Detecting Informationally-Dense Subsets

Chao Zhao, Ali Al-Bashabsheh and Chung Chan

Abstract—**Identifying locally dense subgraphs aims to pinpoint subgraphs characterized by tight internal connectivity. However, existing methods for identifying dense subgraphs based on density can lead to loose internal connections. This paper addresses this issue by introducing a concept of strength to detect strong subsets. Our approach encompasses the existing work that finds a nested chain of densest k-subgraphs as a special case and reveals subgraphs that have tight internal connections overlooked by existing methods. The strong subsets exhibit a laminar structure and can be computed in polynomial time. In contrast to previous works defining locally densest subgraphs without a natural extension to weighted graphs, our method accommodates both weighted and unweighted, directed and undirected graphs, as well as hypergraphs. Furthermore, it extends to a broader notion of information density, surpassing the scope of weighted graphs.**

I. Introduction

Graphs commonly model real-world problems for analysis, with nodes symbolizing individuals and edges depicting relations among them. In a subset of individuals, the denser the edges, the more profound the connections between individuals [2, 3]. The exploration of dense subgraphs finds extensive applications in social network analysis, bioinformatics, market analysis, and various other domains[4–6]. A fundamental definition of density involves the ratio of the number of edges to the number of nodes.

Finding the densest subgraph involves identifying subgraphs with the maximum density, a well-studied problem [7–9]. However, existing studies on finding the densest subgraph and its variants tend to overlook subgraphs within the maximal densest subgraph. Densest subgraphs can be computed efficiently in polynomial time using the max-flow min-cut algorithm [8]. A common variant imposes cardinality constraints on densest subgraphs, such as finding densest k-subgraphs [9–12]. However, these constraints render the problem NP-hard.

Another variant [13] seeks a chain of nested dense subgraphs, where each dense subgraph is a densest k-subgraph with a density strictly smaller than the subgraphs properly contained within it. The maximal densest subgraph is characterized by the one with the smallest cardinality among the identified non-empty dense subgraphs, implying that subgraphs within the maximal densest subgraph remain unaddressed.

Qin et al. [14] introduced the concept of locally densest subgraphs aiming at identifying connected subgraphs that represent dense regions of a graph. However, their definition relies on connectedness, limiting its applicability to unweighted graphs with no natural extension to weighted graphs. Additionally, the maximal connected densest subgraph is always considered the locally densest subgraph, disregarding the internal subgraphs and whether it genuinely possesses tight internal connections.

In this paper, we tackle the challenge of potentially overlooking dense subgraphs or giving loosely connected subgraphs when employing density-based methods for dense subgraph detection. To overcome these challenges, we introduce a notion of strength for detecting dense subgraphs which can be computed in polynomial time and form a hierarchy. Moreover, our approach also applies to detecting dense subsets in models beyond graphs, and few existing methods have such generality.

II. Motivating Examples

Density for graphs. For a simple graph on V with the set $\mathcal{E}$ of edges, the density function is defined as

$$\rho(B) := \frac{|\mathcal{E}(B)|}{|B|}, \qquad \text{for } B \subseteq V : |B| > 0, \tag{1}$$

where $\mathcal{E}(B)$ is the set of edges internal to B.

Consider the graphs in Fig. 1. For Fig. 1a, with $b = 3$,

$$\rho(V) = \frac{7}{6} > 1 = \rho(C_1) = \rho(C_2), \tag{2}$$

and for Fig. 1b, with $b = 4$,

$$\rho(V) = \rho(C_1) = \frac{3}{2} > \frac{5}{4} = \rho(C_2). \tag{3}$$

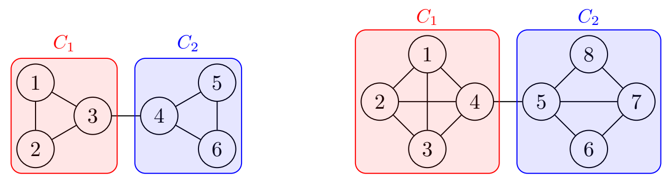
A detailed calculation is in Appendix A.1 (Appendices are in the longer version of this paper given in [1]). Based on the internal and external connections of the subgraphs, any sensible clustering approach would be expected to identify C_1, C_2, and V as dense subgraphs / subsets in both grpahs. However, as we will see, typical methods based on density fail to do so.

A *densest subgraph* is defined as the subgraph with the maximum density. With the notion of densest subgraphs, in Fig. 1a, V can be identified since it has the maximum density. However, C_1 and C_2 are not identified even though they are cliques.

A *densest k-subgraph* [15] is defined as the subgraph that has the maximum density among all the subgraphs of cardinality k (See Appendix A.1). With the notion of densest k-subgraphs, we are able to identify C_1 and C_2 in Fig. 1a since they are densest k-subgraphs for $k = b$. However, in Fig. 1b, C_2 cannot be identified. Additionally, identifying densest k-subgraphs is NP-hard [15, 16].

The method in [13] identifies a nested chain of dense subgraphs, each of which is a densest k-subgraph [17]. However, any subgraphs properly contained in the maximal densest

C. Chan (corresponding author, email: chung.chan@cityu.edu.hk), and C. Zhao are with the Department of Computer Science, City University of Hong Kong. A. Al-Bashabsheh is with the Big Data and Brain Computing (BDBC) center at Beihang University, Beijing, China.



(a) (b)

Fig. 1: (a) The dumbbell graph constructed by connecting two b-clique ($b \geq 3$) by an edge with $b = 3$; (b) The graph constructed by connecting a b-clique and a quasi b-clique ($b \geq 4$) by an edge with $b = 4$.

subgraphs will not be identified, which means, C_1 and C_2 in Fig. 1a and Fig. 1b are not identified.

The notion of *locally densest subgraphs* (See Appendix A.2) proposed in [14] also has limitations when applied on graphs in Fig. 1. One limitation in the definition for locally densest subgraphs is that the connectedness requirement means locally densest subgraph only applies to unweighted graphs and there is no natural extension to weighted graphs. Moreover, on one hand, locally densest subgraphs in a graph are disjoint [14, Lemma 3.3]; on the other hand, any maximal densest connected subgraph is a locally densest subgraph [14, Lemma 4.1]. This means the maximal densest connected subgraph is identified as a locally densest subgraph and any finer dense connected subgraph contained in the maximal densest connected subgraph will not be separately identified.

Then in Fig. 1a, V is the locally densest subgraph, while C_1 and C_2 are not. However, as we can see, C_1 and C_2 are loosely connected by merely one edge, the removal of which will even make the subgraph disconnected. Additionally, C_1 and C_2 are not maximally dense due to that V has both larger density and more nodes than C_1 and C_2.

In Fig. 1b, C_2 is not a densest k-subgraph, hence escaped the detection of methods in [13] and [14].

Density only cares about the total internal edges and does not reflect how the nodes are connected. Given a subgraph that has the maximum density, even if the subgraph is connected, it does not necessarily mean the graph has tight internal connections. To address the issues, we propose to detect dense subgraphs by measuring how strongly the nodes are internally connected. Furthermore, our approach will also applied to a general notion of information density well beyond weighted graphs.

III. Density in Information Networks

In various machine learning applications, such as representation learning, a common practice involves using a network of nodes with associated information or representations [18, 19]. Exploring the density of correlated nodes or nodes with high information content is significant and holds the potential to enhance our understanding of related machine-learning models. In this context, we propose a formulation of information density to address related questions in a theoretical manner.

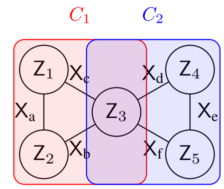


Fig. 2: Graphical illustration of the info-net in Example 1.

A. *Infomation network*

Consider an *information network (info-net)* defined by a discrete random tuple $\mathsf{Z}_V := (\mathsf{Z}_i | i \in V)$, where V is a finite set of nodes and Z_i is the information/representation of node $i \in V$. How to find meaningful informationally dense subsets?

Example 1 (Info-net on Touching Triangles) Given 6 uniformly random independent bits $\mathsf{X}_j, j \in \{\mathrm{a}, \dots, \mathrm{f}\}$, define Z_V for the 5 nodes in $V = \{1, \dots, 5\}$ as $\mathsf{Z}_1 = (\mathsf{X}_\mathrm{a}, \mathsf{X}_\mathrm{c}), \mathsf{Z}_2 = (\mathsf{X}_\mathrm{a}, \mathsf{X}_\mathrm{b}), \mathsf{Z}_3 = (\mathsf{X}_\mathrm{b}, \mathsf{X}_\mathrm{c}, \mathsf{X}_\mathrm{d}, \mathsf{X}_\mathrm{f}), \mathsf{Z}_4 = (\mathsf{X}_\mathrm{d}, \mathsf{X}_\mathrm{e}), \mathsf{Z}_5 = (\mathsf{X}_\mathrm{e}, \mathsf{X}_\mathrm{f})$. The graphical illustration of this info-net is shown in Fig. 2.

When each each random independent bit X_j denotes a publication, and each Z_i denotes a scholar represented by his publications, we get the graph modeling a simple co-authorship network if we view each bit X_j as an edge and each Z_i as a node.

$C_1 := \{1, 2, 3\}$ and $C_2 := \{3, 4, 5\}$ are informationally dense intuitively as they corresponds to 3-cliques. □

Graphs can only represent special cases of info-nets. Detecting communities in info-nets calls for a general definition of information density and community detection methods applicable beyond graphs.

B. *Density in Info-nets*

Undirected info-nets. Consider the info-net

$$\mathsf{Z}_V := (\mathsf{Z}_i | i \in V) \sim P_{\mathsf{Z}_V} \in \mathcal{P}(Z_V), \tag{4}$$

where Z_i is a discrete random variable representing node $i \in V$.

The density can be taken as

$$\rho(B) = \frac{H(\mathsf{Z}_B | \mathsf{Z}_{V \setminus B})}{|B|} \qquad \text{for } B \subseteq V : |B| \geq 1, \tag{5}$$

which measures the amount of information per node internal to B not available from $V \setminus B$.

In Example 1 we show a case of how a simple undirected info-net can model a simple co-authorship network. We analyze the density of subsets in the graph in the following example.

Example 2 With $C_1 := \{1, 2, 3\}$ in the info-net defined in Example 1,

$$\begin{aligned} H(\mathsf{Z}_{C_1} | \mathsf{Z}_{V \setminus C_1}) = H(\mathsf{X}_\mathrm{a}, \mathsf{X}_\mathrm{b}, \mathsf{X}_\mathrm{c}) &= 3 \text{ bits} \\ \rho(C_1) = \frac{H(\mathsf{Z}_{C_1} | \mathsf{Z}_{V \setminus C_1})}{|C_1|} &= 1 \text{ bit per node.} \end{aligned}$$

We skip $C_2 = \{3, 4, 5\}$ since it is symmetric to C_1.

With $C_3 := \{4, 5\}$,

$$\begin{aligned} H(\mathsf{Z}_{C_3} | \mathsf{Z}_{V \setminus C_3}) &= H(\mathsf{X}_\mathrm{e}) = 1 \text{ bit} \\ \rho(C_3) = \frac{H(\mathsf{Z}_{C_3} | \mathsf{Z}_{V \setminus C_3})}{|C_1|} &= \frac{1}{2} \text{ bit per node} < \rho(C_1) \end{aligned}$$

C_1 has a larger density than C_3, which means scholars in C_1 are more productive and have a larger number of average publications. □

IV. Problem Formulation

The study of information involves metrics such as entropy and mutual information. However, the underlying properties go beyond graphical representations. For instance, when entropy is used as the volume function (e.g., in the normalized density in equation (34) in Appendix A.4), a graph alone is insufficient to capture density. Instead, a measure of entropy or mutual information is essential.

To define the density that generalizes the usual density for graphs and applies to information density, we utilize supermodular and submodular functions in our formulation.

A set function $f : 2^V \to \mathbb{R}$ is said to be normalized if $f(\emptyset) = 0$, non-decreasing if

$$f(A) \leq f(B) \qquad \forall A \subseteq B \subseteq V, \tag{6}$$

and submodular if

$$f(B_1)+f(B_2) \geq f(B_1 \cap B_2)+f(B_1 \cup B_2) \; \forall B_1,B_2 \subseteq V. \tag{7}$$

If the inequality (7) is reversed, then f is called supermodular. For example, the entropy function for discrete random variables $\nu(B) := H(\mathsf{Z}_B)$ is submodular (also normalized and non-decreasing), while the conditional entropy $\mu(B) := H(\mathsf{Z}_B \mid \mathsf{Z}_{V \setminus B})$ is supermodular. The cardinality function $\nu(B) := |B|$ is both submodular and supermodular.

Definition 1 The *density function* is defined as

$$\rho(B) := \frac{\mu(B)}{\nu(B)} \qquad \text{for } B \subseteq V : |B| \geq 1 \tag{8}$$

for any (V, μ, ν) network where

- V is a set of $n \geq 1$ nodes;
- $\mu : 2^V \to \mathbb{R}$ is the mass function that is normalized, non-decreasing, and supermodular; and
- $\nu : 2^V \to \mathbb{R}$ is the volume function that is normalized, non-decreasing, and submodular. □

To avoid singularity, we further impose

$$\nu(\{i\}) > 0 \qquad \forall i \in V. \tag{9}$$

In particular, by defining $\mu(B) := H(\mathsf{Z}_B | \mathsf{Z}_{V \setminus B})$ and $\nu(B) := |B|$, the density in (8) reduces to (5).

The dense subgraphs identified based on density are not guaranteed to be internally tightly connected. Even the densest subgraph may exhibit weak internal connections, as illustrated in Section II. Therefore, we introduce the concept of strength to address the limitation, which involves finding the weakest link to break the subsets. Moreover, the concept applies to the supermodular mass function and submodular volume function beyond graphs.

Definition 2 (Strength) Given a family $\mathcal{F} \subseteq 2^V$, define for $C \subseteq V$ the family

$$\mathcal{F}(C) := \{B \in \mathcal{F} | B \subseteq C\}, \tag{10}$$

and the strength

$$\sigma_{\mathcal{F}}(C) := \inf_{B \in \mathcal{F}(C): \nu(C|B) > 0} \frac{\mu(C|B)}{\nu(C|B)}, \tag{11}$$

where μ and ν are, respectively, the mass and volume functions for the density defined in Definition 1. For $k \in \mathbb{N} : k \leq |V| - 1$, define the family

$$\mathcal{F}_k := \{B \subseteq V | |B| \geq k\}, \tag{12}$$

and for notation simplicity, we denote the corresponding strength $\sigma_{\mathcal{F}_k}$ as σ_k. □

In (11), The conditional versions $\mu(C|B) := \mu(C \cup B) - \mu(B)$, and similar for ν.

The strength in (11) can be explained with the *importance* of a set T defined as

$$\mathfrak{Z}_C(T) := \frac{\mu(C) - \mu(C \setminus T)}{\nu(C) - \nu(C \setminus T)}. \tag{13}$$

which measures the decrease of mass normalized by the decrease of volume if T along with the incident edges are removed, and hence captures how a subset T is connected both externally and internally. In (13), when $\emptyset \subsetneq T \subsetneq C$, the less the edges incident on T, the more likely T to be loosely connected externally or internally. Particularly, when $|T| = 1$, (13) reduces to the degree centrality of the node in T; and when $T = C$, (13) equals $\rho(C)$.

$\sigma_k(C)$ evaluates how strong C is by the smallest importance over subsets of C that are in the feasible region, since

$$\sigma_k(C) = \inf_{T \subseteq C: T \neq \emptyset, |C \setminus T| \geq k} \mathfrak{Z}_C(T). \tag{14}$$

When $k = 0$, C is in the feasible region of (14), then $\sigma_0(C)$ will never be larger than $\rho(C)$, or in other words, will never identify the subset with strength larger than $\rho(C)$; while for $k \geq 1$, there is no such limitation. It will be shown in Theorem 1 that when $k \geq 1$, we would be able to identify the desired subsets with strength larger than $\rho(C)$.

Since $\sigma_k(C)$ is a measure of whether C has tight internal connections, we can set a threshold x to identify all subsets that have a strength larger than x. A maximal C with strength larger than x implies that C is locally strong concerning the threshold. We propose the following definition of strong subsets, further confining the comparisons of strong subsets to local instead of global comparisons.

Definition 3 (Strong Subsets) Given a family $\mathcal{F} \subseteq 2^V$, define for all $x \in \mathbb{R}$

$$\mathcal{C}_{\mathcal{F}}(x) := \operatorname{maximal}\{C \in \mathcal{F} | \sigma_{\mathcal{F}}(C) > x\}, \tag{15}$$

where $\operatorname{maximal} \mathcal{F}' := \{B \in \mathcal{F}' | \nexists C \in \mathcal{F}', B \subsetneq C\}$.

The subsets in the collection $\mathcal{C}_{\mathcal{F}} := \bigcup_{x \in \mathbb{R}} \mathcal{C}_{\mathcal{F}}(x)$ are called the strong subsets. □

We will demonstrate how $\mathcal{C}_{\mathcal{F}}$ covers or is related to existing works, and explain why $\mathcal{C}_{\mathcal{F}_1}$ avoids issues of existing work on graphs such as Fig. 1. Moreover, the subsets in $\mathcal{C}_{\mathcal{F}_1}$ has a laminar structure and can be computed in polynomial time.

V. Main Result

The following result elucidates why our formulation for strength with $k \geq 1$ can identify meaningful subsets not found by existing works.

Theorem 1 *For any $C \in \mathcal{F}_k$, when $k \geq 1$,*

$$\sigma_k(C) > \rho(C) \iff \rho(C) > \rho(B) \quad \forall B \subsetneq C : |B| \geq k; \tag{16}$$

and when $k = 0$, $\sigma_0(C) \leq \rho(C)$ for $C \neq \emptyset$. □

PROOF See Appendix B.2. ■

For a subset $C \subseteq V : C \neq \emptyset$, $\sigma_0(C)$ is always upper-bounded by $\rho(C)$, while $\sigma_1(C)$ can exceed $\rho(C)$. In Fig. 1a, $\sigma_1(V)$ is strictly larger than $\rho(V)$, and C_1 is a maximal strong subset with $\sigma_1(C_1)$ larger than $\sigma_1(V)$, and hence C_1 can be captured by $\mathcal{C}_{\mathcal{F}_1}$. Similar for C_2.

Choosing the appropriate value for k of $\mathcal{F}_k$ to restrict the family in the minimization of (11) is crucial. Opting for $k = 1$ allows us to identify subsets with tight internal connections that are overlooked by existing works and mitigates the issue of potentially neglecting weak connectedness within maximal (connected) densest subgraphs. When $k \geq 2$, we can identify overlapping strong communities. For instance, in Fig. 2, both C_1 and C_2 can be identified, despite their overlapping nature.

The subsets in $\mathcal{C}_{\mathcal{F}_0}$ form a nested chain, covering the existing work in [13] as a special case when applied to graphs.

Proposition 1 *The subsets in $\mathcal{C}_{\mathcal{F}_0}$ form a nested chain*

$$B_1 := V \supsetneq B_2 \supsetneq \cdots \supsetneq B_{N'-1} \supsetneq B_{N'} := \emptyset, \tag{17}$$

where $N' = |\mathcal{C}_{\mathcal{F}_0}|$, and $B_{N'-1}$ is the maximal densest subset. □

PROOF See Appendix B.5. ■

This also implies that $\mathcal{C}_{\mathcal{F}_0}$ cannot identify the strong subsets within the maximal densest subsets. As mentioned earlier, for Fig. 1, $\mathcal{C}_{\mathcal{F}_0}$ fails to identify C_1 and C_2.

In contrast, $\mathcal{C}_{\mathcal{F}_1}$ can successfully identify the meaningful communities C_1 and C_2 in Fig. 1. The subsequent result outlines the specific conditions under which these communities can be identified by $\mathcal{C}_{\mathcal{F}_1}$ but not by $\mathcal{C}_{\mathcal{F}_0}$.

Theorem 2 *$\mathcal{C}_{\mathcal{F}_1}$ can contain dense subsets that are not in $\mathcal{C}_{\mathcal{F}_0}$, and the collection of these dense subsets is given by*

$$\mathcal{C}_{\mathcal{F}_1} \setminus \mathcal{C}_{\mathcal{F}_0} = \{C \in \mathcal{C}_{\mathcal{F}_1} \,|\exists C' \in \mathcal{C}_{\mathcal{F}_1} : C' \supsetneq C, \sigma_1(C') > \rho(C')\}. \tag{18}$$

Additionally, $\mathcal{C}_{\mathcal{F}_0} \setminus \mathcal{C}_{\mathcal{F}_1} = \{\emptyset\}$. □

PROOF See Appendix B.6. ■

Theorem 2 presents a universal condition applicable to all cases. It establishes that all non-empty strong communities identified by $\mathcal{C}_{\mathcal{F}_0}$ are also identified by $\mathcal{C}_{\mathcal{F}_1}$, while for a subset C' identified by $\mathcal{C}_{\mathcal{F}_1}$ with $\sigma_1(C') > \rho(C')$, all the strong subsets properly contained in C' cannot be identified by $\mathcal{C}_{\mathcal{F}_0}$.

In both Fig. 1a and Fig. 1b, C_1 and C_2 can be identified by $\mathcal{C}_{\mathcal{F}_1}$ but not by $\mathcal{C}_{\mathcal{F}_0}$. Refer to Appendix A.3 for calculations.

In fact, the strength σ_1 is a similarity measure (see Appendix B.7) and nodes in the same strong subsets in $\mathcal{C}_{\mathcal{F}_1}(x)$ can be regarded as similar w.r.t. the threshold x. The nodes from $C_1 \in \mathcal{C}_{\mathcal{F}_1}(x_2)$ and $C_2 \in \mathcal{C}_{\mathcal{F}_1}(x_2)$ are similar w.r.t. $\min\{x_1, x_2\}$.

Lemma 1 *For all $C_1, C_2 \subseteq V : C_1 \cap C_2 \neq \emptyset$,*

$$\sigma_1(C_1 \cup C_2) \geq \min\{\sigma_1(C_1), \sigma_1(C_2)\}. \tag{19}$$

□

PROOF See Appendix B.7. ■

As a result of Lemma 1, we have the following Theorem.

Theorem 3 *For $0 \leq x_1 \leq x_2$ and $C_i \in \mathcal{C}_{\mathcal{F}_1}(x_i), i \in \{1, 2\}$, we have $C_1 \supseteq C_2$ or $C_1 \cap C_2 = \emptyset$. Furthermore, $C_1 \supsetneq C_2$ implies $x_1 < x_2$.* □

PROOF See Appendix B.8. ■

$\mathcal{C}_{\mathcal{F}_1}$ has an elegant structure according to Theorem 3. The strong subsets in $\mathcal{C}_{\mathcal{F}_1}$ exhibit a laminar structure and form a hierarchy, providing the foundation for its polynomial-time computability.

VI. Computation

A. Compute $\mathcal{C}_{\mathcal{F}_1}(x)$

Consider the following optimization problem

$$f_k(x) := \max_{B \in \mathcal{F}_k} f[B](x), \text{ where} \tag{20}$$

$$f[B](x) := \mu(B) - x\nu(B), \tag{21}$$

and denote $\mathcal{S}_{\mathcal{F}_k}(x)$ as the set of inclusion-wise minimal solution to (20).

We first show how to compute $\mathcal{S}_{\mathcal{F}_k}(x)$, and then show how to make use of $\mathcal{S}_{\mathcal{F}_1}(x)$ to compute $\mathcal{C}_{\mathcal{F}_1}(x)$.

Given $x \geq 0$, $f[B](x)$ is a supermodular function, of which the maximization can be done with submodular function minimization (SFM) methods in polynomial time when the feasible region is a lattice [20]. When $k \geq 1$, the feasible region of (20) is not a lattice, nevertheless, it can be solved by writing it as a two-step optimization problem

$$f_k(x) = \max_{A \subseteq V : |A| = k} f_A(x), \text{ where} \tag{22a}$$

$$f_A(x) := \max_{B \subseteq V : B \supseteq A} f[B](x). \tag{22b}$$

where the feasible region of (22b) is a lattice, which means SFM method can be used for solving (22b) in polynomial time, and there is a unique minimal solution [20, Proposition 10.1].

Let $C_{A,x}$ denote the minimal solution to (22b) and T^*_x denote the solution set to (22a), then $\mathcal{S}_{\mathcal{F}_k}(x)$ is given by the following result, which is an extension of [21, Proposition 2].

Proposition 2 *The set $\mathcal{S}_{\mathcal{F}_k}(x)$ of inclusion-wise minimal solution to (20) is given by*

$$\mathcal{S}_{\mathcal{F}_k}(x) = \text{minimal}\{C_{A,x} | A \in T^*_x\}, \tag{23}$$

which can be computed with $\mathcal{O}(|V|^k)$ submodular function minimizations. □

PROOF It follows from definitions of $\mathcal{S}_{\mathcal{F}_k}$, $C_{A,x}$ and T^*_x. ■

As an important step to show how to compute $\mathcal{C}_{\mathcal{F}_k}(x)$ based on $\mathcal{S}_{\mathcal{F}_k}(x)$, the following result derived from (11) elucidates the dependence of the magnitude relationship between $\sigma_k(V)$ and a given $x \in \mathbb{R}$ based on the optimization over $f[B](x)$.

Proposition 3 *For $x \in \mathbb{R}$ and $k \in \mathbb{N} : k \leq |V| - 1$,*

$$\sigma_k(V) < x \iff f[V](x) < \max_{B \in \mathcal{F}_k : B \neq V} f[B](x), \quad \textbf{(24a)}$$

$$\sigma_k(V) = x \implies f[V](x) = \max_{B \in \mathcal{F}_k : B \neq V} f[B](x), \quad \textbf{(24b)}$$

$$\sigma_k(V) > x \impliedby f[V](x) > \max_{B \in \mathcal{F}_k : B \neq V} f[B](x). \quad \textbf{(24c)}$$

This implies that when $\sigma_k(V) < \infty$, $\sigma_k(V)$ is the x-coordinate of the first turning point of $f_k(x)$ along the x-axis in the increasing direction. Furthermore, if V satisfies condition

$$\forall B' \subseteq V : B' \neq \emptyset, \nexists B \subsetneq B', \text{ s.t. } \nu(B'|B) = 0 \text{ and } \mu(B'|B) = 0, \quad \textbf{(25)}$$

then the converses of (24b) and (24c) hold. □

PROOF See Appendix B.1. ■

The optimization on the right-hand side of (24) share the same objective function as that of (20) but excludes V from the feasible region. The following result shows the connection of the minimal solution set between the two optimizations.

Proposition 4 *For $x \in \mathbb{R}$ and $k \in \mathbb{N} : k \leq |V| - 1$, when $\sigma_k(V) \leq x$, the minimal solution set to (20) is also the solution set to the optimization on the right-hand side of (24), i.e.*

$$\mathcal{S}_{\mathcal{F}_k}(x) = \text{minimal} \operatorname*{arg\,max}_{B \in \mathcal{F}_k : B \neq V} f[B](x), \text{ and} \quad \textbf{(26)}$$

$$f_k(x) = \max_{B \in \mathcal{F}_k : B \neq V} f[B](x); \quad \textbf{(27)}$$

when $\sigma_k(V) > x$, (26) and (27) hold if V satisfy the condition in (25). □

PROOF See Appendix B.9. ■

Proposition 24 and Proposition 4 enables us to make use of $\mathcal{S}_{\mathcal{F}_k}(x)$ for computing $\mathcal{C}_{\mathcal{F}_k}(x)$. When V satisfies the condition in (25), if (27) hold, it means $\sigma_k(V) > x$ by Proposition 4. If not, the following result indicates that each subset in $\mathcal{S}_{\mathcal{F}_k}(x)$ is a strong subset in $\mathcal{C}_{\mathcal{F}_k}(x)$.

Proposition 5 *For $x \in \mathbb{R}$ and $k \in \mathbb{N} : k \leq |V| - 1$, when V satisfies the condition in (25), we have $\mathcal{S}_{\mathcal{F}_k}(x) \subseteq \mathcal{C}_{\mathcal{F}_k}(x)$.* □

PROOF See Appendix B.10. ■

However, when V does not satisfy the condition in (25), for $B \in \mathcal{S}_{\mathcal{F}_k}(x)$, it is not ensured to be a maximal subset with strength larger than x. There might be another $B' \supsetneq B$ with curve $f[B'](x)$ overlapping with $f[B](x)$ and $\sigma_k(B') > x$, but B' will never be watched by $\mathcal{S}_{\mathcal{F}_k}(x)$ since it is not maximal compared with B.

Nevertheless, we have the following result that generalizes Proposition 5 and holds no matter whether V satisfies the condition in (25) or not.

Proposition 6 *For $x \in \mathbb{R}$ and $k \in \mathbb{N} : k \leq |V| - 1$, we have*

$$\tilde{\mathcal{S}}_{\mathcal{F}_k}(x) \subseteq \mathcal{C}_{\mathcal{F}_k}(x), \quad \textbf{(28)}$$

where

$$\tilde{\mathcal{S}}_{\mathcal{F}_k}(x) := \{B \cup \{i \in V | \nu(\{i\}|B) = 0\} \,|\, B \in \mathcal{S}_{\mathcal{F}_k}(x)\}. \quad \textbf{(29)}$$

When V satisfies the condition in (25), $\tilde{\mathcal{S}}_{\mathcal{F}_k}(x) = \mathcal{S}_{\mathcal{F}_k}(x)$. □

PROOF See Appendix B.11. ■

Proposition 6 implies that any subset in $\tilde{\mathcal{S}}_{\mathcal{F}_k}(x)$ is a strong subset. (29) also shows how to construct $\tilde{\mathcal{S}}_{\mathcal{F}_k}(x)$ from $\mathcal{S}_{\mathcal{F}_k}(x)$.

To compute $\mathcal{C}_{\mathcal{F}_1}(x)$, we need to compute the strong subsets besides those in $\tilde{\mathcal{S}}_{\mathcal{F}_k}(x)$ with k=1. The laminar structure indicated by Theorem 3 implies that the remaining strong subsets in $\mathcal{C}_{\mathcal{F}_1}(x) \setminus \tilde{\mathcal{S}}_{\mathcal{F}_1}(x)$ must be subsets of $V \setminus \bigcup \tilde{\mathcal{S}}_{\mathcal{F}_1}(x)$. This motivates the following recurrence relation for computing $\mathcal{C}_{\mathcal{F}_1}(x)$. Though $\mathcal{F}_1(V) = \mathcal{F}_1$ according to (10), we use $\mathcal{F}_1(V)$ here to avoid confusion.

Theorem 4 *For $x \geq 0$, $\mathcal{C}_{\mathcal{F}_1(V)}(x)$ can be calculated with the following recurrence relation*

$$\mathcal{C}_{\mathcal{F}_1(V)}(x) = \begin{cases} \emptyset, & \text{if } V = \emptyset, \quad \textbf{(30a)} \\ \tilde{\mathcal{S}}_{\mathcal{F}_1(V)}(x) \cup \mathcal{C}_{\mathcal{F}_1(U)}(x), & \text{otherwise}, \quad \textbf{(30b)} \end{cases}$$

where $U := V \setminus \bigcup \tilde{\mathcal{S}}_{\mathcal{F}_1(V)}(x)$, and (30a) is the base case. □

In each recursive step except the first step, U is updated and serves as the ground set for the ongoing recursive step. When reaching the base case (30a), it means each node has found its respective strong subset, which further implies that all strong subsets in $\mathcal{C}_{\mathcal{F}_1(V)}(x)$ are detected due to the laminar structure stated in Theorem 3, thereby concluding the recursion.

Concerning $\mathcal{C}_{\mathcal{F}_k}(x)$ with $k \geq 2$, we lack a recurrence relation for its computation because $\mathcal{C}_{\mathcal{F}_k}$ for $k \geq 2$ do not necessarily exhibit a laminar structure. See Appendix C.

B. Compute $\mathcal{C}_{\mathcal{F}_1}$

A straightforward approach is to compute $\mathcal{C}_{\mathcal{F}_1}(x)$ for a suitable x. According to Theorem 3, each of the remaining strong subsets in $\mathcal{C}_{\mathcal{F}_1}$ is either a proper subset of an element in $\mathcal{C}_{\mathcal{F}_1}(x)$ or is the union of at least two elements in $\mathcal{C}_{\mathcal{F}_1}(x)$. For the former case, we can treat each $B \in \mathcal{C}_{\mathcal{F}_1}(x)$ as the ground set and compute $\mathcal{C}_{\mathcal{F}_1(B)}(x)$. In the latter case, we can contract each $B \in \mathcal{C}_{\mathcal{F}_1}(x)$ into a single node and then compute non-singleton strong subsets for the new ground set with the contracted nodes.

This straightforward approach involves $\mathcal{O}(|V|^3)$ submodular function minimizations, as each $\mathcal{C}_{\mathcal{F}_1}(x)$ requires $\mathcal{O}(|V|^2)$ submodular function minimization.

VII. Conclusion

We introduce the notion of strength for detecting strong subsets. Theorem 1 clarifies that the strength σ_k we propose, for $k \geq 1$, is not upper-bounded by density. Consequently, it can be employed to identify strong subsets overlooked by methods based on density. Theorem 2 demonstrates that our formulation not only encompasses existing work on finding nested densest k-subgraphs as a special case but also identifies additional strong subsets with $\mathcal{C}_{\mathcal{F}_1}$. Theorem 3 establishes that $\mathcal{C}_{\mathcal{F}_1}$ have a laminar structure. The strong subsets are commutable in polynomial time based on submodular function minimization, and Theorem 4 provides the recurrence relation for computation. Since our method is rooted in the general notion of information density and aims to discover informationally dense subsets, its applicability extends beyond graphs.

Appendix A
Complementary Explanations

A.1 Dense subgraphs and calculation of motivating examples

Densest subgraph. Define

$$\mathcal{D}_* := \arg\max_{B\subseteq V:|B|\geq 1} \rho(B), \tag{31}$$

which corresponds to the set of node sets of the densest subgraphs. The densest subgraphs can be computed in polynomial time by solving a maxflow problem[8, 22].

It is not enough to only detect the densest subgraphs, as seen in the following example.

Example 3 For integer $b \geq 3$, construct a graph on $V = C_1 \cup C_2$ as a disjoint union of a b-clique on C_1 and another one on C_2 connected by an edge, as shown in Fig. 1a.

$$\rho(C_1) = \rho(C_2) = \frac{|\mathcal{E}(C_1)|}{b} = \frac{b-1}{2},$$
$$\rho(V) = \frac{2|\mathcal{E}(C_1)|+1}{2b} = \rho(C_1) + \frac{1}{2b} > \rho(C_1).$$

Hence, $C_1, C_2 \notin \mathcal{D}_*$, even though $C_1, C_2 \notin \mathcal{D}_*$ are, obviously, dense subsets since they are cliques. □

Densest k-subgraphs. Denote the collection of densest k-subsets for order $k \geq 1$ as

$$\mathcal{D}_k := \arg\max_{B\subseteq V:|B|=k} \rho(B). \tag{32}$$

By this definition, it is obvious that any densest subset is a densest k-subset for a certain k.

In Example 3, $C_1, C_2 \in \mathcal{D}_b$, which means C_1 and C_2 can be identified as dense subsets by the notion of densest k-subgraphs.

However, computing the densest k-subgraphs is NP-hard as it reduces to the maximal clique problem in polynomial time [15, 23]. Additionally, not all densest k-subsets are meaningful. For instance, in Example 3, all subsets in $\mathcal{D}_k$ for $b < k < 2b$ must split either C_1 or C_2.

Beyond densest k-subgraphs. We will show an example where dense subsets are not necessarily densest k-subgraphs.

Example 4 For integer $b \geq 4$, construct a graph on $V = C_1 \cup C_2$ by connecting a b-clique on C_1 and a quasi b-clique on C_2 with an edge removed, as shown in Fig. 1b.

$$\rho(C_1) = \frac{b-1}{2},$$
$$\rho(C_2) = \frac{b-1}{2} - \frac{1}{2b} < \rho(C_1).$$

Hence $C_2 \notin \mathcal{D}_b$ even though it is expected to be identified as a dense subset. □

Density-friendly graph decomposition. The work [13] conducts the density-friendly graph decomposition by finding a nested chain of dense subgraphs and optimizing for $\alpha \geq 0$

$$\max_{B\subseteq V} |\mathcal{E}(B)| - \alpha|B|, \tag{33}$$

of which any solution is a densest k-subgraph. However, any finer dense connected graph will not be the solution. Hence the method cannott identify C_1 and C_2 in Fig. 1a.

A.2 Locally densest subgraphs [[14]]

Definition 4 ([14, Definition 3.1, 3.3]) A graph G on V is called *ρ-compact* if G is connected, and removing any subset S of nodes will result in the removal of at least $\rho \times |S|$ edges in G, where ρ is a nonnegative real number.

Furthermore, a subgraph G_1, on $V_1 \subseteq V$, of G is called a *locally densest subgraph* if G_1 is a maximal $\rho(V_1)$-compact subgraph in G, where $\rho(\cdot)$ is the density function in (1). □

A.3 Calculation of strength on Examples

Example 5 In Fig. 1a, the following shows $C_1, C_2 \in \mathcal{C}_{\mathcal{F}_1}$.

$$\begin{aligned}
\sigma_1(V) &= \min_{B\subsetneq V:|B|\geq 1} \frac{\mu(V|B)}{\nu(V|B)} \\
&= \frac{\mu(V|C_1)}{\nu(V|C_1)} \\
&= \frac{\frac{1}{2}b(b-1)+1}{b} = \frac{b}{2} - \frac{1}{2} + \frac{1}{b}. \\
\sigma_1(C_1) &= \min_{B\subsetneq C_1:|B|\geq 1} \frac{\mu(C_1|B)}{\nu(C_1|B)} \\
&= \frac{\mu(C_1|\{i\})}{\nu(C_1|\{i\})}, \quad i \in C_1 \\
&= \frac{\frac{1}{2}b(b-1)}{b-1} = \frac{b}{2} > \sigma_1(V).
\end{aligned}$$

By symmetry, $\sigma_1(C_2) = \sigma_1(C_1)$. We omitted computation for other subsets of V. □

Example 6 In Fig. 1b, the following shows $C_1, C_2 \in \mathcal{C}_{\mathcal{F}_1}$.

$$\begin{aligned}
\sigma_1(V) &= \min_{B\subsetneq V:|B|\geq 1} \frac{\mu(V|B)}{\nu(V|B)} \\
&= \frac{\mu(V|C_1)}{\nu(V|C_1)} \\
&= \frac{\frac{1}{2}b(b-1)}{b} = \frac{b}{2} - \frac{1}{2}. \\
\sigma_1(C_2) &= \min_{B\subsetneq C_2:|B|\geq 1} \frac{\mu(C_2|B)}{\nu(C_2|B)} \\
&= \frac{\mu(C_2|\{i\})}{\nu(C_2|\{i\})}, \quad i \in C_2 \\
&= \frac{\frac{1}{2}b(b-1)-1}{b-1} = \frac{b}{2} - \frac{1}{b-1} > \sigma_1(V).
\end{aligned}$$

We omitted computation for other subsets of V. □

A.4 Density in info-net

For undirected info-nets, besides the density in (5), we also have the normalized density.

Normalized Density. The density may be normalized to

$$\rho(B) = \frac{H(\mathsf{Z}_B|\mathsf{Z}_{V\setminus B})}{\underbrace{H(\mathsf{Z}_B)}_{\nu(B)=}}, \tag{34}$$

which measures the amount of information internal to B over the amount of information available from B.

In the co-authorship network modelled by undirected info-nets described above, the normalized density of a subset $B \subseteq V : |B| \geq 1$ given by (34) measures the percentage of publications with all the authors in B with respect to all the publications that have an author in B. A larger normalized density of B given by (34) means the scholars in B need less collaboration with scholars outside B contribute as co-authors on publications, or in another word, B is more self-sufficient.

Example 7 With $C_1 = \{1,2,3\}$ as in Example 2,

$$\begin{aligned} H(\mathsf{Z}_{C_1}|\mathsf{Z}_{V\setminus C_1}) &= H(\mathsf{X}_a,\mathsf{X}_b,\mathsf{X}_c) &&= \underline{3 \text{ bits}} \\ H(\mathsf{Z}_{C_1}) &= H(\mathsf{X}_a,\mathsf{X}_b,\mathsf{X}_c,\mathsf{X}_d,\mathsf{X}_f) &&= \underline{5 \text{ bits}} \\ I(\mathsf{Z}_{C_1} \wedge \mathsf{Z}_{V\setminus C_1}) &= H(\mathsf{X}_d,\mathsf{X}_f) &&= \underline{2 \text{ bits}} \\ \rho(C_1) &= \frac{H(\mathsf{Z}_{C_1}|\mathsf{Z}_{V\setminus C_1})}{H(\mathsf{Z}_{C_1})} &&= \underline{\frac{3}{5}} \end{aligned}$$

With $C_3 := \{4,5\}$,

$$\begin{aligned} H(\mathsf{Z}_{C_3}|\mathsf{Z}_{V\setminus C_3}) &= H(\mathsf{X}_e) &&= \underline{1 \text{ bits}} \\ H(\mathsf{Z}_{C_3}) &= H(\mathsf{X}_d,\mathsf{X}_e,\mathsf{X}_f) &&= \underline{3 \text{ bits}} \\ I(\mathsf{Z}_{C_3} \wedge \mathsf{Z}_{V\setminus C_3}) &= H(\mathsf{X}_d,\mathsf{X}_f) &&= \underline{2 \text{ bits}} \\ \rho(C_3) &= \frac{H(\mathsf{Z}_{C_3}|\mathsf{Z}_{V\setminus C_3})}{H(\mathsf{Z}_{C_3})} &&= \underline{\frac{1}{3}} \quad < \rho(C_1) \end{aligned}$$

C_1 has a larger normalized density than C_3, which means scholars in C_1 need less collaboration from scholars outside to contribute as co-authors. □

To avoid zero in the denominator, we can impose a mild condition

$$\underbrace{H(\mathsf{Z}_i)}_{\nu(\{i\})=} > 0 \qquad \forall i \in V.$$

The condition $H(\mathsf{Z}_i) > 0$ for all $i \in V$ can be enforced simply by removing all trivial nodes i from V with no information, i.e., $H(\mathsf{Z}_i) = 0$.

A.5 Assumptions in Definition 1

The assumption

$$\text{for any } i \in V, \nu(\{i\}) > 0 \tag{35}$$

is for avoiding singularity.

As ν is non-decreasing, (35) ensures that density $\rho(B)$ is well-defined, namely, $\nu(B) > 0$, for all non-empty subsets B. Extension to singular subsets without imposing (35) is trivial. For a (V,μ,ν) network that does not satisfy (35), there is a unique maximum 0-volume subset

$$B_0 := \{i | i \in V, \nu(i) = 0\}$$

so that the network (V',μ',ν') defined below by contracting B_0 satisfies (35):

$$V' := V \setminus B_0 \tag{36}$$

$$\mu'(B) := \mu(B|B_0) \tag{37}$$

$$\nu'(B) := \nu(B|B_0) = \nu(B \setminus B_0) \tag{38}$$

for all $B \subseteq V'$.

It suffices to study the contracted network (V',μ',ν'), because if $\mu(B_0) = 0$, then

$$\rho(B) = \rho(B \setminus B_0), \qquad \forall B \subseteq V : |B| \geq 1;$$

if $\mu(B_0) > 0$, then

$$B \supseteq B_0 \qquad \forall B \in \mathcal{D}_k, k > |B_0|;$$

and for the remaining cases where $k \leq |B_0|$, $\mathcal{D}_k$ in (32) is ill-posed.

A.6 Reduction of (8) to other densities

The definition of density in (8) encompasses both the densities in (5) and (34) for info-net, as well as those in graphs.

By defining

$$\mu(B) := H(\mathsf{Z}_B|\mathsf{Z}_{V\setminus B}), \tag{39}$$

$$\nu(B) := H(\mathsf{Z}_B), \tag{40}$$

the density in (8) reduces to (34).

In the unweighted graph, by defining $\mu(B) := |\mathcal{E}(B)|$ and $\nu(B) := |B|$, the density in (8) reduces to (1).

In the weighted graph, by letting $\nu(B) := |B|$, and $\mu(B)$ be the total edge weight in B, then the density in (8) gives the density of B.

Appendix B
Proofs

B.1 Proof of Proposition 3

Proof Let

$$\begin{aligned} & \sigma_k(V) < x && \text{(41a)} \\ \iff & \min_{B\in\mathcal{F}_k:\nu(V|B)>0} \frac{\mu(V|B)}{\nu(V|B)} - x < 0 && \text{(41b)} \\ \iff & \min_{B\in\mathcal{F}_k:\nu(V|B)>0} \mu(V|B) - x\nu(V|B) < 0 && \text{(41c)} \\ \iff & \mu(V) - x\nu(V) < \max_{\substack{B\in\mathcal{F}_k:\\ \nu(V|B)>0}} \mu(B) - x\nu(B) && \text{(41d)} \\ \iff & f[V](x) < \max_{B\in\mathcal{F}_k:\nu(V|B)>0} f[B](x). && \text{(41e)} \end{aligned}$$

In (41), substituting the "<" by "=" or ">", it still holds. E.g., we also have

$$\sigma_k(V) = x \iff f[V](x) = \max_{B\in\mathcal{F}_k:\nu(V|B)>0} f[B](x). \tag{42}$$

To prove (24a), it suffices to show that (41e) holds if and only if

$$f[V](x) < \max_{B\in\mathcal{F}_k:B\neq V} f[B](x). \tag{43}$$

- (41e) implies (43) is because
$$\max_{B\in\mathcal{F}_k:\nu(V|B)>0} f[B](x) \leq \max_{B\in\mathcal{F}_k:B\neq V} f[B](x)$$
due to the fact
$$\{B \in \mathcal{F}_k | B \neq V\} \supseteq \{B \in \mathcal{F}_k | \nu(V|B) > 0\}.$$
- Next we show (43) implies (41e). Let B' maximizes the right hand side of (43), then $B' \subsetneq V$ and
$$f[V](x) < f[B'](x). \tag{44}$$
For any $B'' \in \mathcal{F}_k : (B'' \neq V, \nu(V|B'') = 0)$, by the fact that μ is non-decreasing, we have $f[B''](x) \leq f[V](x)$. Hence $f[B''](x) < f[B](x)$. As a result, B' is an optimal solution of the right-hand side of (41e). Then by (44), (41e) holds. ■

To prove (24b), it is sufficient to show

$$\max_{B\in\mathcal{F}_k:\nu(V|B)>0} f[B](x) \le \max_{B\in\mathcal{F}_k:B\neq V} f[B](x) \tag{45}$$

according to (42). By the fact that μ and ν are non-decreasing, for any $B' \in \mathcal{F}_k : (B' \neq V, v(V|B') = 0)$, we have $f[V](x) \le f[B'](x)$, which implies (45). The statement (24c) follows from the contrapositive of the conjunction of (24a) and (24b).

To show $\sigma_k(V)$ is the x-coordinate of the first turning point of $f_k(x)$ when $\sigma_k(V) < \infty$, notice that in this case there exists $x' \in \mathbb{R}$, s.t. $\sigma_k(V) < x'$. Then by (41e), $\exists B \in \mathcal{F}_k : \nu(V|B) > 0$ s.t. $f[V](x') < f[B](x')$, which means the lines $f[V](x)$ and $f[B](x)$ intersect, therefore $f_k(x)$ has turning points. Let x'' be the x-coordinate of the first turning point of $f_k(x)$. Considering that $f_k(x)$ strictly decreases against x since ν is non-decreasing and positive for nonempty set, there exists $B'' \in \mathcal{F}_k$ s.t. $\nu(V|B'') > 0$ and $f[V](x'') = f[B''](x'') = \max_{B\in\mathcal{F}_k:B\neq V} f[B](x'')$. Then by (42), we have $\sigma_k(V) = x''$, i.e., $\sigma_k(V)$ is the x-coordinate of the first turning point of $f_k(x)$.

Finally, we show that the converses of (24b) and (24c) hold when at least one of the conditions in (25) is fulfilled by V. To show the converse of (24b) holds, it suffices to show that the right-hand side of (24b) implies

$$f[V](x) = \max_{B\in\mathcal{F}_k:\nu(V|B)>0} f[B](x) \tag{46}$$

by (42). In fact, for any $B' \in \mathcal{F}_k : v(V|B') = 0$, we have $\mu(V|B') > 0$ provided that at least one of the conditions in (25) is fulfilled. Then for $x \in \mathbb{R}$, $f[B'](x) = \mu(B') - \nu(B')x \le \mu(V) - \nu(V)x = f[V](x)$, with which we can conclude that the right-hand side of (24b) implies (46). Therefore the converse of (24b) holds. The converse of (24c) can be derived from the contrapositive of the conjunction of (24a) and the converse of (24b).

B.2 Proof of Theorem 1

PROOF For any $C \in \mathcal{F}_k : C \neq \emptyset$, when $k \ge 1$,

$$\sigma_k(C) > \rho(C) \tag{47a}$$

$$\iff \frac{\mu(C)-\mu(B)}{\nu(C)-\nu(B)} > \frac{\mu(C)}{\nu(C)}, \forall B \in \mathcal{F}_k(C) : \nu(C|B) > 0 \tag{47b}$$

$$\iff \frac{\mu(C)}{\nu(C)} > \frac{\mu(B)}{\nu(B)}, \forall B \in \mathcal{F}_k(C) : \nu(C|B) > 0 \tag{47c}$$

$$\iff \rho(C) > \rho(B), \forall B \in \mathcal{F}_k(C) : \nu(C|B) > 0 \tag{47d}$$

$$\iff \rho(C) > \rho(B), \forall B \subsetneq C : |B| \ge k, \tag{47e}$$

where (47c) is by the simple formula that for $a, b, c, d > 0$ with $a > c, b > d$, we have $\frac{a-c}{b-d} > \frac{a}{b} \iff \frac{a}{b} > \frac{c}{d}$; and (47e) is because for any $B' \subsetneq C : |B'| \ge k$ with $\nu(C|B') = 0$, by assumption in Definition 1, $\mu(C|B') > 0$ which implies $\rho(C) > \rho(B)$.

When $k = 0$, $\emptyset \in \mathcal{F}_0(C)$, hence$\sigma_0(C) \le \frac{\mu(C|\emptyset)}{\nu(C|\emptyset)} = \rho(C)$ for $C \neq \emptyset$. ■

B.3 Result on $\mathcal{C}_{\mathcal{F}_k}(x)$ for $x < 0$

Proposition 7 *For $k \le |V| - 1$ and $x < 0$, $\mathcal{C}_{\mathcal{F}_k} = \{V\}$.* □

PROOF For $k \le |V| - 1$, $\sigma_k(V) \ge 0$ by (11). Then for $x < 0$, by (15), and the fact that any subset is a subset of V, we know that V is the unique maximal strong subset with strength w.r.t. σ_k larger than x. ■

B.4 Relation between $\mathcal{C}_{\mathcal{F}_0}(x)$ and $\mathcal{S}_{\mathcal{F}_0}(x)$

Proposition 8 *For $x \in \mathbb{R}$,*

$$|\tilde{\mathcal{S}}_{\mathcal{F}_0}(x)| = 1, \text{ and} \tag{48}$$

$$\mathcal{C}_{\mathcal{F}_0}(x) = \tilde{\mathcal{S}}_{\mathcal{F}_0}(x). \tag{49}$$

□

PROOF When $x < 0$, by Proposition 7, (48) and (49) hold.

When $x \ge 0$, by definition of $\mathcal{S}_{\mathcal{F}_0}(x)$, it is the set of the minimal optimal solution to a submodular function minimization over a lattice, where the minimal solution is unique [20, Proposition 10.1], i.e., $|\mathcal{S}_{\mathcal{F}_0}(x)| = 1$, thereafter (48) holds.

Then by Proposition 5,

$$|\mathcal{C}_{\mathcal{F}_0}(x)| \ge 1. \tag{50}$$

Suppose the inequality in (50) is strict, i.e., $|\mathcal{C}_{\mathcal{F}_0}| \ge 2$. Let $C_1, C_2 \in \mathcal{C}_{\mathcal{F}_0} : C_1 \neq C_2$, and assume $|C_1| \le |C_2|$. Then

$$C_2 \neq \emptyset, \text{ and} \tag{51}$$

$$\sigma_0(C_1), \sigma_0(C_2) > x. \tag{52}$$

We deduce that $C_1 \neq \emptyset$, since $\emptyset \subsetneq C_2$ cannot be maximal in $\mathcal{C}_{\mathcal{F}_0}$.

For any $B \subsetneq C_1 \cup C_2$,

$$\begin{aligned} f[B](x) &\overset{(a)}{\le} f[B \cup C_2](x) \underbrace{- f[C_2](x) + f[B \cap C_2](x)}_{\overset{(b)}{\le} 0} \\ &\overset{(c)}{\le} f[\underbrace{(B \cup C_2) \cup C_1}_{\overset{(d)}{=} C_1 \cup C_2}](x) \underbrace{- f[C_1](x) + f[(B \cup C_2) \cap C_1](x)}_{\overset{(e)}{\le} 0}, \end{aligned}$$

where (a) and (c) is by supermodularity of $f[B](x)$; (b) is by (52) and Proposition 3, and equality holds when $B \subseteq C_2$; (d) is because $B \subseteq C_1 \cup C_2$; (e) is by (24) since $(B \cup C_2) \cap C_1 \subseteq C_1$, and equality holds when $B \supseteq (C_1 \setminus C_2)$.

By assumption, $C_1 \neq C_2$, then the equalities in (b) and (e) cannot hold at the same time. Hence $f[B](x) < f[C_1 \cup C_2](x)$. Then By (3), $\sigma_1(C_1 \cup C_2) > x$.

This implies that C_1 is not maximal with strength w.r.t. σ_0 larger than x, hence C_1 should not be in $\mathcal{C}_{\mathcal{F}_0}(x)$, which is a contradiction.

Hence, only the equality in (50) holds. Then by Proposition 5, (49) holds. ■

B.5 Proof of Proposition 1

PROOF When $x < 0$, By Proposition 7, $\mathcal{C}_{\mathcal{F}_0}(x) = \{V\}$.

Let ρ^* be the maximum density of subsets of V and B^* be the maximal subset with density ρ^*.

When $x \ge \rho^*$, since $f[\emptyset](x) = 0$, $f[B](x) \le f[B](\rho^*) = 0$ for any $B \neq \emptyset$, we have $\emptyset \in \mathcal{S}_{\mathcal{F}_0}(x)$, hence $\mathcal{C}_{\mathcal{F}_0}(x) = \{\emptyset\}$

Let $x' = \lim_{\epsilon \to 0} \rho^* - |\epsilon|$, we deduce that $\mathcal{C}_{\mathcal{F}_0}(x') = \{B^*\}$. This is because B^* is the maximal solution to (20) at $x = \rho^*$, hence B^* is the minimal solution [20] to (20) at $x = x'$.

To show (17), it is sufficient to show that when $0 \le x < \rho^*$, for any $C_1 \in \mathcal{C}_{\mathcal{F}_0}(x_1), C_2 \in \mathcal{C}_{\mathcal{F}_0}(x_2)$ with $x_1 < x_2$,

$$C_1 \supseteq C_2. \tag{53}$$

Suppose to the contrary, $C_1 \not\supseteq C_2$. By $\sigma_0(C_1) > x_2 > x_1$, $\exists C_3 \supseteq C_2$ s.t. $C_3 \in \mathcal{C}_{\mathcal{F}_0}(x_1)$. Considering that $C_1 \in \mathcal{C}_{\mathcal{F}_0}(x_1)$, $|\mathcal{C}_{\mathcal{F}_0}(x_1)| \ge 2$, contradicting with (53). ∎

By Proposition 8, $\mathcal{C}_{\mathcal{F}_0}$ can be computed by computing $\mathcal{S}_{\mathcal{F}_0}$ for all $x \in \mathbb{R}$, i.e.,

$$\mathcal{C}_{\mathcal{F}_0} = \bigcup_{x\in\mathbb{R}} \arg\max_{B\in\mathcal{F}_0} \mu(B) - x\nu(B). \tag{54}$$

In fact, when $\nu(B)$ is a modular function, it is a well-known result in parametric submodular function minimization [20, Proposition 8.4] that (17) holds for the solutions.

When $\mu(B) := |\mathcal{E}(B)|$ and $\nu(B) = |B|$, (54) reduces to the objective function in [13].

B.6 Proof of Theorem 2

PROOF To show (18), it is sufficient to prove that for any $C \in \mathcal{C}_{\mathcal{F}_1}$, if $\exists C' \in \mathcal{C}_{\mathcal{F}_1}$ such that

$$C' \supsetneq C, \tag{55}$$

$$\sigma_1(C') > \rho(C'), \tag{56}$$

then for any $x \in \mathbb{R}$,

$$C \notin \mathcal{S}_{\mathcal{F}_0}(x), \tag{57}$$

since by Proposition 8, (57) means $C \notin \mathcal{C}_{\mathcal{F}_0}$.

(56) implies $\sigma_1(C') > \frac{\mu(C')}{\nu(C')}$, i.e.,

$$f[C'](\sigma_1(C')) < 0. \tag{58}$$

For $x \ge \sigma_1(C')$,

$$f[C](x) \overset{(a)}{\le} f[C](\sigma_1(C')) \overset{(b)}{<} 0 = f[\emptyset](x), \tag{59}$$

where (a) is because ν is non-negative non-decreasing; (b) is by (58). Then (57) holds in this case.

For $x < \sigma_1(C')$, by (24),

$$f[C'](x) > f[C](x), \tag{60}$$

which implies (57) holds in this case.

Hence, (57) holds for $x \in \mathbb{R}$, which implies(18).

Then we show

$$\mathcal{C}_{\mathcal{F}_0} \setminus \mathcal{C}_{\mathcal{F}_1} = \{\emptyset\} \tag{61}$$

in the following.

On one hand, $\emptyset \in \mathcal{C}_{\mathcal{F}_0}$ since $\emptyset \in \mathcal{C}_{\mathcal{F}_0}(\infty)$, and $\emptyset \notin \mathcal{C}_{\mathcal{F}_1}$ since $\emptyset \notin \mathcal{F}_1$.

On the other hand, for $x \in \mathbb{R}$ and $A \in \mathcal{C}_{\mathcal{F}_1}(x) : A \neq \emptyset$, by Proposition 8, $A \in \mathcal{S}_0(x)$. Since $A \neq \emptyset$, then $A \in \mathcal{S}_1(x)$ by definition of $\mathcal{S}_{\mathcal{F}_k}$.

Hence (61) holds. ∎

B.7 Explanation and proof for Lemma 1

Similar with that info-clustering[24], if we define, for $i, j \in V$, the similarity relation $\sim_x$ given $x \in \mathbb{R}$ as

$$i \sim_x j \iff \exists C \subseteq V, \{i,j\} \subseteq C, \text{ and } \sigma_1(C) > x, \tag{62}$$

it can be shown that the above lemma implies $\sim_x$ is an equivalence relation and $\mathcal{C}_x(V)$ is the set of equivalent classes. The desired hierarchical structure follows from the monotonicity of $\sim_x$ that

$$i \sim_x j \Rightarrow i \sim_{x'} j, \forall x' \le x, \tag{63}$$

which follows immediately from definition (62), and means that the equivalent classes can only merge into bigger equivalent classes as x decreases.

Indeed reflexivity ($i \sim_x i$) and symmetry ($i \sim_x j \iff j \sim_x i$) required for an equivalence relation also follows directly from definition (62). The lemma is used to show transitivity

$$i \sim_x j \text{ and} j \sim_x k \Rightarrow i \sim_x k. \tag{64}$$

More precisely, $i \sim_x j$ and $j \sim_x k$ implies there exist $C_1 \supseteq \{i,j\}, C_2 \supseteq \{j,k\}$ with

$$\sigma_1(C_1) \ge x \text{ and } \sigma_1(C_2) \ge x, \text{ and so} \tag{65}$$

$$\sigma_1(C_1 \cup C_2) \ge \min\{\sigma_1(C_1), \sigma_1(C_2)\} \ge x, \tag{66}$$

by Lemma 1 since $C_1 \cap C_2 \supseteq \{k\} \neq \emptyset$, which implies $i \sim_x k$ as desired.

PROOF Since strength is always non-negative, assume

$$\sigma_1(C_1) \ge \sigma_1(C_2) = x \ge 0 \tag{67}$$

It suffices to show that $\sigma_1(C_1 \cup C_2) \ge x$, or equivalently, by (3),

$$f[B](x) \le f[C_1 \cup C_2](x), \forall B \subseteq C_1 \cup C_2 : B \neq \emptyset. \tag{68}$$

Consider the case $B \subseteq C_1$ first,

$$\begin{aligned} f[B](x) &\overset{(a)}{\le} f[C_1](x) \\ &\overset{(b)}{\le} f[C_1 \cup C_2](x) \underbrace{-f[C_1](x) + f[C_1 \cap C_2](x)}_{\overset{(c)}{\le} 0} \end{aligned}$$

which implies (68) as desired, where

- (a) and (c) are by (67) and the contrapositive of (24a);
- (b) is because $f[B](x)$ is supermodular.

Then consider the remaining case $B \not\subseteq C_1$, i.e., $B \subseteq C_1 \cup C_2$ and $B \cap C_2 \neq \emptyset$,

$$\begin{aligned} f[B](x) &\overset{(d)}{\le} f[B \cup C_2](x) \underbrace{-f[C_2](x) + f[B \cap C_2](x)}_{\overset{(e)}{\le} 0} \\ &\overset{(f)}{\le} f[\underbrace{(B \cup C_2) \cup C_1)}_{\overset{(g)}{=} C_1 \cup C_2}](x) \underbrace{-f[C_1](x) + f[(B \cup C_2) \cap C_1](x)}_{\overset{(h)}{\le} 0}, \end{aligned}$$

which implies (68) as desired, where

- (d) and (f) are by the supmodularity of $f[B](x)$;
- (c) and (h) are by (67) and the contrapositive of (24a). As for (h), we utilize the fact that $(B \cup C_2) \cap C_1 \subseteq C_1$, $(B \cup C_2) \cap C_1 \supseteq C_1 \cap C_2$ and $C_1 \cap C_2 \neq \emptyset$;
- (g) is because $B \subseteq C_1 \cup C_2$. ∎

B.8 Proof of Theorem 3

PROOF Suppose to the contrary, $C_1 \not\supseteq C_2$ and $C_1 \cap C_2 = \emptyset$. By Lemma 1,

$$\sigma_1(C_1 \cup C_2) \ge \sigma_1(C_1) > x_i. \tag{69}$$

This contradicts with $C_1 \in \mathcal{C}_{\mathcal{F}_1}(x_1)$, in particular, the inclusion-wise maximality of C_1, since $C_1 \cup C_2 \supsetneq C_1$, where the strict inclusion is because $C_1 \not\supseteq C_2$.

When $C_1 \supsetneq C_2$, suppose to the contrary, $x_1 \geq x_2$. However, $\sigma_1(C_1) > x_1 \geq x_2$, which contradicts with the maximality of C_2. Hence in this case $x_1 < x_2$. ■

B.9 Proof of Proposition 4

PROOF When $\sigma_k(V) \leq x$, by (24a) and (24b),

$$f[V](x) \leq \max_{B\in\mathcal{F}_k:B\neq V} f[B](x), \tag{70}$$

which means V is not the minimal solution to (20), hence (26) is established. (27) follows from (26).

When V satisfy the condition in (25), the converse of (24c) holds, which means V is the minimal solution to (20), hence (26) and (27) does not hold. ■

B.10 Proof of Proposition 5

PROOF It is sufficient to show that for any $B \in \mathcal{S}_{\mathcal{F}_k}(x)$,

$$B \in \mathcal{C}_{\mathcal{F}_k}(x). \tag{71}$$

When V satisfies the condition in (25), the converses of (24b) and (24c) hold. On one hand, by minimality of B in $\mathcal{S}_{\mathcal{F}_k}(x)$, we have

$$f[B](x) > \max_{B'\in\mathcal{F}_k(B):B'\neq B} f[B'](x) \tag{72}$$

which implies $\sigma_k(B) > x$ by the converse of (24c). On the other hand, by the fact that B maximizes (20),

$$f[B](x) \geq \max_{A\in\mathcal{F}_k} f[A](x), \tag{73}$$

which means for any $A \in \mathcal{F}_k : A \supsetneq B$,

$$f[A](x) \leq \max_{A'\in\mathcal{F}_k(A):A'\neq A} f[A'](x), \tag{74}$$

followed by $\sigma_k(A) \leq x$ according to by conjunction of (24a) and the converses of (24b). Hence B is a maximal subset with $\sigma_k(B) > x$, which implies (71) according to (15). ■

B.11 Proof of Proposition 6

PROOF When V satisfies the condition in (25). In this case, or $B \in \mathcal{S}_{\mathcal{F}_k}(x)$, any $B' \supsetneq B$ has to satisfy $\nu(B'|B) > 0$; otherwise $\mu(B'|B) > 0$, then $f[B' \cup B](x) > f[B](x)$ which contradicts with the fact that B is in the solution set to (20). Therefore $\tilde{\mathcal{S}}_{\mathcal{F}_k}(x) = \mathcal{S}_{\mathcal{F}_k}(x)$ according to (28).

Let $B \in \mathcal{S}_{\mathcal{F}_k}(x)$ and $A := B \cup \{i \in V | \nu(\{i\}|B) = 0\}$. Based on Proposition 5, to prove (28), it remains to show that when $A \supsetneq B$, A is the maximal subset with strength larger than x.

First, we show that $\nu(A|B) = 0$. Let $(i_1, i_2, \dots, i_{|A\setminus B|})$ any ordering of nodes in $A \setminus B$, and $A_1 = A \setminus \{i_1\}, A_j = A_{j-1} \setminus \{i_j\}.j = 2, 3, \dots, |A \setminus B|$. By submodularity of ν, $\nu(A) = \nu(\{i_1\}|A_1) + \nu(A_1) \leq \nu(\{i_1\}|B) + \nu(A_1) = \nu(A_1) = \nu(\{i_2\}|A_2) + \nu(A_2) \leq \dots = \nu(B)$. Since ν is non-decreasing, we have $\nu(A|B) = 0$.

Then for any $A' \in \mathcal{F}_k(A) : \nu(A|A') > 0$, we have

$$f[A'](x) < f[A](x). \tag{75}$$

Otherwise, it contradicts with the minimality of B as the optimal solution to (20).

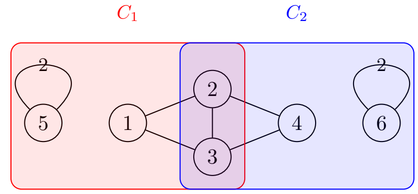


Fig. 3: A graph shows unknown structure of $\mathcal{C}_{\mathcal{F}_2}$ in Example 8.

Next, For any $A' \supsetneq A$, we have $f[A'](x) \leq f[A](x)$, otherwise it contradicts with the optimality of B to (20). Then by (29), $\nu(A'|A) > 0$, which means $A \in \mathcal{F}_k(A')$. Then according to (11),

$$\sigma_k(A') \leq x. \tag{76}$$

(75) and (76) imply that A is the maximal subset with strength larger than x, which establishes (28).

B.12 Proof of Theorem 4

PROOF (SKETCH) On one hand, Proposition 6 shows that the subsets in $\tilde{\mathcal{S}}_{\mathcal{F}_k}(x)$ are in $\mathcal{C}_{\mathcal{F}_k}(x)$. On the other hand, the strong subsets in $\mathcal{C}_{\mathcal{F}_k}(x)$ are non-overlapping due to the laminar structure of $\mathcal{C}_{\mathcal{F}_k}(x)$ as indicated by Theorem 3. Hence we have the recursive relation for computing $\tilde{\mathcal{S}}_{\mathcal{F}_k}(x)$. ■

APPENDIX C
AN EXAMPLE SHOWING UNKNOWN STRUCTURE OF $\mathcal{C}_{\mathcal{F}_2}$

Example 8 We show by the graph in Fig. 3 that $\mathcal{C}_{\mathcal{F}_2}$ do not have similar structure property as that of $\mathcal{C}_{\mathcal{F}_1}$ in Lemma 1.

Let $C_1 = \{1, 2, 3, 5\}$, $C_2 = \{2, 3, 4, 6\}$. We can compute

$$\sigma_2(C_1) = \min_{B\subsetneq C_1:|B|\geq 2} \frac{\mu(C_1|B)}{\nu(C_1|B)} \tag{77}$$

$$= \frac{\mu(C_1|\{5, i\})}{\nu(C_1|\{5, i\})}, i \in \{1, 2, 3\} \tag{78}$$

$$= \frac{3}{2}. \tag{79}$$

By symmetry,

$$\sigma_2(C_2) = \sigma_2(C_1) = \frac{3}{2}. \tag{80}$$

Then consider $C_1 \cup C_2$, we have

$$\sigma_2(C_1 \cup C_2) = \min_{B\subsetneq C_1\cup C_2:|B|\geq 2} \frac{\mu(C_1 \cup C_2|B)}{\nu(C_1 \cup C_2|B)} \tag{81}$$

$$= \frac{\mu(C_1 \cup C_2|\{5, 6\})}{\nu(C_1 \cup C_2|\{5, 6\})} \tag{82}$$

$$= \frac{5}{4}. \tag{83}$$

Although $|C_1 \cap C_2| \geq 2$, we have $\sigma_2(C_1 \cup C_2) < \min\{\sigma_2(C_1), \sigma_2(C_2)\}$, unlike that in Lemma 1 for $\mathcal{C}_{\mathcal{F}_1}$. □

REFERENCES

[1] C. Zhao, A. Al-Bashabsheh, and C. Chan, "Detecting Informationally-Dense Subsets," May 2024. [Online]. Available: https://github.com/ao-hk/Detecting-Informationally-Dense-Subsets

[2] F. Xia, K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, and H. Liu, "Graph learning: A survey," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 109–127, 2021.

[3] E.-M. Mohamed, T. Agouti, A. Tikniouine, and M. El Adnani, "A comprehensive literature review on community detection: Approaches and applications," *Procedia Computer Science*, vol. 151, pp. 295–302, 2019.

[4] A. Gionis and C. E. Tsourakakis, "Dense subgraph discovery: Kdd 2015 tutorial," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 2313–2314.

[5] J. Chen and Y. Saad, "Dense subgraph extraction with application to community detection," *IEEE Transactions on knowledge and data engineering*, vol. 24, no. 7, pp. 1216–1230, 2010.

[6] Y. Dourisboure, F. Geraci, and M. Pellegrini, "Extraction and classification of dense communities in the web," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 461–470.

[7] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[8] A. V. Goldberg, "Finding a maximum density subgraph," 1984.

[9] S. Khuller and B. Saha, "On finding dense subgraphs," in *International colloquium on automata, languages, and programming*. Springer, 2009, pp. 597–608.

[10] T. Lanciano, A. Miyauchi, A. Fazzone, and F. Bonchi, "A survey on the densest subgraph problem and its variants," *arXiv preprint arXiv:2303.14467*, 2023.

[11] U. Feige, D. Peleg, and G. Kortsarz, "The dense k-subgraph problem," *Algorithmica*, vol. 29, no. 3, pp. 410–421, 2001.

[12] R. Andersen and K. Chellapilla, "Finding dense subgraphs with size bounds," in *International workshop on algorithms and models for the web-graph*. Springer, 2009, pp. 25–37.

[13] N. Tatti, "Density-friendly graph decomposition," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 13, no. 5, pp. 1–29, 2019.

[14] L. Qin, R.-H. Li, L. Chang, and C. Zhang, "Locally densest subgraph discovery," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 965–974.

[15] U. Feige, M. Seltser *et al.*, *On the densest k-subgraph problem*. Citeseer, 1997.

[16] S. Khot, "Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique," *SIAM Journal on Computing*, vol. 36, no. 4, pp. 1025–1071, 2006.

[17] K. Nagano, Y. Kawahara, and K. Aihara, "Size-constrained submodular minimization through minimum norm base," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 977–984.

[18] D. Zhang, J. Yin, X. Zhu, and C. Zhang, "Network representation learning: A survey," *IEEE transactions on Big Data*, vol. 6, no. 1, pp. 3–28, 2018.

[19] F. Chen, Y.-C. Wang, B. Wang, and C.-C. J. Kuo, "Graph representation learning: a survey," *APSIPA Transactions on Signal and Information Processing*, vol. 9, p. e15, 2020.

[20] F. Bach *et al.*, "Learning with submodular functions: A convex optimization perspective," *Foundations and Trends® in machine learning*, vol. 6, no. 2-3, pp. 145–373, 2013.

[21] C. Chan, A. Al-Bashabsheh, C. Zhao *et al.*, "Finding better web communities in digraphs via max-flow min-cut," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 410–414.

[22] G. Gallo, M. D. Grigoriadis, and R. E. Tarjan, "A fast parametric maximum flow algorithm and applications," *SIAM Journal on Computing*, vol. 18, no. 1, pp. 30–55, 1989.

[23] M. R. Garey and D. S. Johnson, *Computers and intractability*. freeman San Francisco, 1979, vol. 174.

[24] C. Chan, A. Al-Bashabsheh, Q. Zhou, T. Kaced, and T. Liu, "Info-clustering: A mathematical theory for data clustering," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 1, pp. 64–91, 2016.