Universiteit Leiden

Master Computer Science

From Global Network Attributes to Temporal Motifs: Predicting User Retention in the Sarafu Network

Name: Sadaf Esmaeili Rad
Student ID: S3986160

Date: 28/08/2025

Specialisation: Data Science

1st supervisor: Dr. Frank Takes

2nd supervisor: Dr. Carolina Mattson

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

This thesis investigates a community currency transaction network, focusing on structural and behavioral features that influence sustained user engagement. Specifically, it examines how early network signals relate to long-term user retention, defined here as continued participation over time, within Sarafu, a mobile-based community currency used in Kenya during the COVID-19 period. The analysis is based on a transaction dataset enriched with user demographics. Over time, we studied network properties at macro, meso, and micro levels.
Results show that users who remain active are not necessarily those with the highest transaction volumes or most central global positions. Instead, retention is more common among users, particularly those with high transaction patterns. Overall, the findings suggest that user retention in community currencies depends more on interaction patterns within local communities than on global structural importance.

Acknowledgements

I would like to sincerely thank my supervisors, for their guidance and support throughout this thesis. their clear explanations, thoughtful feedback, and critical perspective taught me a lot and made me more interested in research than I expected at the start.
I also want to thank the professors and staff at the Leiden Institute of Advanced Computer Science. Their teaching and advice over the past two years have played an important role in my academic journey and helped me grow both technically and personally.
Lastly, I am deeply grateful to my family for their constant encouragement during this time. Without their support, finishing this journey would not be possible.

Contents

1 Introduction

Currently the flow of money happens in different ways, from traditional methods of transferring money to digital currencies. Especially in times when the world is facing a crisis such as the COVID-19 pandemic or in communities where the economy is poor, community currencies such as Sarafu may help those regions' economic resilience. Sarafu is a digital community currency used in Kenya during 2020 and 2021. In this community currency system, users transact with each other without needing access to traditional banks, just through their mobile phones. Because the system recorded detailed transaction data alongside user demographics, it provides a valuable opportunity to analyze how money flows within a grassroots economy. More importantly, it allows us to identify influential users in the network who help keep the system active. Understanding why some users stay active while others leave is crucial for evaluating the long-term success of community currencies.

Monetary systems are dynamic in nature, characterized by continuous flows of value and evolving patterns of user interaction. Network science studies such systems by modeling users as nodes and their financial exchanges as edges [1], enabling the analysis of structural patterns, user connectivity, and the circulation of value. This thesis applies that perspective to the Sarafu community currency network. Earlier research has shown that network structure, not just transaction count or value, can play a key role in maintaining system activity [2]. In Sarafu, transactions can be modeled as a directed network, with users as nodes and transactions as edges. This representation makes it possible to examine how users are embedded in the system and how they interact. Circulation in the Sarafu system was found to be highly localized [3], with transactions often concentrated within small, tightly connected groups. These structures, including short cycles and repeated exchanges, have been interpreted as signals of local trust and coordination that contribute to the functioning of the system as a whole. While such patterns reveal how value circulates across the network, their connection to individual user engagement has not been investigated. This thesis builds on that by examining whether specific structural properties and selected temporal motifs help explain which users remain active over time, and which features are most predictive of user retention.

In this thesis retention is defined as the continued participation of users over time, which is essential for the sustainability of digital financial systems. Retained users are those who remain active in the system beyond a certain activity threshold, while non-retained users gradually drop out. Recent studies on mobile money and peer-to-peer platforms have shown that retention is influenced by early engagement, social ties, and users' positions in the network [4]. This study explores various features to identify predictors of long-term engagement. Features are drawn from macro level indicators such as global centrality, which measures a user's prominence in the entire network. At the meso level, community based indicators capture how users are positioned within groups that are more densely connected to each other than to the rest of the network. At the micro level, motifs describe small recurring patterns of interaction between two or three users. In addition, cyclic structures refer to users who are part of closed loops of transactions, which may signal trust and local stability. Finally, demographic attributes are included to examine how retention varies across different user groups. Although most of the analysis is based on static monthly snapshots of the network, limited temporal analysis is included to capture short-term interaction patterns. Together, these features can be instrumental in understanding the underlying real-world systems.
Based on this foundation, the main research question is:
What kinds of features in a community currency network contribute most to user retention?

This question is explored through the following subquestions:

1. How do global centrality measures compare to community-level centralities in predicting user retention?
2. Which static structural patterns and demographic attributes are helpful in identifying retained users?
3. What dynamic network aspects, such as temporal motifs, are useful for predicting user retention?

The remainder of this thesis is structured as follows. Chapter 2 reviews related literature and situates this study within the broader context of research on user retention and network analysis. Chapter 3 describes the Sarafu dataset in detail, including its demographic attributes, temporal scope, and limitations. Chapter 4 explains the preprocessing steps used to define a consistent user base, the logic-based labeling of retention. Chapter 5 consists of the methodology used for network construction, feature extraction. It also introduces the centrality metrics, Network patterns, motifs, and modeling approach. Chapter 6 presents the experimental results, starting with an exploratory analysis of retention trends, structural patterns, meaning features derived from the network's connectivity such as centralities, communities, and cycles, temporal patterns, and demographic distributions. This is followed by predictive modeling using different feature sets, with evaluation of model performance and feature contributions. Finally, Chapter 7 discusses the main findings, concludes the thesis, and outlines directions for future research.

2 Motivation and Related Work

This section outlines the motivation for using a network-based approach to study user retention in Sarafu. Most earlier studies on user retention were based on systems where interactions followed a centralized design. Sarafu, in contrast, is a decentralized platform based on peer-to-peer exchange and voluntary use. This makes timing, trust, and local connections more important. These characteristics suggest that individual-level behavior such as how many transactions a user performs or how frequently they appear cannot fully explain continued participation. Instead, retention in this setting depends on how users are positioned in the transaction network, whom they interact with, whether those interactions repeat, and how their connections relate to others in the system. The methodology described in Section 5 builds on this perspective to investigate whether patterns of interaction are more informative than activity alone in identifying who remains active over time. Next, we review previous work on retention prediction and related approaches that support the design of the methodology used in this study.

Retention prediction has been widely studied across domains such as telecom, and online platforms. In telecom settings, churn prediction models often leverage network data from call records to identify early signs of user disengagement [5, 6]. These studies frame retention as a supervised classification problem, which aligns with this thesis' approach. However, many of them use limited types of features, typically behavioral or structural, not a combination of both. In contrast. Extending prior studies, this thesis examines how combining different aspects of network structure and behavior can provide insights for retention.

Network science offers a range of features to characterize user retention, which are discussed in detail in chapter 5. Global centrality metrics such as degree and betweenness have been applied to detect users who stay in the system, but they may not reflect local influence in informal or low-infrastructure systems [7, 8]. Community detection methods such as Louvain and Leiden are used to identify user groups, with findings showing that users central within their communities are more likely to remain active [9, 10]. Another structural indicator that has shown promise is whether a user belongs to a cyclic or acyclic position in the network. This is typically assessed via strongly connected components (SCCs), maximal directed subgraphs where each node is reachable from every other node via directed paths. Non-trivial SCCs necessarily contain at least one directed cycle. Recent work by Fan shows that cycle-based metrics, such as the cycle ratio, identify aspects of structural importance not captured by centralities like degree or betweenness, indicating that users embedded in cycles may have longer engagement potential [11]. In this thesis, both global and community-level centralities, as well as cyclic structure, are extracted and compared as potential predictors of retention.

Temporal dynamics and fine-grained structural patterns can provide valuable insights about user engagement. Prior research shows that early activity, such as fast feedback or bursty interactions, is not always predictive of long-term participation [12]. Temporal motifs, defined as small, time-ordered interaction patterns such as three-node cycles or reciprocal exchanges, encode both structural and temporal information [13]. They have been used to capture patterns in continuous-time networks, based on the ordering and timing of events. However, motif-based analysis has largely remained descriptive, and their temporal characteristics, particularly the time gaps between motif edges, have rarely been applied in predictive models of user retention. This thesis incorporates both motif participation counts and the average time gap between motif initiation and completion as explicit temporal features for predicting the network. By evaluating this motif completion speed, the study links the pace of micro-level interactions to retention outcomes, an approach not commonly explored in previous work.

3 Data

This chapter gives an overview of the data used in this thesis. It describes the structure and content of the Sarafu dataset, including transaction records and user demographics. It also explains the main selection steps, limitations and the scope of the available data.

3.1 About Sarafu

Sarafu is a digital community currency initiated by Grassroots Economics Foundation, a non-profit organization based in Kenya, to support local trade and economic resilience in areas with limited access to national currency and formal financial services. Sarafu became especially active during the COVID-19 pandemic, enabling users to exchange credits through a mobile interface linked to Kenyan mobile phone numbers, facilitating participation in localized economic exchange networks [3].

3.2 Access and Scope

The dataset used in this thesis is available for research purposes. Due to confidentiality requirements, access to the dataset must be requested through the UK Data Service, and its usage is restricted to scientific and academic purposes. Since the dataset only includes users who registered and interacted with Sarafu during the recorded period, the analysis is limited to this specific user base. Most users in the system belong to low-income or informal communities with limited access to formal financial services. These characteristics should be taken into account when interpreting the results.

3.3 Dataset Overview

The Sarafu dataset consists of two main files, a transaction file and a user file. The transaction file contains records of transfers made between users, with source and target, amount, and timestamp. The user file includes demographic and account information such as gender, location, user role, and business type. Table 1 shows a few example rows from each file. Together, these files cover 422,721 transactions made by 40,767 unique users between January 2020 and June 2021. The total volume transferred is approximately 297 million Sarafu credits.

Table 1: Example rows from transaction and user datasets

Source	**Target**	**Amount**	**Timestamp**
0xEDA5...	0x245f...	18000	2020-01-25 19:13
0xEDA5...	0xC169...	9048	2020-01-25 19:13
User Address	**Gender**	**Area**	**User Role**
0xC169...	Male	Misc Nairobi	Beneficiary
0x245f...	Female	Kilifi	Group Account

3.3.1 User Growth and Activity Trends

To understand how the system was adopted, we look at how transactions and user registrations changed over time. Figure 1 shows the monthly number of transactions and total transaction volume. Both increased sharply starting in April 2020, possibly in response to the COVID-19 pandemic. This suggests that more people used Sarafu during that period and that local groups became more connected.

The bottom panel of Figure 1 shows the number of new registered users. Most users joined early 2020, which likely contributed to the sharp rise in transaction activity shortly afterwards. This pattern reflects a period of rapid onboarding and growing engagement. Registrations declined in 2021, possibly due to reduced promotion or changes in the system.

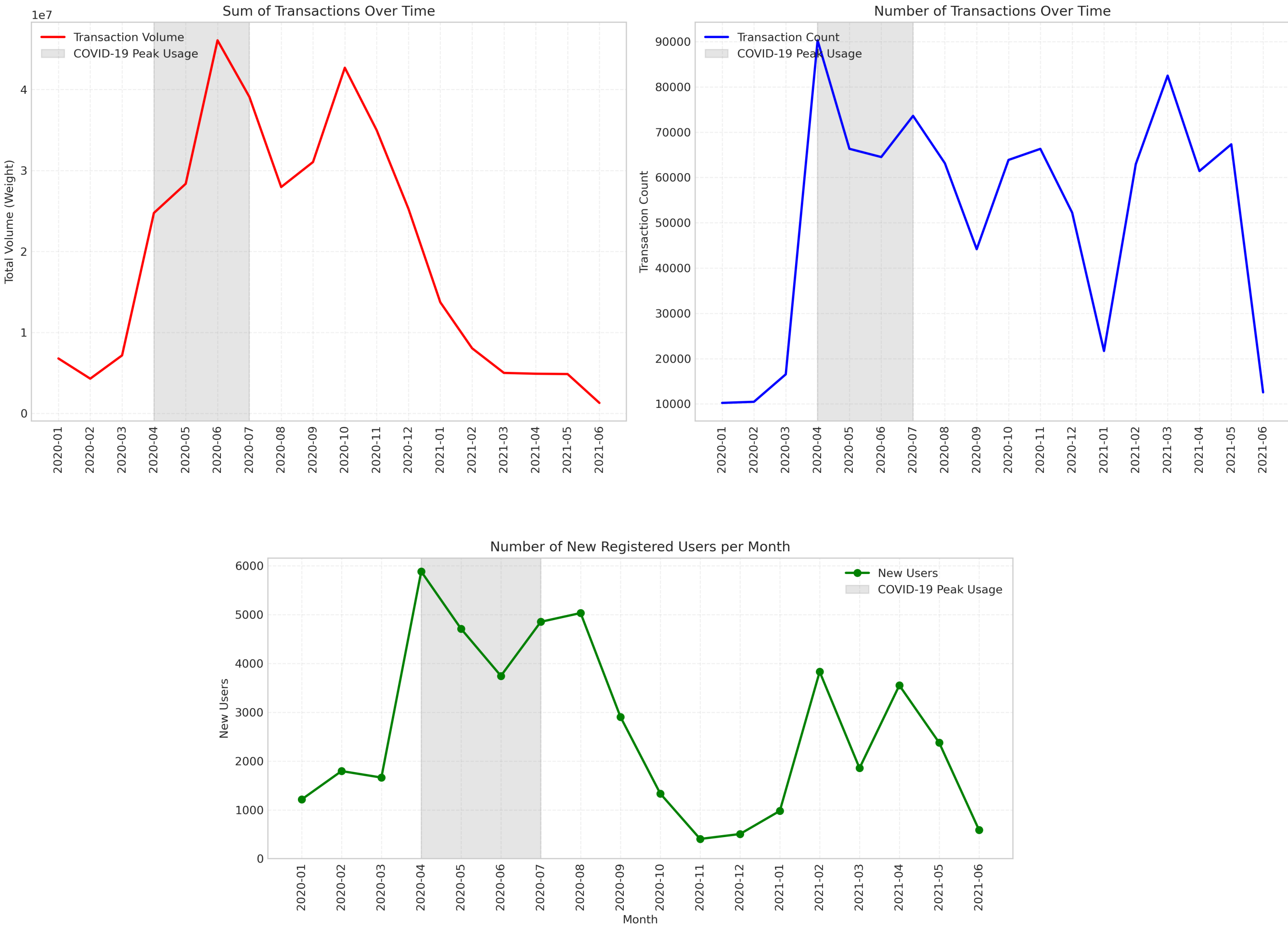
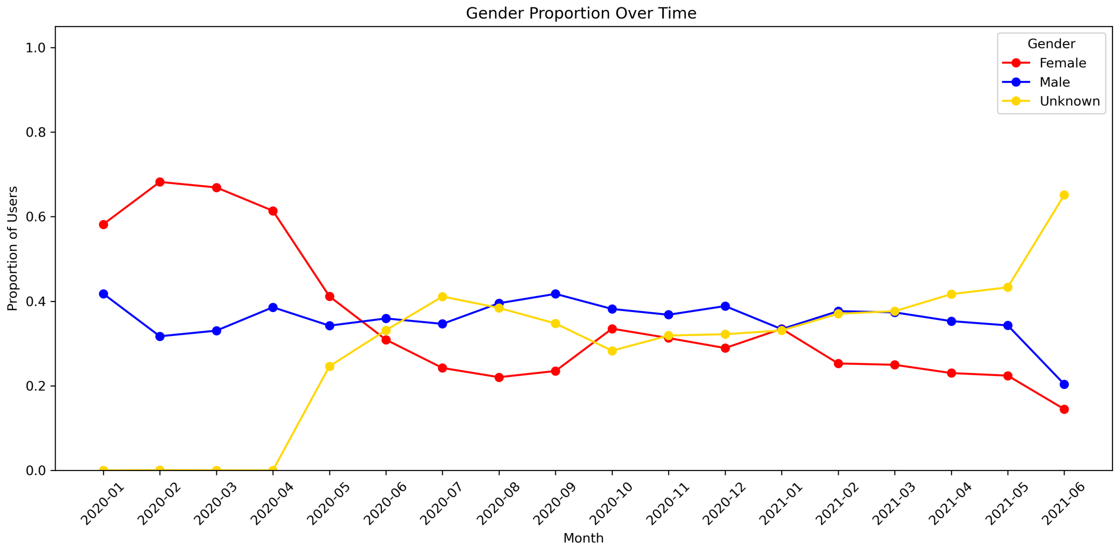


Figure 1: Transaction activity and user growth over time in the Sarafu network.

3.3.2 Demographic Composition of Users

The user dataset includes demographic features that used later in this thesis to compare retained and non-retained users. Figure 2 shows the gender and location distribution of users. At first, more users had a known gender, with female users being the largest group. Over time, the number of users with missing gender data increased, possibly due to faster onboarding or less complete records.
The bottom figure shows that most users came from a few key areas, especially Kinango Kwale and Mukuru Nairobi. These regions remained active throughout the program, while others had fewer users.



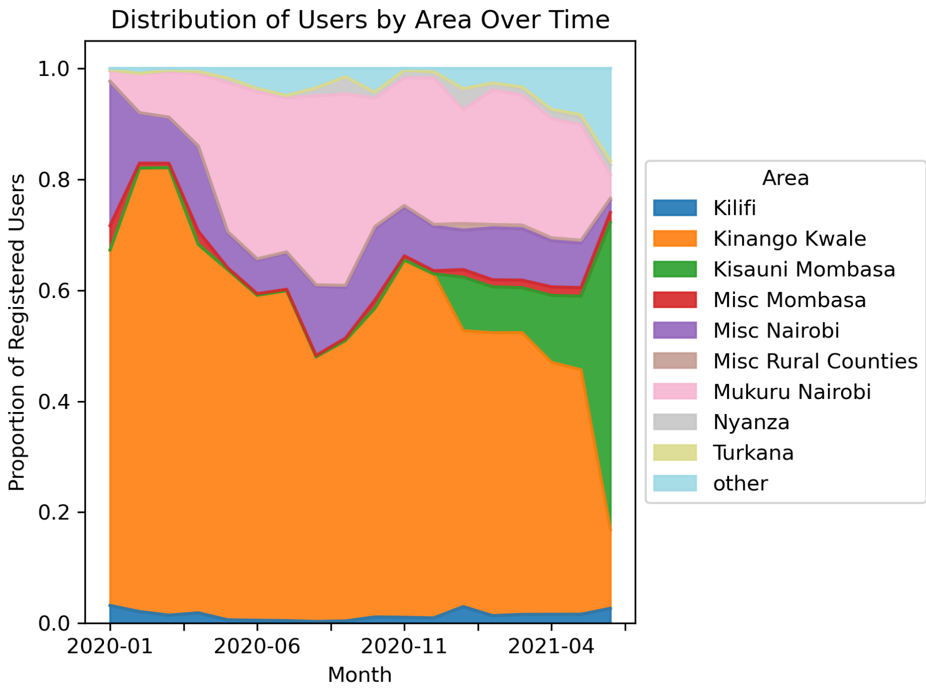


Figure 2: Gender and location distribution of users in the Sarafu dataset.

4 Data Preparation and Labeling

Here we describe the main steps taken before starting the analysis. It explains how the raw data was cleaned and filtered to define the final user group, how retention labels were created based on 2021 activity, and how features were organized into different categories. These steps ensure that the analysis and models were based on consistent and meaningful data.

4.1 Preprocessing Pipeline

The preprocessing aimed to define a consistent and meaningful user base for analysis and retention modeling. Behavioral features which refer to measures of individual activity, such as transaction frequency, regularity, and volume were extracted from user activity during the year 2020. Several filtering steps were applied to ensure that the final user set reflected meaningful and independent participation. First, only transactions of type "STANDARD" were kept, as these represent peer-to-peer exchanges between users. Transfers involving system accounts were removed to avoid bias from automated transactions. Next, users whose first appearance occurred in 2021 were excluded, since they lacked early activity history and could not be used for a fair prediction of future behavior. To focus on engaged users, we then selected only those who sent at least one transaction in 2020. Users who only received transactions but never initiated one were removed, as their behavior did not provide enough insights for feature extraction. This step helped eliminate passive or one-sided accounts from the analysis. The resulting user set, consisting of individuals who actively sent transactions in 2020 and were not system-related, formed the basis for all feature calculations. Table 2 summarizes these filtering decisions.

Table 2: Filtering steps used to define the user set for feature extraction.

User Group	User Count
All unique users in transaction dataset	54,861
Users involved in STANDARD transactions (2020–2021)	40,767
Breakdown of STANDARD users in 2020	
Users who only sent transactions	13,498
Users who only received transactions	544
Users who both sent and received transactions	17,191
Users whose first appearance was in 2021	8,520
System users removed	15
Final user set (senders in 2020: 13,498 + 17,191)	30,689

4.2 Retention Labeling

Retention is defined as continued participation in the Sarafu network during 2021 among users who were active in 2020. In prior work, Mattsson [3] applied a simple threshold on reappearance to label users as retained. While this captures a basic notion of activity, it does not account for the intensity or consistency of participation.

To better reflect different patterns of sustained engagement, this study defines retention using two logic-based strategies inspired by the Recency–Frequency–Monetary (RFM) framework [14], a common approach in customer behavior analysis. In this framework, recency refers to how recently a user transacted, frequency captures how often they transacted, and monetary value reflects the amount they transacted. Following this logic, retention was assessed with three transactional activity metrics, the number of transactions, the number of active months, and the total transaction volume. These measures allow a distinction between one-time reactivation and sustained use.

OR Logic

Under the OR strategy, a user is labeled as retained if at least one of the three activity metrics in 2021 reaches a specified proportion of its 2020 value. This definition is intentionally inclusive and captures users who remained active along a single dimension. Its flexibility, however, may overestimate retention by including users with limited engagement.

AND Logic

The AND strategy enforces more consistent activity across dimensions. It computes a weighted score over the three metrics, assigning weights of 0.5 to transaction count, 0.3 to active months, and 0.2 to volume. Transaction count is weighted most strongly as repeated activity provides a clearer signal of sustained participation than volume alone, which is highly skewed in distribution. Active months are given intermediate weight as they capture temporal spread, while volume receives the lowest weight. A user is retained if the weighted 2021 score reaches a fixed proportion of the corresponding 2020 score. This design results in a smaller but more selective retained group. The chosen weights reflect a preference for frequent and sustained activity rather than transaction volume and are informed by patterns observed in the network during exploratory analysis. While not optimized through a formal model, they align with the skewed distribution of activity metrics such as degree and volume shown in Table 7.

Comparison and Threshold Selection

Both strategies were evaluated across multiple thresholds. Figure 3 shows the resulting distributions. As expected, the OR logic produces around thirty percent retained users, while the AND logic yields a smaller group of about fifteen to twenty percent. The twenty percent threshold was selected as it provided a balance between inclusiveness and stability, with roughly six thousand users classified as retained. At lower thresholds, the OR definition classified many minimally active users as retained, while at higher thresholds of forty to fifty percent the AND definition reduced the retained group to about seven to ten percent. The chosen threshold therefore captures meaningful participation without overstating retention. For all further analysis, the AND definition with a twenty percent threshold was adopted as the primary retention label. The robustness of model performance across these thresholds is examined later in Subsection 6.2.

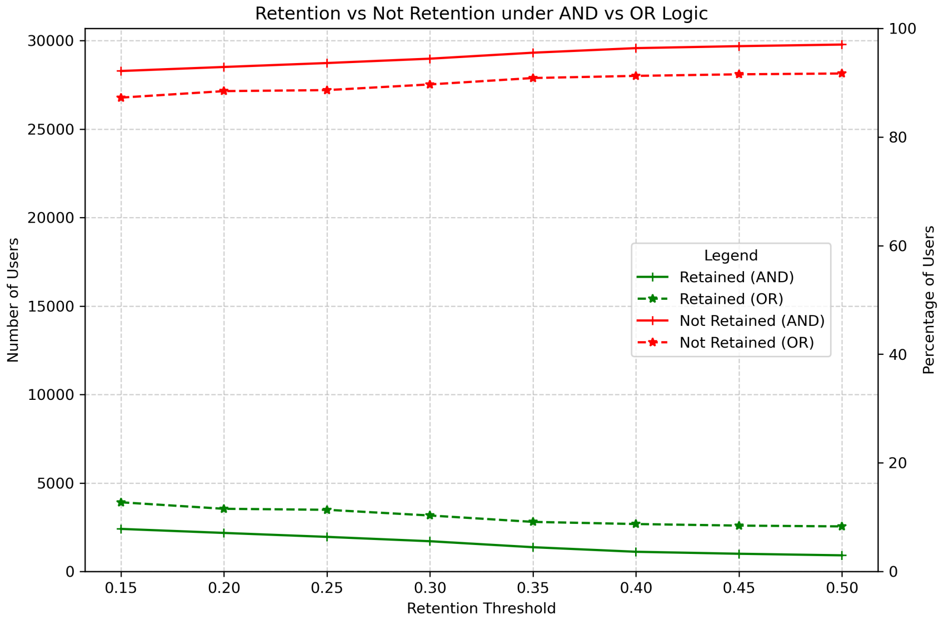


Figure 3: Distribution of retained and not retained users across thresholds, comparing AND and OR strategies.

5 Methodology

This chapter outlines the methodological framework used to analyze user retention in the Sarafu network. It begins with the construction of the transaction network and the categorization of features at multiple analytical levels. The next part introduces centrality measures that describe user positions both globally and within communities. Structural patterns such as cyclic and acyclic patterns are then examined, followed by the detection of temporal motifs that capture localized and repeated interactions. Community detection is presented as a way of identifying groups of closely connected users and comparing global with local structural indicators. Finally, the predictive modeling approach is described, detailing feature-based model design, data preparation, classifier selection, and evaluation metrics.

5.1 Network Construction and Feature Categorization

The methodology begins with the construction of a directed transaction network that captures the flow of the Sarafu community currency. This representation is designed not only to map interactions, but also to serve as the foundation for extracting features that may explain differences in long-term user retention. The expectation is that global position, community structure, and recurring patterns of exchange, when combined with demographic information, will provide insight into why some users remain active while others disengage.

The analysis relies on three complementary representations of the network. First, the complete transaction network of 2020 is modeled as a directed graph $G = (V, C)$, where V denotes the set of user nodes and C the set of directed edges representing transactions. A directed edge from node i to node j indicates that user i transferred tokens to user j at least once. Multiple transactions between the same pair of users are aggregated into a single weighted edge, with weights corresponding to the total transferred volume. This aggregated representation provides the overall structural context of the system. Second, time is explicitly preserved in the transaction records. This temporal representation allows the derivation of features that depend on the ordering and timing of interactions, such as motif completion times, which measure how quickly recurring patterns of exchange close. Third, monthly networks are constructed to analyze activity and structural dynamics at finer temporal resolution, making it possible to track evolving user positions and participation patterns over time.

Following this logic, the final feature set is organized in layers that reflect different analytical levels of features of the network. At the macro level, global centralities describe visibility in the full system. At the meso level, community-based measures capture group connectivity and relative positioning within clusters, then centralities, in detected communities and at the global level. At the micro level, motif counts and completion times capture recurring interaction patterns and their temporal dynamics. Finally, demographic attributes describe individual-level characteristics that may shape participation. The complete set of features is summarized in Table 3. Each feature is defined in more detail later in this chapter.

Table 3: Final feature set organized by analytical level, type, temporal scope, and variable type

Feature Name	Feature Type	Temporal Scope	Variable Type
Macro-level (Global Network)			
In-degree	Structural Connectivity	Static	Numeric
Out-degree	Structural Connectivity	Static	Numeric
Weighted In-degree	Transaction Volume	Static	Numeric
Weighted Out-degree	Transaction Volume	Static	Numeric
Clustering Coefficient	Local Cohesion	Static	Numeric
Betweenness Centrality	Bridging Role	Static	Numeric
Cyclic Status	Structural Pattern	Static	Boolean
Meso-level (Community Structure)			
Community Size	Group Structure	Static	Numeric
In-degree	Structural Connectivity	Static	Numeric
Out-degree	Structural Connectivity	Static	Numeric
Weighted In-degree	Transaction Volume	Static	Numeric
Weighted Out-degree	Transaction Volume	Static	Numeric
Clustering Coefficient	Local Cohesion	Static	Numeric
Betweenness Centrality	Bridging Role	Static	Numeric
Micro-level (Motif-based Patterns)			
3-Node Motif Count	Motif Participation	Static	Numeric
3-Node Motif Completion Time	Motif Timing	Temporal	Numeric
Reciprocal Motif Count	Motif Participation	Static	Numeric
Reciprocal Completion Time	Motif Timing	Temporal	Numeric
Individual-level (Demographics)			
Gender	Demographic Attribute	Static	Categorical
Area	Demographic Attribute	Static	Categorical
User Role	Demographic Attribute	Static	Categorical
Business Type	Demographic Attribute	Static	Categorical

5.2 Community Detection

In network analysis, a community is typically defined as a group of nodes that are more densely connected internally than with the rest of the network [15]. In the context of transaction networks, this structure often reflects recurring interactions within geographically localized or trusted trading circles. Detecting such communities allows for analysis, beyond what global network metrics can reveal.

To identify communities in the Sarafu transaction network, this thesis applied modularity-based algorithms that detect structural communities based solely on network topology. These communities are not geographically defined, nor are they based on user-reported demographic attributes such as area or region. Instead, they are derived purely from patterns of transactional connectivity.
The Leiden algorithm [10] was adopted, as it produced more robust and internally well-connected communities. This choice was motivated by earlier results from Louvain, which led to fragmented and unstable groupings.Given that Leiden includes stochastic elements, it was executed fifty times to obtain stable labels through consensus clustering [16]. A single consensus result was constructed by selecting, for each user, the community label that occurred most frequently across the 50 iterations. This final consensus labeling was then used for all analysis. Each user was assigned to one community, and the resulting group assignments were used to compute community-specific features.

The outcome of the community detection process is shown in Figure 4. The layout reveals considerable variation in the size and shape of communities. A few large clusters dominate the center of the network, each comprising thousands of users and forming dense, visually compact groups. These likely correspond to major transactional hubs. Around them, many smaller communities are visible, some moderately sized, others very small and dispersed. These smaller groups are often found near the periphery and tend to be more loosely structured. The spatial separation between clusters and the uneven distribution of node density make it visually evident how transactional activity is concentrated in certain parts of the network. The figure thus offers an interpretable summary of the Sarafu system's structure, where a small number of communities account for most of the internal connectivity, while many others remain marginal in both size and position.

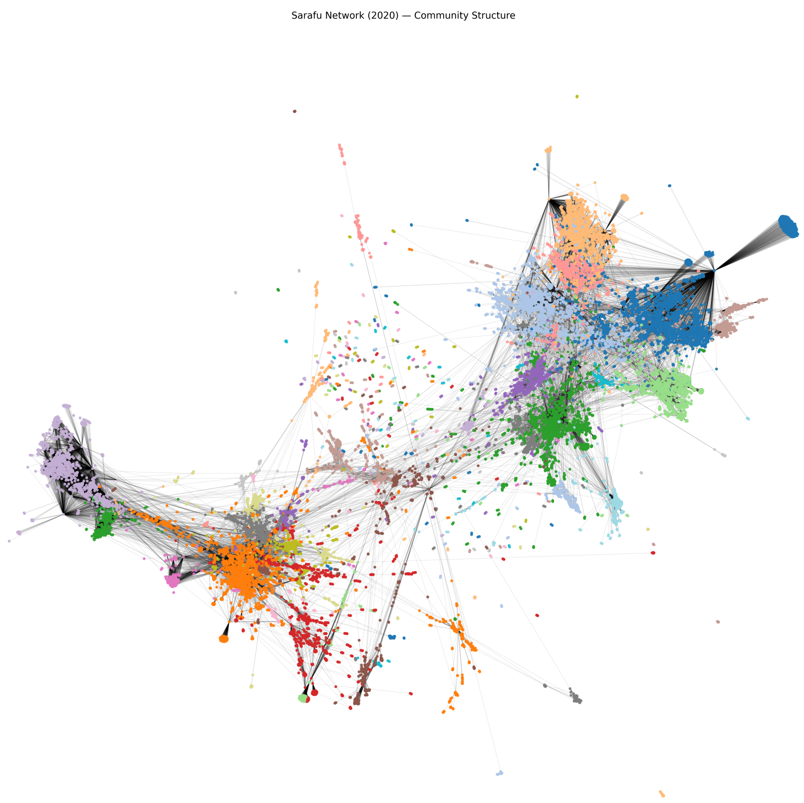


Figure 4: Community structure of the Sarafu transaction network in 2020.

5.3 Centrality Measures

To address the main research question, what kind of features in a community currency network contribute to user retention, this section focuses on structural metrics that describe user positions in the transaction network. These features relate to the first subquestion on how global centrality measures compare to community-level centralities in predicting user retention. Each centrality measure was calculated at both the global and community level. At the global level, they describe a user's position in the entire network. At the community level, they indicate influence within smaller groups of densely connected users. At the local level, summary statistics such as the mean and standard deviation of centralities within each user's community capture variation in the immediate neighborhood. Together, these measures allow a systematic comparison between broad visibility, and group-specific importance. All variables introduced here are summarized in Table 3.

5.3.1 Degree Centrality

Degree centrality counts the number of direct links a user has to others. In a transaction network, this reflects the number of unique users a participant has interacted with. It helps identify users who are highly connected, either by sending to or receiving from many others. This measure was used both globally and within communities. A higher value may indicate greater involvement in the system, although it does not account for the strength of interactions [17].

5.3.2 Weighted Degree Centrality

Weighted degree centrality sums the weights of connections, capturing the total volume exchanged. It reflects how economically involved or influential a user is. This metric is used to distinguish between users who are broadly connected and those who engage in high-volume exchanges with fewer partners [18].

5.3.3 Betweenness Centrality

Betweenness centrality measures how often a user lies on the shortest paths between other users. It identifies users who bridge different parts of the network and help connect otherwise separated groups. Users with high betweenness may not be highly active themselves but play a key role in keeping the network integrated [7, 8]. This feature was used to study whether such bridging roles relate to retention.

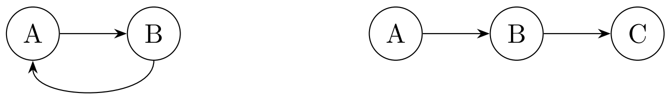
5.3.4 Clustering Coefficient

The clustering coefficient indicates how densely a user's neighbors are connected with one another. In a transaction network, this reflects the extent to which a user participates in local trading circles. Users with high clustering are often embedded in tightly connected groups, while those with low clustering may span across different areas at the network. This distinction is relevant for understanding whether embeddedness in local groups contributes to continued participation [19].

5.4 Cyclic Pattern Detection

Here we explore whether the way users are embedded in structural patterns is related to their long-term engagement. This addresses the second sub-question, which explores whether different static structural patterns such as cyclic and acyclic transaction paths are associated with variation in user retention. These patterns describe whether users are part of repeated, mutual interactions or one-directional, branching exchanges.

Cyclic patterns are formed when a user belongs to a closed loop of transactions, such that value eventually returns to the starting point as shown in Figure 5 (A). These loops may reflect stronger or more persistent ties between users and suggest a higher level of mutual trust. Previous research has linked cycles to stability in transactional networks [20]. In this study, users who belonged to such loops were identified using strongly connected components. Any user in a component of size two or larger, where each participant can reach every other, was labeled as cyclic. These users were then compared to others in terms of retention. Acyclic patterns are structures where value moves outward but does not return, as it illustrated in Figure 5 (B). They often resemble trees or chains and may indicate short-term or weaker interactions. Users not involved in any cycles were considered part of the acyclic group. These structural differences were used to assess whether engagement styles, as captured by network position, relate to continued use. Some studies have shown that users embedded in less connected or tree-like structures are more likely to leave the system [21].



(A) Cyclic pattern (B) Acyclic pattern

Figure 5: Cyclic and acyclic structures

5.5 Motif Detection

To answer the third subquestion, this section examines whether temporal motifs offer informative signals for predicting user retention. Motifs capture short, repeated interaction patterns that combine structural and temporal information at the micro level. Unlike global centrality or cyclicity, motifs reflect localized behavior in pairwise or triadic exchanges. Their occurrence and timing can indicate trust, responsiveness, and sustained engagement between users. Following the formalization introduced by Kovanen [22] and extended in recent work by Zhao [23], two types of motifs are analyzed, reciprocal interactions and a specific form of three-node cycle. These extend the earlier cyclicity measure by incorporating the order and timing of interactions, and their focus is motivated by the high level of reciprocity observed in the Table 7, which shows that many transaction ties occur in both directions.

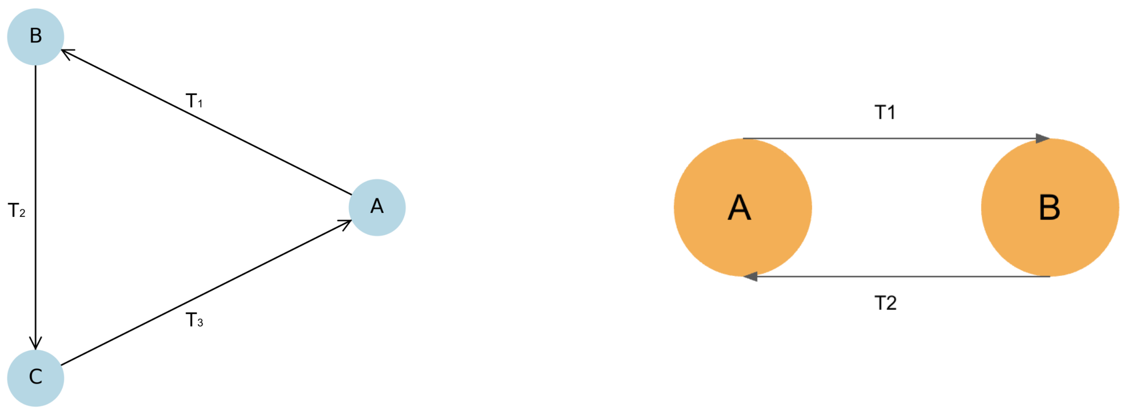
For each user, both the number of motif participations and the average time to complete a motif are included as features. For users who did not participate in any motifs, the count features were set to zero, while the associated time-gap features were left undefined and treated as missing values. These missing values were later handled by imputation during model training at the Subsection 5.6.2. While several additional 3-node motifs exist [24, 25], this study restricts its scope to these two types for two reasons. First, these motifs reflect repeated exchange and reciprocity, which are widely recognized as fundamental behavioral patterns in transactional networks [22, 20]. Second, more complex motifs are combinatorially rare and their interpretability at the user level is limited. Including them would increase computational cost without providing clear explanatory value. Furthermore, motif sampling or mining methods [26] were intentionally not used, to ensure exhaustive detection and comparability across users. This focus allows the analysis to remain tractable and behaviorally grounded while retaining sensitivity to important local structures. The timing features of these motifs enable the analysis to incorporate temporal dynamics, not just structural participation. Both motif types are listed as local temporal features in Table 3, and their predictive contribution is examined further in Chapter 6.

5.5.1 Three-Node Cycles and Their Time Gap

Motif detection relied on exhaustive enumeration of directed subgraphs in the transaction network. For three-node cycles, triplets of users with links A→B, B→C, and C→A were identified, forming a closed 3-cycle [24]. This structure is illustrated in Figure 6 (A). For each user, the number of participations in such cycles was counted. To incorporate temporal dynamics, only time-respecting sequences were considered, meaning that the three transactions had to follow one another in time. The completion time of a cycle was defined as the interval between the first and last transaction in the sequence. If the three transaction times are ordered as $T_1 < T_2 < T_3$, the closure speed is measured as $T_3 - T_1$. These values were then averaged per user to capture both the frequency of participation and the typical time required for cycles to close.

5.5.2 Reciprocal Motifs and Their Time Gap

Reciprocal motifs capture two-way exchanges between pairs of users. All pairs with both A→B and B→A transactions were included. As shown in Figure 6 (B), this motif represents the mutual return of value [27, 28]. The temporal dimension was defined as the interval between a transaction and its first return in the opposite direction, denoted as $T_2 - T_1$ with $T_1 < T_2$. For each user, these reciprocity gaps were averaged to summarize typical response speed. These features are listed as micro-level in Table 3 and analyzed further in Chapter 6.



(A) Three-node Motif (B) Reciprocal motif

Figure 6: Motifs representing closed triadic circulation and bidirectional exchange in the transaction network.

5.6 Predictive Modeling Approach

To answer the main research question and evaluate the influence of the features extracted from the transaction network in Section 5, we trained multiple models using different subsets of features in Table 3. The modeling pipeline consists of six steps, which summarized in Figure 7, and each explained in the following subsections.

Figure 7: Overview of the predictive modeling pipeline

5.6.1 Comparative Model Design

This step analyzes how different types of features contribute to retention prediction. Five model configurations were constructed, each corresponding to a specific subset of variables defined in Section 5 and summarized in Table 3. The setup allows direct comparison of global and community structure as defined in sub-research question 1, and extends the analysis to temporal patterns 3 and demographic attributes 2 through separate configurations. Each model adds a new feature group to the previous one, allowing the examination of whether additional information improves predictive performance. The models listed from I to V define which features each was trained with, and the outcome of each configuration is presented in the following chapter 6.

I. The first model uses only features calculated on the full transaction network, degree, weighted degree, clustering coefficient, and betweenness. It addresses sub-research question 1 by testing whether global network position is related to user retention.

II. The second model includes features computed within Leiden communities. These consist of centrality values within the community subgraph, the average and standard deviation of those values across the user's community, and total community size. It is designed to evaluate whether local network position is more informative than global position for sub-research question 1.

III. This model combines all variables from Model I and Model II. The goal is to compare global and local perspectives jointly and assess whether they offer additional value when used together or reflect the same underlying patterns. This setup continues the analysis of sub-research question 1.

IV. In addition to all features from Model III, this model introduces temporal and cyclic variables. It includes the number and timing of motif instances as well as a binary variable indicating whether the user belongs to a strongly connected component. These additions support sub-research question 3.

V. The final model extends Model IV by adding user-level demographic information, including gender, area name, business type, and reported roles. This configuration addresses sub-research question 2 by testing whether personal attributes improve predictive performance beyond network-derived variables.

5.6.2 Data Preparation and Splitting

The preparation process began by merging all feature sets on the blockchain address, which ensured that every user could be consistently identified throughout the data. Categorical variables such as business type and area were encoded numerically. Missing values in motif counts were set to zero, while missing values in motif time gaps and other continuous features were completed using multiple imputation with an iterative model. This ensured that all users could be included in the analysis. Any remaining missing values were handled internally by XGBoost. Seventy percent of the users were assigned to the training set. The remaining thirty percent was split evenly into validation and test, the distribution of users shown in Table 4, using stratified sampling to keep the balance between retained and non-retained users.

Table 4: Number of users in each dataset split.

Split	Users
Train	21,483
Validation	4,603
Test	4,604

5.6.3 Classifier Selection

XGBoost was selected as the core algorithm due to its robust performance in classification settings and its compatibility with high-dimensional, heterogeneous feature sets [29]. It combines gradient boosting with regularization, which makes it particularly effective in controlling overfitting. Moreover, XGBoost is well-suited for imbalanced datasets, as it allows weighting of individual samples and adjustment of decision thresholds during inference. The choice was further motivated by its resilience to missing values, reduced need for feature scaling, and successful use in related prediction tasks [30]. Its flexibility in handling categorical, numerical, and network-derived features made it an appropriate choice for this application.

5.6.4 Hyperparameter Optimization

The hyperparameters of the XGBoost classifier were optimized using Optuna, a framework based on Bayesian optimization and early stopping. This method adaptively explores the parameter space and discards unpromising trials to reduce computation time [31]. For each model configuration, 30 optimization trials were conducted, and each trial was evaluated using the F1-score on the validation set. The configuration yielding the highest F1-score was selected for final model training. The selected values for each model configuration are summarized in Table 5.
The search space included the following parameters:

- Number of boosting rounds: total number of trees to fit in the ensemble.
- Learning rate: step size shrinkage to control the contribution of each tree and mitigate overfitting.
- Maximum tree depth: limits the complexity of individual trees and governs the model's ability to capture interactions.
- Subsample ratio: proportion of training instances used for constructing each tree, introducing randomness and improving generalization.
- Column sampling rate per tree: fraction of features randomly selected when growing each tree, promoting diversity.

Table 5: Optimized values of tuned hyperparameters and thresholds for Models I–V.

Hyperparameter	Model I	Model II	Model III	Model IV	Model V
Number of boosting rounds	119	210	235	145	158
Learning rate	0.279	0.100	0.052	0.033	0.204
Maximum tree depth	9	5	6	10	4
Subsample ratio	0.859	0.757	0.758	0.911	0.963
Column sampling rate per tree	0.711	0.916	0.748	0.710	0.659
Threshold	0.14	0.21	0.19	0.15	0.13

5.6.5 Handling Class Imbalance and Prediction Calibration

As shown in Figure 3, the number of retained users is much smaller than that of non-retained users. Without adjustment, the classifier would be dominated by the majority class, limiting its ability to detect long-term participants. To address this imbalance, three steps were added after model tuning. First, the loss function was reweighted so that misclassifying a retained user incurred a higher penalty than misclassifying a non-retained user. This adjustment was applied within XGBoost's logistic loss during training, ensuring that the minority class received greater emphasis and that the imbalance between classes was reflected directly in the optimization process. Second, the model outputs were calibrated to ensure that predicted probabilities reflected the actual frequency of retention in the data. This is relevant in imbalanced data, where raw probabilities can overestimate or underestimate retention and make model comparisons less [32]. Finally, the classification threshold was optimized on the validation set. The default value is not suitable under class imbalance, so the threshold was instead chosen from the precision–recall curve at the point that maximized the F1-score. This increased the detection of retained users while keeping precision at an acceptable level.

5.6.6 Evaluation Metrics

Evaluating classification performance in imbalanced settings requires metrics that go beyond overall accuracy. This study employs a combination of threshold-dependent and threshold-independent metrics to gain a comprehensive view of predictive performance. These metrics, summarized in the Table 6.

F1-score is particularly suitable when both false positives and false negatives are critical, as it balances precision and recall. ROC AUC captures the model's overall ranking capability, while the precision-recall curve provides a more accurate view in imbalanced scenarios. Class-wise precision, recall, and accuracy offer detailed insights into model behavior across both majority and minority classes. To ensure the robustness of conclusions, performance metrics were reported separately for the validation and test sets, and the results are shown in Table 11

Table 6: Evaluation metrics used for model assessment

Metric	Type
F1-score	Threshold-dependent
ROC AUC	Threshold-independent
Precision-Recall Curve	Threshold-independent
Precision, Recall, Accuracy (per class)	Threshold-dependent

6 Experiments

This chapter presents the findings of this study. It consists of two parts. The first part investigates structural, temporal, and demographic patterns in the Sarafu network that explained in Chapter 5, and how they relate to user retention. The aim is to explore whether early signals can differentiate retained users from those who disengage. The second part evaluates whether these signals can be combined to build a predictive model that reliably classify users based on their long-term engagement.

6.1 Exploratory Analysis

The primary goal of the exploratory analysis is to examine which types of features appear most relevant for distinguishing retained users. The analysis proceeds in five parts. First, retention outcomes are compared across demographic groups to understand population-level patterns. Second, user assignment to communities is analyzed to assess the relationship between group structure and retention. Third, global and community-level centralities are compared to identify which scales of user prominence are more informative. Fourth, cyclic versus acyclic user patterns are examined to assess structural engagement. Finally, local transaction motifs are analyzed with a focus on both participation and timing. Together, these steps aim to identify meaningful differences between retained and non-retained users.

6.1.1 Network Descreptive Statistics

The main properties of the Sarafu transaction network are summarized in Table 7. These descriptive metrics help to understand the overall structure of the system before looking at more detailed features. The number of nodes and edges shows how many users were active and how many transactions took place. The network is sparse, as expected in large systems, which is reflected in the low density [1]. Average in-degree and out-degree show how many users each person interacted with, while the maximum values point to highly active individuals who may play a more visible role. Assortativity tells us whether users tend to interact with others who are equally active. A negative value means that users with many connections often interact with those who have fewer. Reciprocity measures how often users transact in both directions. This can reflect more stable or repeated exchanges. The number of strongly and weakly connected components shows how fragmented the network is. A large strongly connected component suggests that many users are part of the same core group. Finally, the average shortest path within the largest weakly connected part tells us how quickly value can move between users in the network. These values provide a first view of the system's scale, connectivity, and structure, which helps to interpret later results.

Table 7: Descriptive statistics of the Sarafu transaction network in 2020

Metric	Value
Number of Nodes	31,235
Number of Edges	118,092
Density	0.000121
Average In-Degree	3.78
Average Out-Degree	3.78
Maximum In-Degree	1,417
Maximum Out-Degree	1,170
Degree Assortativity	−0.0823
Reciprocity	0.4867
Number of Strongly Connected Components	15,274
Number of Weakly Connected Components	267
Size of Largest SCC	15,172
Average Shortest Path	2.99

6.1.2 Demographic Analysis

This analysis investigates whether user demographics are associated with variation in retention outcomes. The selected attributes include gender, area, user role, and business type. These attributes are sourced from the user dataset, an example of which is shown in Table 1. As shown in Table 8, group accounts retain at much higher rates than individual users. This means that group accounts were more often active across multiple transactions during the observation period. Individual accounts, on the other hand, were more likely to stop transacting after a short time. Area also shows variation. Users from regions such as Nyanza and Kilifi retained more often than those from Mukuru Nairobi. These differences cannot be explained by external context, but they show that users in some areas stayed active longer. Gender differences are small. Retention among female users is slightly higher than among male users, though the difference is limited. Among business types, users marked as savings or government-related show higher retention than those listed under food, farming, or retail.

Table 8: Retention and non-retention rates by demographic attributes

Category	Group	Retention Rate	Non-Retention Rate
Gender	Female	8.6%	91.4%
	Male	5.8%	94.2%
	Unknown	7.4%	92.6%
Area	Kinango Kwale	8.2%	91.8%
	Mukuru Nairobi	3.4%	96.6%
	Misc Nairobi	6.1%	93.9%
	Nyanza	30.6%	69.4%
	Kilifi	17.9%	82.1%
Role	Beneficiary	7.0%	93.0%
	Group Account	41.0%	59.0%
Business Type	Farming	7.4%	92.6%
	Food	7.9%	92.1%
	Government	22.7%	77.3%
	Savings	32.1%	67.9%
	Shop	5.6%	94.4%

6.1.3 Community Structure

The aim of this subsection is to examine whether the size and distribution of user communities are associated with differences in retention outcomes. This analysis relates to the second research subquestion introduced in Chapter 1, which explores whether meso-level structures, such as communities detected in the transaction network, offer explanatory power in predicting user engagement.

As described in Section 5.2, users were grouped into communities after filtering and constructing the full 2020 transaction network explained in the Section 5.1. This process produced a total of 317 distinct communities. Table 9 provides descriptive statistics of the resulting community assignments. The number of users per group is highly skewed, with a few large communities and many smaller ones. The average community includes approximately 1,859 users, though sizes range from 2 to over 3,600 users. This design choice allowed us to retain all users connected through transactions, not just those in the largest connected component, and avoid bias from removing users in smaller or less active groups. The skewed distribution of community sizes is also visible in the layout shown earlier in Figure 4, where a few large clusters dominate the network and many smaller groups appear around them.

Table 9: Summary statistics for community structure in the Sarafu network

Metric	Value
Number of communities	317
Average community size	1,859
Minimum community size	2
Maximum community size	3,629

To evaluate whether retention is correlated with group size, the retention rate for each community was computed based on the definition in Section 4.2, using the 20 percent AND threshold. Figure 8 shows the distribution of retention rates as a function of community size. A linear regression line was fitted to capture general trends.

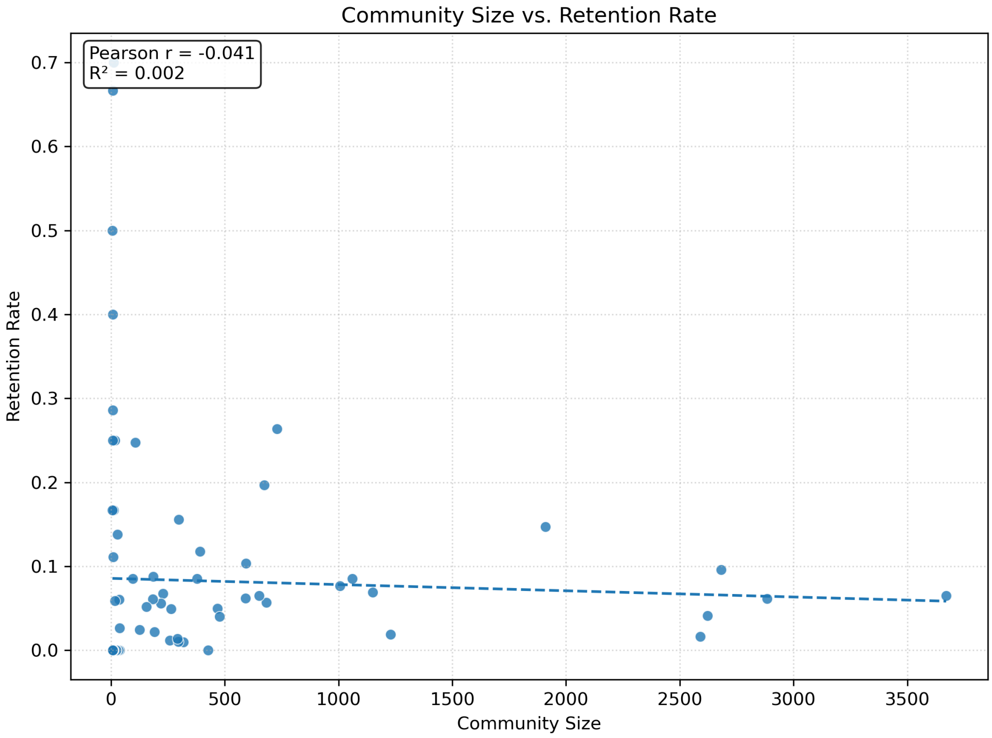


Figure 8: Comparisson of Retention Rate and Community Size

6.1.4 Community Size and Retention

Figure 8 shows that retention does not appear to increase with group size. This relationship is analyzed to test whether larger communities provide more stability. While the regression line indicates a negative slope, the relationship is limited and highly dispersed. In fact, the highest retention rates are observed in some of the smallest communities. To make this explicit, the fitted line yields a Pearson correlation coefficient, a standard measure of linear association between two variables, of $r = -0.041$ and $R^2 = 0.002$. This means group size accounts for almost none of the variation in retention. Each dot represents one community, all shown with equal size and no weighting.

This finding indicates that community size alone does not explain variation in user retention. The wide range of retention outcomes in small communities suggests that local cohesion or internal structure may play a more important role than overall scale. As a result, these findings motivate a closer analysis of intra-community characteristics such as centrality and clustering, which are addressed next.

6.1.5 Global and Community-Level Centralities

This subsection addresses the first subquestion of this study, whether users' prominence within the network structure, measured at both global and community levels, is related to their long-term engagement. The aim of the analysis is to compare retained and non-retained users in terms of four well-established centrality metrics, degree, weighted degree, clustering coefficient, and betweenness. Each measure was calculated globally and within the community assignments, as described in Section 5.2. Summary statistics for each metric by retention label are reported in Table 10.

Degree centrality Retained users have higher mean degree values than non-retained users at both global and community levels. The effect is of similar size across the two levels, as shown in Table 10. At the same time, the standard deviations are large relative to the means, which indicates substantial overlap between retained and non-retained users. This suggests that degree is related to retention at an aggregate level, but by itself it does not reliably separate the two groups.

Weighted degree centrality Retained users exchanged larger total volumes than those who became inactive (Table 10). This holds at both global and community levels. Still, there are cases where users transacted high amounts without staying engaged, showing that large volume on its own does not guarantee continued activity. The results therefore suggest that transaction volume matters for retention, but it cannot fully explain why some users remain active while others leave, and it also supports the chosen logic for the retention definition explained in Section 4.2.

Clustering coefficient Following the differences observed in weighted degree, where repeated transactions aligned more clearly with retention than raw volume, clustering provides a complementary perspective. This measure looks at whether a user's contacts also transacted with each other. In Table 10, average clustering is slightly higher for retained users, both in the full network and within communities, and the variation is nearly identical across groups. Although the absolute difference is small, the pattern is steady. Figure 9 shows that the communities with higher retention often lie above the global average line. This suggests that retained users were not only active with others, but did so in more enclosed circles where repeated interactions could form. The consistency across group sizes and network levels supports the idea that clustering reflects stable group engagement. It also offers a contrast to betweenness, which is discussed next.

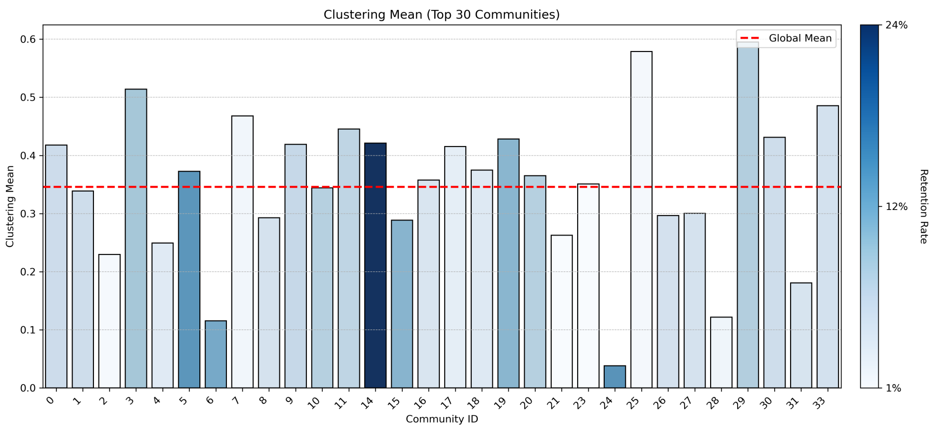


Figure 9: Average clustering coefficient at global and community levels colored by retention rate. The dashed red line indicates the global mean.

Betweenness centrality Compared to clustering, the differences in betweenness are much smaller. Table 10 shows that average values remain close to zero at the global level, regardless of retention outcome. The standard deviation is also large relative to the mean, which makes the metric unstable. This is likely due to the fragmented structure of the network, where most users are not well-positioned to connect different parts of the system. Within communities, the differences are slightly more visible. Figure 10 shows that users in higher-retention communities tend to have somewhat higher betweenness. Still, the overall effect remains small. This suggests that roles based on spanning others are less central to sustained participation in this context, especially when compared to more direct indicators like repeated volume and clustering.

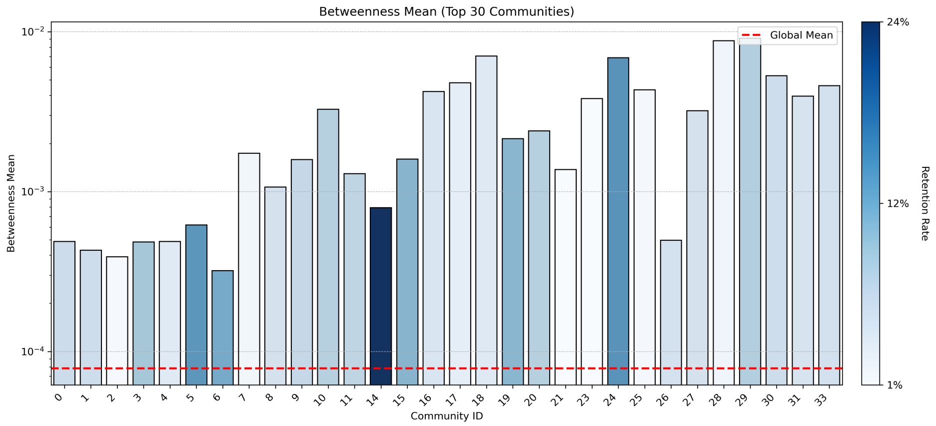


Figure 10: Average betweenness centrality at global and community levels colored by retention rate. The dashed red line indicates the global mean.

Table 10: Centrality values by retention label in log scale (mean ± std)

Source	**Not Retained (mean ± std)**	**Retained (mean ± std)**
Degree (Global)	1.39 ± 0.87	2.02 ± 1.10
Degree (Community)	1.38 ± 0.85	2.01 ± 1.08
Weighted Degree (Global)	6.09 ± 1.72	6.77 ± 2.25
Weighted Degree (Community)	6.02 ± 1.74	6.72 ± 2.24
Clustering (Global)	0.25 ± 0.29	0.34 ± 0.28
Clustering (Community)	0.25 ± 0.30	0.34 ± 0.28
Betweenness (Global)	0.000068 ± 0.00140	0.000195 ± 0.00108
Betweenness (Community)	0.0020 ± 0.0207	0.0047 ± 0.0243

6.1.6 Cyclic and Acyclic Patterns

One of the central hypotheses in this study is that users involved in recurring transactional structures are more likely to stay active than those who interact in one-directional or fragmented ways. Subquestion two specifically asks whether cyclic users are more likely to be retained than acyclic ones. To explore this, users were categorized based on whether they belonged to strongly connected components in monthly directed graphs, following the procedure described in Subsection 5.4. Those within such components were labeled as cyclic, while others were considered acyclic. The first step in the analysis is to examine the monthly distribution of both groups. Figure 11 shows the number of cyclic and acyclic users from January 2020 to June 2021. Across all months, acyclic users are more numerous, particularly during periods of overall network growth in late 2020 and early 2021. However, the number of cyclic users increases at a steadier rate and shows less fluctuation. This suggests that cyclic users may represent a more persistent and structurally embedded segment of the network, while acyclic users are more likely to appear in bursts and drop out.

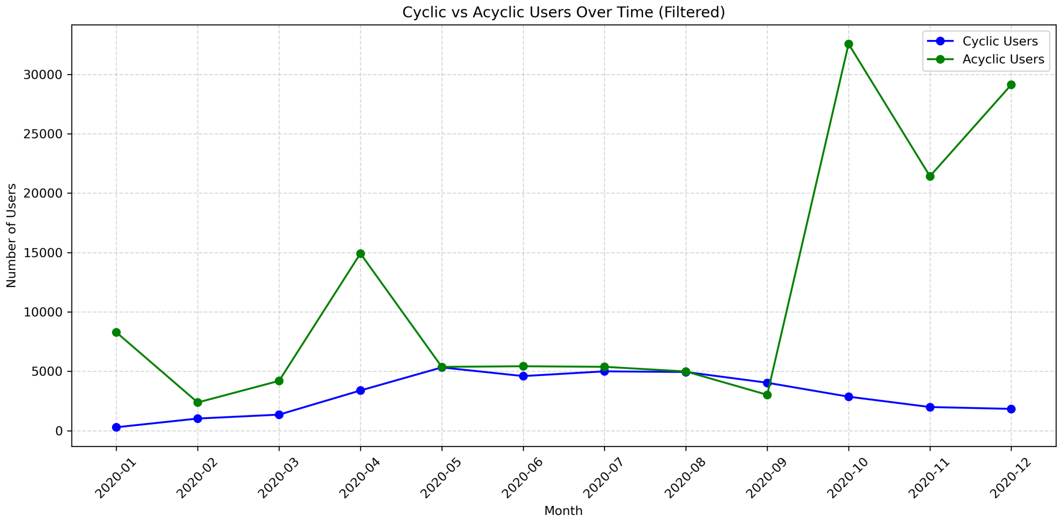


Figure 11: Monthly counts of cyclic and acyclic users from January 2020 to June 2021

To assess whether structural recurrence is linked to user retention, the analysis next compares retention outcomes between the two user types. Figure 12 displays the number of retained and non-retained users in each group. A strong pattern emerges, the vast majority of retained users are cyclic. In the 2020 network over ninety percent of users who meet the retention criteria belong to cyclic structures, while acyclic users make up most of the non-retained population. These findings suggest that users who appear in cyclic formations, where transactions flow back and forth among peers, are more likely to continue participating in the system. The structural implication is that mutual interaction may foster stability, either through trust, repeated exchange, or local , it also prevents tokens getting stuck. This result provides a direct answer to the second subquestion and reinforces the broader conclusion that structural characteristics of the network are closely tied to long-term engagement. The next section continues this line of inquiry by examining more specific recurrent patterns in the form of temporal motifs.

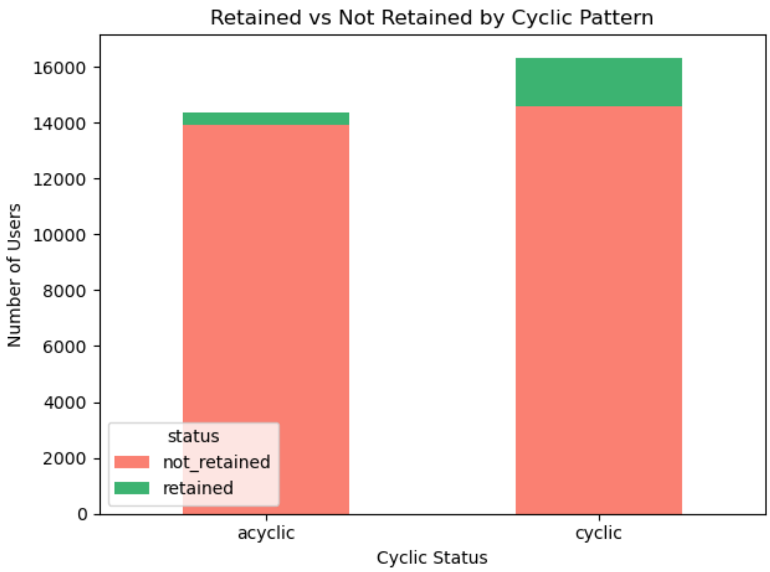


Figure 12: Retention counts by cyclic and acyclic user types

6.1.7 Motif Participation

Subquestion three explores whether recurring structural patterns at the local level, specifically in the form of small transaction motifs, help explain differences in user retention. The goal is to assess whether users who participate in such patterns, either more frequently or with shorter delays, are more likely to remain engaged. Building on the prior analysis of cyclic structures, this section applies motif-based feature extraction to detect two specific patterns, three-node cycles and reciprocal exchanges. Motif enumeration was applied to the directed transaction network, as described in Subsection 5.5. Both motif types were found to follow the overall activity patterns of the Sarafu system. As shown in Figure 13, their monthly occurrence closely matches transaction volume, indicating that structural repetition was part of the system's regular operation rather than an exceptional behavior.

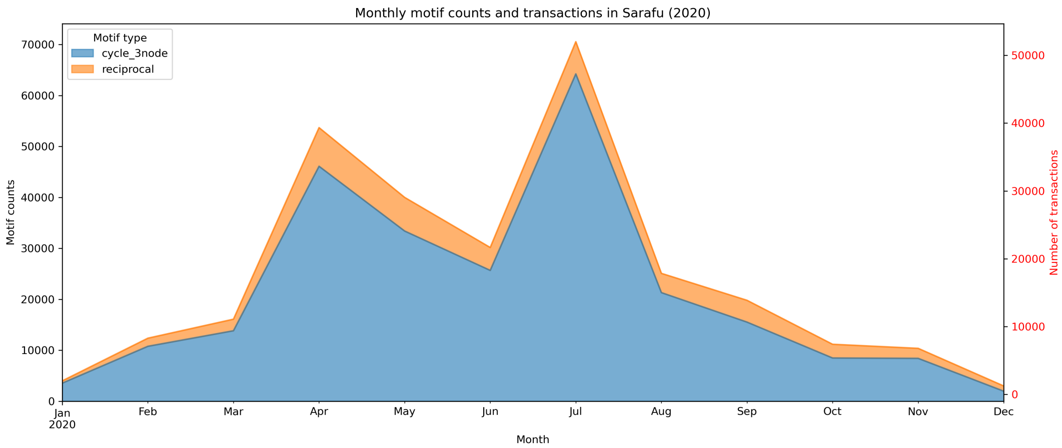
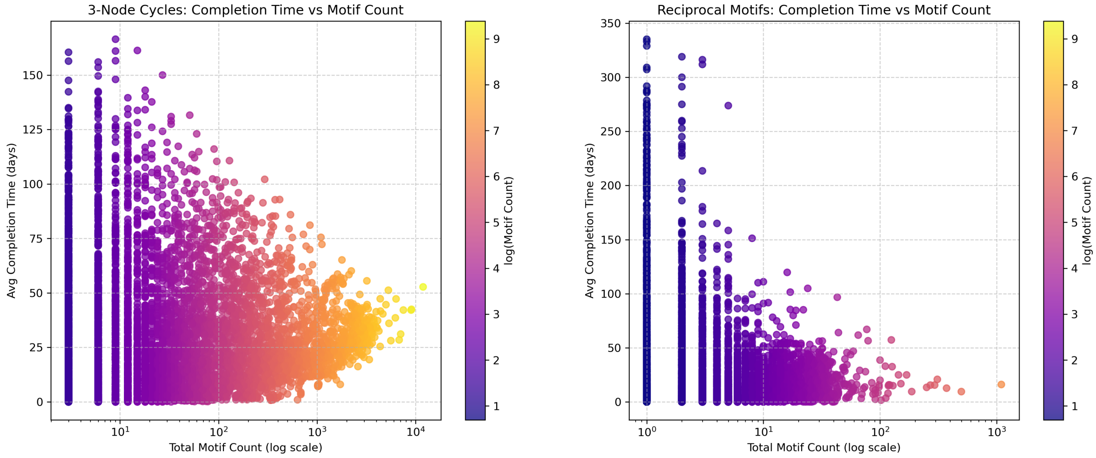


Figure 13: Monthly trends in motif counts and transaction volume in 2020

Figure 14 shows the relationship between how often users participate in motifs and how quickly they complete them. In both panels, there is a negative association between the number of motif instances and the average completion time per user. Users who appear more frequently in motifs tend to complete them in shorter time intervals. This pattern is especially visible for reciprocal motifs, where most users with high counts exhibit short average delays. While the plots do not distinguish between retention groups, they provide a baseline indication that speed and frequency of structural repetition often coincide.



(A) Average completion time versus count of three-node cycles

(B) Average completion time versus count of reciprocal motifs

Figure 14: Relationship between motif frequency and completion time

To examine whether timing differences are also reflected in user retention, Figure 15 shows the distribution of motif completion times for both three-node cycles and reciprocal motifs, separated by retention outcome. For three-node cycles, retained and non-retained users follow very similar patterns, with only a modest tendency for retained users to complete motifs earlier. For reciprocal motifs, non-retained users show a clear peak at very short intervals, suggesting immediate but short-lived reciprocation, whereas retained users display a broader distribution concentrated at relatively short times but extending further. This indicates that while both groups engage in rapid exchanges, sustained and repeated short-interval reciprocation is more typical of users who continue to participate in the system.

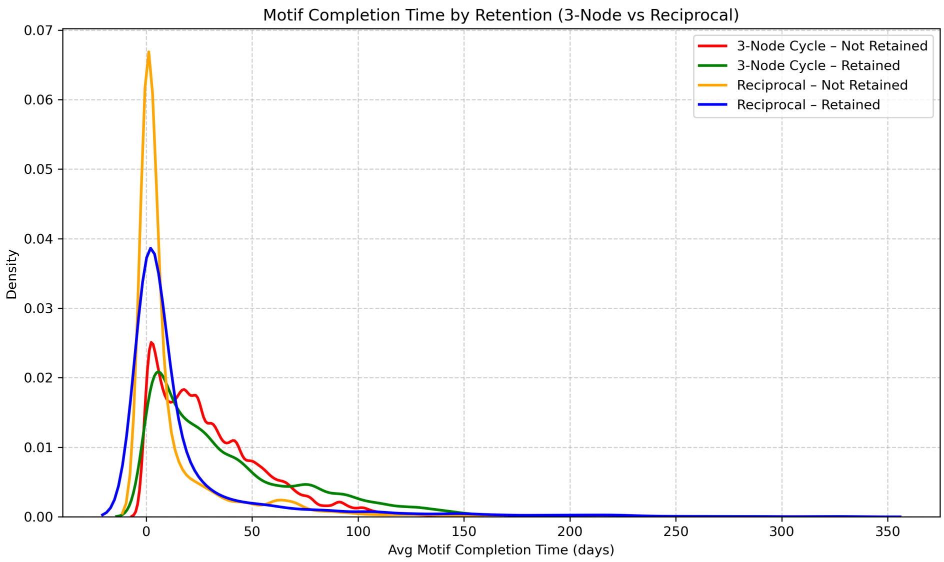


Figure 15: Motif completion time by retention label for both three-node cycles and reciprocal motifs.

These results contribute directly to subquestion three. Users who participate in more motifs tend to complete them more quickly, regardless of retention outcome. When retention groups are compared, this pattern becomes more distinct. Retained users show shorter completion intervals, especially in reciprocal motifs. These findings suggest that frequent and timely mutual exchange is more common among users who remain in the system.

6.2 Model Performance and Feature Contributions

We evaluated the models introduced in Section 5.6.1 to assess whether different types of features can predict long-term user retention in the Sarafu network, as an answer to our main research question. The goal was to identify which aspects of user behavior, observed during early activity in 2020, are most indicative of continued participation. Results from model I to V are shown in Table 11, while robustness across different retention thresholds is illustrated in Figure 16. The corresponding feature rankings are shown in Figures 17 through 21.

We begin with Model I, which used only global centralities. All training and evaluation steps followed the procedure described in Section 5.6. Despite this, the model was unable to reliably detect retained users. The recall remained low and the F1-score showed limited improvement. This outcome reflects the strong imbalance in the data. The limited number of positive cases reduced the model's ability to distinguish between classes when relying only on global features. As shown in Section 6.1.5, these features did not consistently separate retained from non-retained users. The classifier could not compensate for this through resampling or threshold calibration alone that explained in Subsection 5.6.5.

In Model II, the same metrics were computed within communities. The change in feature scope led to higher recall and F1-score, and also a significant improvement in precision. No changes were made to the model design. This shows that differences between retained and non-retained users were more stable at this level.

Model III combined both feature sets. However, results were almost identical to Model II. This shows that the global features did not add new predictive value once the community-level features were included. Since the training pipeline remained unchanged, the lack of further improvement must be due to redundancy between the two inputs. The classifier gave similar predictions with or without the added features, which confirms that they did not help in resolving the imbalance.

The strongest improvement was observed in Model IV, which extended the previous configuration by including cyclic patterns and temporal motifs. As discussed in Sections 6.1.6 and 6.1.7, these features capture repeated interaction and the timing of return flows in the network. Their inclusion led to a clear increase in both recall and F1-score. This result is consistent with earlier findings that retained users were more often part of closed exchange structures and completed them faster. In addition to improving prediction on the main test set, this model also performed well when the threshold in the definition of retention was varied. As shown in Figure 16, its recall and F1-score remained almost stable across different thresholds. This suggests that the added features are not only effective but also less sensitive to how retention is defined.

Model V introduced user-level demographic information in addition to all prior variables. While its performance remained close to the Model IV, the additional attributes did not lead to further improvements. This indicates that structural and temporal behavior explains the observed retention more, and the demographic characteristics may adding noise to the model.

According to all models, the most meaningful performance gains were not the result of increased complexity or feature quantity, but of including features that reflect behavioral timing and local interaction structure. These results point to the importance of meso-level and micro-level regularities in predicting which users continue to engage in a decentralized currency system. Although all models achieved relatively high ROC AUC and accuracy scores, this metric primarily reflects ranking ability and does not account for class imbalance or decision threshold effects. In this context, F1-score and recall were considered more meaningful indicators of predictive value.

Table 11: Comparison table for models using different features and per class

Model	Set	Class	Accuracy	Precision	Recall	F1-score	ROC AUC
I	Validation	0	0.83	0.96	0.86	0.90	0.71
		1	0.83	0.20	0.48	0.29	0.71
	Test	0	0.81	0.95	0.84	0.89	0.70
		1	0.81	0.16	0.39	0.22	0.70
II	Validation	0	0.88	0.96	0.92	0.94	0.78
		1	0.88	0.29	0.44	0.35	0.78
	Test	0	0.89	0.95	0.92	0.94	0.77
		1	0.89	0.30	0.43	0.35	0.77
III	Validation	0	0.90	0.95	0.94	0.94	0.78
		1	0.90	0.32	0.39	0.35	0.78
	Test	0	0.90	0.95	0.94	0.94	0.77
		1	0.90	0.30	0.35	0.33	0.77
IV	Validation	0	0.90	0.96	0.93	0.94	0.81
		1	0.90	0.33	0.44	0.38	0.81
	Test	0	0.90	0.96	0.93	0.94	0.82
		1	0.90	0.34	0.48	0.40	0.82
V	Validation	0	0.90	0.96	0.94	0.95	0.81
		1	0.90	0.36	0.44	0.39	0.81
	Test	0	0.90	0.96	0.93	0.94	0.81
		1	0.90	0.33	0.45	0.38	0.81

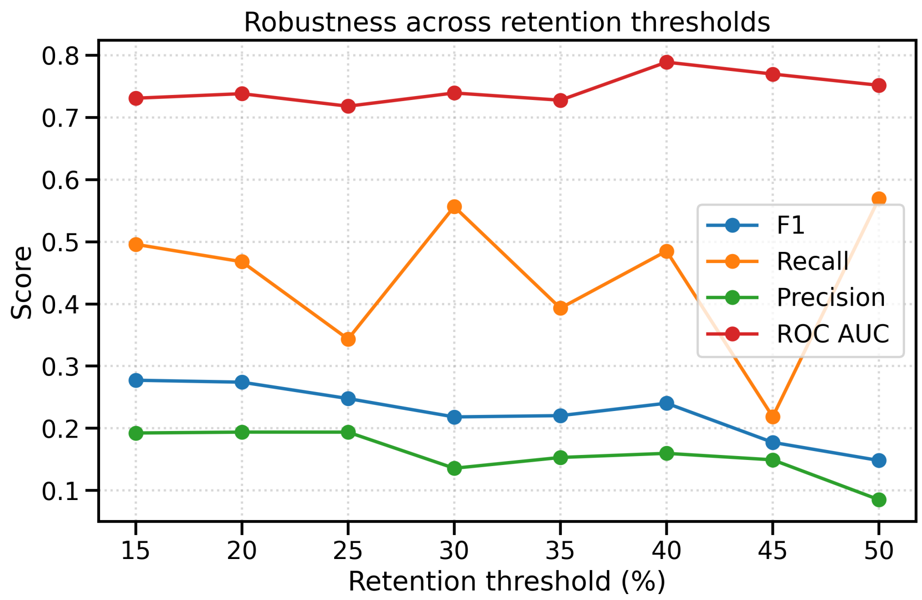


Figure 16: Robustness of Model IV across retention thresholds.

Feature Importance and Interpretability

This part of the analysis examines which features had the strongest influence on model predictions and how their rankings relate to actual predictive value. Figures 17 to 21 show the most important features identified in each model from I to V. While some features that contributed to performance gains also ranked high, others appeared influential despite not improving the model.

In Model I, which used only global features introduced in Section 5.6.1, degree ranked highest, followed by the clustering coefficient and weighted degree, while betweenness scored lowest in Figure 17. Yet, as shown earlier in Section 6.1.5, this model performed poorly. The strong ranking of global features in this restricted setting does not imply strong predictive value. Once additional feature types were introduced in Models II and III, global indicators were quickly displaced in the rankings, confirming that their prominence was relative rather than absolute. This can also be seen in Figure 17, and Figure 18.

The introduction of community-level structure in Model II marked a clear shift. Clustering within communities emerged as the top predictor, consistently ranking highest across configurations where it was included. Degree within communities and group size also ranked high. These features proved more reliable for identifying retained users than any global measure, as illustrated in Figure 18 and Figure 19.

When motif and cyclic features were added in Model IV, their influence on predictions became clear in terms of model performance. Cyclic status consistently appeared among the top features, but motif counts and timing features did not. This difference can be explained by feature interaction and distribution. Motif features are spread across several related variables and their contribution becomes visible in improved performance, but they rank lower individually because their effects are distributed. This is shown in Figure 20. Moreover, global centrality features, particularly weighted degree and betweenness, consistently ranked low once structural and temporal features were included. Their drop confirms that they are easily replaced by more targeted indicators.

Demographic features highlight a mismatch between apparent importance and actual utility. In Model V, variables such as area type, business type, and gender appeared among the top ten. However, their inclusion did not improve the model's ability to identify retained users. The best-performing configuration in fact excluded them. This indicates that their high ranking is not a reflection of direct predictive value, but rather of overlap with patterns already captured by structural indicators. For instance, the prominence of area type may relate to the fact that certain regions dominate the user distribution, as also shown in Figure 2. The ranking for Model V is shown in Figure 21.

Looking at all results, the features that contributed most to prediction were those that captured structural repetition and interaction patterns over time. The importance of some variables in Model V does not imply that they added value; rather, it reflects the way the model redistributed importance among correlated inputs. A comparison between model performance in Table 11 and feature rankings in Figures 17 through 21 shows that interpretability tools must be read carefully, especially in settings with high feature overlap.

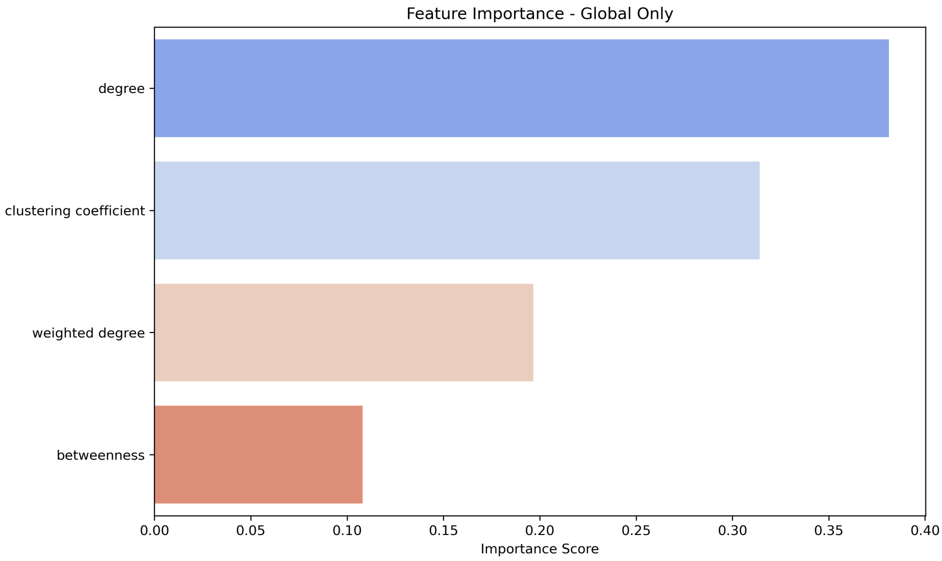


Figure 17: Feature importance scores for Model I.

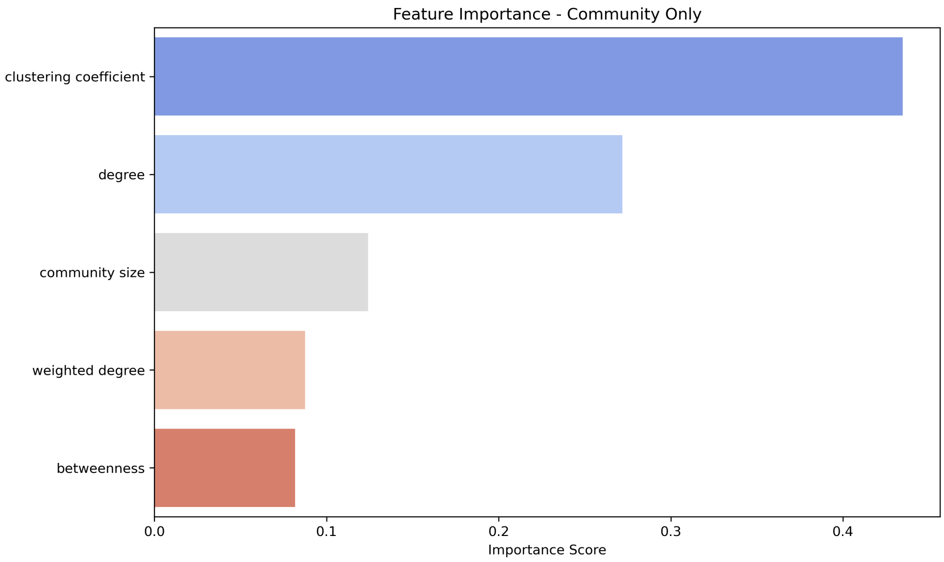


Figure 18: Feature importance scores for Model II.

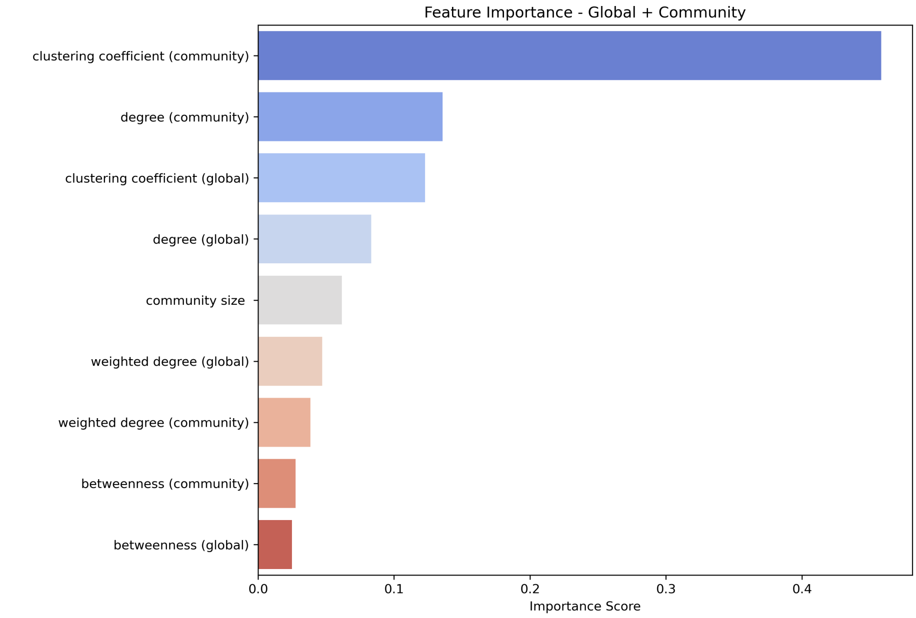


Figure 19: Feature importance scores for Model III.

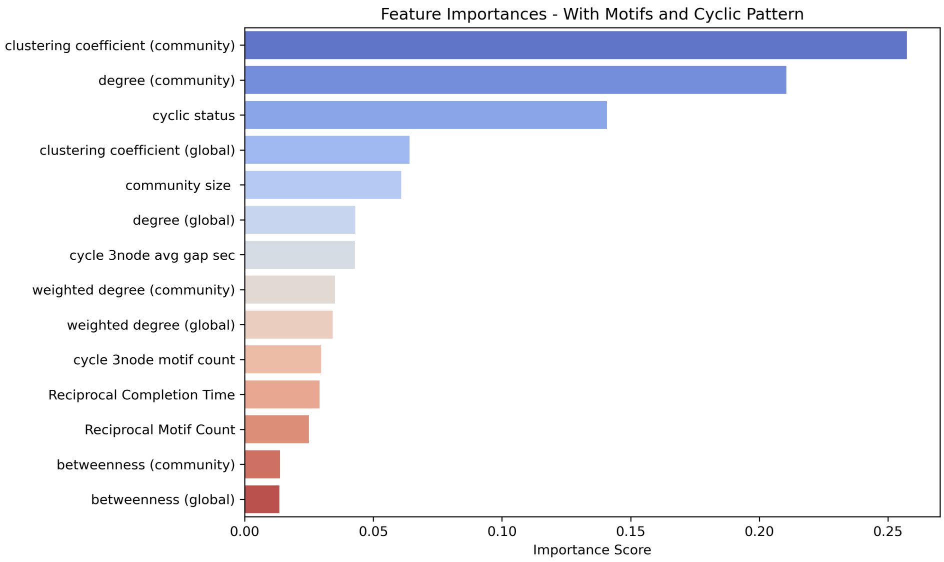


Figure 20: Feature importance scores for Model IV.

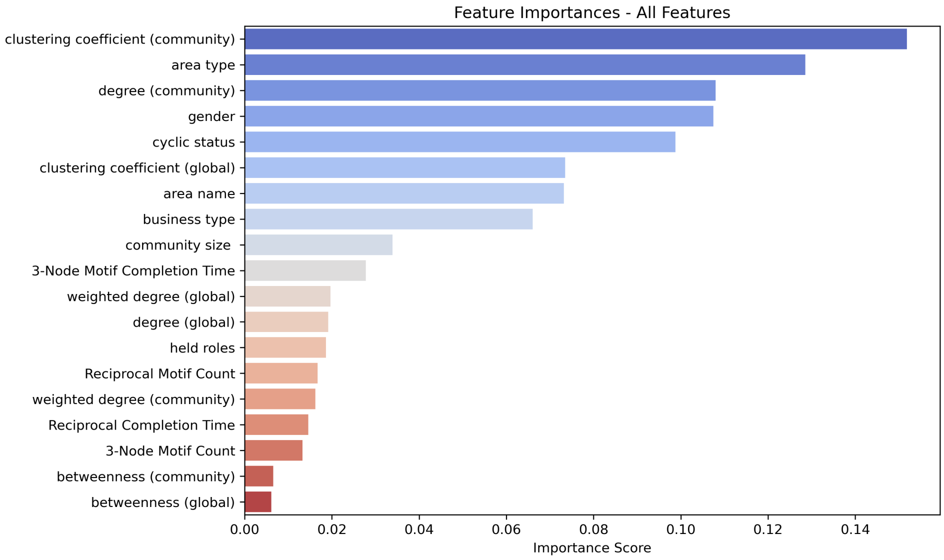


Figure 21: Feature importance scores for Model V.

7 Conclusion

This thesis investigated which types of early network features best explain and predict long-term user retention in the Sarafu community currency system. Building on a directed transaction network constructed from peer-to-peer exchange data, the analysis integrated structural, temporal, and demographic dimensions to evaluate sustained engagement and answer the main research question of which features are most informative for predicting whether users remain active over time.

Rather than relying on absolute activity levels or global prominence, the findings show that retention is primarily shaped by localized interaction patterns. In relation to Subquestion 1, users with high clustering and degree within their community were more likely to remain active, while global centrality indicators were weak predictors, as they often capture one-sided or short-lived connections.

Further, a strong relationship was observed between retention and cyclic positioning, directly answering Subquestion 2. More than ninety percent of retained users belonged to strongly connected components, reinforcing the notion that mutual and recurrent transactions serve as structural anchors in decentralized systems. These effects were further reflected in motif-based features, which respond to Subquestion 3. Users who engaged in reciprocal or triadic exchanges with shorter completion times were more likely to stay, showing that faster and repeated interactions are linked to higher retention.

The predictive models confirmed these insights and supported the expectations set out in the subquestions. Including community-based metrics, cyclic status, and motif timing improved classification performance significantly compared to models based solely on global features. The most informative predictors were those that captured both position within the network and the rhythm of interactions over time, thus providing a clear answer to the main research question.

While all models achieved strong ranking performance based on the ROC AUC, this metric alone did not reflect the ability to identify retained users. In a highly imbalanced setting, recall and F1-score offered a more reliable view of classification effectiveness. This choice of metrics shaped the evaluation focus of the study. Similarly, the interpretation of feature importance required careful consideration. Certain features, such as demographic attributes, appeared among the top-ranked variables in the importance scores. However, when included in the models, they did not lead to better predictive performance. This shows that importance scores alone do not necessarily reflect whether a feature adds value, and they need to be interpreted alongside model outcomes.

Altogether, the results suggest that sustained participation in community currency networks depends less on visibility or intensity of activity and more on embeddedness in mutual, responsive, and locally grounded patterns of exchange. These findings point to practical implications for designing inclusive financial systems, where supporting peer-based interaction may be more effective than targeting users with high global prominence. The study also contributes to a broader understanding of user retention by linking structural, temporal, and behavioural dimensions of engagement.

8 Limitations and Future Work

The findings of this thesis should be interpreted in light of several limitations related to data scope, feature selection and modeling design. A key limitation arises from the fixed observation period used in this study, spanning user transactions from January 2020 to June 2021. Retention labeling was applied using only the first six months of 2021. As such, the analysis captures medium-term engagement patterns, but may not reflect longer-term behavior or the effects of later system interventions. A more extended observation window, if available, could offer deeper insights into user trajectories and refine the definition of sustained participation. This study also limited its motif analysis to two types of temporal motifs, reciprocal pairs and three-node cycles. These patterns were selected based on their behavioral interpretability and relevance in peer-to-peer systems. While their inclusion allowed for tractable and targeted analysis, it excludes other motifs, such as stars, cascades, or higher-order cycles. Additional motif types might capture alternative forms of interaction or support more fine-grained behavioral characterization. Future work could revisit this design choice once coverage and interpretability can be balanced more effectively. Moreover, the predictive modeling was conducted using XGBoost, a tree-based classification model selected for its robustness with structured data and missing values. While performance was evaluated using standard metrics and class balancing was addressed, the dataset remains highly imbalanced, with relatively few retained users. This imbalance can influence the classifier's sensitivity, especially in boundary cases. The choice of algorithm and evaluation design limits the scope for assessing uncertainty across different user groups.

This analysis can be extended in multiple ways. One promising area is the integration of language-based feature representations. A small-scale extension explored the potential of using large language models to encode user-level behavioral profiles, by translating structural features into descriptive text. While these embeddings were not included in the final results, they point to a direction where textual and network-based signals could be jointly modeled. Future work could develop richer user descriptions and apply generative or prompt-based LLM techniques for embedding generation or classification. This line of research could also bridge explainability and prediction, offering a more interpretable model of user engagement. Beyond representation, since the dataset is highly imbalanced and also the features are not simple, alternative modeling approaches may help to improve the evaluation metrics. In particular, future work could explore strategies to directly increase recall and F1-score, such as loss functions tailored to class imbalance, ensemble methods designed for rare-event prediction, or calibrated decision thresholds that adjust for skewed class distributions. This includes models that incorporate time-series patterns, survival analysis, or probabilistic graph models. Finally, a larger or cross-platform dataset could support comparative studies, enabling the identification of general versus system-specific retention factors. Applying the same framework to different community currency networks would clarify whether the structural and temporal predictors found in this thesis are specific to Sarafu or reflect broader dynamics in decentralized economies.

References

[1] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016.

[2] Sajjad Alizadeh and Majid Khabbazian. Solana's transaction network: analysis, insights, and comparison. *EPJ Data Science*, 14:48, 2025.

[3] Carolina E. S. Mattsson, Teodoro Criscione, and Frank W. Takes. Circulation of a digital community currency. *Scientific Reports*, 13(1):5864, 2023.

[4] William Jack and Tavneet Suri. Network structure and financial access: Evidence from mobile money in tanzania. *American Economic Review*, 111(3):940–977, 2021.

[5] Asif Ahmad, M Imran Khan, and Shahbaz Ahmad. Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1):1–24, 2019.

[6] Marijn Oskarsdottir, Cristobal Bravo, Carlos Sarraute, Jan Vanthienen, and Bart Baesens. Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications*, 85:204–220, 2020.

[7] Jichang Zhang and Yi Luo. Degree centrality, betweenness centrality, and closeness centrality in social network. In *2017 2nd International Conference on Systems, Control and Communications (ICSCC)*, pages 26–29. Atlantis Press, 2017.

[8] Deyuan Yang, Jinyan Wang, and Mingzhe Li. Local versus global centrality: a structural approach to node importance in complex networks. *Scientific Reports*, 9:11043, 2019.

[9] Thomas Magelinski, Mihovil Bartulovic, and Kathleen M. Carley. Measuring node contribution to community structure with modularity vitality. *arXiv preprint arXiv:2003.00056*, 2020.

[10] Vincent A. Traag, Ludo Waltman, and Nees Jan van Eck. From louvain to leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233, 2019.

[11] Tianlong Fan, Linyuan Lü, Dinghua Shi, and Tao Zhou. Characterizing cycle structure in complex networks. *Physical Review E*. Preprint originally published on arXiv, 2020.

[12] Thomas Anderl, Joachim Lohmer, and Rainer Lasch. Identifying github trends using temporal analysis of repository activity. In *International Conference on Business Information Systems*, pages 3–16. Springer, 2021.

[13] Hanjo D. Boekhout, Vincent A. Traag, and Frank W. Takes. Investigating scientific mobility in co-authorship networks using multilayer temporal motifs. *Scientific Reports*, 11(1):18175, 2021.

[14] André Luiz Corrêa Vianna Filho, Leonardo de Lima, and Mariana Kleina. A graph-based approach to customer segmentation using the rfm model. *arXiv preprint arXiv:2505.08136*, 2025.

[15] Karsten N. Economou, Cassie R. Norman, and Wendy C. Gentleman. Identifying robust features of community structure in complex networks. *Physical Review E*, 111(4):044303, 2025.

[16] Andrea Lancichinetti and Santo Fortunato. Consensus clustering in complex networks. *Scientific Reports*, 2:336, 2012.

[17] Fairouz Medjahed, Elisenda Molina, and Juan Tejada. Effectiveness of centrality measures for competitive influence diffusion in social networks. *Mathematics*, 13(2):292, 2025.

[18] Shuying Zhao and Shaowei Sun. A study on centrality measures in weighted networks: A case of the aviation network. *AIMS Mathematics*, 9(2):3630–3645, 2024.

[19] Haomin Li and Daniel K. Sewell. Model-based edge clustering for weighted networks with a noise component. *Computational Statistics & Data Analysis*, 2025.

[20] V. Vasiliauskaite, T. S. Evans, and P. Expert. Cycle analysis of directed acyclic graphs. *Physica A: Statistical Mechanics and its Applications*, 596:127097, 2022.

[21] Marijn Óskarsdóttir et al. Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications*, 85:204–220, 2017.

[22] Lauri Kovanen, Martón Karsai, Kimmo Kaski, János Kertész, and Jari Saramäki. Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(09):P09005, 2011.

[23] Qiming Zhao, Yichen Tian, Qian He, Nathan Oliver, Rui Jin, and W.-C. Lee. Temporal network motifs in continuous-time systems: Toward a standard lens. *Journal of Complex Networks*, 2025.

[24] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[25] Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

[26] Ashwin Paranjape, Austin R Benson, and Jure Leskovec. Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 601–610. ACM, 2017.

[27] Ahmet Erdem Sarıyüce. A powerful lens for temporal network analysis: Temporal motifs. *Journal of Computational Social Science*, 8(1):115–139, 2025.

[28] Zhongqiang Liu, Yuxuan Zhang, Himabindu Lakkaraju, Yuyuan Zha, and Yaqing Wang. Temporal motifs for financial networks: A study on mercari, jpmc, and venmo platforms. *arXiv preprint arXiv:2301.07791*, 2023.

[29] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[30] Mark Nielsen. Tree boosting with xgboost, why does xgboost win "every" machine learning competition, 2016.

[31] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, 2019.

[32] Juozas Vaicenavicius, David Widmann, Fredrik Lindsten, Jacob Roll, Carl Mattias Lindqvist, Volodymyr Kuleshov, Fredrik Gustafsson, and Thomas B. Schön. Evaluating model calibration in classification. *Proceedings of the 36th International Conference on Machine Learning*, 97:6415–6424, 2019.