

# A Community Detection Method based on Formal Concept Lattice

Tiantian Wang<sup>1,\*</sup>, Zhenhao Qi<sup>2</sup>, Zhicheng Ding<sup>3</sup>

<sup>1</sup> School of Computer and Software Engineering, Sias University, Zhengzhou, Henan, 451150, China

<sup>2</sup> School of Computer and Information, Hefei University of Technology, Hefei, Anhui, 230601, China

<sup>3</sup> School of Mechanical and Electrical Engineering, Zhengzhou Industrial Application Technology College, Zhengzhou, Henan, 451150, China

\* Corresponding author: Tiantian Wang

**Abstract:** Community detection aims to uncover groups of vertices in a network that are more closely connected to each other than to the rest of the graph. While many existing algorithms rely purely on structural information, additional insights can be gained by incorporating conceptual relationships among nodes. This paper presents a community detection framework based on classical concept lattices. The method constructs a lattice from the given context, employs a greedy covering strategy to identify representative communities, and applies a merging mechanism to refine the results. Experiments on synthetic and real-world networks show that the proposed approach produces stable and interpretable communities, demonstrating the potential of concept lattice theory as a tool for network analysis.

**Keywords:** Community Detection; Concept Lattice; Greedy Covering; Network Analysis.

## 1. Introduction

Community detection, also referred to as graph clustering, aims to identify groups of vertices that exhibit higher internal connectivity compared to their connections with the rest of the network. This process reveals the meso-scale structure commonly found in social, biological, and technological systems. The Louvain algorithm offers near-linear scalability and is capable of handling graphs with millions of edges [1], while Infomap approaches community detection from an information-theoretic perspective by minimizing the description length of random walks [2]. Surveys in this area have outlined the evolution of the field, highlighting a transition from purely structural heuristics toward methods that incorporate node features and higher-order network patterns [3, 4].

Modern community detection algorithms generally fall into three categories. The first type consists of topology-based methods that rely entirely on network connectivity, such as modularity optimization and its refinements like the Leiden algorithm, which ensures well-connected communities even at large scales [5]. The second category includes feature-based methods, which utilize node attributes; for instance, the community-affiliation block model by Yang and Leskovec jointly learns both group membership and feature affinity [6]. The third category integrates structural and attribute information. These hybrid methods range from graph convolutional networks, which propagate features along network edges [7], to frameworks that jointly evaluate structural and semantic alignment with known community labels [8, 9]. However, even with these advancements, empirical evaluations on both synthetic benchmarks and real-world networks continue to expose several challenges. These include the resolution limit [10], sparsity of node features [11], and sensitivity to noise [12], all of which contribute to reduced detection accuracy.

Most existing community detection methods are designed for large-scale networks with well-defined structures and

abundant information. Their effectiveness, however, often decreases when the available data is noisy, incomplete, or derived from partial snapshots. To address these challenges, this study explores the use of classical formal concept analysis (FCA) [13, 14] as the foundation for a concept-guided community detection strategy.

FCA organizes object-attribute relationships into concept lattices, providing a structured way to represent knowledge and identify clusters of objects that share common features. This framework has been widely studied in data mining [15], knowledge representation [16, 17], and conflict analysis [18, 19]. For instance, Zhi et al. [20] introduce dual concepts and investigate their lattice-theoretic properties, thereby enriching the algebraic basis of FCA. These developments highlight the versatility of concept lattices in capturing meaningful patterns across diverse application domains.

The potential of FCA has also been demonstrated in areas such as conflict analysis and predictive modeling. Zhi et al. [21] examine conflicts in multi-level organizations, while Qi et al. [22] extend concept-based reasoning to distributed settings. In learning tasks, Wu et al. [23] establishes extent-based correlations between feature concepts and multi-label concepts, leading to a learning model that outperforms conventional methods on ten benchmark datasets. To address missing labels, Liu et al. develops a multi-level stochastic concept-clustering model, which hierarchically integrates similar features and selects appropriate granularity based on mutual information, yielding superior results across twelve datasets and six evaluation metrics.

Despite substantial progress from early topological heuristics to more advanced hybrid approaches, community detection still faces challenges when communities overlap, boundaries are ambiguous, or information is incomplete. To address these issues, we propose a lattice-based framework that combines the structural strengths of network topology with the interpretability of concept lattices. By regulating internal similarity and explicitly incorporating boundary information, the method offers a principled way to identify

communities that are both structurally coherent and semantically meaningful.

The main contribution of this work is the introduction of a concept lattice-based framework for community detection, where formal concept analysis is employed to capture network structure through object-attribute relationships. Building on this foundation, we design an algorithm that constructs concept lattices and derives communities using a greedy covering and merging strategy. Furthermore, we establish a theoretical link between concept lattice structures and community detection, ensuring both interpretability and consistency beyond conventional graph-based heuristics. In summary, this study connects formal concept analysis with community detection, offering a transparent and theoretically grounded alternative to existing methods.

The remainder of the paper is structured as follows. Section 2 introduces the basic concepts of formal concept analysis. Section 3 presents the proposed community detection framework. Section 4 concludes the paper and discusses directions for future research.

$$C = \{C_1, C_2, \dots, C_k\}, \quad \bigcup_{i=1}^k c_i = v, \quad C_i \cap C_j = \emptyset \text{ for } i \neq j,$$

where each  $C_i$  represents a detected community.

In an undirected graph, each edge  $(u, v) \in E$  indicates a symmetric relationship between nodes  $u$  and  $v$ , reflecting mutual interaction or similarity. By contrast, a directed graph uses ordered pairs to represent asymmetrical relationships. When edges carry numerical weights—such as strength, cost, or probability—the graph is considered weighted. Otherwise, it is unweighted, with all edges treated equally. In this study, all graphs are assumed to be unweighted and undirected.

The objective of a community detection algorithm is to assign similar nodes to the same community, resulting in higher intra-community edge density than inter-community edge density. To evaluate the quality of the resulting partitions, various metrics have been proposed. In this work, we use two

$$H(C) = - \sum_{i=1}^n \frac{|C_i|}{v} \log \left( \frac{|C_i|}{v} \right), \quad H(Y) = - \sum_{j=1}^s \frac{|Y_j|}{v} \log \left( \frac{|Y_j|}{v} \right)$$

and  $v$  is the total number of nodes. NMI ranges from 0 to 1, with higher values indicating stronger agreement between the predicted and true partitions. This metric is most applicable when reference labels are available.

Feature Consistency (FC), on the other hand, measures the internal homogeneity of features within each community. It is based on a formal context, defined as a tuple  $(P, M, R)$ , where  $P$  is a set of objects,  $M$  is a set of features, and  $R : P \times M \rightarrow \{0, 1\}$  is a binary relation indicating whether object  $x \in P$  possesses feature  $a \in M$ .

Definition 3. Let  $(P, M, R)$  be a formal context, and let  $C = \{C_1, C_2, \dots, C_n\}$  denote the detected communities, where each  $C_i$  is a subset of objects. The FC score for community  $C_i$  is defined as:

$$FC(C_i) = \frac{1}{|M|} \sum_{a \in M} \left| \frac{1}{|C_i|} \sum_{x \in C_i} R(x, a) \right|$$

This value reflects the average absolute presence frequency of each feature within the community. The overall FC across all communities is then computed as:

$$FC = \frac{1}{n} \sum_{i=1}^n FC(C_i).$$

A higher FC value indicates that the features within each community are more consistent, which is especially important

## 2. Preliminaries

This section introduces the foundational concepts of community detection and three-way concept analysis.

While community detection is typically formulated on graphs, this study adopts a binary object-feature matrix as the underlying data representation. For comparative purposes, we also describe the method used to convert a binary matrix into a graph structure in later sections.

### 2.1. Community Detection

Community detection is a core task in network science and data mining. It aims to identify groups of nodes, known as communities, that are more densely connected to each other than to nodes outside the group.

Definition 1. Let  $G = (V, E)$  be an undirected graph, where  $V = \{v_1, v_2, \dots, v_n\}$  is the set of nodes and  $E \subseteq V \times V$  is the set of edges. A community partition of  $G$  is a collection of disjoint subsets of nodes such that:

representative indicators: Normalized Mutual Information (NMI), a supervised metric that compares detected communities to ground truth labels, and Feature Consistency (FC), an unsupervised metric that evaluates the homogeneity of node features within each community.

$$NMI(C, Y) = (2I(C, Y)) / (H(C) + H(Y))$$

where  $I(C, Y)$  denotes the mutual information between the two partitions, and  $H(C)$ ,  $H(Y)$  are their respective entropies:

$$I(C, Y) = \sum_{i=1}^n \sum_{j=1}^s \frac{|C_i \cap Y_j|}{v} \log \left( \frac{v \cdot |C_i \cap Y_j|}{|C_i| \cdot |Y_j|} \right),$$

in applications where feature semantics are meaningful. While FC is useful for assessing feature-aware community detection, its effectiveness may be limited in the presence of sparse or noisy feature data.

### 2.2. Formal Concept Analysis

Formal concept analysis builds upon the mathematical foundation of formal contexts, with the concept lattice serving as its core structure.

Definition 4. [23] Let  $K = (P, M, R)$  be a formal context, where  $P$  is a set of objects,  $M$  is a set of attributes, and  $R \subseteq P \times M$  is a binary relation indicating whether object  $x \in P$  possesses attribute  $a \in M$ . For any object subset  $X \subseteq P$  and attribute subset  $A \subseteq M$ , the positive and negative Galois operators are defined as follows:

$$X^* = \{a \in M \mid R(x, a) = 1 \text{ for all } x \in X\}, \quad A^* = \{x \in P \mid R(x, a) = 1 \text{ for all } a \in A\},$$

Definition 5. Based on the above operators, two composite mappings are defined as follows:

$$X^\circ = A, \quad A^\circ = X.$$

Given a formal context  $K = (P, M, R)$ , a pair  $(X, A)$  is called a three-way concept if  $X^\triangleright = A$  and  $A^\triangleleft = X$ . In this case,  $X$  is the extent and  $A$  is the intent of the three-way concept.

The set of all concepts, ordered by the partial order “ $\leq$ ”, where:

$$(X_1, A_1) \leq (X_2, A_2) \Leftrightarrow X_1 \subseteq X_2, A_1 \supseteq A_2$$

forms a complete lattice. This structure is referred to as the three-way concept lattice, denoted as  $L(K)$ .

### 3. Proposed Community Detection Algorithms

This section presents the proposed method TWCLCD. Given a formal context  $K = (P, M, R)$ , the objective is to partition the object set  $P$  into disjoint communities  $C = \{C_1, \dots, C_n\}$  such that members within each community share highly consistent feature patterns, while avoiding overly large groups that may obscure meaningful structure.

#### 3.1. Algorithm Design

To quantitatively assess the internal consistency of a candidate object set, we define a similarity degree that measures the extent to which the objects exhibit shared feature patterns.

**Definition 6.** Let  $K = (P, M, R)$  be a formal context, and let  $(X, A) \in L(K)$  be a three-way concept. The similarity degree  $S(X)$  of the object set  $X$  is defined as:

$$S(X) = (|\alpha(X)| + |\beta(X)|) / |M|$$

where  $\alpha(X) = \{a \in M \mid R(x, a) = 1 \text{ for all } x \in X\}$  denotes the set of features shared by all objects in  $X$ , and  $\beta(X) = \{a \in M \mid R(x, a) = 0 \text{ for all } x \in X\}$  denotes the set of features universally absent. The similarity degree reflects the proportion of features on which all objects in  $X$  fully agree.

In the first stage, the algorithm gradually extracts candidate object sets from the formal concept lattice  $L(K)$  to serve as initial communities. A concept  $(X, A)$  is retained if it meets the criteria  $S(X) \geq \tau_s$  and  $|X| \leq \tau_{mx}$ , where  $\tau_s$  represents the minimum similarity threshold and  $\tau_{mx}$  specifies the maximum allowed community size. From the set of all admissible candidates, the one with the highest score is chosen based on the function  $\frac{1}{2} \cdot (S(X) - \tau_s) / (1 - \tau_s) + \frac{1}{2} \cdot |X| / \tau_{mx}$ , which assigns equal importance to similarity and community size.

In the second stage, communities whose size is smaller than the lower limit  $\tau_{mn}$  are considered for merging. For each such community  $C_i$ , its associated feature set is compared against those of the other communities using Jaccard similarity, calculated as  $J(A, B) = |A \cap B| / |A \cup B|$ , where  $A$  and  $B$  are the feature sets of two respective communities.

For each small community  $C_i$ , the algorithm searches for the community  $C_j^*$  with the greatest Jaccard similarity. If this similarity exceeds the merging threshold  $\tau_{merge}$  and the combined size does not surpass  $\tau_{mx}$ , the two communities are merged. The procedure is repeated until all communities satisfy the minimum size requirement  $\tau_{mn}$ .

The thresholds used in the algorithm are determined adaptively from the dataset itself. The similarity threshold  $\tau_s$

is initialized as the median similarity score across all concepts in the lattice, i.e.,  $\tau_s = \text{median}\{S(X) \mid (X, (A, B)) \in L(K)\}$ . If no community satisfies this condition, the threshold is progressively relaxed. Size constraints are obtained from the empirical distribution of community sizes:  $\tau_{mx}$  is chosen as the 80th percentile to prevent excessively large clusters, while  $\tau_{mn}$  is taken as the 20th percentile to identify undersized communities. The merging threshold  $\tau_{mg}$  is defined as the median Jaccard similarity between each small community and its closest larger counterpart; to avoid overly aggressive merges,  $\tau_{mg}$  is fixed to at least 0.2.

Overall, the framework operates in two adaptive phases. First, an initial set of communities is constructed using greedy covering under similarity and size constraints. Then, smaller residual communities are iteratively merged into larger ones based on feature similarity. Together, these phases yield a community structure that is both size-balanced and semantically consistent.

In summary, proposed method first constructs a high-similarity, size-controlled skeleton of communities and then refines the result by merging small residual communities based on feature similarity, as implemented in Algorithms 1 and 2.

### 4. Conclusion

This paper introduced the proposed method, a community detection framework that integrates formal concept lattices with an adaptive greedy covering strategy and a merging mechanism. Unlike conventional graph-based approaches, the method jointly exploits structural and feature information, using concept lattice structures to guide and complete the detection process.

The main contributions of this study can be summarized as follows:

- We develop a practical and interpretable lattice-based covering algorithm that establishes a clear link between formal concept analysis and community detection.
- The method provides a systematic procedure for constructing communities from concept lattices and refining them through similarity-based merging, ensuring both structural consistency and interpretability.
- The detected communities are derived from explicit object–attribute relationships in the lattice, allowing each decision to be traced and explained.

Despite its strengths, the proposed method has several limitations that merit further study. First, on datasets with clean structures and clear community boundaries, modularity-based algorithms such as Leiden may still perform better, since they are well-suited for purely topological scenarios. Second, the framework currently assumes binary input data, requiring discretization when continuous or fuzzy attributes are present, which may lead to loss of semantic information. Finally, although some components are adaptive, key parameters such as similarity thresholds and merging criteria still require manual tuning, which can affect performance across different datasets.

---

**Algorithm 1. Adaptive Greedy Cover**

---

Input: Binary context  $K = (P, M, R)$

Output: Preliminary community list  $C$

Step 1: Construct the formal concept lattice  $L(K)$ .

Step 2: For each concept  $(X, (A,B))$  in  $L(K)$ , compute its similarity score  $S(X)$ .

Step 3: Initialize the set of uncovered objects  $U$  with all objects in  $P$ .

Step 4: While  $U$  is not empty, repeat the following:

Step 4.1: Identify the set of candidate communities  $P'$  under the current threshold  $\tau_s$ .

Step 4.2: If  $P'$  is empty and  $\tau_s > 0$ , relax the threshold by updating  $\tau_s \leftarrow \tau_s / r$  and recompute  $P'$ .

Step 4.3: If  $P'$  is still empty, assign each remaining object in  $U$  as a singleton community, update  $C$  accordingly, and exit the loop.

Step 4.4: Otherwise, select the best candidate community based on the composite score

$$\text{Score} = \frac{1}{2} \cdot (S(X) - \tau_s) / (1 - \tau_s) + \frac{1}{2} \cdot (|X| / \tau_{\max}).$$

Step 4.5: Add the selected community to  $C$  and remove its objects from  $U$ .

Step 5: Return the final community list  $C$ .

---

---

**Algorithm 2. Merge Small Communities**

---

Input: Preliminary community set  $C$

Output: Refined community set  $C$

Step 1: Initialize a flag variable  $\text{changed} \leftarrow \text{true}$ .

Step 2: While  $\text{changed} = \text{true}$ , perform the following steps:

Step 2.1: Set  $\text{changed} \leftarrow \text{false}$ .

Step 2.2: For each community  $C_i$  in  $C$  with size smaller than  $\tau_{\min}$ :

(a) Derive the feature set  $A_i = \{a \in M \mid R(x,a) = 1 \text{ for some } x \in C_i\}$ .

(b) For each other community  $C_j$  in  $C$  (excluding  $C_i$ ):

i. If  $|C_i| + |C_j| \leq \tau_{\max}$ , compute its feature set  $A_j = \{a \in M \mid R(x,a) = 1 \text{ for some } x \in C_j\}$ .

ii. Calculate the Jaccard similarity  $s = |A_i \cap A_j| / |A_i \cup A_j|$ .

iii. Track the community  $C_{j^*}$  with the highest similarity value  $s^*$ .

(c) If  $s^* > \tau_{\text{merge}}$ , merge  $C_i$  into  $C_{j^*}$ , remove  $C_i$  from  $C$ , set  $\text{changed} \leftarrow \text{true}$ , and break out of the inner loop.

Step 3: Repeat Step 2 until no changes occur ( $\text{changed} = \text{false}$ ).

Step 4: Return the refined community set  $C$ .

---

To address these limitations and extend the scope of the proposed method, future work may consider the following directions:

- Extending the framework to handle fuzzy or set-valued contexts, enabling the integration of continuous and uncertain attributes without information loss.
- Leveraging data-driven or meta-learning techniques to automatically determine thresholds and parameters, improving robustness and reducing the need for manual adjustments.
- Enriching the lattice construction with multi-source attributes to better capture heterogeneous information and produce more semantically coherent communities.

By pursuing these directions, the proposed method can be further enhanced in terms of scalability, adaptability, and interpretability, contributing to broader applications of concept lattice-based reasoning in real-world network analysis.

## Acknowledgments

This work is supported by the Sias University Scientific Research Funding Project (grant No. 2024XKB021, No. 2024XKD040).

## References

- [1] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, *J. Sci. Commun.* 163 (2000) 51-59.
- [2] Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. *Proceedings of the national academy of sciences* 105(4), 1118–1123 (2008).
- [3] Fortunato, S.: Community detection in graphs. *Physics reports* 486(3-5), 75–174(2010).
- [4] Fortunato, S., Hric, D.: Community detection in networks: A user guide. *Physics reports* 659, 1–44 (2016).
- [5] Traag, V., Waltman, L., Van Eck, N.: From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9, 5233 (2019).
- [6] Yang, J., Leskovec, J.: Defining and evaluating network communities based on ground-truth. In: *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, pp. 1–8 (2012).
- [7] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv: 1609.02907* (2016).
- [8] Lancichinetti, A., Fortunato, S.: Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 80(1), 016118 (2009).
- [9] Peel, L., Larremore, D.B., Clauset, A.: The ground truth about metadata and community detection in networks. *Science advances* 3(5), 1602548 (2017).
- [10] Xie, J., Kelley, S., Szymanski, B.K.: Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys (csur)* 45(4), 1–35 (2013).
- [11] Zhang, X., Liu, H., Li, Q., Wu, X.M.: Attributed graph clustering via adaptive graph convolution. In: *28th International Joint Conference on Artificial Intelligence, IJCAI 2019*, pp. 4327–4333 (2019). *International Joint Conferences on Artificial Intelligence*.
- [12] Stacey, R.G., Skinnider, M.A., Foster, L.J.: On the robustness of graph-based clustering to random network alterations. *Molecular & Cellular Proteomics* 20 (2021).
- [13] Wille, R.: Restructuring lattice theory: An approach based on hierarchies of concepts. In: Rival, I. (ed.) *Ordered Sets*, pp. 445–470. Springer, Dordrecht (1982).
- [14] Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer (1997).
- [15] Ren, R., Wei, L.: The attribute reductions of three-way concept lattices. *Knowledge-Based Systems* 99, 92–102 (2016).
- [16] Zhang, H.-R., Min, F.: Three-way recommender systems based on random forests. *Knowledge-Based Systems* 91, 275–286 (2016).
- [17] Yu, H., Li, Q., Cai, M.: Characteristics of three-way concept lattices and three-way rough concept lattices. *Knowledge-Based Systems* 146, 181–189 (2018).
- [18] Lang, G., Luo, J., Yao, Y.: Three-way conflict analysis: a unification of models based on rough sets and formal concept analysis. *Knowledge-Based Systems* 194, 105556 (2020).
- [19] Lang, G., Yao, Y.: Formal concept analysis perspectives on three-way conflict analysis. *International Journal of Approximate Reasoning* 152, 160–182 (2023).
- [20] Zhi, H., Qi, J., Qian, T., Wei, L.: Three-way dual concept analysis. *International Journal of Approximate Reasoning* 114, 151–165 (2019).
- [21] Zhi, H., Li, J., Li, Y.: Multilevel conflict analysis based on fuzzy formal contexts. *IEEE Transactions on Fuzzy Systems* 30(12), 5128–5142 (2022).
- [22] Qi, Z., Zhi, H., Ding, W.: Distributed conflict analysis across varying analysis levels based on fuzzy formal contexts. *Information Sciences* 707, 122038 (2025).
- [23] Wu, J., Tsang, E.C., Xu, W., Zhang, C., Yang, L.: Correlation concept-cognitive learning model for multi-label classification. *Knowledge-Based Systems* 290, 111566 (2024).