Master in Data Science

Complex and Social Networks Laboratory

Finding and assessing community structure

Domenico Azzarito
domenico.azzarito@estudiantat.upc.edu

Gabriele Villa
gabriele.villa@estudiantat.upc.edu

Project Report

Academic Year 2025/2026

Contents

1 Introduction

The analysis of community structure is a fundamental task in the study of complex networks. A community is typically defined as a group of nodes that are more densely connected to each other than to the rest of the network. Identifying these structures helps to uncover the functional organization of real-world systems, ranging from social interactions to biological networks.
In this laboratory, we address the problem of finding and assessing communities using the R environment and the `igraph` package. Several algorithms have been proposed in the literature to solve this problem, each based on different heuristics such as modularity optimization, random walks, or edge betweenness. Consequently, different methods often yield different partitions for the same network.
The main objective of this report is to evaluate and compare the quality of clusterings produced by various algorithms, including Louvain, Label Propagation, and Walktrap. To assess the reliability of these methods, we propose a comparison criteria based on the Jaccard Index. We apply this methodology to four different networks: the Zachary's Karate Club, a synthetic scale-free network, the Enron email network, and a network of choice. Specifically, we measure how well the detected communities match a reference clustering, which is either the known ground truth or the best-performing partition according to significance scoring functions.

2 Results

In this section, we present the comprehensive results of our community detection analysis across all four networks. For each network, we report the significance scores obtained by each algorithm, the global and local Jaccard similarity with respect to the reference clustering, and visual comparisons of the detected communities against the ground truth.

2.1 Zachary's Karate Club Network

The Karate Club network consists of 34 nodes and 78 edges, representing social interactions among members of a university karate club. The known ground truth partition divides members into two factions.

Table 1: Significance scores for the Karate Club network.

Algorithm	Modularity	Conductance	Clust. Coef.	Expansion	Agg. Score
Louvain	0.411	0.241	0.606	3.254	**0.833**
Label Prop.	0.363	0.160	0.580	2.128	0.704
Walktrap	0.411	0.255	0.614	3.471	0.815
Edge Betw.	0.362	0.419	0.451	5.765	0.000

Table 2: Global Jaccard similarity for the Karate Club network.

Algorithm	Global Jaccard
Louvain	0.710
Label Prop.	**0.816**
Walktrap	0.706
Edge Betw.	0.559

Table 3: Local Jaccard similarity for the Karate Club network. Each cell shows the best matching cluster and the corresponding Jaccard index.

Algorithm	**GT Cluster 1**	**GT Cluster 2**
Louvain	Cluster 1 | 0.678	Cluster 3 | 0.738
Label Prop.	Cluster 1 | 0.714	Cluster 3 | 0.927
Walktrap	Cluster 2 | 0.688	Cluster 1 | 0.722
Edge Betw.	Cluster 1 | 0.563	Cluster 4 | 0.556

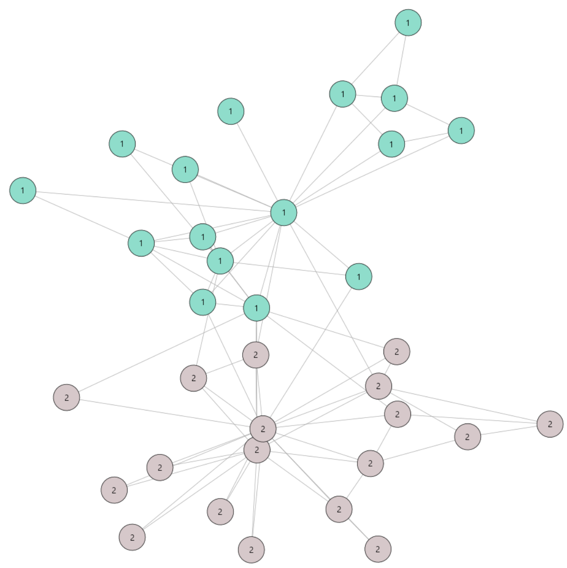
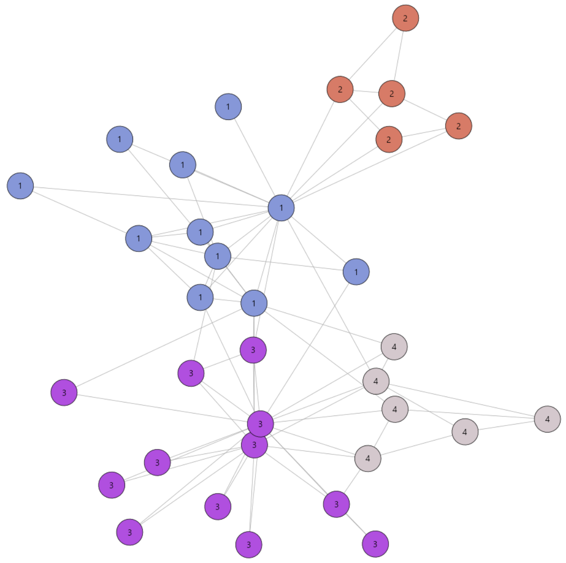
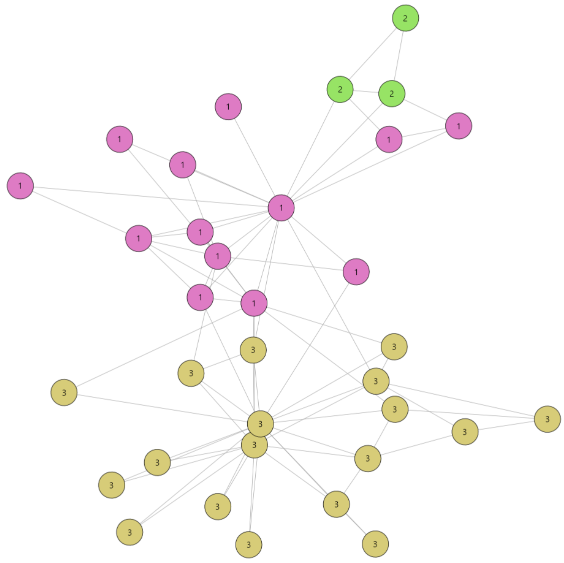


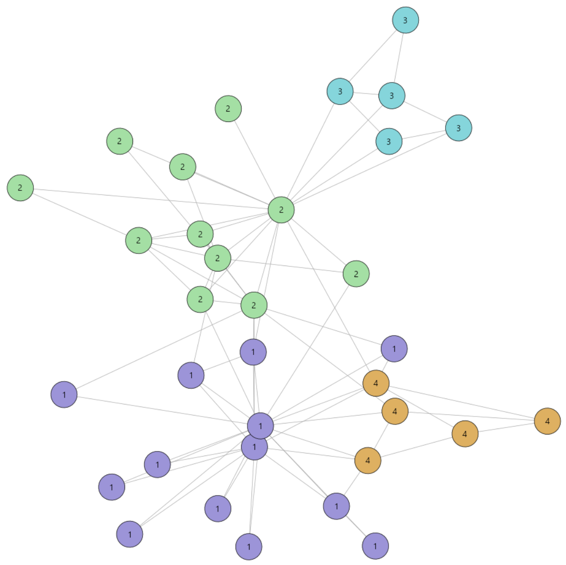
Figure 1: Ground truth community structure for Zachary's Karate Club. The network is divided into two factions based on the actual social fission that occurred.



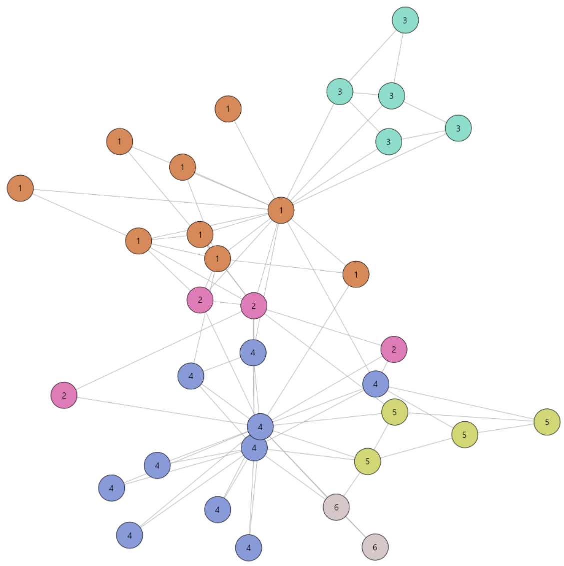
(a) Louvain



(b) Label Propagation



(c) Walktrap



(d) Edge Betweenness

Figure 2: Community detection results for Zachary's Karate Club using four different algorithms. Node colors and numeric labels correspond to detected communities.

2.2 Synthetic Barabási-Albert Blocks Network

The synthetic network was generated with 200 nodes, 800 edges, and 4 planted communities using the preferential attachment model with block structure. The mixing matrix ensures high intra-community connectivity (probability 1.0) and low inter-community connectivity (probability 0.1).

Table 4: Significance scores for the synthetic BA blocks network.

Algorithm	Modularity	Conductance	Clust. Coef.	Expansion	Agg. Score
Louvain	0.467	0.277	0.200	2.207	0.495
Label Prop.	0.295	0.174	0.171	1.393	**0.500**
Walktrap	0.449	0.290	0.212	2.310	0.487
Edge Betw.	0.464	0.296	0.202	2.300	0.436

Table 5: Global Jaccard similarity for the synthetic BA blocks network.

Algorithm	Global Jaccard
Louvain	**0.921**
Label Prop.	0.563
Walktrap	0.879
Edge Betw.	0.914

Table 6: Local Jaccard similarity for the synthetic BA blocks network. Each cell shows the best matching cluster and the corresponding Jaccard index.

Algorithm	GT Cluster 1	GT Cluster 2	GT Cluster 3	GT Cluster 4
Louvain	Cluster 2 | 0.961	Cluster 4 | 0.953	Cluster 1 | 0.955	Cluster 4 | 0.902
Label Prop.	Cluster 2 | 0.864	Cluster 5 | 0.744	Cluster 1 | 0.573	Cluster 4 | 0.913
Walktrap	Cluster 4 | 0.913	Cluster 3 | 0.878	Cluster 2 | 0.867	Cluster 1 | 0.864
Edge Betw.	Cluster 2 | 0.957	Cluster 3 | 0.818	Cluster 1 | 0.983	Cluster 4 | 0.879

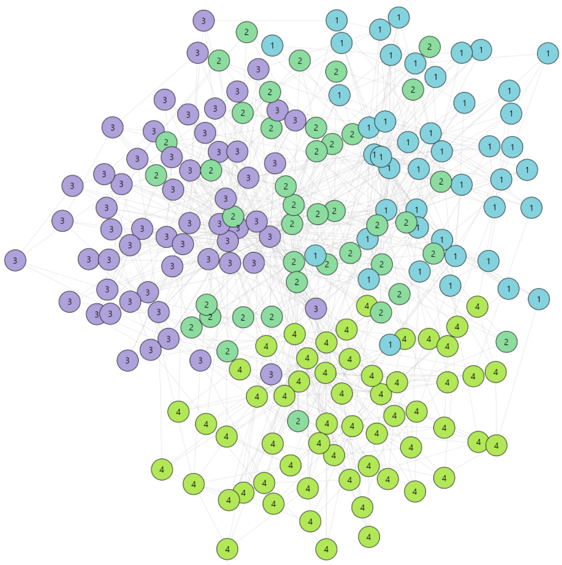
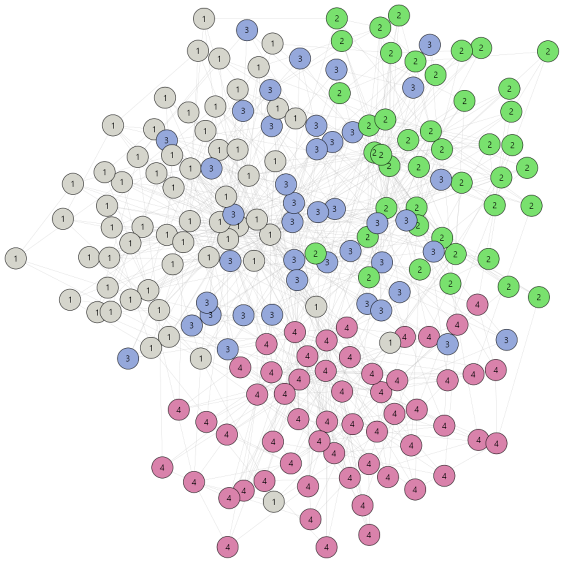
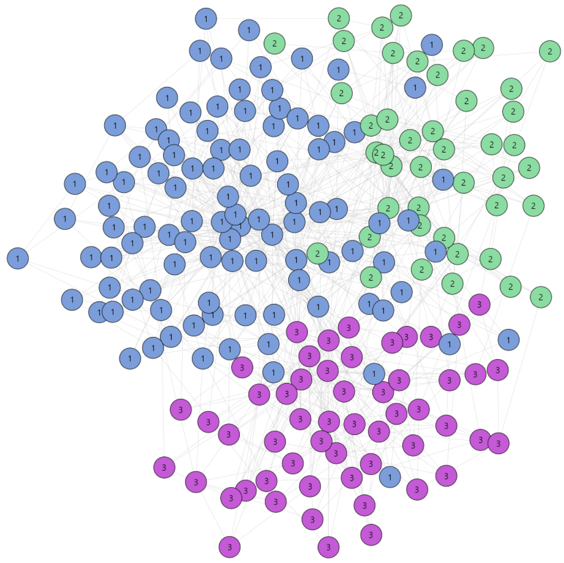


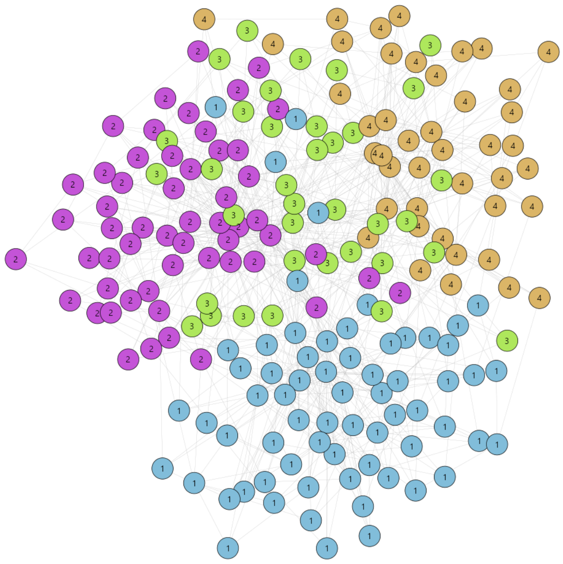
Figure 3: Ground truth community structure for the synthetic Barabási-Albert blocks network. The network contains 4 planted communities with high intra-community connectivity.



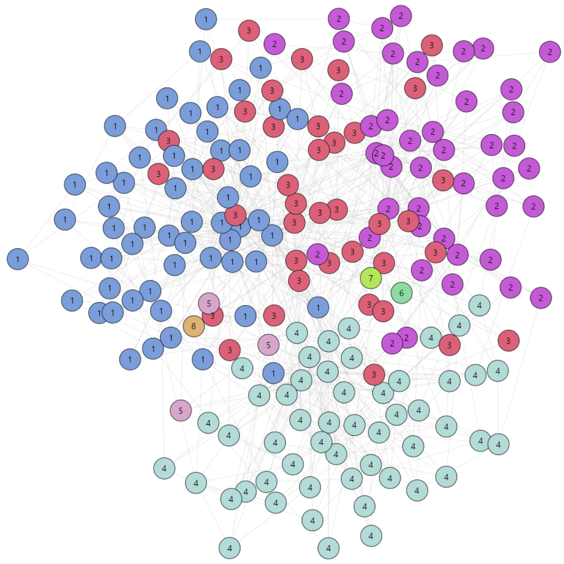
(a) Louvain



(b) Label Propagation



(c) Walktrap



(d) Edge Betweenness

Figure 4: Community detection results for the synthetic BA blocks network using four different algorithms. Node colors and numeric labels correspond to detected communities.

2.3 ENRON Email Network

The ENRON network, after preprocessing into a weighted undirected simple graph, contains 184 nodes and 2,085 edges representing email communications among Enron employees. No ground truth is available for this network.

Table 7: Significance scores for the ENRON network. Louvain is selected as the inferred ground truth.

Algorithm	Modularity	Conductance	Clust. Coef.	Expansion	Agg. Score
Louvain*	0.347	0.464	0.622	10.727	**0.605**
Label Prop.	0.012	0.011	0.379	0.198	0.500
Walktrap	0.295	0.442	0.599	9.652	0.573
Edge Betw.	0.139	0.696	0.661	13.696	0.345

*Selected as inferred ground truth based on highest aggregated score.

Table 8: Global Jaccard similarity for the ENRON network against Louvain (inferred GT).

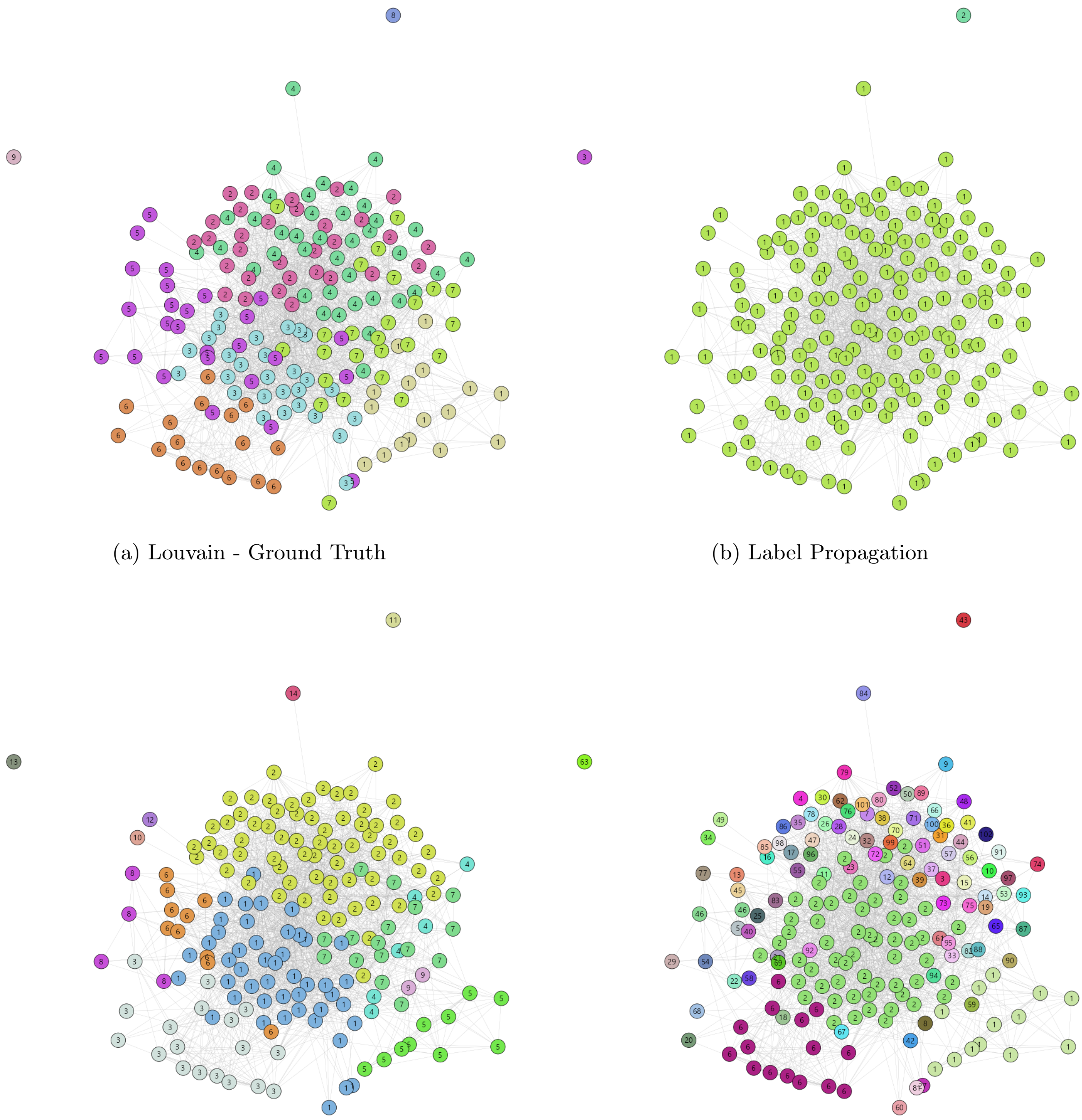
Algorithm	Global Jaccard
Label Prop.	0.185
Walktrap	**0.558**
Edge Betw.	0.288

Table 9: Local Jaccard similarity for the ENRON network (Part 1: GT Clusters 1–5). Each cell shows the best matching cluster and the corresponding Jaccard index.

Algorithm	GT Cluster 1	GT Cluster 2	GT Cluster 3	GT Cluster 4	GT Cluster 5
Label Prop.	Cluster 1 | 0.255	Cluster 2 | 0.196	Cluster 3 | 0.302	Cluster 2 | 0.255	Cluster 3 | 0.281
Walktrap	Cluster 5 | 0.625	Cluster 2 | 0.400	Cluster 1 | 0.609	Cluster 2 | 0.493	Cluster 6 | 0.385
Edge Betw.	Cluster 1 | 0.813	Cluster 2 | 0.115	Cluster 2 | 0.400	Cluster 2 | 0.067	Cluster 2 | 0.079

Table 10: Local Jaccard similarity for the ENRON network (Part 2: GT Clusters 6–9). Each cell shows the best matching cluster and the corresponding Jaccard index.

Algorithm	GT Cluster 6	GT Cluster 7	GT Cluster 8	GT Cluster 9
Label Prop.	Cluster 3 | 0.184	Cluster 3 | 0.169	Cluster 3 | 1.000	Cluster 4 | 1.000
Walktrap	Cluster 3 | 0.944	Cluster 7 | 0.640	Cluster 11 | 1.000	Cluster 13 | 1.000
Edge Betw.	Cluster 6 | 0.882	Cluster 2 | 0.141	Cluster 43 | 1.000	Cluster 63 | 1.000



(a) Louvain - Ground Truth

(b) Label Propagation

(c) Walktrap

(d) Edge Betweenness

Figure 5: Community detection results for the ENRON email network, where node colors and numeric labels correspond to detected communities.

2.4 UKfaculty Network

The UKfaculty network from the `igraphdata` package represents personal friendships among faculty members at a UK university. After preprocessing into a weighted undirected simple graph, the network contains 81 nodes and 577 edges. No ground truth community structure is available for this network.

Table 11: Significance scores for the UKfaculty network. Walktrap is selected as the inferred ground truth.

Algorithm	Modularity	Conductance	Clust. Coef.	Expansion	Agg. Score
Louvain	0.403	0.187	0.913	16.819	0.492
Label Prop.	0.344	0.192	0.942	14.891	0.439
Walktrap*	0.382	0.156	0.919	13.778	**0.837**
Edge Betw.	0.399	0.195	0.866	17.309	0.232

*Selected as inferred ground truth based on highest aggregated score.

Table 12: Global Jaccard similarity for the UKfaculty network against Walktrap (inferred GT).

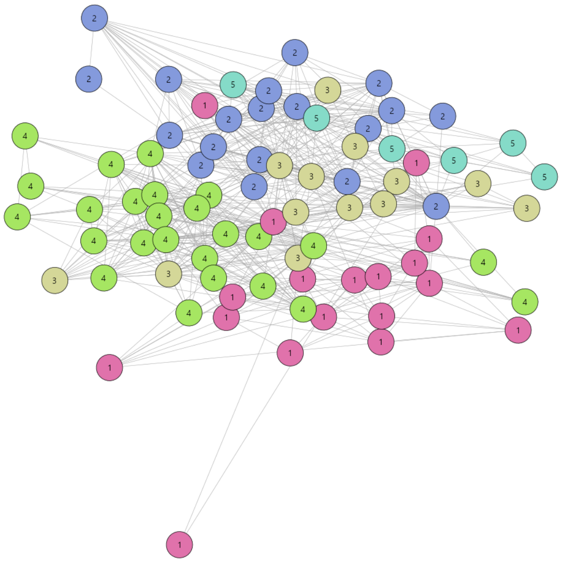
Algorithm	Global Jaccard
Louvain	**0.820**
Label Prop.	0.775
Edge Betw.	0.636

Table 13: Local Jaccard similarity for the UKfaculty network (Part 1: GT Clusters 1–3). Each cell shows the best matching cluster and the corresponding Jaccard index.

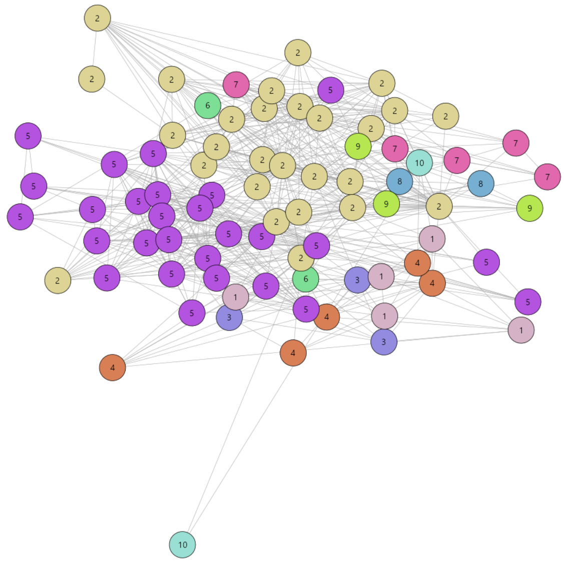
Algorithm	GT Cluster 1	GT Cluster 2	GT Cluster 3
Louvain	Cluster 1 | 0.833	Cluster 4 | 1.000	Cluster 3 | 0.615
Label Prop.	Cluster 1 | 0.487	Cluster 3 | 0.926	Cluster 8 | 0.375
Edge Betw.	Cluster 1 | 0.882	Cluster 3 | 0.893	Cluster 2 | 0.176

Table 14: Local Jaccard similarity for the UKfaculty network (Part 2: GT Clusters 4–6). Each cell shows the best matching cluster and the corresponding Jaccard index.

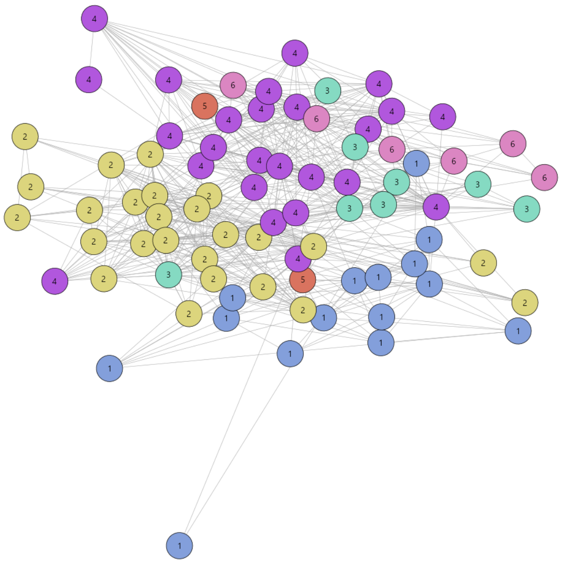
Algorithm	GT Cluster 4	GT Cluster 5	GT Cluster 6
Louvain	Cluster 2 | 0.760	Cluster 1 | 0.111	Cluster 5 | 1.000
Label Prop.	Cluster 2 | 0.924	Cluster 6 | 1.000	Cluster 7 | 0.833
Edge Betw.	Cluster 2 | 0.541	Cluster 1 | 0.056	Cluster 2 | 0.152



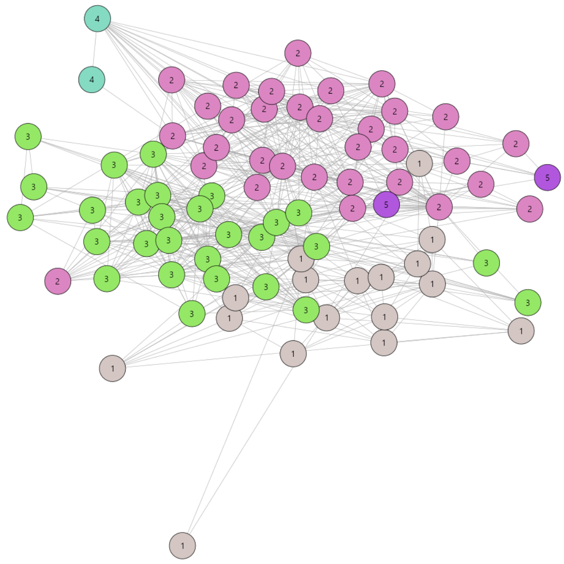
(a) Louvain



(b) Label Propagation



(c) Walktrap - Ground Truth



(d) Edge Betweenness

Figure 6: Community detection results for the UKfaculty network, where node colors and numeric labels correspond to detected communities.

3 Discussion

The results collected across the four distinct network topologies offer a multifaceted view of the community detection problem. By contrasting networks with known ground truth against those without, and synthetic structures against real-world social graphs, several key patterns emerged regarding the reliability of scoring functions, the impact of network topology, and the robustness of the algorithms.

3.1 Analysis of the Karate Network

In Table 1, Louvain achieves the highest aggregated score (0.833), followed closely by Walktrap (0.815). Edge Betweenness performs notably worse, with the lowest aggregated score due to the highest conduc-

tance and expansion values, as well as the lowest modularity and clustering coefficient. This suggests that, from a purely topological perspective, Louvain finds the most mathematically optimized partition by maximizing density within groups.
Notably, in Table 2, Label Propagation achieves the highest global Jaccard similarity (0.816), despite having a lower aggregated significance score than Louvain. This indicates that Label Propagation's partitions more closely match the actual social fission, particularly for Faction 2, where it reaches a local Jaccard index of 0.927.
This discrepancy results from the different granularity of the partitions, as seen in Figure 1. The Ground Truth consists of only two large factions. Louvain and Walktrap identify 4 communities; they effectively subdivide the two main factions into smaller, denser sub-groups. While this increases modularity, it lowers the Jaccard similarity because the algorithms over-segment the network compared to the binary ground truth. In contrast, Label Propagation identifies 3 communities, maintaining a structure that is closer to the original two factions.
In conclusion, while Louvain is the best algorithm for structural optimization, Label Propagation proves to be the best method for recovering the specific ground truth of this social network, likely because its majority-voting mechanism captures the fluidity of the social split better than strict modularity maximization.

3.2 Analysis of the Synthetic Barabási-Albert Blocks Network

In Table 4, the aggregated scores are remarkably similar across algorithms, with Label Propagation marginally leading (0.500). This suggests that all methods perform comparably on internal validity metrics for this well-structured synthetic network.
However, Table 5 reveals a strikingly different picture: Louvain dramatically outperforms Label Propagation in recovering the planted community structure (Global Jaccard: 0.921 vs. 0.563). Edge Betweenness also achieves excellent recovery (0.914), while Walktrap follows closely (0.879). The local analysis in Table 6 indicates that Label Propagation struggles particularly with GT Cluster 3, achieving only 0.573. This divergence between internal scores and ground truth recovery can be attributed to the topology of the network. Since the graph was generated using the Barabási-Albert model, it possesses a scale-free degree distribution characterized by the presence of hubs (high-degree nodes). Label Propagation relies on local majority voting; consequently, hubs can exert disproportionate influence, propagating labels across community boundaries even when inter-community connectivity is low.
This effect is clearly visible in Figure 4. While Louvain clearly distinguishes the four planted blocks, Label Propagation produces a fragmented partition where the community structures are mixed and less defined. Thus, despite its favorable internal scores, Label Propagation proves unstable in scale-free environments, whereas Louvain demonstrates superior robustness in recovering the underlying block structure.

3.3 Analysis of the ENRON Network

In Table 7, Louvain achieves the highest aggregated score (0.605) and is therefore selected as the inferred ground truth for Jaccard similarity evaluation. The Louvain clustering identifies 9 communities in the network. In contrast, Label Propagation shows notably different behavior with very low modularity (0.012); while its conductance and expansion are also low, this suggests it produces fewer, larger communities, effectively merging the majority of the network into a single block.
Table 8 shows that Walktrap exhibits the highest agreement with the Louvain-based reference (Global Jaccard: 0.558), suggesting that these two methods capture similar community structures despite their different underlying heuristics. Edge Betweenness achieves only moderate similarity (0.288), while Label

Propagation performs poorly (0.185), confirming its inability to recover the granular structure identified by Louvain.

Finally, the local analysis in Tables 9 and 10 reveals that while medium-sized communities show variation in recovery quality, the smallest communities, specifically GT Clusters 8 and 9, are perfectly recovered by all algorithms (Jaccard index 1.000). This perfect score occurs because these clusters correspond to single isolated or pendant nodes. Since these nodes lack complex connectivity patterns, they are trivially identified as distinct units by all algorithms, regardless of the method used.

3.4 Analysis of the UKfaculty Network

As shown in Table 11, Walktrap achieves the highest aggregated score (0.837), indicating it creates the most well-balanced partition in terms of conductance and cluster isolation. Consequently, it is selected as the inferred ground truth for this network, identifying 6 distinct communities.

A notable divergence appears when observing Louvain. While it achieves the highest modularity (0.403), its aggregated score is significantly lower (0.492) due to a high expansion value (16.819). This suggests that while Louvain maximizes internal density, the cuts it performs result in a higher number of edges pointing outside the communities compared to Walktrap. Edge Betweenness performs the worst overall (Aggregated Score: 0.232), primarily due to the highest expansion (17.309) and the lowest clustering coefficient (0.866), confirming its struggle to find cohesive groups in this topology.

Regarding the similarity analysis presented in Table 12, Louvain exhibits the highest agreement with the Walktrap-based reference (Global Jaccard: 0.820). This strong correlation implies that the community structure of the UKfaculty network is robust, as two different heuristics, random walks and modularity optimization, converge on very similar partitions. In contrast, Edge Betweenness shows notably lower similarity (0.636), producing a partition that deviates significantly from the consensus found by the other methods.

The local analysis in Tables 13 and 14 provides deeper insight into the differences between Walktrap and Louvain. GT Clusters 2 and 6 are perfectly recovered by Louvain (Jaccard = 1.000), indicating these are well-defined, distinct social groups. However, GT Cluster 5 shows very poor correspondence (0.111). Detailed inspection reveals that Louvain merges this small community with Cluster 1 and 2. This is a classic example of the resolution limit of modularity optimization: Louvain tends to merge small, tight cliques into larger communities to maximize the global modularity score, whereas Walktrap's random walker gets trapped in the small substructure, effectively preserving it as a separate entity.

4 Methods

In this section, we present the comprehensive methodological framework adopted for the analysis. We detail the selected algorithms, the significance evaluation metrics, the implementation of the similarity criteria, and the specific data preprocessing steps applied to each network.

4.1 Community Detection Algorithms

We selected four algorithms from the `igraph` package to cover a diverse range of theoretical approaches to community detection:

- **Louvain** (`cluster_louvain`): A heuristic method based on multi-level modularity optimization. It is fast but stochastic (order-dependent).
- **Label Propagation** (`cluster_label_prop`): A near-linear time algorithm based on node majority voting. It is inherently stochastic.

- **Walktrap** (`cluster_walktrap`): Hierarchical clustering based on random walks, assuming short walks stay within communities. Treated here as deterministic.

- **Edge Betweenness** (`cluster_edge_betweenness`): A divisive algorithm based on the Girvan-Newman method, removing edges with high centrality. It is deterministic.

4.2 Data Preprocessing and Network Generation

Before applying the algorithms, specific preprocessing steps were required:

- **Synthetic Network Generation:** We generated a scale-free network with community structure using the `barabasi_albert_blocks` model. We set the parameters to $N = 200$ nodes and $m = 4$ communities. To ensure a clear structure, we defined a mixing matrix B with high internal probability (1.0) and low external probability (0.1) between blocks.

- **ENRON Network:** The original dataset is a multigraph. We simplified it into a weighted undirected graph using the `simplify` function, combining multiple edges by summing their weights (edge attribute `weight = "sum"`). This preserves the intensity of communication between users.

- **UKfaculty:** Loaded from the `igraphdata` package. We will discuss more about its pre-processing in the following subsections.

4.3 Significance Evaluation Framework

To assess clustering quality without a ground truth, we utilized four internal validity indices from the `clustAnalytics` package:

- **Modularity** (Higher is better)

- **Clustering Coefficient** (Higher is better)

- **Expansion** (Lower is better)

- **Conductance** (Lower is better)

These metrics serve as community scoring functions that capture different structural properties of the network. The choice of these metrics is, in fact, not casual: modularity is based on the network model, the clustering coefficient is based on internal connectivity, expansion is based on external connectivity and conductance combines internal and external connectivity.

4.3.1 Score Normalization and Aggregation

Since these metrics have different scales and directions of optimization, we implemented a robust normalization strategy. For each metric, raw scores were transformed into a $[0, 1]$ range using Min-Max normalization. Crucially, for metrics where "lower is better" (Conductance, Expansion), we inverted the normalized score so that 1 always represents the best performance and 0 the worst. The final **Aggregated Score** for each algorithm is computed as the arithmetic mean of these four normalized metrics.

4.4 Jaccard Similarity Assessment

We developed a custom implementation to quantify the similarity between a computed partition and a reference partition (Ground Truth).

4.4.1 Metric Definition

Our procedure involves three stages:

1. **Local Jaccard Matrix:** We compute a matrix J where J_{ij} is the Jaccard index between cluster i of the first partition and cluster j of the second.
2. **Best Matching:** For each cluster in the reference partition, we identify the cluster in the target partition that maximizes the Jaccard index.
3. **Global Weighted Similarity:** We compute the weighted mean of these maximum local Jaccard values, weighting each cluster by its size (fraction of total nodes).

4.4.2 Handling Stochasticity

A fundamental challenge in evaluating community detection algorithms arises from the inherent stochasticity of certain methods. Both Louvain and Label Propagation algorithms exhibit non-deterministic behavior: Louvain's results depend on the order in which nodes are processed during the local moving phase, while Label Propagation relies on random tie-breaking when multiple labels have equal frequency in a node's neighborhood. Consequently, a single execution of these algorithms may yield results that are not representative of their typical performance, potentially leading to misleading conclusions.
To address this challenge rigorously, we implemented a Monte Carlo simulation framework with $t = 100$ independent iterations. This approach provides statistically robust estimates of algorithm performance while quantifying the variability inherent in stochastic methods.

Score Averaging Protocol For each stochastic algorithm A and each significance metric S_k (where $k \in \{\text{modularity, conductance, clustering coefficient, expansion}\}$), we execute the algorithm t times and compute:

$$\bar{S}_k^{(A)} = \frac{1}{t} \sum_{i=1}^{t} S_k^{(A,i)} \tag{1}$$

where $S_k^{(A,i)}$ denotes the score obtained in the i-th iteration. This averaging procedure smooths out run-to-run fluctuations and provides a more reliable estimate of expected performance. The aggregated score is then computed on these averaged values.

Jaccard Stability Assessment The evaluation of Jaccard similarity for stochastic algorithms requires particular care, as both the partition structure and cluster assignments may vary across runs. Our protocol proceeds as follows:

1. **Iteration Loop:** For each iteration $i \in \{1, \dots, t\}$:
   - Generate a fresh partition $\mathcal{P}^{(i)}$ by running the stochastic algorithm
   - Compute the full Jaccard similarity matrix $J^{(i)}$ between the ground truth partition $\mathcal{P}_{\text{GT}}$ and $\mathcal{P}^{(i)}$
   - For each ground truth cluster C_k^{GT}, identify the best-matching cluster $M_k^{(i)}$ and record both the match identity and Jaccard value $J_k^{(i)}$
   - Compute the weighted global Jaccard similarity $J_{\text{global}}^{(i)}$
2. **Global Jaccard Aggregation:** The reported global Jaccard similarity is the arithmetic mean across all iterations:

$$\bar{J}_{\text{global}} = \frac{1}{t} \sum_{i=1}^{t} J_{\text{global}}^{(i)} \tag{2}$$

3. **Local Jaccard Aggregation:** We group iterations by their match identity and select the match with the highest mean Jaccard value:

$$\bar{J}_k = \max_{m \in \mathcal{M}_k} \left\{ \frac{1}{|I_m|} \sum_{i \in I_m} J_k^{(i)} \right\} \tag{3}$$

where $\mathcal{M}_k$ is the set of all distinct matches observed for cluster k, and $I_m = \{i : M_k^{(i)} = m\}$ is the set of iterations where match m occurred.

This methodology ensures that the reported local Jaccard values reflect the most consistent and highest-quality matches, rather than being diluted by occasional poor alignments that may occur due to the stochastic nature of the algorithms.

4.5 Reference Clustering Strategy

The evaluation of clustering quality through Jaccard similarity fundamentally requires a reference partition against which algorithmic outputs can be compared. We developed a principled dual-strategy approach that adapts to the availability of ground truth information:

- **Known Ground Truth:** For *Karate* and *Synthetic* networks, we used the provided class labels.
- **Inferred Ground Truth:** For *ENRON* and *UKfaculty*, where no ground truth exists, we first ran the significance evaluation. The algorithm with the highest Aggregated Score was identified as the "best available clustering" and subsequently used as the reference Ground Truth to evaluate the similarity of the other methods.

Several aspects of this strategy merit discussion:

- **Circularity Concern:** Using algorithmic output as ground truth introduces potential circularity. However, by evaluating algorithms against multiple diverse metrics before selection, we mitigate the risk of selecting a partition that merely optimizes a single criterion.
- **Exclusion from Comparison:** When an algorithm's output serves as the inferred ground truth, it is excluded from the Jaccard similarity comparison (as it would trivially achieve perfect similarity with itself).
- **Partition Reference:** For stochastic algorithms selected as the best performer, we fix a single partition instance (using a set random seed) as the reference. This ensures consistency across all subsequent comparisons.

4.6 Choice of the Fourth Network: UKfaculty

We selected the UKfaculty network from the `igraphdata` package to satisfy the requirement of a network with no known community structure. This network represents personal friendship connections among faculty members at a UK university, captured through self-reported survey data. The original network is a directed multigraph with 81 nodes and 577 edges, where multiple edges between node pairs indicate repeated friendship declarations.

Preprocessing methodology: To apply community detection algorithms, we performed two essential transformations. First, we converted the directed graph to undirected, combining reciprocal edges into single undirected edges. This transformation is justified because friendship relationships are inherently

symmetric: if person A considers person B a friend, this represents a mutual social connection regardless of whether B explicitly declared A as a friend. Second, we simplified the resulting multigraph aggregating multiple edges into weighted edges where weights represent the strength of the relationship.
This network is particularly suitable for community detection analysis because:

1. Its moderate size (81 nodes) ensures computational tractability while remaining non-trivial
2. Faculty members naturally form communities based on departmental affiliations and research interests, providing an implicit ground truth for validation
3. The weighted structure resulting from edge aggregation captures relationship strength, which community detection algorithms can leverage to identify more cohesive clusters.