion of small groups in technology-mediated learning environments: A systematic literature review. *Educational Research Review* 35 (2022), 100427.
- [107] Fan Zhang, Haicheng Guo, Dian Ouyang, Shiyu Yang, Xuemin Lin, and Zhihong Tian. 2023. Size-constrained community search on large networks: An effective and efficient solution. *IEEE Transactions on Knowledge and Data Engineering* 36, 1 (2023), 356–371.
- [108] Yining Zhao. 2025. Codebase for the paper "How Cohesive Are Community Search Results on Online Social Networks?: An Experimental Evaluation". <https://yining.github.io/PIANO/>
- [109] Yining Zhao, Sourav S Bhowmick, Nastassja L. Fischer, and SH Annabel Chen. 2025. How Cohesive Are Community Search Results on Online Social Networks?: An Experimental Evaluation [Technical Report]. arXiv:2504.19489 [cs.IR] <https://arxiv.org/abs/2504.19489>
- [110] Philip G Zimbardo, Robert Lee Johnson, Vivian McCann, and Carol Carter. 2003. *Psychology: core concepts*. Allyn and Bacon Boston.