Semantic-Driven Community Detection in Complex Networks Using Fuzzy Logic and Multi-Criteria Decision Making

Mohamed El-Moussaoui[1]*, Mohamed Hanine[2], Sulieman S. Alshuhri[3], Amal Alomran[4], Amir Mohamed Talib[5]
LTI Laboratory-École Nationale des Sciences Appliquées, Chouaib Doukkali Université, El Jadida, Morocco[1, 2]
College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia[3, 4, 5]

***Abstract*—The study introduces a novel approach to community detection in complex networks by integrating fuzzy logic with multi-criteria decision-making techniques. Unlike traditional methods that rely primarily on topological metrics, the proposed approach incorporates semantic attributes to identify meaningful community structures. Fuzzy logic addresses the inherent uncertainty and ambiguity in processing these attributes, enabling a flexible detection process that is not dependent on network topology. To enhance scalability, this study customizes the k-means clustering algorithm to accommodate small- and large-scale network structures. Experimental results show that the proposed fuzzy logic-based approach achieves competitive performance compared with conventional algorithms. Additionally, the proposed approach demonstrates robustness by generating well-balanced communities with competitive execution times compared with the evaluated methods. These findings highlight the potential benefits of incorporating semantic attributes and fuzzy reasoning into community detection in complex networks.**

***Keywords*—Community detection; fuzzy logic; Multi-Criteria Decision Making (MCDM); social network analysis; semantic communities**

I. Introduction

Real-world networks such as biological, citation, transportation, and social networks have been modelled throughout research advancements to be represented as graphs, where nodes and edges represent, respectively, the network components and the interconnecting relationships between them. These representations offer practical utility in highlighting groups of nodes when their connections are observed more internally than externally. Thereby, the identified groups are called communities, reflecting as much as possible a degree of similarity among their nodes [1, 2]. However, behind this conceptual presentation, challenging problems such as network heterogeneity, complex interactions, and computational effort make the task challenging for researchers.

In the literature, existing approaches to community detection address the clustering problem from different perspectives. For example, using modularity-based techniques, local and global subsetting analysis, and optimization tasks often prove insufficient when applied to a directed network structure [1, 3-8], one of many recurring problems in the clustering paradigms. This article makes a methodological contribution by combining data mining [10, 11] with fuzzy logic [14, 15] in the field of community detection, improving the processing of large datasets and enhancing uncertainty extraction performance to handle the semantic properties of nodes. Thus, the proposal in this study represents an approach that captures uncertainty and linguistic ambiguity in the semantic properties of nodes, supporting the detection of existing semantic communities.

The manuscript is structured as follows: Section II reviews existing approaches to community detection. Section III presents the proposed methodology, while Section IV describes the experimental configuration and findings. Finally, Section V discusses the results obtained and offers conclusions.

II. Related Work

Various approaches have been proposed over the last decade for identifying community structures. Girvan and Newman introduced a seminal edge-betweenness-based approach for community detection in social networks. Their method iteratively removes edges with high betweenness to reveal community boundaries, with modularity commonly used to evaluate the resulting partitions [4, 5, 6]. Other similar approaches have explored spectral clustering and the minimum cut formulations [8], while others have relied on statistical inference [7]. Additional methods used unsupervised clustering based on similarity measures [14]. However, these approaches face some limitations, particularly when handling edge directionality [6, 7]. Similarly, some studies [10, 16] do not explicitly account for edge directionality, limiting their applicability to directed networks, particularly when the number of communities parameter is unknown a priori.

Recently, increasing attention has been devoted to combining structural and semantic information to improve community detection [8, 13]. The identification of communities based on semantic attributes has attracted increasing attention, particularly for weighted and directed networks. This has motivated the development of methods that jointly consider both topological and semantic information to address semantic-community detection. For instance, methods integrating multi-criteria decision-making [17, 18, 19] and fuzzy logic [14, 15] provide mechanisms for modeling heterogeneous and uncertain node attributes, thereby complementing purely topology-based representations with semantic information.

*Corresponding author.

This study introduces a fuzzy logic-based approach that integrates semantic attribute analysis with multi-criteria decision-making to identify semantically coherent communities in directed networks. The proposed method uses preference functions to quantify the relevance of node properties and incorporates these preferences into the community detection process.

III. DEFINITION AND APPROACH METHOD

This section first presents the definitions of the key concepts used in this study, followed by an introduction to the fundamental principles of the proposed fuzzy approach.

Definition 1. Let G be a graph denoted by (V, E), composed of n nodes and m edges. V represents the set of vertices of G, where v_i denotes the i^{th} node. Each node v_i incorporates the set A_i representing qualitative and quantitative attributes of node v_i, for $1 \leq i \leq N$. Where, E denotes the set of edges of G, and e_i represents an edge connecting two nodes v_i and v_j.

Definition 2. Let M=(l, m, u) be a fuzzy number [9, 11] associated with the membership function f(x). Where l, m, and u represent the parameters of the membership degree.

The two possible forms of the membership function f(x) are considered in this work, as defined in Eq. (a) and (b).

$$f(x) = \begin{cases} 0 & if\ x < l \\ \frac{x-l}{m-l} & if\ \ l \leq x \leq m \\ \frac{u-x}{u-m} & if\ \ m < x \leq u \\ 0 & if\ x > l \end{cases} \tag{a}$$

$$f(x) = \begin{cases} 0 & if\ x < l \\ \frac{x-l}{m_1-l} & if\ \ l \leq x < m_1 \\ 1 & if\ m_1 \leq x < m_2 \\ \frac{u-x}{u-m_2} & if\ \ m_2 \leq x < u \\ 0 & if\ x \geq u \end{cases} \tag{b}$$

Definition 3. Let M be a fuzzy number as defined in Definition 2, with an associated membership function $\mu_M(x)$ derived from f(x), where $\mu_M(x) \in [0,1]$. When a fuzzy number has a single peak m, it is referred to as a triangular fuzzy number. When it has two peaks m_1 and m_2 with $m_1 \leq m_2$, it is referred to as a trapezoidal fuzzy number. To compute precise representative values of the fuzzy number, the maximization and minimization processes are denoted respectively by $\mu_{max}(x)$ and $\mu_{min}(x)$.

$$\mu_{max} = \begin{cases} x\ if\ 0 \leq x \leq 1 \\ 0\ otherwise \end{cases} \quad \text{and} \quad \mu_{min} = \begin{cases} 1 - x\ if\ 0 \leq x \leq 1 \\ 0\ otherwise \end{cases}$$

Based on these operations, three score values can be derived for the fuzzy number M: $\mu_R(M)$, $\mu_L(M)$, and $\mu_T(M)$, which correspond respectively to the right score, left score, and total score. Their mathematical formulations are given in Eq. (1), (2), and (3).

$$\mu_R(M) = \sup_x\{\mu_M(x) \wedge \mu_{max}(x)\} \tag{1}$$

$$\mu_L(M) = \sup_x\{\mu_M(x) \wedge \mu_{min}(x)\} \tag{2}$$

$$\mu_T(M) = \frac{(\mu_R(M)+1-\mu_L(M))}{2} \tag{3}$$

Definition 4. Let us consider a preference function that models the degree of preference among associated fuzzy sets defined in Definition 3. It characterizes how different evaluation scenarios influence the assignment of preferences.

Four preference cases ψ are considered: 1) alignment with the preference ψ(1), 2) independence from the preference ψ(2), 3) favorable preference ψ(3), and 4) contradiction with the preference ψ(4). Each of which is defined mathematically in Table I as follows:

TABLE I. MATHEMATICAL REPRESENTATION OF THE PREDEFINED PREFERENCE CASES

Preferences	Triangular fuzzy number	Trapezoidal fuzzy number
ψ(1) : Alignment	e	$\begin{cases} 0\ \text{if}\ 0 \leq e \leq m_1 \\ \frac{e-m_1}{m_2-m_1}\ \ if\ \ m_1 \leq e \leq m_2 \\ 1\ if\ \ m_2 \leq e \leq 1 \end{cases}$
ψ(2) : Independence	$\begin{cases} \frac{e}{m_1}\ \ if\ \ 0 \leq e \leq m_1 \\ \frac{1-e}{1-m_1} if\ m_1 \leq e \leq 1 \end{cases}$	$\begin{cases} 0 & if\ 0 \leq e \leq m_1 \\ \frac{e-m_1}{m_2-m_1} & if\ \ m_1 \leq e \leq m_2 \\ 1 & if\ m_2 \leq e \leq m_3 \\ \frac{m_4-e}{m_4-m_3} & if\ m_3 \leq e \leq m_4 \\ 0 & if\ m_4 \leq e \leq 1 \end{cases}$
ψ(3): Favorable	$\begin{cases} \frac{m_1-e}{m_1} & if\ \ 0 \leq e \leq m_1 \\ \frac{e-m_1}{1-m_1} & if\ \ m_1 \leq e \leq 1 \end{cases}$	$\begin{cases} 1\ \ if\ 0 \leq e \leq m_1 \\ \frac{m_2-e}{m_2-m_1}\ \ if\ m_1 \leq e \leq m_2 \\ 0\ \ if\ m_2 \leq e \leq m_3 \\ \frac{e-m_3}{m_4-m_3}\ \ if\ \ m_3 \leq e \leq m_4 \\ 1\ \ if\ m_4 \leq e \leq 1 \end{cases}$
ψ(4): Contradictory	$1 - e$	$\begin{cases} 1 & if\ 0 \leq e \leq m_1 \\ \frac{m_2-e}{m_2-m_1} & if\ m_1 \leq e \leq m_2 \\ 0 & if\ m_2 \leq e \leq 1 \end{cases}$

A. Proposed Architecture

This section describes the architecture of the proposed fuzzy approach [12, 19].

The architecture of the proposed approach includes three phases (Fig. 1): The Environment preparation phase (Phase 1), the Fuzzy analysis phase (Phase 2), and the social network analysis phase (Phase 3).

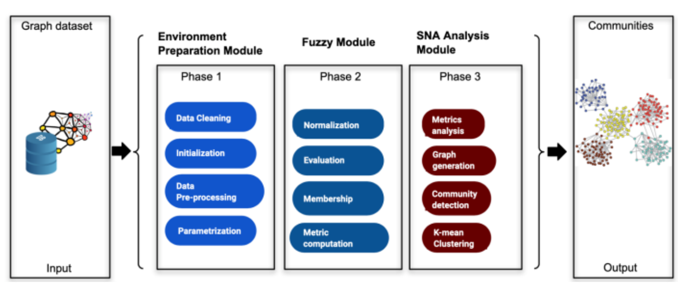


Fig. 1. The architecture of the proposed fuzzy approach-based three main modules within the three phases.

- Input Layer: A network is not only a collection of interconnected nodes; it also incorporates layers of semantic data that can significantly enrich the understanding of the network's structure. This layer ensures the network dataset is prepared for further processing in subsequent phases.
- Environment Preparation Layer: The environment preparation layer plays a critical role in integrating the network dataset by performing data cleaning and preprocessing, enhancing data clarity, and extracting semantic information from node attributes.
- Fuzzy Layer: Phase 2 involves a series of processes powered by the fuzzy conversion technique, including normalization and data evaluation, along with the conversion of quantitative data into qualitative insights.

The fuzzy conversion technique is applied in this phase to derive an initial appreciation matrix the identification of decision elements. The appreciation matrix is then normalized into values within the interval [0, 1]. Thus, the preference function is constructed to capture the analyst's preferences. Then, the appreciation matrix undergoes a final normalization, leading to the decision matrix. This step produces a new weighted matrix of the network.

- Social Network Analysis Layer: A new network is constructed at this stage, based on the generated decision matrix, which will be integrated into the community detection algorithm.
- Community Detection Layer: This phase, dedicated to the community detection algorithm, utilizing the enhanced k-means clustering algorithm to define community structure within the network.

B. Workflow of the Proposed Approach

The workflow of the proposed approach emphasizes the extraction and analysis of node properties. The application of the fuzzy approach involves six processes. The following Fig. 2 illustrates the exhaustive steps of the approach:

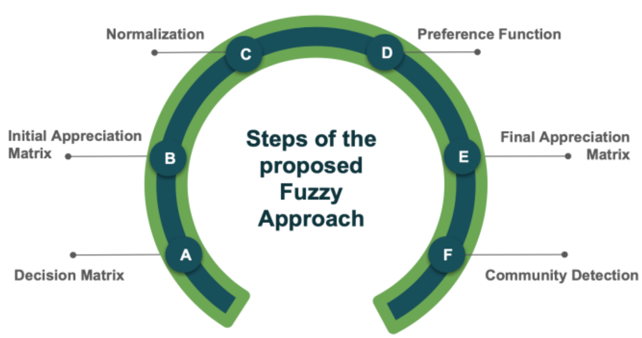


Fig. 2. The six steps of the proposed fuzzy approach.

1) Decision matrix: Where the node attribute criteria are formulated into families of attributes, denoted F = {f1, f2, f3, ... , fk}, with k=|F|. Where, F serves as the network analyst's preferences, reflecting decisions that have either positive or negative impacts on the community detection process.

2) Initial appreciation matrix: The initial appreciation matrix is built based on the conversion of node properties from qualitative into quantitative values. In fact, this transformation based on linguistic terms and their corresponding membership function M, helps determine the precise numerical values by using Eq. (3) of the fuzzy conversion technique, providing as a result the final appreciation matrix A′:

$$(A'_i)_{1\le i\le N} = ((f_i(a_j))_{1\le j\le N}) \text{ with } f_i \in F \text{ and } a_j \in A$$

3) Normalization: The normalization process ensures the numerical values of the initial appreciation matrix are within the interval [0,1], using Eq. (4). The output of this step is then a normalized appreciation matrix E={e_{ij}}, with e_{ij} defined as follows:

$$e_{ij} = \frac{a'_{ij}}{\sum_j a'_{ij}} \tag{4}$$

4) Preference function: The preference cases as introduced in Definition 4 are defined by the fuzzy number M_i = (m_1, m_2, ..., m_q), where each M_i corresponds to the q inflection points throughout the interval [0, 1].

The preference function depends on the fuzzy number M_i, the preference cases ψ(i), and the family of attributes as follows:

$$P_i = P_{\psi(i)}^{M_i}(f_i)$$

5) Final appreciation matrix: For a given f_i, the final appreciation matrix denoted $P_i(e_{ij})$ introduces the decision matrix D_{ij} [22] defined by the following Eq. (5).

$$D_{ij} = \frac{P_i(e_{ij})}{\sum_i P_i(e_{ij})} \tag{5}$$

6) Community detection: In this step, the weight vector θ_i is determined for each vector f_i using the AHP process [22, 26]. In fact, the weight θ_i associated with f_i is defined by comparing the values $\sum_i^k w_{ik}$ using Eq. (6), which measures the importance of each element of f_i.

First, the reciprocal square matrix is constructed [20]. The k-means clustering algorithm is then enhanced [21, 24], as described in Algorithm 1.

$$\theta_i = \frac{(\prod_{j=1}^k w_{ij})^{1/k}}{\sum_{i=1}^k \left(\prod_{j=1}^k w_{ij}\right)^{1/k}} \tag{6}$$

This introduces the utility function, aggregating the vector of weights θ_i, and the computed decision matrix D_{ij}:

$$U_i = \sum_{j=1}^k \theta_i \times D_{ij} \tag{7}$$

The introduced utility function U, denoted U={u_1, u_2, ..., u_N}, is considered as a data point of the evaluated node properties within the k-dimensional space. Here, k is the dimension of f_i with i∈[0, k].

The proposed approach extends conventional k-means by incorporating semantic information derived from the fuzzy-logic and MCDM stages. It considers semantic utility values and attribute relevance during cluster formation. This procedure is further adapted to accommodate networks of different scales while preserving the iterative k-means framework.

Algorithm 1: Fuzzy community detection algorithm based on enhanced K.

Require:
  Dataset $U = \{u_1, u_2, ..., u_N\}$
  Number of clusters k
  Utility values derived from the fuzzy-MCDM process
Ensure:
  Community set $C = \{C_1, C_2, ..., C_k\}$
Initialize:
  Select k initial centroids $\theta_1, \theta_2, ..., \theta_k$ using the utility values obtained from fuzzy-MCDM process
Repeat
  1- Initialize $C_1, C_2, ..., C_k \leftarrow \emptyset$
  2- For each data point u_i, $i \in \{1..N\}$ do
    $J^* \leftarrow \text{argmin}_{j \in \{1..k\}} \|u_i - \theta_j\|^2$ {Find index of closest centroid}
    Assign u_i to cluster C_{J^*}
  3- For each cluster C_j, $j \in \{1..k\}$ do
    if C_j is not empty then
      $\theta_j \leftarrow (1/|C_j|) * \sum_{u_i \in C_j} u_i$
  4- Repeat until convergence of θ_j
Return $C = \{C_1, C_2, ..., C_k\}$

The obtained utility function U determines the input of the k-means algorithm; it represents the initial centroids for the community detection algorithm, with a starting number k' of centroids, where k'=unique(k).

The sum of square distance (Δ) between each centroid and the utility function u_i is computed via Eq. (8). Each utility value u_i is then assigned to the nearest centroid c_j, and the community formation process continues iteratively.

$$\Delta(C) = \sum_{i=1}^{N} \sum_{j=1}^{k'} \|u_i - c_j\|^2 \quad (8)$$

The community detection algorithm continues adjusting the centroids until convergence stability is observed, forming well-defined clusters within the network.

IV. Experimental Evaluation

The proposed approach was implemented in the R programming language on an operating system equipped with an Intel Core i5 CPU running at 2.6 GHz and 8 GB of memory. The proposed approach was evaluated on synthetic networks of different sizes and compared with four well-known community detection algorithms.

A. Datasets Preparation

The adopted network structure is the Erdős- Rényi random graph [25], which allows the generation of six graphs with different sizes, ranging from 50 to 10000 nodes.

Nine properties were randomly assigned to all nodes of the generated networks. Specifically, five of them are qualitative values (f_1, f_2, f_3, f_4, f_5) and four quantitative values (f_6, f_7, f_8, f_9) to each node, ensuring diversified categories.

The synthetic semantic attributes used in the experiments provide a controlled environment for evaluating the influence of semantic information of the proposed approach. However, they do not fully capture the complexity and heterogeneity of the semantic attributes in real-world networks. Table II and Table III represent the environment prerequisites of the experimentation.

TABLE II. The Six Generated Erdős- Rényi Graphs, with the Average Degree

Network scale	Nbr. Of nodes	Nbr. Of edges	Average degree
50-nodes	50	61	2,44
100-nodes	100	246	4,92
300-nodes	300	2.224	14,82
1000-nodes	1.000	24.967	49,93
3000-nodes	3.000	225.498	150,33
10000-nodes	10.000	2.499.749	499,9498

TABLE III. Categories of the Assigned Random Attributes and their Associated Abbreviations F_I, where $I \in [1, 9]$. For Qualitative Attributes, the Initial Most Probable Value (MPV) is Specified. Not Applicable for Quantitative Attributes

F	Attribute label	Attribute category	Initial MPV (VL, L, M, H, VH)
f_1	GD	Qualitative	(0.1, 0.2, 0.2, 0.2, 0.2)
f_2	RE	Qualitative	(0.1, 0.2, 0.3, 0.4, 0.9)
f_3	PE	Qualitative	(0.2, 0.3, 0.4, 0.5, 0.9)
f_4	LU	Qualitative	(0.2, 0.3, 0.5, 0.7, 0.9)
f_5	TP	Qualitative	(0.3, 0.4, 0.5, 0.6, 0.7)
f_6	MS	Quantitative	Not applicable
f_7	RS	Quantitative	Not applicable
f_8	WS	Quantitative	Not applicable
f_9	DG	Quantitative	Not applicable

At this step, fuzzy conversion is performed using Definition 3 to determine the Most Probable Value (MPVs) for each qualitative property. In practice, the values are mapped to the interval [0,1] to represent the corresponding linguistic terms. Then, for each numerical value, the corresponding fuzzy sets are determined, and the MPVs are computed for the defined linguistic terms. The following Table IV represents the computed final MPVs for the qualitative properties.

At this stage, the obtained MPV values are used to define the membership function for each family f_i. Fig. 3 illustrates the resulting membership-function shape, which is used to determine the preference function associated with each vector of f_i. The universe of discourse corresponds to the computed MPVs for each vector associated with f_i.

TABLE IV. COMPUTATION OF THE FINAL MPVS FOR THE QUALITATIVE ATTRIBUTES F1, F2, F3, F4 AND F5

F	Desc.	Linguistic terms				
		VL	L	M	H	VH
f_1	Fuzzy Sets	(0 ,0.1 ,0.2)	(0.15,0.2,0.3)	(0.15,0.2,0.3)	(0.15,0.2,0.3)	(0.15,0.2,0.3)
	MPVs	**0.1**	**0.217**	**0.217**	**0.217**	**0.417**
f_2	Fuzzy Sets	(0, 0.1, 0.2)	(0.15,0.2,0.3)	(0.25,0.3,0.4)	(0.35,0.4,0.5)	(0.85,0.9,0.9)
	MPVs	**0.1**	**0.217**	**0.317**	**0.417**	**0.883**
f_3	Fuzzy Sets	(0, 0.2, 0.3)	(0.25,0.3,0.4)	(0.35,0.4,0.5)	(0.45,0.5,0.6)	(0.85,0.9,0.9)
	MPVs	**0.167**	**0.317**	**0.417**	**0.517**	**0.883**
f_4	Fuzzy Sets	(0, 0.2, 0.3)	(0.25,0.3,0.4)	(0.45,0.5,0.6)	(0.65,0.7,0.8)	(0.85,0.9,0.9)
	MPVs	**0.167**	**0.317**	**0.517**	**0.717**	**0.883**
f_5	Fuzzy Sets	(0, 0.3, 0.4)	(0.35,0.4,0.5)	(0.45,0.5,0.6)	(0.55,0.6,0.7)	(0.85,0.9,0.9)
	MPVs	**0.233**	**0.417**	**0.517**	**0.617**	**0.750**

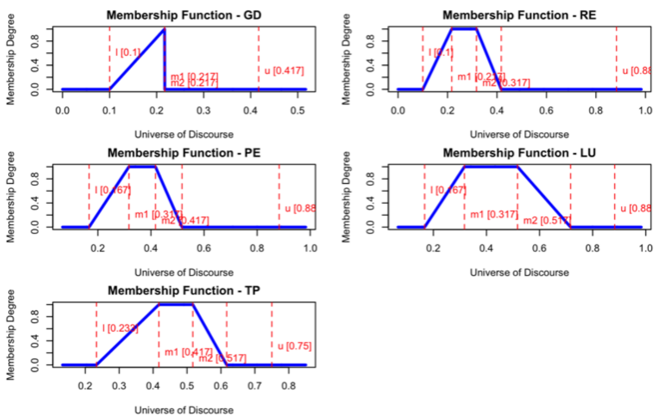


Fig. 3. The graphical illustration of the membership degree for the five qualitative attributes f_1 = GD, f_2 = RE, f_3 = PE, f_4 = LU, f_5 = TP.

Based on the membership degrees illustrated above, the preference function associated with f_1 is modeled using triangular membership $\psi(1)$. In contrast, the preference functions associated with f_2, f_3, f_4, and f_5 are modeled using trapezoidal membership functions $\psi(2)$. The corresponding fuzzy numbers M_i are computed as follows: $M_i = \{M_1, M_2, M_3, M_4, M_5\}$, with M_1=(0, 0.22, 0.22), M_2=(0.22, 0.32, 0.42, 0.88), M_3=(0.17, 0.32, 0.47, 0.88), M_4=(0.17, 0.52, 0.72, 0.88), M_5=(0.23, 0.52, 0.62).

Therefore, the decision matrix is constructed using the preference function P_i defined in Eq. (5), which is then used to determine the utility function U in Eq. (7.

Following the multi-criteria decision-making (MCDM) process, the preferences associated with each vector f_i, are determined to construct the corresponding pairwise comparison matrix [23]. Consequently, the resulting weights vector Θ is given as follows: $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9\}$, with θ_1=0.029, θ_2=0.019, θ_3=0.042, θ_4=0.027, θ_5=0.060, θ_6=0.400, θ_7=0.175, θ_8=0.223, and θ_9=0.025.

At this step, the utility function is used to determine the weights for each vector f_i, and to initialize the centroids for the modified k-means algorithm (Algorithm 1). The next section presents the results obtained by applying the proposed fuzzy approach to the selected network datasets.

B. Results and Observations

This section presents the experimental results obtained using the proposed fuzzy logic approach on six networks, as listed in Table II. A comparative analysis is then presented to highlight the main findings and performance differences across the evaluated networks. The proposed community detection approach begins by determining the optimal value k using the Elbow method [24].

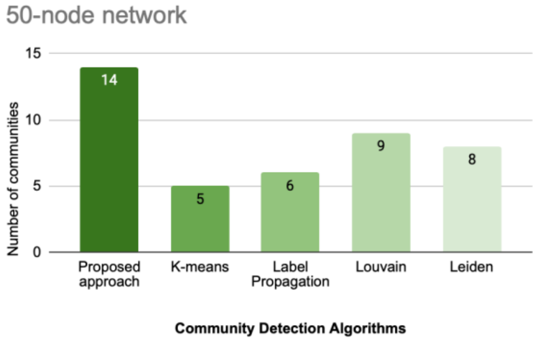
To assess the stability and computational behavior of the proposed approach, each experiment was repeated 10 times. Table V reports the mean execution time and standard deviation, together with the average number of detected communities.

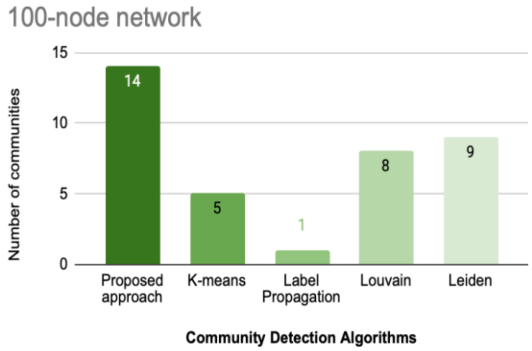
TABLE V. COMPUTATIONAL PERFORMANCE OVER 10 INDEPENDENT RUNS

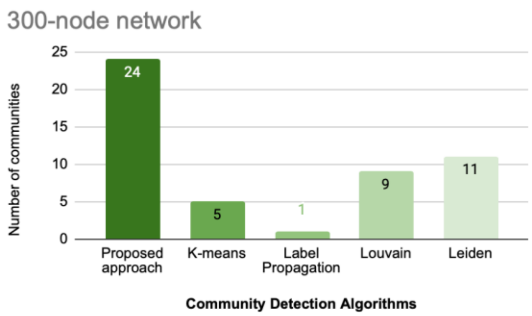
Network size	Mean Execution time (s)	Standard Deviation SD (s)	Number of communities (Mean ± SD)
50-nodes	0.000434	0.000142	10 ± 0.00
100-nodes	0.000320	0.000045	9 ± 0.00
300-nodes	0.000443	0.000054	5 ± 0.00
1000-nodes	0.001109	0.000799	5 ± 0.00
3000-nodes	0.002156	0.000295	5 ± 0.00
10000-nodes	0.015916	0.001404	5 ± 0.00

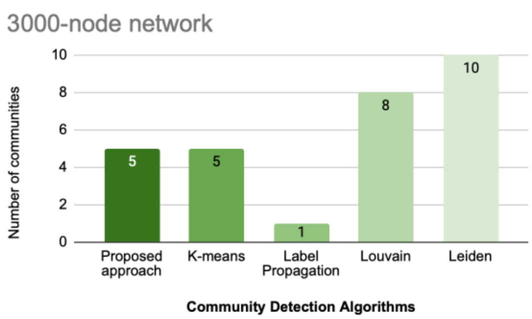
The results obtained show low execution times across all evaluated network sizes, increasing from 0.000434 seconds for the 50-node network to 0.015916 seconds for the 10000-node network. The relatively small standard deviation indicates stable execution time across repeated runs. Moreover, the number of detected communities remains unchanged across 10 runs for each network size, resulting in a standard deviation of zero. This provides stable computational behavior while maintaining consistent community partitions across repeated executions.

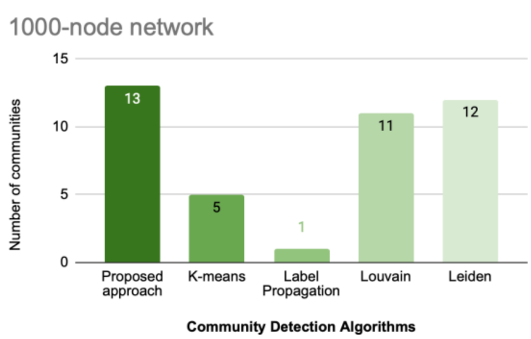
Fig. 4 illustrates a bar chart showing the number of detected communities obtained by the proposed approach and the selected classical community detection algorithms. Therefore, the proposed approach achieves competitive performance across networks of different sizes compared with the evaluated baseline algorithms. Fig. 5 further highlights the comparative execution times over the different algorithms across the considered network sizes.











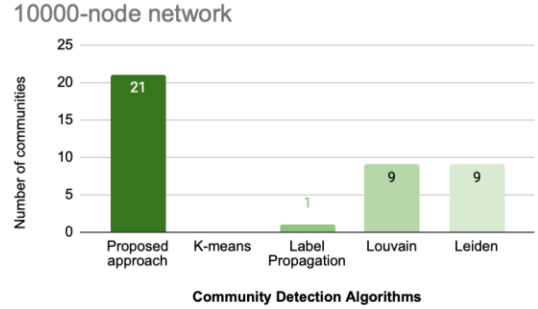


Fig. 4. Graphical representation of the fuzzy approach application, highlighting both the total number of detected communities and the comparison insights with the traditional clustering algorithms over different network sizes.

The results demonstrate that the proposed fuzzy logic approach achieves competitive execution times across different network sizes. For small networks (50-node and 100-node networks), the proposed approach achieves an execution time of 0.22 seconds, compared with 0.45-0.5 seconds for the standard k-means algorithm. While Label Propagation, Leiden, and Louvain also demonstrate competitive performance on small networks, their execution time increases as the network size grows. For the large network (10000 nodes), the proposed approach maintains stable performance with an execution time of approximately 0.40 seconds. Whereas the k-means algorithm times out. While Louvain, Leiden, and Label Propagation algorithms require longer execution times. Overall, these results demonstrate the computational efficiency and scalability of the proposed approach, particularly as network size increases.

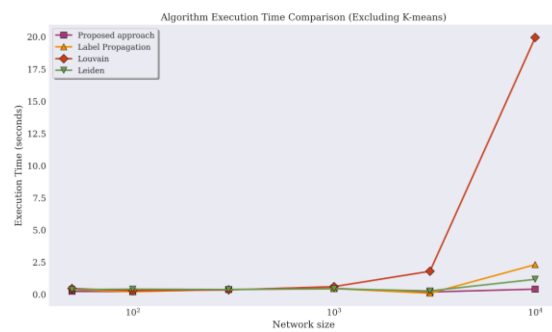


Fig. 5. Time consumed for the comparison of community detection algorithms across different network sizes (excluding the k-means algorithm due to computational instability at 10000-node network size).

The results demonstrate that the proposed fuzzy logic approach achieves competitive execution times across different network sizes. For small networks (50-node and 100-node networks), the proposed approach achieves an execution time of 0.22 seconds, compared with 0.45-0.5 seconds for the standard k-means algorithm. While Label Propagation, Leiden, and Louvain also demonstrate competitive performance on small networks, their execution time increases as the network size grows. For the large network (10000 nodes), the proposed approach maintains stable performance with an execution time of approximately 0.40 seconds. Whereas the k-means algorithm reaches a timeout. While Louvain, Leiden, and Label Propagation algorithms require longer execution times. Overall, these results demonstrate the computational efficiency and scalability of the proposed approach, particularly as network size increases.

V. Conclusion

This study presents a fuzzy logic–based approach for community detection that incorporates the semantic attributes of nodes in complex networks. Unlike classical community detection algorithms that primarily rely on topological metrics (e.g., centrality, clustering coefficients, or modularity), the proposed approach focuses on semantic properties of nodes to better capture complementary information that may be overlooked by topology-based approaches. By integrating fuzzy logic with multi-criteria decision-making techniques [19], the proposed approach addresses the uncertainty and ambiguity associated with semantic attributes. Furthermore, the proposed approach extends the standard k-means algorithm by adapting it to large-scale networks.

Experimental results show that the proposed method achieves competitive performance across the evaluated community-quality metrics compared with five well-known community detection algorithms on large-scale networks. The approach identifies meaningful and balanced community structures within a reasonable execution time. Overall, these findings highlight the efficiency and adaptability of the proposed fuzzy logic approach in environments characterized by uncertainty, high-dimensional attributes, and large network sizes.

Despite these promising results, some limitations remain to be addressed in future work. The current evaluation is based on synthetic network datasets, and further validation on longer and more diverse real-world attributed networks is required. Future work will extend the evaluation using standard community-quality metrics, including Modularity, NMI, ARI, and Silhouette Score, where applicable, while investigating adaptive fuzzy membership and MCDM weighting mechanisms, richer semantic representations, and real-world network structures.

References

[1] Dey, A.K., Tian, Y. and Gel, Y.R., Community detection in complex networks: From statistical foundations to data science applications. Wiley Interdisciplinary Reviews: Computational Statistics, 14(2), p.e1566. (2022)

[2] Ruggeri, N., Contisciani, M., Battiston, F. and De Bacco, C., Community detection in large hypergraphs. Science Advances, 9(28), p.eadg9159. (2023)

[3] Li, J., Lai, S., Shuai, Z., Tan, Y., Jia, Y., Yu, M., Song, Z., Peng, X., Xu, Z., Ni, Y. and Qiu, H. A comprehensive review of community detection in graphs. Neurocomputing, 600, p.128169. (2024)

[4] Newman, M.E. and Girvan, M., Finding and evaluating community structure in networks. Physical review E, 69(2), p.026113. (2004)

[5] Newman, M.E. and Girvan, M., Mixing patterns and community structure in networks. In Statistical mechanics of complex networks (pp. 66-87). Berlin, Heidelberg: Springer Berlin Heidelberg. (2003)

[6] George, R., Shujaee, K., Kerwat, M., Felfli, Z., Gelenbe, D. and Ukuwu, K., A comparative evaluation of community detection algorithms in social networks. Procedia Computer Science, 171, pp.1157-1165. (2020)

[7] Latifah, R. and Adriani, M., Combining topological and topical features for community detection. In 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS) (pp. 327-334). IEEE. (2016)

[8] Tsitsulin, A., Palowitch, J., Perozzi, B. and Müller, E., Graph clustering with graph neural networks. Journal of Machine Learning Research, 24(127), pp.1-21. (2023)

[9] Peddi, P.B.R. and Abdullah, L., Nearest Triangular Fuzzy Quantity Approximation of Trapezoidal Fuzzy Number with Vagueness in Lower Decision Level for Ranking Fuzzy Numbers. Malaysian Journal of Mathematical Sciences, 19(2). (2025)

[10] Mondal, R., Ignatova, E., Walke, D., Broneske, D., Saake, G. and Heyer, R., Clustering graph data: the roadmap to spectral techniques. Discover Artificial Intelligence, 4(1), p.7. (2024)

[11] Stoklasa, J., Luukka, P. and Stoklasová, J., Generalized dissemblance index as a difference of first moments of fuzzy numbers–A new perspective on the distance of fuzzy numbers. Information Sciences, 660, p.120118. (2024)

[12] Javed, M.A., Younis, M.S., Latif, S., Qadir, J. and Baig, A., Community detection in networks: A multidisciplinary review. Journal of Network and Computer Applications, 108, pp.87-111. (2018)

[13] Shahin, M., Bertl, M., Iman, M.R.H., Ghasempouri, T., Sharma, R., Shah, S.A. and Draheim, D., Significant factors extraction: A combined logistic regression and apriori association rule mining approach. In Computer Science On-line Conference (pp. 295-311). Cham: Springer Nature Switzerland. (2024)

[14] Saha, U., Hossain, S.M.M. and Sarker, I.H. Predicting depression level based on human activities and feelings: A fuzzy logic-based analysis. Data Science and Management. (2024)

[15] Krömer, Pavel, et al.: Fuzzy classification by evolutionary algorithms. IEEE International Conference on Systems, Man, and Cybernetics. IEEE, (2011).

[16] Afridi, M.K., Azam, N. and Yao, J., Variance based three-way clustering approaches for handling overlapping clustering. International Journal of Approximate Reasoning, 118, pp.47-63. (2020)

[17] Li, M., Yang, Z., Zhou, X., Fang, Y., Li, K. and Li, K., Clustering on attributed graphs: From single-view to multi-view. ACM Computing Surveys, 57(7), pp.1-36. (2025)

[18] Ding, Y.: Community detection: Topological vs. topical. Journal of Informetrics 5(4), 498–514 (2011)

[19] Amorocho-Daza, H., Sušnik, J., van der Zaag, P. and Slinger, J.H., A model-based policy analysis framework for social-ecological systems: Integrating uncertainty and participation in system dynamics modelling. Ecological Modelling, 499, p.110943. (2025)

[20] Paramanik, A.R., Sarkar, S. and Sarkar, B., OSWMI: An objective-subjective weighted method for minimizing inconsistency in multi-criteria decision making. Computers & Industrial Engineering, 169, p.108138. (2022)

[21] Gao, H., Han, J., Liu, Y., Zhang, P., Yang, B., Zu, Y., Liu, F. and Qian, Y., K-Means Community Detection Algorithm Based on Density Peaks. Entropy, 28(2), p.152. (2026)

[22] Pant, S., Kumar, A., Ram, M., Klochkov, Y. and Sharma, H.K., Consistency indices in analytic hierarchy process: a review. Mathematics, 10(8), p.1206. (2022)

[23] Cavallo, B. and Ishizaka, A., A comparative study on precision of direct evaluations, the pairwise comparisons method and the best-worst method. Omega, 130, p.103175. (2025)

[24] El-Moussaoui, M., et al., A k-mean classification study of eight community detection algorithms: application to synthetic social network datasets. AI and IoT for Sustainable Development in Emerging Countries: Challenges and Opportunities. Cham: Springer International Publishing, 557-572. (2022)

[25] Bucheli, V.A., Pabón, O.S. and Ordoñez, H., Evolutionary algorithms guided by Erdős–Rényi complex networks. PeerJ Computer Science, 10, p.e1773. (2024)

[26] Hanine, M., Boutkhoum, O., Agouti, T. and Tikniouine, A., A new integrated methodology using modified Delphi-fuzzy AHP-PROMETHEE for Geospatial Business Intelligence selection. Information Systems and e-Business Management, 15(4), pp.897-925. (2017).