Grenze International Journal of Engineering and Technology, June Issue

Homophily-based Community Formation in Social Networks using Fuzzy Logic and Decision-Making

Manoj Kumar Srivastav[1], Somsubhra Gupta[2] and Subhranil Som[3]
[1,2]School of Computer Science, Swami Vivekananda University, Barrackpore, West Bengal, India,
Email: mksrivastav2015@gmail.com, gsomsubhra@gmail.com
[3]Department of Computer Science, Bhairab Ganguly College, Kolkata, West Bengal, India,
Email: subhranil.som@gmail.com

***Abstract*—Social networks have become integral to modern life, significantly influencing friendship formation and interpersonal interactions. This study focuses on homophily, which means people tend to connect with others who share similar interests. Homophily plays a key role in forming and shaping communities. The proposed methodology introduces an approach that integrates fuzzy logic and decision-making to discuss a framework for formation of homophily in user networks. Pairwise similarity scores are calculated using a user-interest matrix and the Jaccard index, while fuzzy logic categorizes these scores into levels like low, medium, and high. Decision-making thresholds are applied to form communities and assign users based on aggregated homophily scores. This method helps analyze how social networks are structured and suggests ways to create new connections within communities.**

***Keywords*— Social networks, Homophily, Fuzzy logic, Decision-making, User-interest matrix, Jaccard index, Community formation, Similarity scores.**

I. INTRODUCTION

Social networks such as Facebook, Instagram, and Twitter (now X) have become integral to modern communication, serving as platforms for individuals to connect, share information, and form diverse communities. These platforms provide spaces for a wide range of interactions, where relationships are often influenced by shared interests and common attributes.

Over time, users tend to connect with those who share similar interests, giving rise to the concept of homophily, which refers to the tendency of individuals to form connections with others who are similar to themselves in various ways. Homophily plays an important role in shaping communities within social networks, as it facilitates the formation of ties based on shared preferences, behaviors, and characteristics. Understanding and modeling homophily in online interactions is still challenging, even with social network analysis (SNA) methods. User behavior is complex, and social media relationships are dynamic and often overlap. Traditional models don't capture these complexities, making it difficult to analyze homophily accurately. This highlights the need for more flexible methods to understand homophily and community formation. Overlapping relationships occur when users are connected to multiple people or groups with shared characteristics but also have different interests. For example, a user may belong to several interest groups that overlap in some areas but differ in others.

Grenze ID: 01.GIJET.11.2.649

In this study, fuzzy logic is used to classify similarity scores between users into overlapping categories like "low," "medium," and "high." Instead of labeling users as either similar or not, fuzzy logic allows them to belong to more than one category at once. This reflects the complexity of real-world relationships, where people share multiple interests and belong to different communities. Using fuzzy membership functions, similarity scores can overlap, giving a more detailed view of connections. For example, a user could be both "medium" and "high" in similarity to another, based on overlapping interests. This improves community detection and helps better understand how users form connections with shared and diverse interests. This study proposes a framework that integrates fuzzy logic and decision-making techniques to analyze and foster homophily in social networks. The framework employs a user-interest matrix to calculate pairwise similarity scores using the Jaccard index, and then applies fuzzy logic to categorize these scores into different levels (low, medium, and high). By incorporating decision-making thresholds based on aggregated homophily scores, the methodology facilitates more accurate community detection and enhances the understanding of how relationships are structured in social networks. This approach offers a valuable tool for identifying new potential connections within communities, based on shared interests and behaviors.

A. Contributions of the Research

This research contributes to the field of social network analysis in several ways:

- Introducing a fuzzy logic-based method to classify user similarity scores into overlapping categories.
- Establishing decision-making thresholds for community formation based on aggregated homophily scores.

B. Organization of the Paper

The remainder of this paper is structured as follows:

- Section 2 reviews related works in social network analysis, focusing on the concept of homophily and the application of fuzzy logic.
- Section 3 details the proposed methodology, including similarity calculations, fuzzy logic classification, and the decision-making process for community formation.
- Section 4 presents the experimental results, followed by a discussion on community formation and the effectiveness of the proposed framework.
- Section 5 concludes the study by summarizing key findings and suggesting potential areas for future research.

II. LITERATURE REVIEW

A. Homophily in Social Networks

Homophily, the tendency of individuals to connect with others who share similar attributes, has long been a central concept in social network analysis (SNA)[13]. Early research, such as that by McPherson, Smith-Lovin, and Cook [8], demonstrated that individuals are more likely to form connections with others who are similar in terms of demographics, interests, and behaviors. These studies have highlighted homophily's significant role in shaping social network structures, particularly in community formation. It leads to the emergence of subgroups or clusters of like-minded individuals within larger networks.

Khanam et al. [3] provide a comprehensive survey of homophily in social network analysis, emphasizing its importance in understanding user behavior and community formation. Their research illustrates how homophily influences not only direct connections but also the broader network structure through indirect relationships and overlapping communities. The formation of social ties through homophily occurs via both choice and induced mechanisms.

Kossinets and Watts [4] distinguish between these two types: choice homophily, where users actively seek connections with similar others, and induced homophily, which arises due to structural constraints and opportunities present in the network. This distinction is crucial for understanding the organic formation and evolution of communities over time.

B. Fuzzy Logic Applications in Social Network Analysis

Recent advances in fuzzy logic applications have introduced more sophisticated ways to model the inherent uncertainty and the gradual nature of social relationships. Traditional binary similarity measurements have been replaced with more nuanced methods that capture varying degrees of connection strength and interest overlap between users. Bisgin et al. [2] demonstrated that fuzzy logic approaches are particularly useful in modeling social media relationships, providing a more accurate representation of shared interests and behavioral patterns.

Fuzzy logic offers an advanced way to represent uncertainty and overlapping categories in social networks. In the context of homophily, fuzzy logic classifies user similarity scores into multiple overlapping categories, such as low, medium, and high, rather than using the traditional binary classification of similar or dissimilar.The integration of fuzzy logic with community detection in social networks has been explored by several researchers. Langari et al. [5] proposed a fuzzy clustering approach that detects communities in social networks, where fuzzy membership functions classify users into multiple communities based on their connections. This approach allows users to belong to more than one community, reflecting the overlapping nature of social groups.

C. Community Detection and Formation

Community detection has greatly benefited from the integration of fuzzy logic and decision-making frameworks. Lewis [6] discusses how modern network analysis has moved beyond merely identifying patterns to understanding the underlying mechanisms of community formation.

D. Decision-Making in Community Formation

The application of decision-making frameworks in community formation offers a systematic way to understand how and why users join specific groups. Bisgin et al. [1] explored how decision-making processes influence community emergence in online social networks, highlighting the impact of both explicit user preferences and implicit behavioral patterns.

Incorporating decision-making rules into community formation has been a growing area of research. Zhang et al. [10] demonstrated how fuzzy decision-making rules could classify users into communities based on aggregated similarity scores, enhancing the accuracy of community detection. Moreover, Wu and Xu [9] proposed a decision-making framework for dynamic networks, where relationships evolve over time. Their study integrated fuzzy logic and decision rules, allowing the framework to adapt to the changing nature of social connections, making it suitable for dynamic environments such as social media platforms.

III. METHODOLOGY

A. Fuzzy logic approach

Step 1: Membership value calculation

Objective : Assign membership values(μ) to different fuzzy sets(e.g., Low, Medium, High) for a given similarity score

Defining fuzzy set

Low similarity (μ_{Low}) :

$$\mu_{Low}(S) = \begin{cases} 1 - 2S, & 0 \leq S \leq 0.5 \\ 0 & S > 0.5 \end{cases}$$

Medium Similarity (μ_{medium})

$$\mu_{medium}(S) = \begin{cases} 2S - 0.6, & 0.3 \leq S \leq 0.6 \\ 1.4 - 2S, & 06 < S \leq 0.8 \\ 0 & \text{otherwise} \end{cases}$$

High Similarity(μ_{High})

$$\mu_{High}(S) = \begin{cases} 2S - 1.4\,, & 0.7 \leq S \leq 1 \\ 0 & S < 0.7 \end{cases}$$

Step 2: Fuzzification

Objective: Convert the crisp similarity score (S) into fuzzy membership values.

Process:

- For a given S, calculate membership in all fuzzy sets using the defined membership functions.

- ➢ Represent S as overlapping memberships across the sets

Step 3: Inference: Apply fuzzy rules to determine the degree of homophily based on membership values.
Fuzzy Rules:

- ➢ "IF Similarity is Low, THEN Homophily is Weak."
- ➢ "IF Similarity is Medium, THEN Homophily is Moderate."
- ➢ "IF Similarity is High, THEN Homophily is Strong."

Step 4: Defuzzification

Objective: Convert fuzzy outputs into a single crisp homophily score.

Defuzzification Methods:

1. **Centroid Method** (Weighted Average):
   Combine the outputs of all fuzzy rules:

$$H_{\text{crisp}} = \frac{\sum(\mu_i \cdot c_i)}{\sum \mu_i}$$

- μ_i: Membership value for a fuzzy set.
- c_i: Centroid (crisp value) for that fuzzy set.

Figure 1

B. Dataset overview

This methodology provides a systematic approach to analyzing homophily in social networks by combining similarity measures and fuzzy logic. It allows for the classification of user relationships, aiding in community detection and decision-making for social network analysis.https://www.kaggle.com/datasets/arindamsahoo/social-media-users]

TABLE 1: DATASET OVERVIEW [12]

UserID	Gender	DOB	Interests	City	Country
1	Female	15-10-1958	'Movies', 'Fashion', 'Fashion', 'Books'	Sibolga	Indonesia
2	Female	21-07-2004	'Gaming', 'Finance and investments', 'Outdoor activities', 'Travel'	Al AbyÄ r	Libya
3	Female	07-02-2000	'DIY and crafts', 'Music', 'Science', 'Fashion'	WÄ dÄ« as SÄ«r	Jordan
999	Female	04-08-1983	'Photography', 'Sports', 'Gaming'	Haitang	China
1000	Female	17-08-1990	'Photography', 'Gardening', 'Gaming', 'Business and entrepreneurship'	Brunswick	Australia

TABLE II: USER INTEREST MATRIX

$$\begin{array}{c|ccccccccccc} & I1 & I2 & I3 & I4 & I5 & \dots & I25 & I26 & I27 & I28 & I29 \\ \hline U1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 1 \\ U2 & 0 & 0 & 1 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 \\ U3 & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 & 0 & 0 \\ U4 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ U99 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ U100 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 0 \end{array}$$

TABLE III: SIMILARITY SCORE AND FUZZY CATEGORY

Sample Results (first 10 rows):

	User1	User2	Jaccard Similarity	Fuzzy Category
0	u0	u1	0.000	Low
1	u0	u2	0.167	Low
2	u0	u3	0.000	Low
3	u0	u4	0.000	Low
4	u0	u5	0.000	Low
5	u0	u6	0.250	Medium
6	u0	u7	0.000	Low
7	u0	u8	0.000	Low
8	u0	u9	0.000	Low
9	u0	u10	0.000	Low

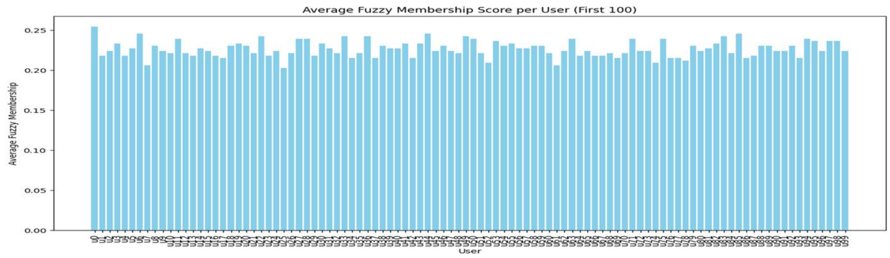

Figure 2: Average Fuzzy Memberships Score Per User

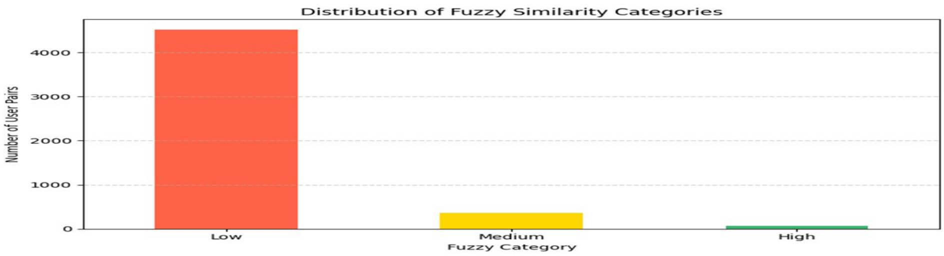

Figure 3: Distribution of Fuzzy Similarity Categories

C. Decision-Making approach for Homophily

The decision-making approach for homophily aims to identify and analyze user relationships in a social network based on shared interests. By leveraging similarity measures and fuzzy logic, this method helps determine community formation and user interactions within a user-interest matrix.

Steps in the Decision-Making Approach

Step 1: Compute Similarity Scores

To assess user relationships, we calculate the Jaccard similarity between users based on their interests.

The Jaccard similarity between two users, **ui** and **uj**, is calculated using the formula:

$$\mathrm{Sim}(ui, uj) = \frac{|I(ui) \cap I(uj)|}{|I(ui) \cup I(uj)|}$$

Where:

- **I(ui)** is the set of interests of user **ui**.
- **I(uj)** is the set of interests of user **uj**.
- **|I(ui) ∩ I(uj)|** is the number of shared (common) interests.
- **|I(ui) ∪ I(uj)|** is the total number of unique interests from both users.

This value ranges from 0 to 1, where 0 means no common interests and 1 means identical interests.

Figure 3

Step 2: Fuzzy Logic-Based Categorization

Once the similarity scores are computed, they are categorized into fuzzy sets to classify the strength of user relationships. The categories are defined as follows:

TABLE IV: FUZZY LOGIC BASED CATEGORIZATION

Similarity Score	Fuzzy Category
S≤0.2	Low Similarity
0.2<S≤0.5	Medium Similarity
S>0.5	High Similarity

Step 3: Apply Fuzzy Rules for Decision Making

Fuzzy inference rules are applied based on the similarity scores:

- ➢ If similarity is High, then users are strongly connected and likely belong to the same community.
- ➢ If similarity is Medium, then users have a weak connection and may be in overlapping communities.
- ➢ If similarity is Low, then users are not connected and belong to different communities.

Step 4: Decision Outcome for Community Formation
Based on the fuzzy logic classification, the decision-making approach determines user grouping and network formation:
Users with high similarity are grouped into the same cluster.
Users with medium similarity may belong to overlapping clusters.
Users with low similarity remain separate.
Example of the Decision-Making Approach
Consider three users: u1, u2, and u3 with the following interest sets:
I(u1)={A,B,C,D}
I(u2)={A,B,E,F}
I(u3)={G,H,I}
Step 1: Compute Jaccard Similarity

- **Similarity between u1 and u2:**
$$\mathrm{Sim}(u1, u2) = \frac{|\{A,B\} \cap \{A,B,C,D,E,F\}|}{|\{A,B\} \cup \{A,B,C,D,E,F\}|} = \frac{2}{6} = 0.33$$
- **Similarity between u1 and u3:**
$$\mathrm{Sim}(u1, u3) = \frac{|\{A,B,C,D,E,F\} \cap \{G,H,I\}|}{|\{A,B,C,D,E,F\} \cup \{G,H,I\}|} = \frac{0}{9} = 0.00$$
- **Similarity between u2 and u3:**
$$\mathrm{Sim}(u2, u3) = \frac{|\{A,B,E,F\} \cap \{G,H,I\}|}{|\{A,B,E,F\} \cup \{G,H,I\}|} = \frac{0}{7} = 0.00$$

Step 2: Categorize the Similarity Scores
Sim(u1,u2)=0.33 → Medium Similarity
Sim(u1,u3)=0.00 → Low Similarity
Sim(u2,u3)=0.00 → Low Similarity
Step 3: Apply Fuzzy Rules
u1 and u2 → Medium Similarity → Weak connection may
u1 and u3 → Low Similarity → No connection.
u2 and u3 → Low Similarity → No connection.
Step 4: Decision Outcome
u1 and u2 might form a weak connection.
u3 remains isolated.
Community formation groups u1 and u2, while u3 is separate.

IV. RESULT AND DISCUSSION

A. Jaccard Similarity Computation

Jaccard similarity was used to measure how many interests users have in common. It was calculated for all pairs among the first 100 users. The score shows the ratio of shared interests to the total number of unique interests. For example, users u0 and u2 had 1 shared interest out of 6 total, giving a similarity of 0.167. These scores help to understand how closely users are connected based on their interests.

B. Fuzzy Categorization of Similarity

To better understand the similarity scores, fuzzy logic was used to divide them into three categories: Low (≤ 0.2), Medium (0.2 to 0.5), and High (> 0.5). This makes it easier to describe how strong the connection is between two users. For instance, the similarity between u0 and u2 is 0.167, which is Low, while u0 and u6 have a similarity of 0.250, which is Medium. Most user pairs fall in the Low category, showing a wide variety of interests.

C. Application of Fuzzy Rules

Fuzzy rules were applied to draw conclusions from the similarity categories. High similarity means strong connections and users likely belong to the same group. Medium similarity shows weak connections with possible overlaps in groups.
Low similarity means no connection and users are likely in different groups. For example, u1 and u2 had Medium similarity, while u1 and u3 had Low similarity, showing weaker or no links between them.

D. Decision-Making for Community Formation

Based on fuzzy similarity, users with High similarity are grouped into close communities. Users with Medium similarity may belong to overlapping or loose groups. Those with Low similarity stay separate. Since most users had Low similarity, it shows that the network has diverse interests. However, even Medium similarity cases can be starting points for forming groups using flexible methods like fuzzy or hierarchical clustering.

VI. Conclusion

This study proposed a fuzzy logic-based decision-making framework for homophily-based community formation in social networks. By leveraging the Jaccard similarity index to compute pairwise similarity scores and applying fuzzy logic for classification, the approach effectively captures the gradual transitions in similarity among users. Unlike traditional hard-threshold methods, fuzzy logic allows for more flexible and realistic classification of user relationships into low, medium, and high similarity categories.The experimental results demonstrated that users with high similarity scores tend to form stronger communities, while users with medium similarity scores may belong to overlapping clusters, and those with low similarity scores remain largely unconnected. This confirms that homophily plays a significant role in shaping social network structures.

A. Key Contributions

- Developed a fuzzy logic-based framework to analyze homophily in social networks.
- Established a decision-making approach for dynamic community formation based on similarity scores.
- Provided a systematic method for classifying user relationships, enhancing community detection accuracy.

B. Future Work

- Scalability Assessment: Implementing the approach on larger, real-world social networks to validate performance.
- Dynamic Homophily Modeling: Adapting the method to account for evolving user interests over time.
- Integration with Machine Learning: Exploring hybrid models combining fuzzy logic with machine learning for enhanced prediction and analysis.

By refining the decision-making framework, this methodology offers a valuable tool for understanding and structuring online social communities, contributing to the broader field of social network analysis and recommendation systems.

References

[1] H. Bisgin, N. Agarwal, and X. Xu, "Investigating homophily in online social networks," in Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, ON, Canada, 2010, pp. 533–536.

[2] H. Bisgin, N. Agarwal, and X. Xu, "A study of homophily on social media," World Wide Web, vol. 15, no. 2, pp. 213–232, 2012, doi: 10.1007/s11280-011-0143-3.

[3] K. Z. Khanam, G. Srivastava, and V. Mago, "The homophily principle in social network analysis: A survey," Multimedia Tools and Applications, vol. 82, pp. 8811–8854, 2023, doi: 10.1007/s11042-021-11857-1.

[4] G. Kossinets and D. J. Watts, "Origins of homophily in an evolving social network," American Journal of Sociology, vol. 115, no. 2, pp. 405–450, 2009, doi: 10.1086/599247.

[5] R. K. Langari, S. Sardar, S. A. A. Mousavi, and R. Radfar, "Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks," Expert Systems with Applications, vol. 141, p. 112968, 2020, doi: 10.1016/j.eswa.2019.112968.

[6] K. Lewis, "How networks form: Homophily, opportunity, and balance," Emerging Trends in the Social and Behavioral Sciences, pp. 1–15, 2015, doi: 10.1002/9781118900772.etrds0164.

[7] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 6, pp. 2094–2107, 2014, doi: 10.1109/JSTARS.2014.2329330.

[8] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," Annual Review of Sociology, vol. 27, pp. 415–444, 2001, doi: 10.1146/annurev.soc.27.1.415.

[9] H. Wu and Z. S. Xu, "Fuzzy logic in decision support: Methods, applications, and future trends," International Journal of Computers Communications & Control, vol. 16, no. 1, Art. no. 4044, 2021, doi: 10.15837/ijccc.2021.1.4044.

[10] X. Zhang, D. Chen, and J. Mi, "Fuzzy decision rule-based online classification algorithm in fuzzy formal decision contexts," IEEE Transactions on Fuzzy Systems, vol. 31, no. 9, pp. 3263–3277, 2023, doi: 10.1109/TFUZZ.2023.3250955.

[11] M. Zhang and Y. Chen, "Link prediction based on graph neural networks," in Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS), Montréal, Canada, 2018, pp. 5171–5181.
[12] A. Sahoo, "Social media users," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/arindamsahoo/social-media-users
[13] M. K. Srivastav, S. Gupta, and S. Som, "Analyzing interest-based homophily in online social networks using community detection methods," Educational Administration: Theory and Practice, vol. 30, no. 1, pp. 6258–6278, 2024, doi: 10.53555/kuey.v30i1.9576.
[14] M. K. Srivastav, S. Gupta, and S. Som, "Modeling friendship dynamics in social networks: A graph-theoretic and computational approach," Journal of Information Systems Engineering and Management, vol. 10, no. 37s, 2024. [Online]. doi: 10.52783/jisem.v10i37s.6382.