Fork-based user migration in Blockchain Online Social Media

Cheick Ba
cheick.ba@unimi.it
CS Department - University of Milan
Milan, Italy

Andrea Michienzi
andrea.michienzi@di.unipi.it
CS Department - University of Pisa
Pisa, Italy

Barbara Guidi
barbara.guidi@unipi.it
CS Department - University of Pisa
Pisa, Italy

Matteo Zignani
matteo.zignani@unimi.it
CS Department - University of Milan
Milan, Italy

Laura Ricci
laura.ricci@unipi.it
CS Department - University of Pisa
Pisa, Italy

Sabrina Gaito
sabrina.gaito@unimi.it
CS Department - University of Milan
Milan, Italy

ABSTRACT

Nowadays, Online Social Media (OSM) are among the most popular web services. Traditional OSM are known to be affected by serious issues including misinformation, fake news, censorship, and privacy violations, to the point that a pressing demand for new paradigms is raised by users all over the world. Among such paradigms, the concepts around the Web 3.0 are fueling a new revolution of online sociality, pushing towards the adoption of innovative and groundbreaking technologies. In particular, the decentralization of social services through the blockchain technology is representing the most valid alternative to current OSM, enabling the development of rewarding strategies for value redistribution, and fake news detection. However, the so-called Blockchain Online Social Media (BOSMs) are far from being mature, with different platforms that continually try to redefine their services in order to attract larger audiences, thus causing blockchain forks and massive user migrations, with the latter dominating the dynamics of the current OSM landscape, too.

In this paper, we deal with the evolution of BOSMs from the perspective of user migration across platforms as a consequence of a fork event. We propose a general user migration model applicable to BOSMs to represent the evolution patterns of fork-based migrations, the multi-interaction structural complexity of BOSMs, and their growth characteristics. Within this framework, we also cope with the task of predicting how users will behave in the case of a fork, i.e. they will remain on the original blockchain or they will migrate to the new one. We apply our framework to the case study of the Steem-Hive fork event, and show the importance of considering both social and economic information, regardless of the learning algorithm considered. To the best of our knowledge, this is the first study on blockchain fork and its related user migration.

WebSci '22, June 26–29, 2022, Barcelona, Spain

ACM ISBN 978-1-4503-9191-7/22/06…$15.00
https://doi.org/10.1145/3501247.3531597

CCS CONCEPTS

• **Networks** → **Online social networks**; **Social media networks**; • **Computing methodologies** → **Machine learning**; • **Computer systems organization** → **Peer-to-peer architectures**.

KEYWORDS

Blockchain Online Social Media, User Migration, Temporal Networks

ACM Reference Format:
Cheick Ba, Andrea Michienzi, Barbara Guidi, Matteo Zignani, Laura Ricci, and Sabrina Gaito. 2022. Fork-based user migration in Blockchain Online Social Media. In *14th ACM Web Science Conference 2022 (WebSci '22), June 26–29, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3501247.3531597

1 INTRODUCTION

Online Social Media (OSM) have become an important part of the life of more than half of the World's population[1], and nowadays they are among the most used web applications. People use social media for many purposes, including to share their personal information, to keep in touch with friends and family, to gather information about the latest events in the world, and more. The current OSM landscape is characterized by competitions to get larger audiences, the introduction of novel and disruptive services leading to the death of the oldest ones, and massive customer migrations that continuously reshape the social web scenario. Users often tend to migrate, i.e. move to different social media platforms due to specific events, such as the emergence of new platforms or changes to previous platforms. Thanks to the emergence of technologies related to the Web 3.0, decentralization through blockchain dominates the landscape of new OSM platforms, proposing creative solutions to the well known problems of OSM, and introducing innovative key aspects. In this context, Blockchain Online Social Media (BOSMs) have been proposed and are still raising. In BOSMs, the blockchain technology enables the possibility to redistribute the wealth generated by their users by the means of a reward granted to the users that help the platform grow. These rewarding systems are usually based on the attention economy and/or token economy [3, 14]. Several new BOSMs are proposed, motivated with the common trait of decentralizing control [9, 22], adopting different strategies, such as encouraging a constant social and economic dedication, or reward the creation of pieces of content with outstanding quality.

[1] https://wearesocial.com/digital-2021

Due to the lively competition among OSM platforms, a new phenomenon that goes under the name of user migration, is manifesting in these scenarios. There are numerous causes to the user migration phenomenon, including the ethics of the company offering the social service, or the sheer quality of the service offered. User migration affects both centralized and decentralized Social Media, and this aspect is related not only to social services but also to the infrastructure of Social Media. In the scenario of BOSMs, the phenomenon of user migration can be observed and measured with a high temporal resolution when a new BOSM is generated after a fork event.

As concerns user migration, literature proposes several works on this topic. However, none of these works is focused on the study of the evolution of the subgraphs of users induced by a migration. And most importantly, none of them considers the peculiar characteristics of the user migration that manifests after a fork event of a BOSM. In such a system, migration can be studied, with some advantages thanks to the blockchain technology, which represents an invaluable and unprecedented source of reliable longitudinal data.

The contribution of this paper is to deal with the evolution of BOSMs from the perspective of the user migration among platforms. Specifically, we focus on the impact of a shocking event - a hard fork leading to a user migration - on the structural properties of the social and economical networks supported by the blockchains, and to what extent social and economical structural features can be predictive of the choice of a single user migration. We propose a framework to model the user migration process that is general and therefore applicable to any process of this type. It is based on a representation through an attributed temporal multidigraph, which allows us to measure the effects of the fork on the evolution of the social and economic networks derived from the underlying blockchains. Furthermore, we deal with the prediction of migrating users by exploiting some user characteristics - individual and structural - or activities. As a case study, we apply our framework to the social blockchain Steem, used by Steemit, a leading BOSMs, and the blockchain Hive, introduced after a fork event happened on Steem. To the best of our knowledge, this is the first study on the fork of a blockchain and the corresponding migration of the users of the services relying on it. Furthermore, it is the first work that deals with the prediction of which users will migrate to the new platform right at the time of the fork. It shows that even with the only information on the network structure, without including textual or context data such as the trend of the cryptocurrency, it is possible to predict user migration. Finally, it shows how a multilayer approach improves the performance of the predictors compared to settings that consider the different types of interaction separately.

The paper is organized as follows. Section 2 presents the concepts that are most relevant to our paper, and Section 3 the related work. Section 4 describes how we model the activity of a social blockchain in a fork scenario. Section 5 describes the dataset used in our paper. Section 6 presents our results concerning the difference in the structural evolution of the interaction networks supported by the two blockchains; and the feasibility of predicting which users are willing to migrate after a fork. Finally, Section 7 concludes the paper, pointing out possible future works.

2 BACKGROUND

Over the last decade the world of Internet services underwent a set of profound changes, shifting the attention from monolithic centralized services, to open, decentralized and distributed approaches. This revolution, often referred to as the Web 3.0, aims at building a new type of Web that is based on the blockchain technology, applying it to several systems. The application of Web 3.0 principles in social media resulted in BOSM; in the economy, the application led to Decentralized Finance (DeFi); whereas the application in governance resulted in Decentralized Autonomous Organizations (DAOs); just to cite a few well-known examples.

A blockchain is one of the possible implementations of a distributed ledger [6]. Its characteristic trait is that single pieces of information, usually called *transactions*, are grouped together into *blocks*, and each block is cryptographically linked to its predecessor as the mails of a chain. To add a new block to the chain, specialized nodes, called *miners*, have to compete or cooperate according to a consensus protocol.

Initially, the technology has been applied mainly to store economic transaction among a network of untrusted nodes, such as in Bitcoin. However the technology quickly evolved and gained the crucial trait of programmability, which has led to the introduction of smart contracts. In short, a smart contract is a piece of code whose execution outcome is agreed upon all the nodes of the network. Thanks to the introduction of smart contracts, the new generation of blockchain made possible to support the development of Decentralized APPlications (Dapps), which are applications that run on top of the hosting blockchain. As the ideas of the Web 3.0 are gaining a foothold, Dapps represent one of the aspects that is fueling the decentralization process of online services through the blockchain.

2.1 Blockchain forks

When miners need to modify their behavior, they make a *fork*. There are two types of fork: soft, and hard. A soft fork happens when the change to the protocol governing the blockchain is retro-compatible with the previous version, while hard forks are more drastic. Indeed, in the case of a soft fork, usually, all the miners keep adding blocks on the same chain, while in the case of a hard fork, miners running different versions of the protocol will see each other blocks as invalid, and therefore they might create two distinct *branches* of the blockchain. Since hard forks are more dramatic, miners will usually agree upon a specific time at which they should upgrade the protocol, helping to minimize everyone's loss. Hard forks are events for which a migration phenomenon may happen, depending on the motivation that caused the hard fork. In addition, to introduce small modifications to the consensus protocol, soft forks can be used in very specific scenarios, such as to freeze account funds or revert certain transactions.

2.2 Blockchain Online Social Media

BOSMs, thanks to their nature, address some common problems of traditional OSM, such as the so-called *Single Point of Failure*. From the users' perspective, BOSMs are very resistant to *censorship*. *Content Authenticity* is another problem cursing OSM platforms, for which there is no clear solution. In a BOSM, this problem is

partially addressed by the fact that data is append-only. Blockchain can also add value and functionality to the social platform by implementing a *Rewarding System* for beneficial contribution. The rewarding system can be made such to promote positive behaviors in all the aspects of the platform, but its primary focus is geared towards the rewarding of exquisite content and its mindful evaluation [8, 19]. Rewards are generally issued as cryptocurrency tokens adding a new dimension compared to traditional OSMs. In fact, in traditional OSMs, users interactions are only "social". Users post and share content on the platform, other users interact using comments or votes to express like or dislike. In BOSMs, users can also interact through "economical" or "financial" interactions, as users can share the cryptocurrency tokens by asset transfer actions, i.e. they can move a certain amount of tokens from a source account to a destination account.

However, the blockchain technology is affected by issues concerning the consensus protocol, such as the 51% attack [23]. Another limitation common to decentralized social system is the eternal dilemma of content moderation, for which a clear solution has not yet been found. Lastly, since the blockchain is an append-only structure, it is hard to modify it in case some illegal content is put on it.

In terms of platforms, the first proposal of BOSM is **Steemit** [10, 11], launched in 2016, which was the first to introduce a rewarding system in a social network [1, 15, 18]. It is hosted on Steem, which natively supports social applications thanks to its set of 38 transaction types, that allow users interaction, both "social" and "financial" interactions. After a controversial series of events that led to a hostile takeover of the Steem blockchain by a single entity, some of its users created Hive through a hard fork, alongside **Hive Blog** as its main interface. Details on both social media platforms and the supported user interactions will be presented in Section 5.

Other platforms such as **Sapien**[2] and **Minds**[3] are implemented on Amazon Web Services and with their ERC-20 token hosted on the Ethereum blockchain. In terms of academic proposals, **BCOSN** [13] focuses primarily on privacy issues. The blockchain is used as a trusted server to provide central control services.

3 RELATED WORK

In this Section, we only deal with the literature on user migration as it is the focus of the paper. Users often tend to migrate, i.e. move to different social media platforms. Among the main reasons, there is the emergence of new platforms, with novel interesting features. But we often find scenarios where users decide to leave a social media due to changes introduced in the platform such as moderation or rule variations. In other cases, conflicts or disagreements in the community lead to the migration of groups of users.

One of the earlier data-driven studies on user migration is [17], that analyzes migration patterns across multiple platforms. Account across different social media platforms are matched by relying on self-published accounts or usernames in Blogcatalog. The study shows the presence of different migration patterns in terms of attention. A reference point in user migration studies is [21] that focuses on permanent migration of activity. They also examine cross-platform migration, by matching accounts between Reddit and Reddit alternatives with an algorithmic approach. They, then, divide users into migrants (those who move all the activities to another platform and remain there), tourists (those who change platform only temporarily), and dual citizens (active in both). It is a macroscopic analysis of user activity that relies on user surveys to understand user motivations. Other works study user migrations between communities in the same platform. In [24], they construct a weighted network that treats a subset of Facebook groups as vertices, while weighted edges represent the amount of user migrations among them, showing the presence of non-random migration patterns. Whereas [4] studies user migration between COVID-19-related subreddits, by analyzing migration both at the microscale (attention migration, shift of activity from post to post) and macroscale (shift of activity of entire groups). They show the presence of migration through the aggregation of activity values, too.

None of these works is focused on the study of the evolution of the subgraphs of users induced by a migration. While some works try to study motivations, none of them tries to predict user migration. And most importantly, none of them is looking at user migration in a BOSM, which happens after a fork event. In such a system, user migration can be studied, with some advantages. First, a fork event effectively generates two platforms, allowing the study of cross-platform migration. Moreover, unlike the other scenarios, account matching is trivial, as user accounts are duplicated. Finally, the blockchain technology, at the basis of BOSMs, represents an invaluable and unprecedented source of reliable longitudinal data.

4 MODELING BOSMS, FORK AND MIGRATION

Blockchain online social media offer their users a rich set of actions and functions to support different kinds of interaction, namely *interaction actions*. Interaction actions - such as comments, likes, reacting and following - generate different types of relationships among users. In some interactions, the relationship between two users is explicit, that is the case of a user A following B; while some others are implicit, such as a user A who likes or leaves a comment on a post made by B. Moreover, all the interaction actions happen at a precise point in time. Finally, besides traditional OSM, BOSMs provide interaction actions not merely "social", rather economical or financial. In fact, users can share cryptocurrency tokens by asset transfer actions. A more detailed example will be presented in the case study, in Section 5.

In general, interaction actions can be modeled as a set of tuples $I = \{(u, v, t, r)\}$ where u and v are users, who explicitly or implicitly interact through an action of type r at time t. We leverage the temporal information associated to each tuple in I to build a sequence of directed multigraphs [12]. Specifically, due to the different types of relation expressed by r, we consider an evolving edge-labeled multidigraph $\mathcal{G}$ represented by a sequence $< G_1, ..., G_T >$ where each $G_t = (V_t, E_t, R, w_t)$ is a weighted edge-labeled multidigraph, and T is the maximum timestamp in I [16]. Each graph of the sequence is defined by the following elements:

- V_t: the set of users u which belong to at least one interaction action in I which has occurred before or at the timestamp t;

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

- E_t: the set of triple (u, v, r) with $u, v \in V_t$ and $r \in R$, which represents a specific type of action taking value on the set R of actions offered by the blockchain;
- $w_t : E_t \rightarrow \mathbb{R}$: a weighting function which, given the triple (u, v, r), returns the number of interaction actions of type r involving u and v and occurring before or at the timestamp t.

Finally, it is worth noting that throughout our analysis, we focus only on additive interaction actions, i.e. actions which can only increase the state of a multidigraph. For example, the "follow" action is additive as once a directed link is added to the graph, it cannot be removed unless we also consider the dual operation "unfollow". This way, in our setting the number of nodes, edges and the values returned by w_t always increase, up to the last timestamp T.

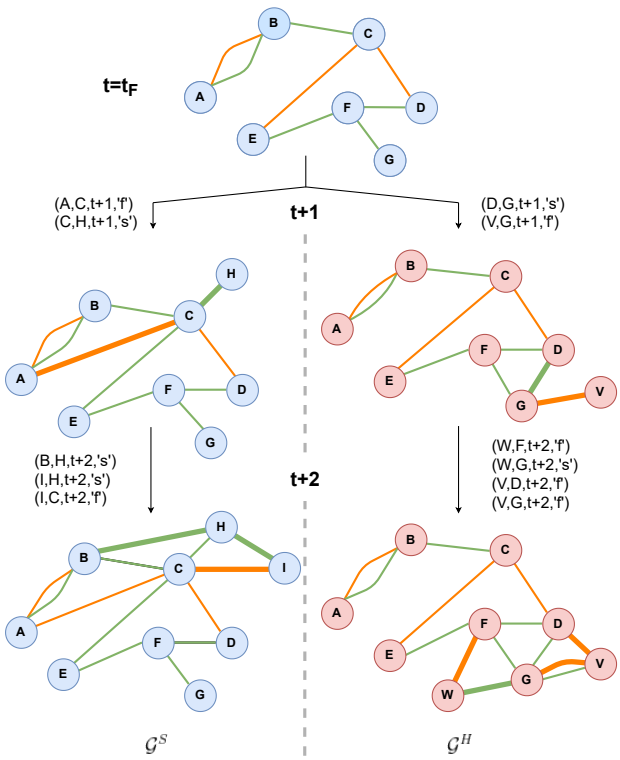

Figure 1: Example of construction of $\mathcal{G}^S$ and $\mathcal{G}^H$ before and after the blockchain fork. The multidigraph on top represents the state of the network at fork time t_F. Then, we report the bifurcation, and the two sequences $\mathcal{G}^S$ and $\mathcal{G}^H$ evolve independently. Alongside the arrows, we display the interaction actions, occurring during a time window, which generate the links in the corresponding multidigraph. Social links are shown in green, and financial links in orange. Bold links indicate the new added interactions. The sequence on the left describes the evolution of the original blockchain - Steem, while the sequence on the right is related to Hive.

Given the above representation, modeling user migration is quite straightforward. As depicted in Figure 1, both the original blockchain - Steem in our case study - and the new one - Hive - result in two distinct evolution multidigraphs: $\mathcal{G}^S$ and $\mathcal{G}^H$, respectively, with a common ancestor representing the multidigraph at fork time t_F. Despite the modeling, the construction of the sequences of multidigraphs is more challenging, since we may cope with two scenarios:

- *internal user migration*: the set of migrant users remains on the same platform but they move to a different "place" in the platform, e.g. a migration from a subreddit A to a subreddit B in Reddit, or a change of group in Facebook. In this case, the identification of the migrant users is immediate, since they maintain the same identity (username or user ID).
- *across-platform user migration*: users migrate to a different platform. In this scenario, it is difficult to identify the migrants - especially in the case of game-changing events, like a fork - due to the lack of explicit signals, such as account deletion or migration communication. In these cases, profile-matching or entity-linkage techniques may be applied to connect accounts on different platforms to the same identity.

In BOSMs, the user migration due to a fork is part of the second scenario, but with a crucial difference: after the fork, the blockchain supporting the original BOSM is completely copied, so that just after the fork both platforms have the same set of users. In this case, and in particular in our case study, profile-matching techniques are not required, since the profiles related to the same identity are explicitly linked, i.e. they are cloned. However, the issue related to the identification of the migrants still persists, as the accounts of migrants are still in the blockchain supporting the original platform, as well as the users who remain on the original platform are also in the new platform.

To identify migrants we exploit the activity of users on both platforms. Specifically, a user u migrates from platform S to H after a fork occurring at t_F, if after t_F s/he does at least one action on H; while a user u remains on the original platform if s/he keeps performing actions on the platform S and no actions on H after the fork event. We call *migrant* the first type of user, and *resident* the latter. It is to note that in the remainder of the paper a third category - *inactive users*, i.e. people who are inactive or have abandoned both platforms - has been only considered in the feature construction for the prediction task. The above categorization of the users is at the basis of the construction of the node sets $V^S_{t_F}$ and V^H_1, we detail in Section 5.

Finally, the above representation and modeling methodology is applicable not only to other blockchain forks, but also to other user migration processes whose data are known. In fact, the construction of the sequence of multidigraphs only requires the set of tuples I. In the absence of blockchain data, data availability and profile matching are the obstacles and put a limit on the applicability of the representation. Specifically, how to collect high-resolution temporal data from the old and the new platform and to carry out account matching are the main issues to be faced when applying the proposed model and methodology to out-of-blockchain contexts.

5 DATASET

Our study on evolution patterns characterizing the user migration due to a fork event relies on data about interaction actions got from the blockchain **Steem** - the original blockchain - and **Hive** -

Table 1: List of social and financial operations used in the study. Each operation is characterized by its name, its type and a full description. We do not differentiate between Hive and Steem because the name and the meaning of the operations in the table are common.

Operation	Group	Description
`comment`	social	A user publishes content or comment on an post.
`vote`	social	An account upvotes or downwotes a content. Users can vote on both posts and comments.
`custom_json`	social	A general-purpose operation designed to add new functionalities without the need for new operations. Social functionalities include: i) **"follow"** to receive updates on what other users are posting; ii) "unfollow" to stop following other users; iii) "mute" to block users from the feed in the case of harassing or unwanted content; and iv) **"resteem/reblog"** to share content with all the followers.
`transfer`	financial	Transfer an asset from one account to another.
`transfer_to_vesting`	financial	Convert an asset to a vesting share and give it to another account.
`delegate_vesting_shares`	financial	Borrow vesting shares to another account, so that it gain the rights to vote contents.
`set_withdraw_vesting_route`	financial	Withdraw vesting shares and transfer the amount to another account.
`transfer_to_savings`	financial	Place assets into time locked savings balances.
`transfer_from_savings`	financial	Transfers assets from the time locked savings balances.

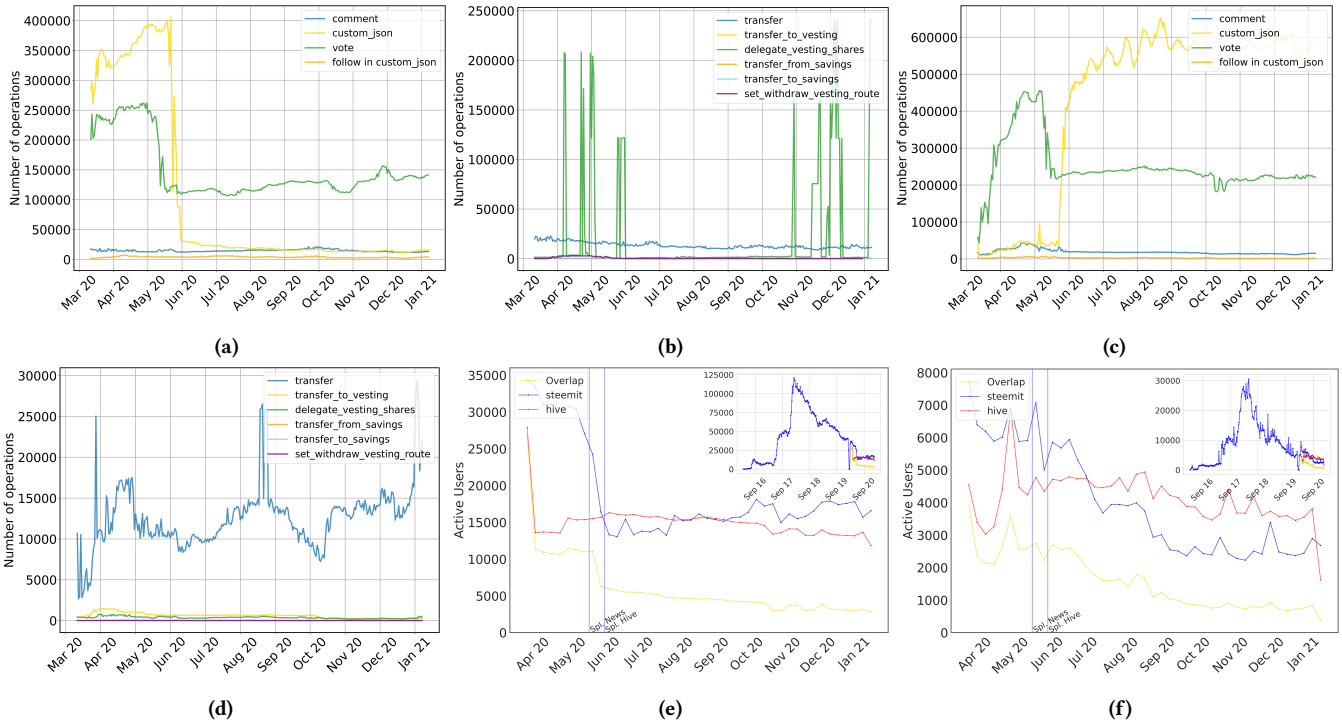

Figure 2: From (a) to (d) daily volume of interaction actions in Steem and Hive blockchains after the fork, grouped by category. In order: (a) daily volume of the social operations on Steem, (b) daily volume of financial operations on Steem, (c) the daily volume of social operations in Hive, and (d) daily volume of financial operations in Hive. In (e) and (f): number of unique users in Steem, Hive and their overlap, i.e. active users in both platforms. In particular: (e) unique users performing social actions, (f) unique users for financial actions. In the inset, unique users over the entire observation period, from 2016 to 2021.

the new descendant blockchain. Indeed, a hard fork, happened on the 20th of March 2020, originating the new blockchain Hive from Steem, after a 51% attack. The attack originated in February 2020 when TRON, a company that owns a gambling-oriented blockchain,

acquired a strong interest in Steem, and thanks to a reservoir of tokens, acquired a strong power in the platform. Fearing that these actions would lead to a hostile takeover, some of the most dedicated users tried to freeze the tokens acquired by TRON through a soft fork. However TRON, thanks to the help of some cryptocurrency exchangers, managed to acquire temporary immense voting power on the platform, up to the point where it could elect its trustworthy witnesses, owning more than 51% of them. With its witnesses in place, TRON managed to revert the effects of the soft fork, thus proving that even a DPoS blockchain is not immune to centralization. The reply to this hostile takeover by the old witnesses of Steem was a hard fork, happened on the 20th of March 2020, originating Hive. Until the hard fork, Hive shares the same blocks, therefore to prevent further hostilities on the platform, Hive witnesses froze or confiscated all the funds owned by the perpetrators of the hostile takeover. Among other innovations, Hive introduced a delayed voting influence mechanism to address possible future 51% attacks, which gives the community time to react preemptively.

The two blockchains, Steem and Hive, support two social media platforms: **Steemit** and **Hive blog**. Steemit, is the original social media platform launched in 2016. In Steemit, users post and share multimedia content, and users can interact with the content through comments, votes. User can follow other users to receive updates when other users post new content. Steemit was one of the first to implement a reward system: users that create the most popular posts are rewarded with cryptocurrency tokens. These tokens can then be traded with other users, for goods or services. Similar characteristics can be observed in Hive.blog, the social media platform born after the fork. In particular, similar interaction actions available, both on the "social" and "economical" side.

All interaction actions are recorded in transactions stored in the supporting blockchains; and they are a subset of all the available actions, generally called *operations*. In fact, both Steem and Hive have released more than 50 operations, whose complete list can be consulted in the official documentations [5, 7]. Among them, we are interested only in those actions that represent an interaction between two users, either explicit or implicit. A complete description of the operations generating interaction actions is reported in Table 1. As shown in the Table, we also distinguish between two main groups of interactions: *i)* financial and *ii)* social operations. Financial operations are those operations designated for rewards and token management, and asset and share transfer; whereas social operations are those that users usually do on traditional social media platforms, like posting, rating, voting, sharing and following.

The details about blocks and operations for both platforms have been gathered through official public APIs, whose structure and usage are similar. We recall that data between the two blockchains are identical up to the fork event, i.e. 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. So, we collected operations from the very first block on Steem blockchain, produced on 24th March 2016, up to January 2021. For Hive, we start from the first block after the fork (20/03/2020), and up to January 2021. Overall, from Steem, we extracted 993641075 operations related to social interaction actions and 72370926 operations related to financial actions; from Hive we have a total of 206224132 social actions and 4041060 financial actions. All the usernames have been pseudo-anonymized as soon as they have been collected and stored in their pseudo-anonymized version.

An overview of the data obtained from Steem and Hive after the fork is presented from Figure 2a to Figure 2d. In detail, as shown in Figure 2a, in Steem, we observe a stable or even increasing trend for vote and `custom_json` operations during the first two months after the fork, but at the beginning of June 2020, two abrupt changess in the volume of operations occurred. On Steem, we observe a steep decrease of the `custom_json` operation, while in Hive (Figure 2c) we observe the opposite, with an increase in the volume of `custom_json` operations. Specifically, on Steem, the volume dropped by 10X in a week (from 350K to 30K operations), while on Hive the volume rose by the same factor. Moreover, after this abrupt increase the overall volume of `custom_json` operations on Hive has reached higher values w.r.t. Steem's volumes before the drop of June. As for vote, the trend in Hive is similar to Steem on the whole observation period, with a sudden decrease of the volume at the beginning of June 2020. The vote trend is different in the bootstrap phase of Hive, where vote operations have continuously increased until May 2020. Moreover, it is to note that after the drop of June, the volume of vote operations in Hive is double the volume in Steem. Conversely, the volume of comment operations and follow operations are quite stable on both blockchains, and marginally influenced by the June's events.

As for financial actions, in Figure 2b and Figure 2d we report the daily volume of each operation belonging to the financial group. Each blockchain is characterized by a specific financial action. In particular, in Steem `delegate_vesting_shares` operations reach the highest daily volumes and are characterized by an unstable trend with a few spikes in the first (April - June 2020) and last (November 2020 - January 2021) months. Such trait might indicate anomalous behaviors in the voting operations since, through `delegate_vesting_shares`, users can "borrow" their voting power to other accounts. On the contrary, in Hive, we do not observe spikes in `delegate_vesting_shares`, and the `transfer` operations are the most common one. In the case of transfer operations, the average volume in Steem and Hive is comparable, i.e. from 10K to 15K daily transfer operations. The remaining operations are quite marginal on both blockchains and have stable trends.

Construction of the evolving multidigraphs: After the gathering, we process the blockchain data so as to cast the sequence of interaction actions returned by the blockchain into the representation framework described in Section 4. In particular, given an interaction action (u, v, t, r), we consider the timestamp associated with the block containing the interaction operation of type r as the time t of the interaction. Hence, we can build each multidigraph G_i of the evolving multidigraph by selecting a one-month temporal window between two consecutive graphs G_i and G_{i+1}. Specifically, we aligned each evolving graph G_i to the 20th day of each month, at 2:00 PM. This allowed us to start the first snapshot post-fork for both the sequences exactly at fork time, for a better comparison of network characteristics. Finally, we selected and grouped the interactions based on categorization defined in Table 1, so that r takes values on the set $\{social, financial\}$.

As also displayed in Figure 1, the construction of $\mathcal{G}$ proceeds incrementally. Given $G_i \in \mathcal{G}$, we define G_{i+1} by first setting $G_{i+1} =$

G_i. Then, we iterate over the interaction actions (u, v, t, r) such that $i < t \leq i + 1$. If the labeled edge (u, v, r) is not in G_{i+1}, we insert it and assign to it a weight equal to 1; otherwise, if $(u, v, r) \in G_{i+1}$, we only increment by one its weight.

As a final note, `custom_json` is an operation that can be used for multiple functionalities, and in the analysis we only considered one of the actions assignable to this operation, i.e. the "follow" action.

6 RESULTS

In this Section, we report the main findings on the structural effects of the fork on the Steem and Hive social and financial networks by taking into account the entire dataset, from 2016 to 2021, resulting into 48 pre-fork and 9 post-fork temporal snapshots. Then, we deal with the problem of predicting whether or not a user will migrate after the fork, within a machine learning framework.

6.1 Structural effects of the fork

The March 20th fork represents a game-changing event in the history of Steem and Hive due to both its exceptional nature and the way it happened, i.e. a reaction of part of the Steem users towards some design choices and hostile behaviors in the original blockchain. In this study, we deal with the impact of this important event on the interaction networks generated by the interaction actions in Steem and Hive, taking separately into account social and financial relationships. Through the representation of the evolution of the blockchains described in Section 4, we aim at identifying to what extent the fork event has made the Steem and Hive interaction networks different.

Evolving graphs (all users) . We first analyze the evolving interaction multidigraphs $\mathcal{G}^H$ for Hive and $\mathcal{G}^S$ for Steem by inspecting different structural properties on each element - a multidigraph - of the sequence $\mathcal{G}^{(\cdot)}$. We analyze them at regular time steps, by following the construction methodology presented in Section 5. Therefore, we have 48 snapshots - multidigraphs - describing Steem before the fork date - *pre-fork*, while for the snapshots after the fork, we rely on data from Steem and Hive, thus obtaining 9 snapshots after the fork - *post-fork* - for both platforms. In Table 2, we show a summary of the network properties measured pre-fork and post-fork. Since our focus is the comparison between Hive and Steem interaction networks, we focus on the properties in the post-fork period, reported in the last two columns of the table. For each platform, we report the average and standard deviation of each property, both on the financial and social networks, separately.

Starting from diameter measures, we can observe similar values, with Hive showing only a slightly smaller diameter. This may suggest the fork had not shrinkage effects on the diameters of both social networks. Similarly, Hive has a bigger largest connected component in both social and financial networks. Other properties computed, such as average clustering coefficient, reciprocity, and degree assortativity are similar across both platforms. The values of degree assortativity suggest a lack of degree assortativity and reciprocity values are also low with respect to other measurements on major online social networks [20]. As for reciprocity, we also observe a further decrease from 0.22 to 0.19, that suggests the creation, on both blockchains, of many non reciprocal links after the fork. In fact, by construction, the sequences of multidigraphs after

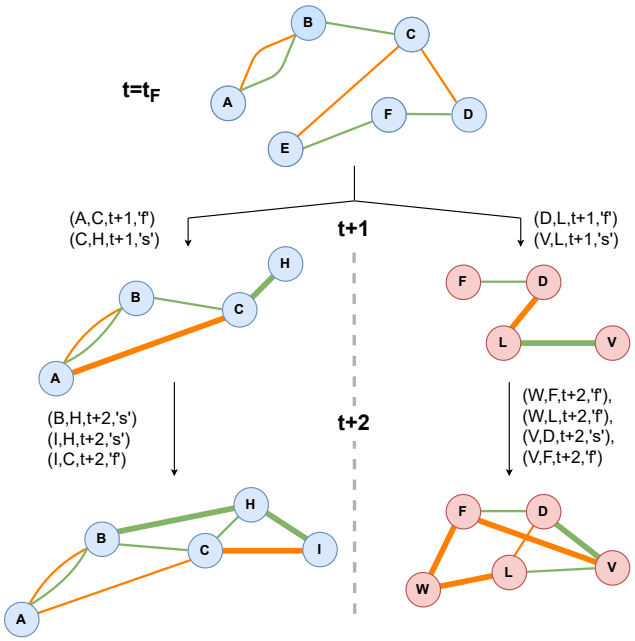

Figure 3: Construction of the sequence of multidigraphs for active users. The multidigraph on top corresponds to the subgraph induced by the set of active nodes on the top graph in Figure 1 - the node G is inactive. Then, for each blockchain, we only maintain resident and migrant nodes, respectively. These multidigraph represents the starting point of the building procedure depicted in Figure 1. Node E will be active after $t + 2$.

the fork keep the information of the previous snapshots, so even a small variation of the indices might indicate a significant change in the structure.

Active users. Finally, we compare Steem and Hive in terms of active users. We measured the number of active users, in different time periods, in both Steem and Hive. We also retrieve the intersection of user actives in both platforms, to get a grasp of the overall overlap. We show the obtained information in Figure 2e for social interactions and in Figure 2f for financial ones. We can see an overall drop in active users, in both social and financial networks. However, there was an already decreasing trend in the number of users, as we can see from the inset of figures, that cover the entire period. The trend continues on both Hive and Steem. Specifically, on the social side, we observe that Steem still has a higher number of active users. We can also note that the overlap - the yellow line in the figures - also drops quickly. Along the time, users stop being active in both platforms, deciding where to focus their efforts. We note a few differences in the financial side. First, the number of active users in the financial side of Hive surpasses Steem. Also, while we still have a drop in overlap, the drop is slower than the one we observe for social actions.

Active users induced subgraph. In addition to the generated evolving networks, we also study in more detail the behavior of active

Table 2: Network statistics. Statistics are computed on the evolving multidigraphs every 30 days.

Metrics	Steem (Pre-fork)		Steem (Post-fork)		Hive (Post-fork)	
	Social	Financial	Social	Financial	Social	Financial
Density (x10^{-4})	17.81 ± 58.0983	0.72 ± 1.3204	1.15 ± 0.0324	0.02 ± 0.0055	1.17 ± 0.0069	0.03 ± 0.0004
Diameter	6.06 ± 1.2784	10.06 ± 11.2655	5.89 ± 0.3333	9.00 ± 0.0000	5.67 ± 0.7071	9.00 ± 0.0000
Degree Assortativity	-0.09 ± 0.0362	-0.13 ± 0.0566	-0.06 ± 0.0002	-0.09 ± 0.0030	-0.06 ± 0.0001	-0.10 ± 0.0008
Reciprocity	0.22 ± 0.0302	0.15 ± 0.0454	0.19 ± 0.0003	0.18 ± 0.0048	0.19 ± 0.0002	0.18 ± 0.0003
Average Local Clustering	0.38 ± 0.0382	0.39 ± 0.0340	0.37 ± 0.0031	0.40 ± 0.0037	0.37 ± 0.0035	0.41 ± 0.0035
Perc Largest Component	58.88 ± 5.2383	17.42 ± 8.1788	57.74 ± 0.5297	12.23 ± 1.5715	58.15 ± 0.1312	15.00 ± 0.0184

users in the period before the fork. In our set of users of interest, we include users active before the fork (3 months before), while including new users that would appear in the following nine months, namely the set U. The obtained set of users is then monitored throughout the period after the fork, by extracting the subgraph induced by the set of selected users in each snapshot of the sequence. More specifically, as summarized in Figure 3, we identified the subgraph induced by U on $G_{t_F} \in \mathcal{G}^S$: it represents the starting point for the construction of the evolution sequences for active users. In the case of Steem, we only keep resident nodes and their links from the induced subgraph and proceed with the procedure described in Section 4. In the case of Hive, from the induced subgraph we only consider migrant nodes and apply the same procedure on Hive data only. In Table 3 we show the network properties for the induced sequences. Compared to the previous networks, we can see that Hive still has lower diameter values. Hive also exhibits bigger largest component, in both financial and social networks. We see slightly higher values of reciprocity, but they are still far from reciprocity values typical of online social networks. Finally, the degree assortativity is not significant in the subgraphs as well.

6.2 User migration

As shown by the above results, the fork-based user migration has been a relevant event that has involved a substantial amount of users. So, for each user, we would like to understand if their choice to adopt a new platform could be explained or even predicted by some user's characteristics or activity; and, in that case, which are the early signals indicating that s/he will move to a new platform.

This problem can indeed be formulated as a machine learning task, specifically a binary node classification task.

Definition 6.1 (User migration prediction task). Given the graph G_t and considering the successive timestamps t', where $t' > t$, we define the user migration prediction task as the prediction of a node migration in one of the successive time steps.

The objective is to predict the two classes (Migrant or Resident) based on several user/node features. The assumption is that user features, at the network structure level, could be predictive of a future user migration. Note that features can be extracted from both layers of the evolving multidigraph: the financial and the social layers; thus obtaining two additional scenarios.

We can define the first case as financial user migration prediction task, whereas for social actions only, we can define a social user migration prediction task.

Definition 6.2 (Financial user migration prediction task). Given a graph G_t and considering the successive timestamps t', where $t' > t$, we define financial user migration prediction task as the prediction of a node migration, on the financial layer, in one of the successive time steps.

Definition 6.3 (Social user migration prediction task). Given a graph G_t and considering the successive timestamps t', where $t' > t$, we define social user migration prediction task as the prediction of a node migration, on the social network layer, in one of the successive time steps.

As in the first task, for both tasks, we predict the label Migrant or Resident based on the user/node features, extracted on the financial or social layers, respectively.

Features and labels. The features we considered are the most common node-level features utilized in many network-based prediction tasks, and that encode information of a node and its neighborhood. Specifically, for each user in G_t, we compute in-degree and out-degree, weighted in-degree, Pagerank, neighborhood average degree, and local clustering coefficient. Alongside the structural information, we also include information on the status of nodes in the neighborhood. We define two additional features:

- Percentage of inactive neighbors: the number of neighbors whose status is inactive at time t, divided by the total number of neighbors.
- Percentage of resident neighbors: the number of neighbors whose status is resident at time t divided by the total number of neighbors.

These features can be computed on both the financial and social layers. Given the defined features, the objective is to predict a potential migration in the future. The labels for the two classes are Migrant and Resident.

Experimental Setting. Our prediction context is the migration from Steem to Hive. Hence, for the following experiments, we focus on the Steem evolving multidigraph of active users, and its financial and social subgraphs. More precisely, we select the snapshot at fork time, $t_F = 2020/03/20$, at 2:00 PM. Then, we obtain the labels describing the future cases, Migrant or Resident. However, the two classes observed are imbalanced. In the social layer, there is more

Table 3: Network statistics on the induced subgraphs of active graphs. The multidigraphs were induced by considering users active three months before the fork and those active after the fork. Statistics are measured on each snapshot of the evolving multidigraphs.

	Steem (Post-fork)		Hive (Post-fork)	
Metrics	Social	Financial	Social	Financial
Density (x10^{-4})	42.20 ± 8.7483	6.16 ± 0.1424	46.21 ± 2.8060	5.80 ± 0.3213
Diameter	7.00 ± 1.5811	8.11 ± 1.4530	5.78 ± 0.6667	7.33 ± 2.0000
Degree Assortativity	-0.07 ± 0.0013	-0.20 ± 0.0008	-0.08 ± 0.0018	-0.20 ± 0.0009
Reciprocity	0.25 ± 0.0005	0.32 ± 0.0038	0.25 ± 0.0031	0.32 ± 0.0023
Average Local Clustering	0.39 ± 0.0045	0.40 ± 0.0057	0.40 ± 0.0052	0.41 ± 0.0043
Perc Largest Component	89.22 ± 5.4459	88.60 ± 1.9379	92.17 ± 0.7496	86.23 ± 2.2581

severe imbalance, as residents are 3/4x more than migrants (66.9 %, 33.1%). While in the monetary layer the two categories are closer, there are more migrants (56.1%) than residents (43.9%) .

The main options to deal with sample imbalance consist of undersampling, so discarding examples from the most numerous classes, or oversampling, which is generating new examples starting from the existing minority class. One of the pivotal advantages of oversampling is that we would not discard any of the available data. Among the many oversampling techniques, the most used is SMOTE [2]. Oversampling allows us to balance the example for both classes.

We perform experiments in a 5-fold cross-validation setting. For each fold, we apply oversampling on the training portion of the fold. Note that oversampling is applied only to the training portion of the data. Then, we train a model and compute a set of evaluation metrics. The metrics are averaged over the five folds. For the evaluation, we compute the main evaluation metrics for classification tasks: weighted F1, accuracy, precision, recall and AUC. The metrics are computed on the testing portion of each fold and then averaged. For the classification task, we rely on standard machine learning methods: Logistic Regression, Random Forest, Support Vector Machine with linear kernel and a Gradient Boosting classifier.

Results. In this Section, we are dealing with three migration prediction tasks: social user migration, financial user migration and user migration.

The experimental results for the social user migration prediction task 6.3 are presented in Table 4. As we can see, the structural features, together with the simple information on the activity of the neighbors, are able to provide a prediction on the migration of a node, even if the performances are modest across the different models. Among them, we observe that Random Forest and Gradient Boosting are leading the tested models in F1, Accuracy, with Gradient Boosting performing better in terms of precision, while RF shows a better recall. The other two models tested lag behind in terms of performance, with lower scores across the board.

Similar results can be observed for the financial user migration prediction task. In Table 5, we report the obtained evaluation metrics. Overall, we can see better performances for all models. Indeed, as in the previous experiment, we can see that Random Forest and Logistic Regression are performing better than the other models. We can infer that financial information may be more informative

Table 4: Social migration prediction. Features were computed on the Steem multidigraph, limiting on edges with type " social". Metrics (Weighted F1, Accuracy, Precision, Recall, AUC) are the average over a 5-fold cross-validation.

Models	F1	Acc.	Prec.	Rec.	AUC
Random Forest	0.66	0.65	0.78	0.70	0.61
Logistic Regression	0.58	0.56	0.79	0.51	0.59
Linear SVM	0.58	0.56	0.79	0.52	0.59
Gradient Boosting	0.62	0.60	0.79	0.59	0.61

Table 5: Financial user migration prediction task. Features computed on the Steem multidigraph, limiting on edges with type " financial". Metrics (Weighted F1, Accuracy, Precision, Recall, AUC) are the average over a 5-fold cross-validation.

Models	F1	Acc.	Prec.	Rec.	AUC
Random Forest	0.69	0.69	0.78	0.78	0.63
Logistic Regression	0.60	0.59	0.75	0.62	0.57
Linear SVM	0.61	0.59	0.75	0.63	0.57
Gradient Boosting	0.65	0.64	0.79	0.67	0.62

for the prediction of future user activities. We may hypothesize a possible explanation for that: as detailed in Section 5, the 51% attack has been conducted by gaining a large amount of voting power, and the reaction to the attack acted on the same direction. Since the voting power is strictly related to financial operations, such as exchanging assets for shares and borrowing shares to gain more rights to vote, the structure of the resulting financial interaction networks has been influenced by the dynamics leading to the hard fork, and the resulting migration of one of the factions.

In Table 6 we show the results for the user migration prediction task. In this task, we are combining features from both the social and financial layers, fully leveraging both the evolving graphs. The concatenated features provide additional information for the prediction of the user migration. Overall, we can see an improvement

Table 6: User migration prediction. The features are a concatenation of those computed on the Steem financial network and Steem social network, respectively. Metrics (Weighted F1, Accuracy, Precision, Recall, AUC) are the average over a 5-fold cross-validation.

Models	F1	Acc.	Prec.	Rec.	AUC
Random Forest	0.71	0.71	0.72	0.78	0.71
Logistic Regression	0.66	0.66	0.67	0.77	0.65
Linear SVM	0.66	0.67	0.66	0.78	0.66
Gradient Boosting	0.70	0.70	0.70	0.78	0.69

in the metrics all across the board. Specifically, models that were performing the best improve their performances over the previous migration prediction tasks. In addition, we can observe that the additional information aids the models that were not performing well, like SVM and Logistic regression, that see an improvement over all the metrics. The obtained results suggest the need for modeling more layers to fully understand user's behavior.

Finally, we perform a feature importance analysis to highlight the most predictive features, and, in our specific temporal setting, to identify the early signals of willingness to migrate. The features along with their importance ranked in descending order are displayed in Figure 4. The most important features are related to both social and financial layers. Specifically, the clustering coefficient in the social layer, and the neighbor degree in both social and financial ones are among the most important features. The analysis confirms the importance of taking into account information derived by both types of interaction action for the user migration prediction task.

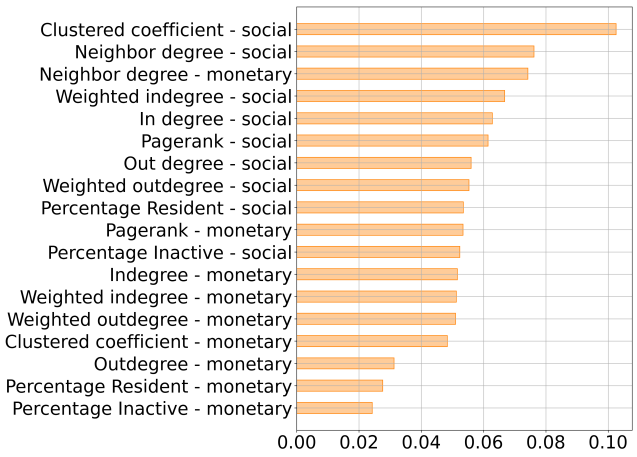

Figure 4: Feature importance for the best performing model, i.e. Random Forest, on the user migration prediction task. Importance values are based on the mean accumulation of the impurity decrease within each tree of the Random Forest.

7 CONCLUSION AND FUTURE WORKS

In this work, the topic of user migration among OSM has been addressed. User migration has been a relevant process in the past with the migration of users from a platform to another, new and more interesting one. But it might become more and more massive with the crisis of traditional platforms and the emergence of new social media paradigms, among which the most interesting are blockchain social media with their promises to be able to overcome the many well-known issues of traditional OSM. Also BOSMs have issues as they are not yet mature platforms, often subject to internal changes that can lead to blockchain forks and related user migration between the overlying services.

Despite the importance of these processes, research on user migration in general, and on blockchain forks in particular, is still at an early stage. Among the the many obstacles to research on this topic, there is certainly the difficulty in collecting representative datasets, as they must be longitudinal and need users matching. In this sense, BOSMs represent an invaluable source of data in this field.

To the best of our knowledge, this is the first study on blockchain fork and user migration in BOSM. It contributes with a general user migration model applicable to other BOSMs; it shows that it is possible to predict user migration even on the basis of the network structure only, as in the Steem-Hive case study. The methodology, the tools and the results herein provided are applicable in the case of possible hard fork, but they do not offer practical solutions to prevent a hard fork. In fact, platform administrators, if a hard fork is a very likely event, should look at both social interactions and economical transactions to identify the set of users who likely will abandon the old platform to join the new blockchain. To this aim, our findings about prediction have highlighted that, in a stratified context where social and economical relationships mixed together, both dimensions are important in describing and forecasting users' behaviors during and after a shocking event in the network. Actually, we have focused on proprieties of the networks to predict user migration, however a further step would be to understand the motivations that lead a user to migrate or not. To this aim, an integration of the features extracted from the textual content produced by users with the structural features might highlight the reasons of the migration.

We hope that this work will pave the way for other studies on blockchain fork and user migration in order to better understand these so important, but still largely unknown processes. Besides user migration, the representation for the blockchain data modeling might be applied to a few phenomena characterizing the Web 3.0, for instance 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, 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*. ACM, 127–132.

[2] Nitesh Chawla, Kevin Bowyer, Lawrence Hall, and W. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)* 16 (06 2002), 321–357.

[3] Thomas H Davenport and John C Beck. 2001. *The attention economy: Understanding the new currency of business*.

[4] 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 *Workshop on Cyber Social Threats at the 15th International AAAI Conference on Web and Social Media (ICWSM 2021)*. AAAI.

[5] Steemit developer documentation. 2021. Broadcast Ops. https://developers.steem.io/apidefinitions/broadcast-ops

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

[7] Hive Developer Documentation. 2021. API Docs - API Definitions. https://developers.hive.io/apidefinitions/

[8] 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.

[9] Barbara Guidi. 2020. When Blockchain meets Online Social Networks. *Pervasive and Mobile Computing* 62 (2020), 101131.

[10] Barbara Guidi, Andrea Michienzi, and Laura Ricci. 2020. Steem Blockchain: Mining the Inner Structure of the Graph. *IEEE Access* 8 (2020), 210251–210266.

[11] 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.

[12] Petter Holme and Jari Saramäki. 2012. Temporal networks. *Physics Reports* 519, 3 (2012), 97–125. Temporal Networks.

[13] Le Jiang and Xinglin Zhang. 2019. BCOSN: A blockchain-based decentralized online social network. *IEEE Transactions on Computational Social Systems* 6, 6 (2019), 1454–1466.

[14] Alan E Kazdin. 2017. The token economy. In *Applications of conditioning theory*. 59–80.

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

[16] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. 2014. Multilayer networks. *Journal of Complex Networks* 2, 3 (07 2014), 203–271.

[17] Shamanth Kumar, Reza Zafarani, and Huan Liu. 2011. Understanding User Migration Patterns in Social Media. In *AAAI*.

[18] 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 (WebSci19)*. 145–154.

[19] 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* (2021), 1–23.

[20] Seth A Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. 2014. Information network or social network?: the structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web (WWW14)*. ACM, 493–498.

[21] 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 *ICWSM*.

[22] T Poongodi, R Sujatha, D Sumathi, P Suresh, and B Balamurugan. 2020. Blockchain in social networking. *Cryptocurrencies and Blockchain Technology Applications* (2020), 55–76.

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

[24] 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. *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)* (2018), 357–362.