Published: 13 June 2025
Online AM: 10 January 2024
Accepted: 12 December 2023
Revised: 01 November 2023
Received: 30 March 2023

RESEARCH-ARTICLE

Analyzing User Migration in Blockchain Online Social Networks through Network Structure and Discussion Topics of Communities on Multilayer Networks

CHEICK TIDIANE BA, University of Milan, Milan, MI, Italy

MANUEL DILEO, University of Milan, Milan, MI, Italy

ALESSIA GALDEMAN, University of Milan, Milan, MI, Italy

MATTEO ZIGNANI, University of Milan, Milan, MI, Italy

SABRINA GAITO, University of Milan, Milan, MI, Italy

Distributed Ledger Technologies: Research and Practice, Volume 4, Issue 2 (June 2025)
https://doi.org/10.1145/3640020
EISSN: 2769-6472

Analyzing User Migration in Blockchain Online Social Networks through Network Structure and Discussion Topics of Communities on Multilayer Networks

CHEICK TIDIANE BA, Computer Science Department, University of Milan, Milan, Italy and Queen Mary University London, London, United Kingdom
MANUEL DILEO, ALESSIA GALDEMAN, MATTEO ZIGNANI, and SABRINA GAITO, Computer Science Department, University of Milan, Milan, Italy

User migration (i.e., the movement of large sets of users from one online social platform to another one) is one of the main phenomena occurring in modern online social networks and even involves the most recent alternative paradigms of online social networks, such as blockchain-based online social networks. In these platforms, user migration mainly occurs through hard forks of the supporting blockchain (i.e., a split of the original blockchain and the creation of an alternative blockchain), to which users may decide to migrate. However, our understanding of user migration and its mechanisms is still limited, particularly regarding the role of densely connected user groups (communities) during migration and fork events. Are there differences between users who stay and those who decide to leave, in terms of network structure and discussion topics? In this work, we show, through network-based analysis centered on the identification of communities on multilayer networks and text mining that (a) the "position" of a group within the network of social and economic interactions is connected to the likelihood of a group to migrate (i.e., marginal groups are more likely to leave); (b) group network structure is also important, as users in densely connected groups interacting through monetary transactions are more likely to stay; (c) users who leave are characterized by different discussion topics; and (d) user groups interacting through monetary transactions show interest in migration-related content if they are going to leave. These findings highlight the importance of social and economic relationships between users during a user migration caused by fork events In general, in the larger context of online social media, it motivates the need to investigate user migration through a network-inspired approach based on groups and specific subgraphs while leveraging user-generated content, at the same time.

CCS Concepts: • **Networks → Online social networks**; **Social media networks**; • **Computer systems organization → Peer-to-peer architectures**; • **Computing methodologies → Topic modeling;**

Additional Key Words and Phrases: User migration, blockchain online social networks, multi-layer network, community detection, topic modeling, text retrieval

ACM Reference Format:
Cheick Tidiane Ba, Manuel Dileo, Alessia Galdeman, Matteo Zignani, and Sabrina Gaito. 2025. Analyzing User Migration in Blockchain Online Social Networks through Network Structure and Discussion Topics of Communities on Multilayer Networks. *Distrib. Ledger Technol.* 4, 2, Article 10 (June 2025), 19 pages. https://doi.org/10.1145/3640020

This work was supported in part by project SERICS (PE00000014) under the NRRP MUR program funded by the EU–NGEU.
Authors' addresses: C. T. Ba, Computer Science Department, University of Milan, Via Celoria, 18, Milan, Italy and Queen Mary University London, London, UK; e-mail: cheick.ba@unimi.it; M. Dileo, A. Galdeman, M. Zignani, and S. Gaito, Computer Science Department, University of Milan, Via Celoria, 18, Milan, Italy; e-mails: {manuel.dileo, alessia.galdeman, matteo.zignani, sabrina.gaito}@unimi.it.

ACM 2769-6472/2025/06-ART10
https://doi.org/10.1145/3640020

1 INTRODUCTION

Today's online social media landscape is still very dynamic, with new platforms and services frequently entering the market to compete for the attention of ever-expanding audiences against established global platforms such as Facebook, Instagram, X (formerly known as Twitter), and Weibo. Those established platforms are experiencing problems that are degrading the user experience, including censorship, the spread of false information, privacy concerns, data leakage, and misinformation. As a result, many users are abandoning popular platforms in favor of different ones that promise to resolve or at least mitigate the aforementioned problems. The term for this phenomenon is *user migration* and has emerged as one of the main issues that modern centralized online social media are facing. User migration is not limited to well-established and centralized social platforms, but it is a widespread phenomenon even involving the most recent alternative paradigms of online social media. **Blockchain-based Online Social Networks (BOSNs)** are becoming the most promising solution among them. BOSNs are the result of the application of the Web3 paradigm (i.e., the design of platforms and software systems built on blockchain technologies) to the field of online social media. Essentially, a BOSN leverages a blockchain to support all the functionalities that facilitate social interactions, along with some advantages and peculiar features, such as a reward system based on cryptocurrencies, which aims at promoting worthy behaviors such as the production of high-quality content or the reporting of misinformation and fake news. User migration in BOSNs is often a consequence of a hard fork (i.e., a bifurcation of the main branch of the original blockchain) that allows users to create new platforms originating from the original BOSN. While user migration is an important issue, our understanding of user migration and its mechanisms is still limited, mainly due to a lack of precise and high-resolution data on the process.

In this article, we deal with user migration as a result of a hard fork in blockchain-based social media. Specifically, we analyze user migration through the lens of network science from a mesoscale perspective [24]. In fact, our main goal is to highlight the role of groups, identified through community detection algorithms [35], during the user migration process. It is well established that groups or densely connected regions of a social network may exert pressure through peer influence on the choices of their members. So, groups may play a fundamental role when users have to decide whether to migrate or not. Note that groups may arise not only through social interactions but also around economic interest: in BOSNs, social interactions are complemented by monetary interactions supported by the underlying blockchain. Another interesting aspect of migration in social media is that users can discuss and organize the fork on the same platform they are about to leave. Group discussions between those who want to stay and those who want to leave can reflect their future intentions, and vice versa, it could influence users' future decisions. These aspects led us to the following research questions:

- *RQ1:* Are there differences in the network structure of groups of users who stay and those who decide to leave for a new platform?
- *RQ2:* Is there any interplay between what groups of users write before the fork and their migration choice?

To answer, we focus on the ecosystem of social platforms based on the Steem blockchain, whose main member is Steemit, and Hive, the blockchain originating from a hard fork of the Steem blockchain on March 20, 2020. The two platforms were chosen as a case study for their substantial user bases as well as the explicit linkage of profiles, resulting from data copying during the fork event, that eliminates the need for profile-matching techniques—an attractive feature compared to other cross-platform migration scenarios. We gathered data from both publicly available blockchains and represented the interactions among their accounts by a multilayer temporal network, to distinguish between the networked structure determined by social interactions and the one resulting from monetary transactions. Then, we identified groups—communities—on both layers by applying one of the state-of-the-art community detection algorithms for multilayer networks. By inspecting users' activity on both blockchains, we also identified users who have migrated after the hard fork. Then, by combining the information about groups and migrating users, we analyzed how groups are composed in terms of migrant and resident users, and which are the relationships among the groups. In addition, we leverage user discussion data to understand the interplay between group discussions before the fork and user migration choices, analyzing two components related to text content: *hashtags* (i.e., words preceded by a hash mark (#) and used to categorize content), as well as *topics* extracted using topic modeling methods. Our analysis, applied on both layers, has highlighted the following main findings:

— How groups—communities—are embedded into the network of the communities is crucial in determining whether their members will migrate or not. Specifically, marginal groups, loosely connected to the core of the community network, are more likely to contain members who will migrate to another platform. Moreover, the density of a group (i.e., how it is tight-knit) has a stronger impact on the decision to migrate or stay in the monetary layer rather than in the social one. It may be the first evidence that peer influence exerts more efficiently through economic interactions (RQ1).
— What users discuss highlights important characteristics of communities that plan to stay and those that will migrate. In fact, there are differences in the usage of hashtags and topics of interest between migrant and resident users. A key takeaway is that migrant communities show a stronger interest in migration-related hashtags as well as migration-related topics when the community is born from monetary interactions (RQ2).

The article is structured as follows. The concepts and works on user migration in online social networks that are most pertinent to our work are introduced in Section 2. The modeling of the monetary and social exchanges recorded in the blockchains, as well as the techniques for identifying and describing groups and users who migrate, are covered in Section 4. In Section 5, we provide an overview of the dataset, and in Section 6, we present our findings regarding how the characteristics of groups and the network structure among groups influence user migration. The article concludes in Section 7, which also suggests potential directions for future research.

2 BACKGROUND AND RELATED WORKS

Blockchain Online Social Networks. The field of online services has changed significantly over the past decade, with attention shifting away from monolithic centralized systems and toward open, decentralized, and distributed alternatives [31]. This revolution is commonly referred to as Web3, which refers to the development of platforms and software systems based on blockchain technology to develop a decentralized Web [43]. A blockchain is one of the possible implementations of a distributed ledger [12]. Single fragments of information, known as *transactions*, are collected together into *blocks*, and each block is cryptographically connected to its predecessor as the links of a chain. To add a new block to the chain, specialized nodes known as *miners* must compete or collaborate following a consensus mechanism. Initially, Web3 principles were applied in the economy, leading to decentralized finance (DeFi). A well-known example is Bitcoin, where blockchain technology was applied to store economic transactions among a network of untrusted nodes. However, Web3 principles have been applied to many more fields: to cite a few well-known examples, the application in governance resulted in decentralized autonomous organizations (DAOs), or in social networks, it resulted in BOSNs. Because of their nature, BOSNs alleviate certain frequent issues with regular online social networks, such as the so-called *single point of failure*. From the point of view of the users, BOSNs are particularly resilient to *censorship*. One of the most enticing features of blockchain technology in this sector is its ability to bring value and usefulness to social platforms by establishing a *rewarding system* for good contributions. The incentive system can be designed to encourage positive behavior in many elements of the platform, but its major focus is on the awarding of outstanding content and its thoughtful evaluation [15, 34]. Rewards are typically provided as cryptocurrency tokens, which add a new dimension to regular online social networks. In fact, user interactions in typical online social networks are simply "social": users upload and share material on the platform, and other users interact via comments or votes to indicate likes or dislikes. Users in BOSNs may also interact through "economic" or "financial" interactions, as they are able to exchange cryptocurrency tokens through asset transfer operations (i.e., moving a particular quantity of tokens from a source account to a destination account). Nevertheless, blockchain technology also has some disadvantages: it is affected by issues concerning the consensus protocol, such as the 51% attack [38]. Another challenge encountered by decentralized social systems regards content moderation, a question that has yet to be fully answered. In terms of platforms, the first proposal of BOSN is *Steemit* [18, 20], a platform launched in 2016 and one of the most widespread decentralized online social platforms. Following a contentious set of events, some of its users developed Hive via a hard fork, with *Hive Blog* as its primary interface. Other platforms such as *Sapien*[1] and *Minds*[2] are built on Amazon Web Services and use the Ethereum blockchain to host their own ERC-20 tokens.

Among these platforms, Steemit has gathered the interest of researchers for its characteristics. Steemit is regarded as a pioneer for the Web3 ecosystem since it introduced the seminal concepts of the rewarding system in a social network [27, 32] and delegated proof-of-stake (DPoS) consensus algorithm for block validation in social networks apps. Users on Steemit publish and share multimedia content, and they are able to interact with the content using comments and votes. Users can follow other users to receive notifications when new content is posted by other users. Steemit was among the first to establish a reward system, with users receiving cryptocurrency tokens for creating the most successful articles. These tokens can then be exchanged for products or services with other users, and both "social" and "financial" interactions are recorded on the Steem blockchain. Recent work on Steemit includes studies on social network structure [21, 27] and communities [19], economic aspects [41, 44], and the bursty dynamics of the link creation process [3]. Other works focused on the interplay of economic aspects and network structure. For instance, Li and Palanisamy [32] and Guidi et al. [22] have analyzed the rewarding system in Steemit from a network perspective. In contrast, Ba et al. [4] discussed the interplay between cryptocurrency price and the link creation process. Tang et al. [40] model voting and currency transfer data to investigate user collusion behavior. Other works leverage user-generated content for text mining and bot detection [13, 26, 28]. Li et al. [33] rely on Steem data to conduct an individual-level measurement study comparing decentralization in blockchains with different consensus protocols, introducing a comparison framework and a new metric, revealing insights into decentralization levels, and suggesting individual-level centralization risks in Steem.

User Migration and Hard Fork. Online social platforms have been offering services to attract and foster huge and active communities, but for a variety of reasons, some of its users have chosen to move to other platforms. *User migration* is the term used to describe this occurrence. User migration is a "universal" process that occurs in both centralized and decentralized online social networks. However, in BOSNS, it is usually tied to fork events—a key feature of blockchain-based systems. A fork event is when miners modify the consensus protocol. There are two types of fork: the first one is the *soft fork*, which occurs when miners make modifications to the consensus protocol, but those changes are still backward compatible with the prior consensus protocol. As a result, miners will continue to add new blocks to the same chain. This type of fork is used to make minor changes to the consensus protocol, freeze account money, or reverse specific transactions. The other type is the *hard fork*, which will cause miners to reject blocks certified using a different protocol. If the decision between protocols is not made, two distinct branches will be formed. For example, Steemit experienced a hard fork event, where some of its users created the Hive blockchain via a hard fork, effectively creating a new social networks platform with its own interface—Hive Blog—and cryptocurrency system, and causing a user migration. Indeed, events of this kind pose a significant risk to the survival of the platforms, due to the division of user, miner, and developer populations [45].

Despite being an increasingly common event in both traditional online social networks and the blockchain social network environment, user migration is not well understood, mostly because of a shortage of accurate

[1]https://www.sapien.network/
[2]https://www.minds.com/

and high-resolution data on the process. Indeed, there are a few studies on cross-platform user migration, such as that of Kumar et al. [30], which have investigated user movement patterns by matching user accounts with external sources of information such as BlogCatalog. Cross-platform migration was further investigated in the work of Newell et al. [36], where the authors conducted a macroscopic study of user behavior based on user surveys to identify migration reasons, focused on permanent user migration from Reddit to alternative websites, where users are matched using an algorithmic approach. Other works concentrate on a more particular sort of user migration, such as people moving across communities on the same platform. For example, Senaweera et al. [39] demonstrated the occurrence of nonrandom migration patterns using graph-based modeling, which considers Facebook groups as vertices and weighted edges indicate the number of individuals traveling across them, whereas Davies et al. [9] identified and quantified migration in COVID-19-related subreddits both at the microscopic time scale (attention migration, the shift of activity from post to post) and at a macroscopic time scale (shift of activity of entire groups). Although these studies provide useful insights into the phenomenon of user migration, only two of them take into account textual content. Kumar et al. [30] considered the production of posts and comments as user activities but did not analyze the textual content, whereas Newell et al. [36] took into account only the content of a small subset of annotated comments. In a concurrent study, He et al. [23] analyzed the migration of more than 136,000 X users to Mastodon following Elon Musk's acquisition of X. The study explores the impact on Mastodon's ecosystem, factors influencing migration, and differences in user behavior and topics between the two platforms. Classifiers are developed, revealing that tweet content, number of URLs, likes, and tweet length are effective metrics for predicting user migration. Even though it is still an open problem, work on user migration has been limited mostly due to the difficulty in obtaining suitable longitudinal data. Another critical issue is user account matching, which involves following individuals across platforms with various usernames. Blockchain technology facilitates this sort of study since access to its data provides researchers with a dependable supply of longitudinal data. In addition, unlike other platforms, account matching in BOSNs is a simple operation since user accounts are replicated upon a hard fork. This has led to some studies on fork-based user migration in BOSNs. Ba et al. [2] focused on user migration in Web3 social platforms by evaluating the effects of user migration on the graph structure of the interactions and assessing the predictability of migrating, whereas Galdeman et al. [16] studied the influence of hubs on the user migration decision on their direct neighbors and found that users directly interacting with hubs tend to migrate. Ba et al. [1] proposed a machine learning pipeline, utilizing graph neural networks on directed temporal multilayer graphs to predict user migration in blockchain online social media. The study introduces a data-level balancing technique to handle class imbalance, demonstrating that graph neural networks are effective in predicting user migration.

Content-Based Social Network Analysis. Integrating both topological and content-based approaches has proven effective in several studies addressing various aspects of online social network analysis. For example, Garimella et al. [17] examined user browsing histories by leveraging the link structure of online news networks and users' explicit content choices, contributing to a better understanding of polarization in online news consumption. Villa et al. [42] introduced an approach that applies community detection strategies to distinct representations that incorporate both topology and content aspects of the COVID-19 conversation graph on X to detect echo chambers. Kumar et al. [29] utilized the Reddit hyperlink network and employed Word2Vec-based user and subreddit embeddings to analyze community interactions and conflicts on the platform. Dileo et al. [13]evaluated the impact of textual content on link prediction in BOSNs using novel deep learning on graph architectures.

This work analyzes the user migration across online social networks presenting a novel methodology for taking into account not only network structure but also communities and the content of a large-scale collection of user-generated posts, comments, and hashtags. Building upon the research conducted in this work, we extend the findings of a previous work [5], which studied the Steemit user migration from a mesoscopic point of view, observing how communities are characterized by different migration behaviors.

3 RESEARCH QUESTIONS

Previous studies suggest that network structure has a key role in user migration processes. However, the role played by densely connected groups of users—communities—during user migration and fork events is still not clear. From a network perspective, we would like to understand the interplay between group network structure and migration. Moreover, another important aspect of migration in social networks is that users are able to discuss and coordinate their choices on the same platform from which they will eventually leave. Discussion between those who want to stay and those who want to leave may indicate their future intent, and vice versa, it may impact users' decision making: we would like to exploit user discussion data to better understand the relationship between what communities of users write before the fork and their migration choice. The previous considerations can be summarized through the following research questions:

RQ1: Are there differences in the network structure of groups of users who stay and those who decide to leave for a new platform?

RQ2: Is there any interplay between what communities discuss before the fork and their migration choice?

4 METHODOLOGY

In this section, we present our proposed methodology. We begin by describing how to model the dataset for the task at hand, how to extract the network structure, and how to construct user migration-related labels before identifying communities. Finally, we provide the approach we intend to use to answer our research questions.

4.1 Modeling BOSN and User Migration

Graph-Based Modeling. Users on a BOSN platform publish and share multimedia content through a web interface, which allows them to browse content authored by other users as well. Users can take actions that influence the social media platform; moreover, the platforms also provide users with actions for social interaction with both other users and the content they post. For example, users can commit actions such as commenting or voting on posts authored by others, or they can follow and receive notifications of new content from an intriguing content creator. In essence, we have many different *types* of interactions supported, with the traditional "social" interactions coexisting with economic or financial operations tied to the transfer of cryptocurrency tokens. Furthermore, each activity is timestamped (i.e., it can occur at any given moment). Each interaction in this scenario can be represented as a tuple (u, v, t, r), where u and v are users who interact through an action of type r at time t. These tuples can be leveraged to create a multilayer network [25] using the sequence of all the users' interactions. We denote this network as $G_{T_{fork}} = (V, E, R)$, where

- V is the set of users u who have participated in at least one interaction action in the set $I = \{(u, v, t, r)\}$ which has occurred before or at the timestamp T_{fork};
- E is the set of triple (u, v, r) with $u, v \in V$ and $r \in R$, representing a specific type of action among the ones in the set R of actions supported by the blockchain.

The resulting multilayer network encodes the structure of the interactions among users prior to T_{fork} (i.e., the hard fork date), where the layers correspond to different types of actions available to support user interaction. In particular, here, we separate social and economic/financial interactions into two separate groups, thus decreasing the number of layers in $G_{T_{fork}}$ to two: the "social" layer and the "monetary" layer.

Given this setting, we are able to model a fork event, and the subsequent cross-platform migration, where users might migrate to another platform.

In this work, as done by other related works on user migration (e.g., [9, 30, 36, 39]), we consider the concept of migration as the shift of user activity from a platform to a new one. These approaches classify users according to their activity on the platform. For example, in the work of Kumar et al. [30], a user is a migrant if it was active on platform A, then stops its activity on A as he becomes active on a second platform B. Similarly, Senaweera et al. [39] classify users as active if they are active in more social groups or as inactive when they are active in a single group. However, the concept of migration is not univocally defined. In this work, we follow a similar approach to these previous works.

Specifically, given two platforms, S and H, and a fork event at time T_{fork}, we consider *migrant* a user who performs at least one action on the new platform H after time T_{fork}; *resident* a user staying on the original platform S, without performing any actions on the new platform H after time T_{fork}; and *inactive* as users who are inactive or abandoned both platforms. Specifically, for migrants, we adopt a definition where the choice of moving to a new platform is more related to the users' reaction toward the causes of the hard fork, leaving space for possible combined usage of both platforms. This is contingent upon the observation that the hard fork on March 20 was generated from a reaction of the Steemit community to the takeover, whereas the preceding hard forks were primarily associated with changes in the algorithm for wealth distribution among users [8].

Community Detection. Our primary purpose is to understand the importance of groups during the user migration process. We use community detection methods to discover groups via the network structure. Among the potential state-of-the-art options for identifying communities in a multilayer network [35], we decided to use *Infomap* [37], a community detection algorithm based on the concept of random walks. We chose this algorithm mainly due to its scalability [35]. In Infomap, the community detection process starts by assigning a codeword to each node, using a prefix-free code such as Huffman: in this way, a random walk on the network can be represented as a concatenation of codes. The basic assumption is that once a random walker joins a denser area of the graph—such as a group or community—it will most likely stay there for a long period. This occurs because each node is more linked to nodes in the same region than to distant nodes. Infomap assigned a different codebook to each region, called *module*, to shorten the codewords that refer to nodes in the same region. Therefore, communities in a network can be identified by finding the partition that minimizes the code length. Since we are modeling data as a multilayer network, we can use the multilayer version of Infomap [10]. The multilayer Infomap functions in a similar way to the single-layer version. The key adjustment is that the same user will be represented in each layer, and each of its copies will be connected through interlayer edges. The random walker can then use the interlayer edges to continue its path through the edges of another layer. It is worth noting that the same user may belong to multiple communities depending on the layer, but information from all layers is considered during the community assignments. As a result, in our setting, where we consider two layers, for social and economic interactions, we can define *social communities* as the community assigned to each node in the social layer, and similarly *monetary communities* for the monetary layer.

4.2 Community Structure and User Migration

To answer our research question regarding the interplay of group network structure and user migration (RQ1), we need to understand the role of groups in user migration. As our starting point, we highlight the relations among communities by constructing a community graph $G^C = (V^C, E^C)$ as an attributed network, with nodes representing communities, and links representing connections between users in different communitie—that is, we draw a link between communities c_1 and c_2 if there is a link between a user in c_1 and a user in c_2, weighted by the number of links connecting nodes in c_1 and nodes in c_2. This construction can be applied using both social and monetary communities, resulting in a *social community graph* and a *monetary community graph*. Relying on community graphs, we are able to perform multiple types of analyses: for each community c_i, we can derive attributes such as the number of inactive, resident, and migrant members, which can be leveraged alongside network structure to obtain various insights. A recap of the construction of community graphs is presented in Figure 1.

Visualizing the Interplay. To answer RQ1, we first analyze the community graphs by concentrating on the *connectivity* among communities, as a function of the migration status of the community members. Moreover, to characterize how a community is unbalanced toward a specific category of users (migrants or residents),

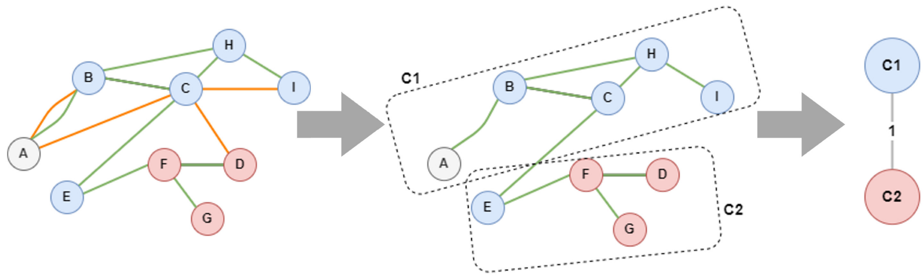


Fig. 1. An example outlining the construction of community graphs. The starting point is the multilayer network where nodes are colored according to their migration decision and can be connected by social links (green) and monetary links (orange). After we detect communities ($c1, c2$), we can derive the corresponding community graph: each community is represented by a node in the community graph, and edges among community nodes are derived from links in the multilayer network. The colors of community nodes are determined by the migration decisions of their members. More precisely, colors represent the majority between migrant and resident nodes: communities with more residents will go toward sky blue, whereas more migrants will lead to red nodes, and white will be used for nodes with a balanced mix of both. Edge width is proportional to the weight of the intercommunity edge.

we compute the community *entropy* $H(c_i)$, defined as $H(c_i) = \sum_{q=1}^{m} p_q(c_i) \log_2 p_q(c_i)$, where $p_q(c_i)$ denotes the fraction of users in the community c_i with label $q \in \{resident, migrant\}$. Additionally, we consider the subgraph induced by the nodes in a community c_i and compute its *density* $D(c_i) = \frac{|E|}{|V|*(|V|-1)}$, where $|E|$ is the number of edges in the subgraph and $|N|$ the number of nodes. In addition, the density and entropy of the communities are then examined as a function of the migration labels.

Measuring the Interplay. Finally, to confirm the insights from the visual examination, we conduct additional quantitative analysis. We consider the number of inactive, residents, and migrants in each community, as well as its density and entropy, and measure the correlation between the selected community features, focusing on density and entropy in relation to other features. It should be noted that some communities may have too few nodes or even no links within them, because nodes may belong to the same community due to information coming from another layer, without being directly connected—a side effect of multilayer Infomap. Finally, we may have communities made up of only inactive nodes or with a majority of inactive nodes. We discard this kind of community from our analysis.

4.3 Community Discussion and User Migration

To answer RQ2, we need to leverage user-generated content in combination with network structure. In social networks, users can communicate through posts or comments, and we are interested in detecting and analyzing the discussion around migration. We decided to analyze two components derived from user posts: hashtags and content topics. *Hashtags* (i.e., words preceded by a hash mark (#)) are used to categorize content and facilitate a search for it. When available, hashtags can be interpreted as a user-made categorization of content. Indeed, by finding hashtags related to user migration, we can effectively detect and analyze the discussion around migration. To directly analyze the text content and its interplay with migration, we propose an analysis revolving on *content topics*: topics are not given—but they can be extracted using topic modeling methods.

In the following, we propose the methodology to analyze hashtags and text content topics.

Hashtags and Communities. For the analysis of hashtags, we represent for each user u a *user hashtag vector*—that is, a vector representing the counts of hashtags (words) is denoted as $\mathbf{v}_u = [v_1, v_2, \ldots, v_K]$, where $\mathbf{v}_u[k]$ represents the count of the k-th hashtag. To compare multiple communities, we define the *community hashtag vector* of a community of users C as the average of their vectors using the pointwise addition, so

$\mathbf{v}_C = \frac{1}{|C|}\sum_{u \in C} \mathbf{v}_u$. The *community hashtag distribution* $\mathbf{D}_C$ can be obtained by normalizing by the sum—that is, $\mathbf{D}_C[k] = \frac{\mathbf{v}_C[k]}{\sum_{i=1}^{|K|}\mathbf{v}_C[i]} \forall k \in 1\ldots K$. As we have many hashtags, in the analysis we can focus on a subset (i.e., some related to migration and some unrelated). Then, to compare multiple communities, we can plot them as a heatmap, focused on the selected hashtags.

Topics and Communities. For the analysis of text, we propose a methodology centered around topics. First, we extract topics relying on a modeling technique called **Latent Dirichlet Allocation (LDA)** [7]. With LDA, given a number of topics k, the model is trained to group the articles in the requested number of topics, based on their content. By doing this, a selection of the most relevant topics can be performed. We can apply a topic model to the entire collection of text documents (i.e., posts and comments) to visualize which are the topics in the discussion and their most important words. Indeed, once an LDA model θ has been trained on the entire collection of documents, it can return for each document d its *document topic distribution* $\theta(d)$. For each user, we can consider the subset of posts written by the user u as $\mathcal{D}_u$. By doing so, we can describe a user through a *user topic vector*, computed as the average of its document topic distribution $\mathbf{v}_u = \frac{1}{|\mathcal{D}_u|}\sum_{d \in |\mathcal{D}_u|}\theta(d)$. The obtained vector represents how much the user is interested in each topic. Then, to compare different communities of users, we can compute a *community topic vector* of a community of users C as $\mathbf{v}_C = \frac{1}{|C|}\sum_{u \in C}\mathbf{v}_u$. The comparison between community topic vectors can be performed visually through a heatmap plot.

5 DATASET

The proposed methodology allows the study of structure and content characterizing the user migration following a fork event. As a case study, we rely on data from the blockchain *Steem* (the original blockchain) and *Hive* (the new descendant blockchain). We selected Steemit and Hive as our case study for two main *reasons. First, the platforms are relevant in the field.* Steemit holds a prominent position among blockchain social networks, with more than 1 million registered users active in both social and economic dimensions; similarly, the Hive platform, generated after the fork event, engaged a substantial portion of active users who remained active even long after the fork [2]. Second, *profile-matching techniques are not required.* As the data is copied across the blockchains at fork time, we have that profiles related to the same identity are explicitly linked; this is a crucial point, as the results can be obtained independent of user profile matching heuristics, unlike in other cross-platform migration scenarios. The two blockchains, Steem and Hive, support two social network platforms: *Steemit* and *Hive Blog*. Everything began in February 2020 when TRON, a company that owns a gambling-oriented blockchain, led by Justin Sun, acquired Steem.[3] Since the beginning, Steemit's founder allocated a reservoir of tokens that were supposed to be used solely for the development of the Steem ecosystem and to be nonvoting in governance issues:[4] however, after the acquisition, there were no guarantees, and therefore some of the most active users tried to freeze the tokens acquired by TRON through a soft fork.[5] Nevertheless, TRON was able to temporarily amass a significant amount of voting power on the platform with the aid of some cryptocurrency exchangers, reaching the point where it was able to elect its selected witnesses because it owned more than 51% of them. With its witnesses in place, TRON managed to revert the effects of the soft fork.[6] In response to the hostile takeover, the old witnesses of Steem announced a hard fork,[7] which happened on March 20, 2020, originating Hive.[8] Because Hive shares the same blocks before the hard fork, Hive witnesses froze or confiscated all funds owned by the perpetrators of the hostile takeover to prevent issues on the new platform. Hive, among other

[3]https://news.bitcoin.com/steemit-for-sale-tron/
[4]https://steemit.com/steem/@softfork222/soft-fork-222
[5]https://www.coindesk.com/tech/2020/02/24/justin-sun-bought-steemit-steem-moved-to-limit-his-power/
[6]https://www.coindesk.com/tech/2020/03/03/steem-community-mobilizes-popular-vote-in-battle-with-justin-sun/
[7]https://www.coindesk.com/tech/2020/03/17/steem-community-plans-hostile-hard-fork-to-flee-justin-suns-steemit/
[8]https://cointelegraph.com/news/hive-hard-fork-is-successful-steem-crashes-back-to-earth

innovations, introduced a delayed voting influence mechanism to address potential future 51% attacks, giving the community time to respond in advance.

In the following, we describe the data used for the analysis.

Interaction Data. Using data from the blockchain Steem and its new derivative blockchain Hive, we investigate the impact of the mesoscale properties of network layers on user migration. All actions are stored in the supporting blockchain as transactions. All interactions are saved as operations, and a complete list is available for both platforms [11, 14]. In this work, we focus on actions that represent an interaction between two users, either explicit or implicit. Specifically, we consider two main groups: financial and social operations. *Financial operations* are those operations designated for the management of tokens, rewards, and asset transfer. In contrast, *social operations* are those that users are able to do on traditional social network platforms, like posting, rating, voting, sharing, and following. All blocks and the corresponding operations can be gathered through official APIs for both platforms, whose structure and usage are similar. For the construction of the graph, we gathered operations from the very first block on the Steem blockchain, produced on March 24, 2016, up to the fork event—that is, to block 41818752, with timestamp *2020-03-20T14:00:00*. For migration status, we examine data after that timestamp, and up to January 2021. We recall that data between the two blockchains are identical up to the fork event—that is, to block 41818752, with timestamp *2020-03-20T14:00:00*. From there, Hive and Steem have different data, as they have become two different blockchains. Overall, from the Steem blockchain, we extract 993,641,075 operations describing social interactions and 72,370,926 operations describing economic interactions; from the Hive blockchain, we get a total of 206,224,132 social operations and 4,041,060 financial actions.

Text Data. As we are interested in the discussion, we leverage the textual content produced by users before the fork. In Steem and Hive, users' posts and comments are stored as *comment operations* on the underlying blockchains. The content of the post can be accessed as the *body*, and metadata information is also accessible including the *hashtags*. Please note that in Steemit, hashtags are called *tags*.[9] As a starting point, we consider the operations from Ba et al. [4], a total of 93,832,667 *comment operations* that include both posts and comments. For this analysis, we are interested in fork-related discussion. Since everything started after the acquisition, we can focus on a limited period: for this work, we focus on the period from February 20, 2020, to March 20, 2020. Comment operations (both posts and comments) total 831,403. We selected only (a) posts not comments going down to 234,396 posts, and (b) among them, we consider only posts written in English, for a total of 140,638 (the language is detected by the Python library lang-detect). For the corpus of posts, preprocessing and cleaning are applied to the data to delete noisy, inconsistent, or incomplete items from the collection. Specifically, we applied the following operations: removal of HTML tags, URLs, punctuation, multiple whitespaces, numbers, stopwords, words shorter than three characters, and stemming. For this subset of documents, we find 284,932 unique terms, whereas the average token length of posts is 104.5. Then, we consider the associated metadata information of the corpus of published posts, to derive the collection of hashtags. We performed some preprocessing on hashtag data as well: we filtered out the hashtags with fewer than two characters, then we merged some hashtags of interest that share the same semantics. Specifically, we grouped all the hashtags that contain "fork" (e.g., softfork, hardfork), "hive," "Justin," and "tron." We obtain 39,626 unique hashtags, and on average, we observe 5.19 hashtags per post.

6 RESULTS

Applying the methodology outlined in Section 4, on the Steem/Hive dataset, we create a multilayer network with two layers: social and monetary, labeling each node based on its activity after the fork. A summary of network statistics and labels is presented in Table 1.

[9] https://steemit.com/faq.html#What_are_tags

Table 1. Statistics for the Multilayer Network $G_{T_{fork}}$, Grouped by Social and Monetary Layers

	Social layer	Monetary layer
Nodes	1,352,114	1,247,587
Edges	217,926,899	5,056,317
Inactive	1,287,321	1,218,535
Resident	43,339	12,757
Migrant	21,454	16,295

Overall, the social layer has more active users and links: this is consistent with the operations under consideration; in fact, social operations are far more common than monetary transactions. However, despite having fewer links, the monetary layer involves a comparable number of users (i.e., the volume of nodes is roughly the same as in the social layer). Finally, we observe that the social and monetary layers differ when the proportion of resident and migrant users is considered. In fact, in the social layer, most active users are residents (i.e., one-third of the active users have migrated to Hive), whereas in the monetary layer, we see an opposite trend where user migration has had a greater impact, namely the majority of users have decided to migrate to Hive to conduct their financial transactions.

6.1 The Interplay of Community Structure and Migration

In this section, we answer RQ1 on the interplay between group network structure and user migration. We created the monetary community graph and social community graph using the methodology described in Section 4. The monetary layer's community graph has 76 communities and 252 intercommunity edges, whereas the social layer's community graph has 105 communities and 205 intercommunity edges. We visualize the obtained community graphs, with nodes—communities—colored according to the proportion of migrants and residents among their members in Figure 2.

Visualizing the Interplay. Taking into account only node coloring and their *connectivity*, we see that migrant communities tend to be on the periphery of the community graph, with few or no intercommunity links. This characteristic can be found in both monetary and social layers. We can also see that the community graphs have a more central part, which is made up of very connected communities with the majority of members being residents. In contrast, only a few communities with a majority of migrants are linked to the central core of the community graphs. The isolation of migrant nodes and communities is the first important indication of the significance of network structure: for a community, being marginal may be a trait that leads to the majority of its members migrating. We proceed in our analysis of group structure by taking *density* and *entropy* into account as community features. For the evaluation of the impact of community density on migration, we focus on the size of the community nodes, for the social community graph in Figure 2(a) and for the monetary community graph in Figure 2(b). A visual examination of the network representation of the social community graph reveals that there is no clear distinction in terms of density among resident and migrant communities: we find both migrant and resident communities among the densest communities. However, in the monetary layer, we can see that the resident communities—more specifically, those in the network's central region—have the highest density values. For visual analysis of entropy values, we vary the size of communities of the social community graph in Figure 2(c) and at the monetary community graph in Figure 2(d), according to their entropy. We observe that entropy values are pretty similar in the social layer. Entropy values are high across all communities in the social layer, and we are unable to distinguish any specific differences. However, we see a more varied situation on the monetary layer. In this layer, we can observe that the communities in the central part are characterized by low entropy values. Moreover, they tend to be composed of a majority of resident users and are more likely

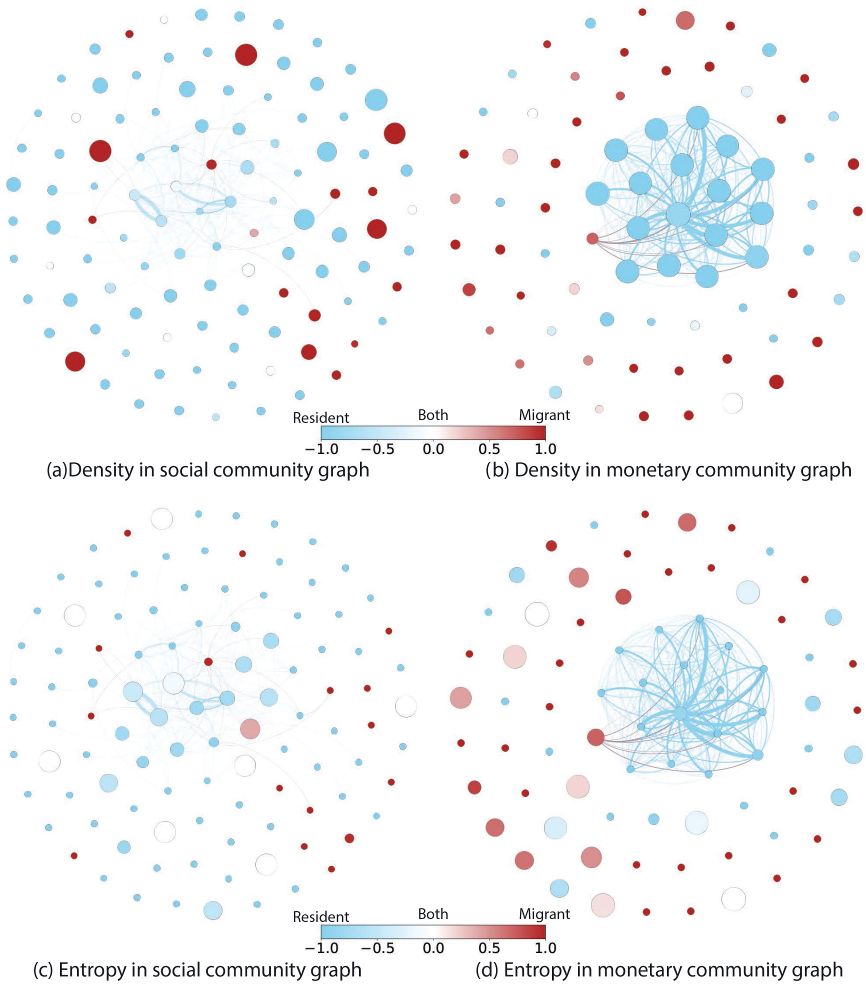


Fig. 2. Community graphs, for social layer (left) (105 communities, 205 intercommunity edges) and monetary layer (right) (76 communities, 252 intercommunity edges). In (a) and (b), community node size is proportional to its density. In (c) and (d), node size is proportional to its community entropy. We use colors to represent the majority between migrant and resident nodes: communities with more residents will go toward sky blue, whereas more migrants will lead to red nodes, and white is for nodes with a balanced mix of both. Edge width is proportional to the weight of the intercommunity edge. The position of the nodes is determined by the visualization library Gephi [6] by leveraging connectivity in a force-based layout.

to connect with other resident communities. In addition, we observe high entropy values in isolated communities, but there is no distinction between resident and migrant communities. So, even though there are some differences in network structure between the considered layers, overall, entropy does not help characterize the two groups.

Measuring the Interplay. We then move on to the quantitative analysis of the interplay between the network structure (density and entropy) and the migration decision (inactive, resident, migrant). We computed correlation statistics between the selected community features, taking into account the communities on the social and monetary layers. In Table 2, on the left side, we report correlation measures for the social communities.

Table 2. Correlations on Community Properties in the Social Layer on the Left and in the Monetary Layer on the Right

	Density	**Entropy**
Inactive	−0.187 (0.057)	0.176 (0.073)
Resident	−0.123 (0.211)	0.025 (0.797)
Migrant	−0.075 (0.448)	0.357 (0.005)

	Density	**Entropy**
Inactive	−0.296 (0.009)	0.164 (0.157)
Resident	0.583 (0.0)	−0.209 (0.07)
Migrant	−0.275 (0.016)	−0.060 (0.608)

p-Values are reported in parentheses.

We can observe that for the social communities, density has a slightly negative correlation with the number of resident and inactive users, whereas there is no correlation with the number of migrant users, which is consistent with the earlier network-based visual inspection. In terms of entropy, we observe a significant positive correlation (p-value ≤ 0.005) with the volume of migrants. On the right side of Table 2, we show the measurements computed with the communities in the monetary layer. Density has a moderately positive correlation with the number of residents but a negative correlation with the presence of migrant nodes. These observations are in line with the network-based visual analysis, which revealed that density characterized monetary communities made up of residents, whereas migrants tend to be more loosely connected. Similarly, entropy shows a slight negative correlation with the number of resident nodes. So even at the quantitative level, we can confirm that group density can characterize users at a mesoscopic level. On the contrary, entropy does not seem to be helpful in the characterization of the groups.

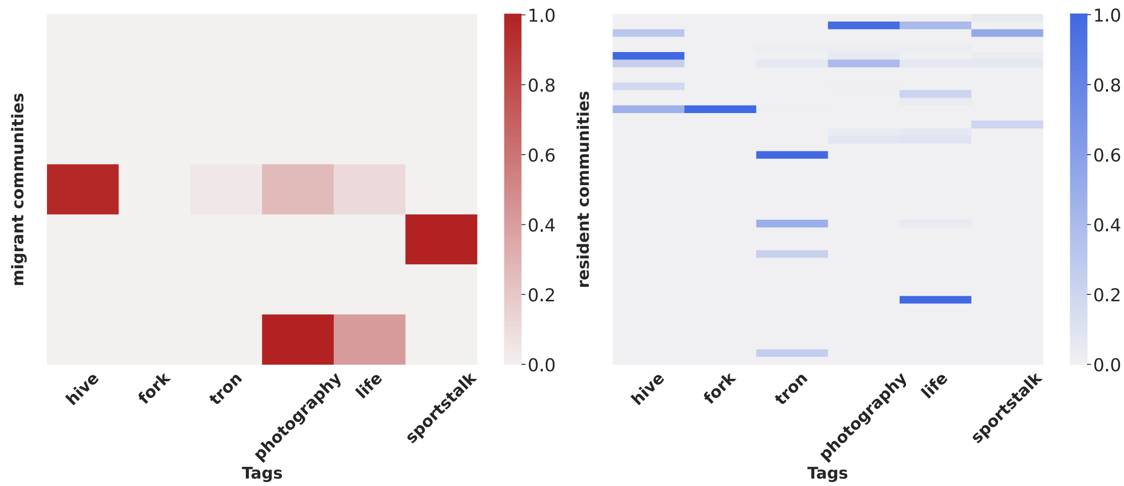
Therefore, regarding the interplay of group network structure and user migration (RQ1), we can conclude that (a) the "position" of a group within the network of social and economic interactions is related to the likelihood of a group to migrate (i.e., marginal groups are more likely to leave), (b) users in densely connected groups interacting through monetary transactions are more likely to stay, and (c) user migration affects the network built on social interactions and the network based on monetary transactions differently.

6.2 The Interplay of Community Discussion and Migration

In this section, we answer our research question on the interplay of group discussions and user migration (RQ2). We first present the results obtained by the analysis of hashtags, followed by the analysis of topics.

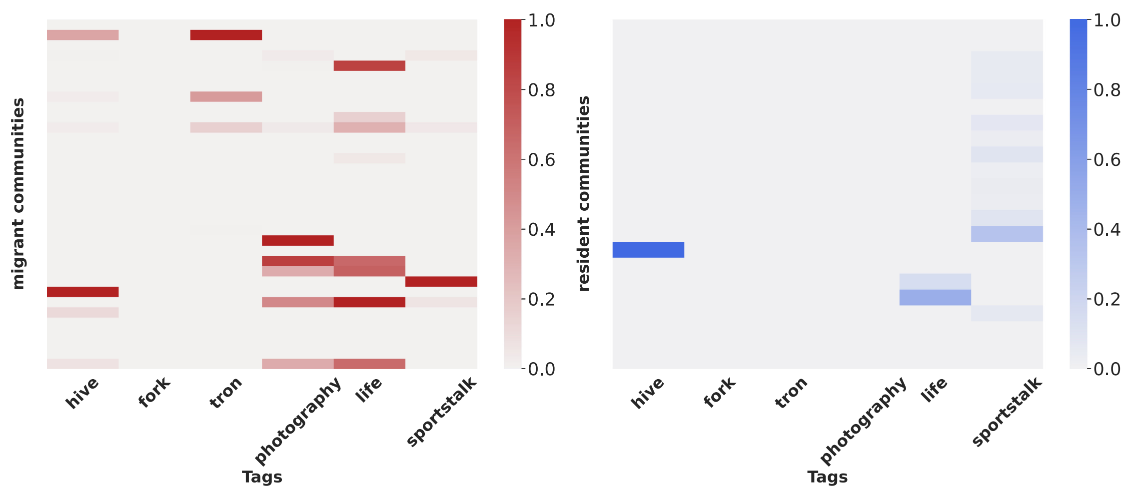
Hashtags and Communities. To study the interplay with community structure, we compare hashtag usage across communities. We rely on communities obtained considering either social interactions or financial interactions to compute the community hashtags distributions and compare them through a heatmap plot, as described in Section 4. For easier comparison, we group communities into those with a majority of resident users and the ones with a majority of migrant users. We first consider the communities on the social layer in Figure 3, where each row represents a community on the social layer and its community hashtag distribution. Overall, there is a difference in hashtag distribution between migrant and resident users, but there is no clear distinctive trait. Moreover, migration-related hashtags are not limited to migrant communities, as discussion involves resident users as well: on the social migrants' side (see Figure 3(a)), we can see how some of the communities barely use the selected migration-related tags. We can observe a community discussing using the *hive* hashtag much more than other ones, whereas other communities seem to be focused on other hashtags as often. Moreover, the usage of migration-related hashtags is not limited to migrants: when we consider communities with a majority of residents (see Figure 3(b)), we can observe a few communities using migration-related hashtags a lot, especially *hive* and *Tron*.

We then analyze hashtag distribution, focusing on communities on the monetary layer. We report the hashtag distributions in Figure 4. Overall, here we find more differences between migrant and resident users, as we see higher usage of migration-related hashtags on the migrant side. In fact, when we consider the migrant



(a) Migrant communities (b) Resident communities

Fig. 3. Social community hashtag distribution. Heatmaps represent communities on the social layer and their most used hashtags. On the x-axis a selection of hashtags and on the y-axis the communities, in migrant communities (a) and resident communities (b). Values in each cell correspond to the frequency (count) of a hashtag in a community, min-max normalized by hashtag.



(a) Migrant communities (b) Resident communities

Fig. 4. Monetary community hashtag distribution. Heatmaps represent communities on the social layer and their most used hashtags. On the x-axis a selection of hashtags and on the y-axis the communities, in migrant communities (a) and resident communities (b). Values in each cell correspond to the frequency (count) of a hashtag in a community, min-max normalized by hashtag.

communities (see Figure 4(a)), we observe quite a few communities using migration-related hashtags often, especially *hive* and *tron*. However, resident monetary communities (see Figure 4(b)) tend to use migration-related hashtags rarely, except for one community. Overall, the communities exhibit very different hashtag distributions, but there is not a clear trend distinguishing migrant communities from resident ones.

Table 3. Top 10 Stemmed Keywords for Each Topic Detected with the LDA Topic Model

Label	**Top 10 keywords (stemmed)**
Platform	steem, commun, steemit, vote, power, post, blockchain, tron, hive, justin
Monetary	token, crypto, blockchain, user, invest, bitcoin, platform, coin, account, cryptocurr
Food	jpeg, food, coffe, cook, restaur, fresh, tast, fruit, weight, rice
Nature	walk, beauti, time, flower, like, todai, activ, home, natur, place
Appics	appic, amazonaw, content, east, categori, author, hashtag, caption, permlink, profileimageurl
Positive	like, peopl, time, know, thing, want, life, think, love, feel
Investments	open, deal, forecast, market, rate, price, expect, coronaviru, post, year
Games	game, video, plai, imag, link, steemhunt, post, view, youtub, screenshot
Dapps	post, upvot, photo, themarkymark, actifit, dtube, steem, contest, vote, follow
Chinese	chines, center, mandarin, btdx, ccenter, dtube, http, jesu, class, muslim

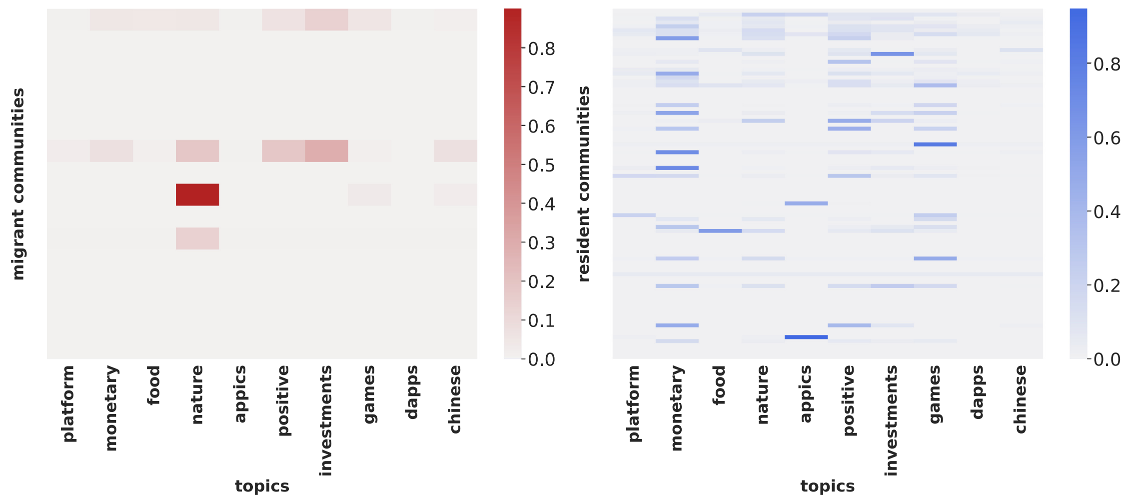
Topics and Communities. In this section, we present the results obtained by applying the methodology proposed in Section 4 for the analysis revolving on *content topics*. We first observe the obtained topics and their most important words in Table 3.

We can see how topics are varied, from topics of general interest such as food, nature, and so on, to other topics that are more focused on the economic and technical aspects of the platform and the blockchain world. For easier comprehension, for each topic, we assigned a label based on its most important words. Most labels are self-explanatory, but we briefly go over each label for a better understanding of the following analyses. The *Platform* topic is characterized by terms related to the platform and others related to migration. The *Monetary* topic is characterized by cryptocurrency-related terms; similarly, the *Investments* topic is characterized by keywords related to finance (price, forecast, open). Topics like *Food*, *Nature*, and *Positive* tend to have terms of general interest. *DApps* stands for Decentralized APPlications—that is, applications that run on top of the hosting blockchain, and the corresponding *DApps* topic reunites discussion over some of them. For instance, *Dtube*[10] is a video-sharing platform with a cryptocurrency-based reward system, whereas *Actifit* is another Dapp for fitness enthusiasts.[11] *Appics* is another DApp, similar to Steemit,[12] that relies on the Steem blockchain. Some gaming content is reunited in the *Games* topic. Finally, it seems that although non-English posts are removed, there is a significant community discussing China-related topics. Overall, the choice of 10 topics produced coherent topics. Hence, we can proceed with the analysis of the interplay between topics and user migration.

We now consider the topic distributions that characterize the communities. We apply the methodology to compute *community topic vectors* (see Section 4). We first start with communities on the social layer: the obtained *social community topic vectors* are shown in Figure 5. As a general observation, we can see that the migrant social communities detected tend to have lower values overall, whereas on the resident side, we find more communities and can observe more interest peaks in the values. When we focus on topics, we can see that on both sides, we do not find communities mainly interested in the platform and migration-related topics. Among the topics of interest for the resident groups, the *Monetary* topic emerges, as it seems the focus in many communities. Additionally, we can see that other topics tend to be of interest across multiple communities such as *Nature*, *Positive*, *Investments*, and *Games*, whereas interest in the remaining topics seem to be limited to only a few communities.

Finally, we consider the *monetary community topics vectors* in Figure 6. The first observation is that overall, there is a difference in topics of interest between migrant and resident users: on the migrant side, we can observe

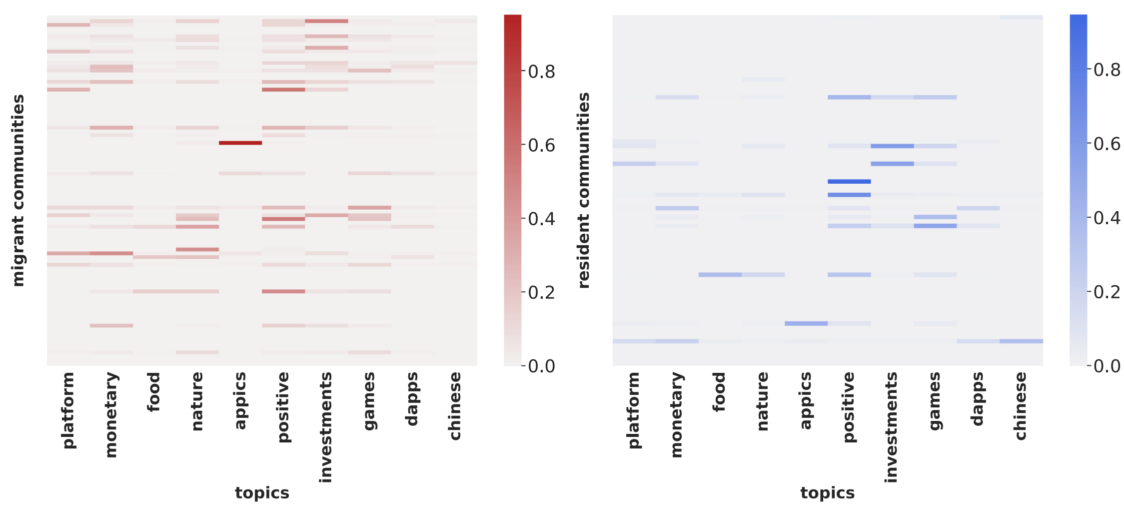
[10] https://d.tube/
[11] https://actifit.io/
[12] https://www.appics.com/



(a) Migrant communities

(b) Resident communities

Fig. 5. Social community topics vectors. Heatmaps represent communities on the social layer and their topics of interest. On the x-axis topics and on the y-axis the communities, in migrant communities (a) and resident communities (b). Values represent the interest in each topic.



(a) Migrant communities

(b) Resident communities

Fig. 6. Monetary community topics vectors Heatmaps represent communities on the social layer and their topics of interest. On the x-axis topics and on the y-axis the communities, in migrant communities (a) and resident communities (b). Values represent the interest in each topic.

more often peaks in the values, whereas on the resident side, values are generally more distributed across all topics. When we focus on topics, we can see that there is a strong difference concerning the *Platform* topic: communities on the migrant side often have high values in this topic. There is a greater interest by migrant users on the platform and migration-related discussions. On the contrary, we can see how *Monetary* topic peaks are actually more frequent on the migrants' side as well; a similar observation can be made for *Nature*. When it comes down to other topics, the difference is less evident: on both sides, we can see that *Food*, *Dapps*, *Chinese*, and *Appics* are actually limited to only a few communities,whereas *Positive*, *Investments*, and *Games* tend to be more spread out and of interest to more communities, on both sides.

Therefore, regarding the interplay of group discussions and user migration (RQ2), we can conclude the following: (a) between migrant and resident users, there is a difference in hashtag distributions as well as topics of interest; (b) social communities migration-related hashtags and migration-related topics involve both migrant and resident users; and (c) vice versa, on the monetary layer, we see a clearer interest by migrant users in migration-related hashtags and topics.

7 CONCLUSION

In this work, we addressed the open problem of user migration due to hard fork events occurring in BOSN. A hard fork signifies a crucial event capable of jeopardizing the survival of a platform. Therefore, any additional insight into the events during such an occurrence can make a significant difference in navigating their impact. This issue holds significant importance, considering the proliferation of blockchain-based projects, as it not only implicates users but also involves platform developers, stakeholders, and policymakers. Specifically, we investigate user decision making to stay (resident) or leave (migrant) the platform by leveraging network structure and user-generated text content. Our findings on the impact of network structure, such as the crucial role of density, show that structural information, derived from user interactions, should be considered for analysis and user migration prediction tasks. In fact, our network structure analysis revealed that marginal groups, in terms of network structure, are more prone to leaving the platform in such scenarios. It is evident that any BOSN facing similar circumstances and aiming to minimize user loss should prioritize efforts to enhance the connectivity of these users with the broader user base. Our findings on differences related to text content and user discussion show how the groups could be also characterized by the content they post and share, something that can be useful not just for predicting user migration but also for the analysis and understanding of causes and dynamics during conflict or turning-point events in a social platform. For instance, our observations revealed that users deeply engaged in the economic aspects were also actively participating in discussions related to user migration. This highlights the significance of this user subset for platform managers, not just due to their involvement in transactions and trust in the currency. Consequently, it becomes imperative to allocate time and resources to monitor and strategically engage with these users. Prioritizing user retention strategies, which may include incentives or loyalty programs, should be a focal point in maintaining their sustained involvement on the platform. In general, understanding user migration is of high importance for both traditional social platforms and BOSN platforms that are trying to retain their users as they grow, as well as for new alternative platforms trying to emerge. Indeed, data analysis not only offers valuable and actionable information for developers and platform managers but also contributes to informing the development of future or new platforms. This includes the adoption of contingency plans, platform monitoring tools, and customer retention strategies right from the outset. These findings could be useful to other blockchains, as they show the importance of designing proper consensus protocols to handle turning-point events. Besides user migration, the representation for the blockchain data modeling might be applied to a few phenomena characterizing the Web3, such as the trading networks generated by NFT (not-fungible token) exchanges, or other kinds of social and financial interaction mediated or fueled by DApps, such as games or thematic social networks.

REFERENCES

[1] Cheick Tidiane Ba, Alessia Galdeman, Manuel Dileo, Matteo Zignani, and Sabrina Gaito. 2023. User migration prediction in blockchain socioeconomic networks using graph neural networks. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good (GoodIT '23)*. ACM, 333–341. https://doi.org/10.1145/3582515.3609552

[2] Cheick Tidiane Ba, Andrea Michienzi, Barbara Guidi, Matteo Zignani, Laura Ricci, and Sabrina Gaito. 2022. Fork-based user migration in blockchain online social media. In *Proceedings of the 14th ACM Conference on Web Science*.

[3] Cheick Tidiane Ba, Matteo Zignani, and Sabrina Gaito. 2021. Social and rewarding microscopical dynamics in blockchain-based online social networks. In *Proceedings of the Conference on Information Technology for Social Good (GoodIT '21)*. ACM, 127–132.

[4] Cheick Tidiane Ba, Matteo Zignani, and Sabrina Gaito. 2022. The role of cryptocurrency in the dynamics of blockchain-based social networks: The case of Steemit. *PLoS ONE* 17, 6 (2022), e0267612.

[5] Cheick Tidiane Ba, Matteo Zignani, and Sabrina Gaito. 2022. The role of groups in a user migration across blockchain-based online social media. In *Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops '22)*. IEEE, 291–296.

[6] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Web and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

[7] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* (March 2003), 993–1022.

[8] Raffaele Ciriello, Roman Beck, and Jason Thatcher. 2018. The paradoxical effects of blockchain technology on social networking practices. In *Proceedings of the 39th International Conference on Information Systems*. 1–17.

[9] Cai Davies, James R. Ashford, Luis Espinosa-Anke, Alun David Preece, Liam D. Turner, Roger M. Whitaker, Mudhakar Srivatsa, and Diane H. Felmlee. 2021. Multi-scale user migration on Reddit. In *Proceedings of the Workshop on Cyber Social Threats at the 15th International AAAI Conference on Web and Social Media (ICWSM '21)*.

[10] Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall. 2015. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X* 5, 1 (2015), 011027.

[11] Steem. 2021. Steem Developer Portal: Broadcast Ops. Retrieved January 12, 2024 from https://developers.steem.io/apidefinitions/broadcast-ops

[12] Massimo Di Pierro. 2017. What is the blockchain? *Computing in Science & Engineering* 19, 5 (2017), 92–95.

[13] Manuel Dileo, Cheick Tidiane Ba, Matteo Zignani, and Sabrina Gaito. 2022. Link prediction with text in online social networks: The role of textual content on high-resolution temporal data. In *Discovery Science*. Lecture Notes in Computer Science, Vol. 13601. Springer, 212–226.

[14] Hive. 2021. Hive Developer Portal—API Definitions. Retrieved January 12, 2024 from https://developers.hive.io/apidefinitions/

[15] Pierluigi Freni, Enrico Ferro, and G. Ceci. 2020. Fixing social media with the blockchain. In *Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good*. 175–180.

[16] Alessia Galdeman, Matteo Zignani, and Sabrina Gaito. 2023. User migration across Web3 online social networks: behaviors and influence of hubs. In *Proceedings of the IEEE International Conference on Communications*.

[17] Kiran Garimella, Tim Smith, Rebecca Weiss, and Robert West. 2021. Political polarization in online news consumption. In *Proceedings of the International AAAI Conference on Web and Social Media*. 152–162. https://doi.org/10.1609/icwsm.v15i1.18049

[18] Barbara Guidi. 2020. When blockchain meets online social networks. *Pervasive and Mobile Computing* 62 (2020), 101131.

[19] Barbara Guidi and Andrea Michienzi. 2021. Interaction communities in blockchain online social media. In *Proceedings of the 2021 3rd International Conference on Blockchain Computing and Applications (BCCA '21)*. IEEE, 89–96.

[20] Barbara Guidi, Andrea Michienzi, and Laura Ricci. 2021. A graph-based socioeconomic analysis of Steemit. *IEEE Transactions on Computational Social Systems* 8, 2 (2021), 365–376. https://doi.org/10.1109/TCSS.2020.3042745

[21] Barbara Guidi, Andrea Michienzi, and Laura Ricci. 2020. Steem blockchain: Mining the inner structure of the graph. *IEEE Access* 8 (2020), 210251–210266.

[22] Barbara Guidi, Andrea Michienzi, and Laura Ricci. 2022. Assessment of wealth distribution in blockchain online social media. *IEEE Transactions on Computational Social Systems*. Early Access, December 20, 2022.

[23] Jiahui He, Haris Bin Zia, Ignacio Castro, Aravindh Raman, Nishanth Sastry, and Gareth Tyson. 2023. Flocking to Mastodon: Tracking the great Twitter migration. In *Proceedings of the 2023 ACM on Internet Measurement Conference (IMC '23)*. ACM, 111–123. https://doi.org/10.1145/3618257.3624819

[24] Petter Holme and Jari Saramäki. 2019. *Temporal Network Theory*. Vol. 2. Springer.

[25] Roberto Interdonato, Matteo Magnani, Diego Perna, Andrea Tagarelli, and Davide Vega. 2020. Multilayer network simplification: Approaches, models and methods. *Computer Science Review* 36 (2020), 100246.

[26] Kristina Kapanova, Barbara Guidi, Andrea Michienzi, and Kevin Koidl. 2020. Evaluating posts on the Steemit blockchain: Analysis on topics based on textual cues. In *Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good*. 163–168.

[27] Moon Soo Kim and Jee Yong Chung. 2019. Sustainable growth and token economy design: The case of Steemit. *Sustainability* 11, 1 (2019), 167.

[28] Tae-Hyun Kim, Hyo Min Shin, H. Hwang, and Seungwon Jeong. 2020. Posting bot detection on blockchain-based social media platform using machine learning techniques. *arXiv abs/2008.12471* (2020).

[29] Srijan Kumar, William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2018. Community interaction and conflict on the web. In *Proceedings of the 2018 Conference on World Wide Web*. 933–943.

[30] Shamanth Kumar, Reza Zafarani, and Huan Liu. 2011. Understanding user migration patterns in social media. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*. 1204–1209.

[31] Lucio La Cava, Sergio Greco, and Andrea Tagarelli. 2021. Understanding the growth of the Fediverse through the lens of Mastodon. *Applied Network Science* 6, 1 (2021), 1–35.

[32] Chao Li and Balaji Palanisamy. 2019. Incentivized blockchain-based social media platforms: A case study of Steemit. In *Proceedings of the 10th ACM Conference on Web Science*. 145–154.

[33] Chao Li, Balaji Palanisamy, Runhua Xu, and Li Duan. 2023. Cross-consensus measurement of individual-level decentralization in blockchains. In *Proceedings of the 2023 IEEE 9th International Conference on Big Data Security on Cloud (BigDataSecurity '23), the IEEE International Conference on High Performance and Smart Computing (HPSC '23), and the IEEE International Conference on Intelligent Data and Security (IDS '23)*. 45–50. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS58521.2023.00018

[34] Liqun Liu, Weihan Zhang, and Cunqi Han. 2021. A survey for the application of blockchain technology in the media. *Peer-to-Peer Networking and Applications* 14 (2021), 3143–3165.

[35] Matteo Magnani, Obaida Hanteer, Roberto Interdonato, Luca Rossi, and Andrea Tagarelli. 2019. Community detection in multiplex networks. *arXiv preprint arXiv:1910.07646* (2019).

[36] Edward Newell, David Jurgens, Haji Mohammad Saleem, Hardik Vala, Jad Sassine, Caitrin Armstrong, and Derek Ruths. 2016. User migration in online social networks: A case study on Reddit during a period of community unrest. In *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM '16)*. 279–288.

[37] Martin Rosvall, Daniel Axelsson, and Carl T. Bergstrom. 2009. The map equation. *European Physical Journal Special Topics* 178, 1 (2009), 13–23.

[38] Sarwar Sayeed and Hector Marco-Gisbert. 2019. Assessing blockchain consensus and security mechanisms against the 51% attack. *Applied Sciences* 9, 9 (2019), 1788.

[39] Malith Senaweera, Ruwanmalee Dissanayake, Nuwini Chamindi, Anupa Shyamalal, Charitha Elvitigala, Sameera Horawalavithana, Primal Wijesekara, Kasun Gunawardana, Manjusri Ishwara Ellepola Wickramasinghe, and Chamath Keppitiyagama. 2018. A weighted network analysis of user migrations in a social network. In *Proceedings of the 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer '18)*. 357–362.

[40] Hongting Tang, Jian Ni, and Yanlin Zhang. 2022. Identification and evolutionary analysis of user collusion behavior in blockchain online social medias. *IEEE Transactions on Computational Social Systems*. Early Access, October 27, 2022.

[41] M. Thelwall. 2018. Can social news websites pay for content and curation? The SteemIt cryptocurrency model. *Journal of Information Science* 44 (2018), 736–751.

[42] Giacomo Villa, Gabriella Pasi, and Marco Viviani. 2021. Echo chamber detection and analysis: A topology- and content-based approach in the COVID-19 scenario. *Social Network Analysis and Mining* 11, 1 (Aug. 2021), 78.

[43] Shermin Voshmgir. 2020. *Token Economy: How the Web3 Reinvents the Internet*. Vol. 2. Token Kitchen.

[44] Rongen Zhang, Junyoung Park, and Raffaele Ciriello. 2019. The differential effects of cryptocurrency incentives in blockchain social networks. In *Proceedings of the Pre-ICIS Workshop on Blockchain and Smart Contract (SIGBPS '19)*.

[45] Yujie Zheng, Mariana Andrade, and Waifong Boh. 2023. Examining the survival of forked blockchain projects: A multi-stakeholder perspective. In *Proceedings of the Pacific Asia Conference on Information Systems (PACIS '23)*. https://aisel.aisnet.org/pacis2023/214

Received 30 March 2023; revised 1 November 2023; accepted 12 December 2023