The Hub-bit: traits, role and influence of central nodes during a user migration

Alessia Galdeman, Cheick Ba, Alessia Caputo, Matteo Zignani, Sabrina Gaito
Dipartimento di Informatica “Giovanni degli Antoni”, Universitá degli Studi di Milano, Milan, Italy
Email: *lastname.firstname@unimi.it*

I. Introduction

It is a well-established fact that the evolution of many networks is characterized by turning-point events, such as network splits, partitioning or even user migration, i.e. users moving to a new competitor platform or system. In particular, split events, also called hard forks, are very frequent in blockchain-based systems like Bitcoin, but also on blockchain-based online social networks. In this context, split events generate in case of disagreement during data validation processes or, in more striking cases, they result from a conflict within the community. What makes hard fork so interesting is that these split events may lead to a complete duplication of entire social networks, along with detailed data on the process. Finally, after a fork event, users can decide to *a*) be active on both the original and the new social networks, *b*) stay only on the original platform, or *c*) definitively migrate on the new platform and abandon the old one.

In this complex scenario, different factors may influence the user's decision to choose between the aforementioned options. And, while it is hard to gain insight on a user's motivations, it would be important to investigate the possible impact of the network structure and the role played by important nodes during and after user migration. Hence, our research goal is twofold:

- How central nodes behave before, during and after a hard fork and the resulting user migration, and what is their role?
- Does the decision made by the central nodes whether to migrate or not influence their neighboring nodes?

To this end, we studied a fork event on Steemit, one of the most widespread blockchain-based online social network. In this online social network, the fork event has led to the birth of a new blockchain, Hive, which is now supporting different social platforms. In this scenario, users maintain the same username across the two blockchains and are able to be active on both social networks. These characteristics sustain the study of the decision made by users, and, in particular, by central nodes. In fact, using data extracted from their blockchains, we analyzed the decision of central nodes, or *hubs*, looking at their activity on both platforms. Specifically, we investigated the potential influence of these important nodes on decision making across the social media platforms. Our study has highlighted that:

- the majority of hubs has decided to stay active on both platforms, so that they exploit their status quo in the original platform and explore the opportunities offered by the new one;
- the distribution of the decisions is influenced by how we identify hubs, i.e. if we use in-degree or out-degree as measure of centrality; and
- the neighbors of hubs tend to choose more often the new platform Hive with respect to the overall distribution of the decisions of the user platform.

II. Background

Blockchain based OSNs: Blockchain technology has led to blockchain online social networks (BOSN). In these platforms, the underlying blockchain provides data storage and data validation. The validation process enables the production and exchange of cryptocurrencies. Among the proposed BOSNs, one of the most interesting is Steemit [1]. Steemit is a blockchain social network launched in March 2016, hosted on the Steem blockchain. Like other BOSNs, it relies on a cryptocurrency, called STEEM, that can be exchanged for goods or services. Moreover, the cryptocurrency fuels a reward mechanism, which supports the network growth by repaying users for their activity on the platform.

In these networks, we can observe network splits. In this scenario, users duplicate the network on a different blockchain: then, they can start interacting on the newly generated platform. Such a split has happened on Steemit as well. After a dispute inside the network, a group of users copied the blockchain data on a new blockchain called Hive. Alongside it, they created a new interface - Hive blog - and cryptocurrency system; thus, effectively creating a new social media platform, active from the 20th March, 2020. Note that users are provided with the same username on both platforms, which means that they can still be active on both the original and the new platform.

III. Dataset

All users' activities of both Hive and Steemit are tracked down by the actions that they perform, called operations, captured with a granularity of three seconds. The blockchains supporting the two platforms store user operations as transactions. Guidi *et al.* [2] categorized the several types of operations (more than 50) into three macro types: *social, financial and management*. Here, we are interested in interactions between users, so we focus only on social operations, for instance follow, rating, sharing, posting, and financial ones, whose goal concerns reward, transfer sharing, token management, and asset.

Through specific API, we collected all the users' operations from June 3, 2016 up to January 21, 2021. Concerning the Steem blockchain, the obtained data collection consists of $993,641,075$ social operations and $72,370,926$ financial operations; while Hive registers a total number of $206,224,132$ social operations and $4,041,060$ financial actions.

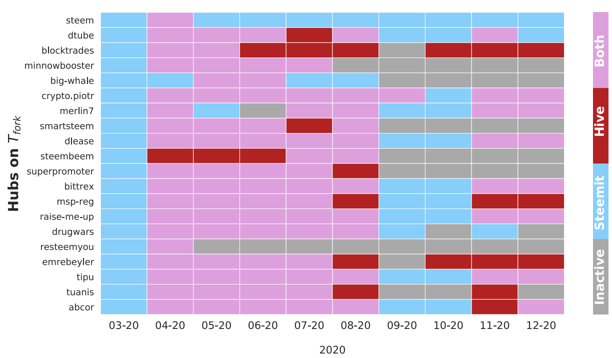

Fig. 1: **States of out-degree hubs in the monetary layer:** each row of the heatmap tracks the states of an out-degree hub of the monetary layer. As shown in the legend, pink slots indicate that the user was active on both platforms in the specified snapshot. On the other hand, red and light blue slots stand respectively for users who migrate to Hive and users who stay on Steemit. Finally, grey slots indicate that the user becomes inactive, i.e. s/he does not perform any action.

IV. Methodology

We modeled each action collected by the APIs as a tuple $I = (u, v, t, r)$, that describes an interaction between users u and v of type r at time t . As cited before, we grouped the transaction types, so that $r \in \{social, financial\}$. Now, we can build a sequence of edge-labeled multigraphs $< G_1, .., G_{T-1} >$, where each $G_i, i = 1, \ldots T-1$ represents a 1-month window aligned to the day of the hard fork. Specifically, each snapshot covers transactions with a timestamp from the 21th of month $i-1$ to the 20th of month i[1]. We denote as $G_{T_{fork}}$ the snapshot that ends on the day of the hard fork which caused the rise of Hive. Moreover, the sequence of graphs is incremental. This implies that once an edge and its end-point nodes are added to the graph, they cannot be removed in the following snapshots but only updated in their edge weight. Specifically, an edge $e = (u, v, r, w) \in G_i$ indicates that nodes u and v had w interactions of type r with $t \leq i$, so once e is added, its weight w can only grow over time.

Our main goal is to study central nodes, in particular their decisions and the influence they have on their neighbors. We distinguish central nodes depending on their out-degree or in-degree. Note that the two measures can highlight different meanings of being a hub. In fact, a high in-degree identifies nodes that are popular in the social layer or which receive a lot of actions in the monetary one. On the other hand, through the out-degree, we identify those nodes that perform a lot of financial or social actions. Thus, we study the behavior of different sets of central nodes, depending on the degree type and also the layer they belong to (social or financial).

Our first analysis concerns the state of central nodes. The *state* indicates whether a node in a given one-month window is active only in Hive or Steemit, in both or neither of them. Note that we define as active a node that performed at least one social or financial action in the fixed time window. This study provides insights into the temporal behavior of high degree nodes. Moreover, the monthly granularity of the state evolution may also highlight some anomalies that characterize a specific month. Next, we focus on the influence of the selected hubs on their neighbors. First, we divide the nodes according to their final decision, defined by the last active state, so that the decision of each hub is $d \in \{$*Hive, Steemit, both*$\}$. Then, we pick the set of users with $d =$ *Hive*, denoted as *migrants*, and the users with $d =$ *Steemit*, called *residents*. Then, we sort the node sets by in-degree and out-degree, separately and select the top 20 nodes. In this way, we obtain four different sets of nodes that represent: *(i)* the 20 highest in-degree nodes that migrate to Hive, *(ii)* the 20 highest in-degree nodes that stay on Steemit, and their counterpart considering the out-degree. For each hub, we perform the following steps:

- we collect the in-neighbors, because we are interested in the followers;
- we observe the state of each neighbor;
- we compute n = *number of active neighbors* and m = number of migrant neighbors (nodes that migrates to Hive);
- we compute a random sample following a binomial distribution $b(n, p)$, where p stands for $P(state(node) =$ *migrant*$|node \in$ *active nodes set*$)$; and
- we compare the distribution of the random sample with m, the actual number of migrant neighbors of the fixed hub, in order to quantify how the results are different w.r.t. a null model.

V. Results

Hub's state discovery: As described in Section IV, we observe the behavior of the out-degree hubs in the monetary layer. Specifically, we look at the state each hub has on each time window during the 9 months of observation. Fig. 1 summarizes the trend of the hubs. We can observe two main insights: first, there is an initial period of indecision, where most of hubs are active on both platforms, and then around August/September ($6th$ and $7th$ column of the chronomap) we observe a general trend of taking sides or become inactive. The second observation concerns the final decision of the hubs: half hubs decide to stay on both platforms, while the other half equally divides between Steemit and Hive. The distribution of decisions for the other settings is shown in Table I. We note that the out-degree hubs are equally distributed over the

[1] $i = 1$ corresponds to March 2020.

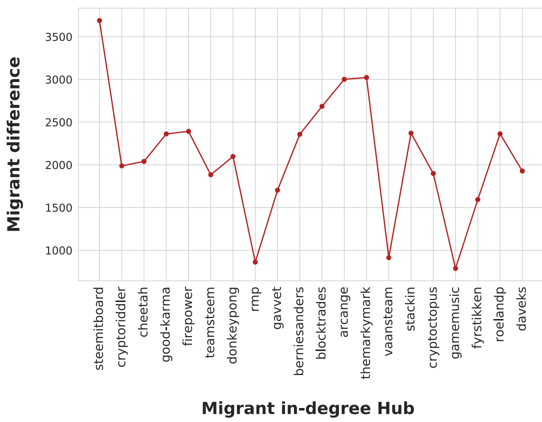

Fig. 2: **Migrant difference for resident in-degree hubs**: The plot shows, for each resident in-degree hub, the difference between the actual number of migrant neighbors and the average number of migrants in the binomial sample $b(n, p)$ with n = number of active neighbors and p = probability of an active node being migrant = 0.194.

two platforms even in the social layer, where 40% of hubs have stayed on both platforms, and the remaining 60% divided between migrants and residents. On the other hand, the in-degree hubs present different distributions for the social and financial layers. In the social scenario the majority of nodes have stayed on both platforms (65%) and the remaining nodes have preferred to migrate. On the opposite, in the financial layer the percentages are more uniform. *To sum up, the most common decision for hubs is to not decide, staying active on both Steemit and Hive.*

TABLE I: **Distribution of decisions for each experiment:** the macro columns distinguish between the nodes with highest out-degree and the ones with highest in degree. Then, we define the percentage of nodes with different final decision in both social and financial layers.

	Out-degree		In-degree	
	Social	Financial	Social	Financial
both	0.4	0.50	0.65	0.40
Hive	0.3	0.25	0.20	0.25
Steemit	0.3	0.25	0.15	0.35

Hub's neighborhood decisions: The next goal of our analysis is to discover the influence that hubs have on their followers. By applying the method in Section IV, we obtain, for each hub of each type, a random distribution parameterized with its number of followers and a probability of being migrant (0.194) and the actual number of migrant followers m. First, we computed $\alpha_h = P(X \leq m)$, but all the results were the same for all hub h, i.e. 1. This means that the hub's neighbors tend to choose Hive more often than the general behavior, and this is true for all types of hubs, not only migrants. Thus, we plot the difference between the expected value of the binomial and m, the actual number of migrants. An example is depicted in Fig.2, where we consider the migrant hubs for in-degree. We observe that the difference is always over 0, confirming the value of α_h. Moreover, for some hubs the difference value mark a peak (see users *steemitboard, arcange* and *themarkymark*). To sum up, *the hubs' decision does not really have an influence on the neighborhood. However, hubs' neighbors are more likely to migrate with respect to other users.*

VI. Discussion

In conclusion, this works aims to observe the decisions of central nodes and the influence on their neighbors, in the context of a blockchain-based social network's split event. We focused on the fork event involving Steemit, and leading to the birth of a new social network, Hive. Since the latter has maintained the same usernames as Steemit, we were able to track the user migration. We modeled both platforms and the fork as a sequence of edge-labeled multigraph, composed of two layers: social and financial ones.

On this data source, we observe the variety of decisions of hubs defined by in-degree and out-degree, on both social and financial layers, highlighting that the most common decision is staying active on both platforms. Then, we focus on the decisions of hubs' neighbors, studying if they are influenced by the state of their hub. Results show that hubs' neighborhood tends to migrate more frequently than the general distribution, regardless of the hub being migrant or resident.

Future works in this context may concern the *centrality transferability*, i.e. the analysis on how the centrality of nodes is correlated across different layers. In this case, the different layers could come from the stratification of the social or financial layers, dedicating a layer for each operation type. Another possible extension concerns the *witness* hubs. In fact, Steemit and Hive use the *Delegated Proof of Stake* [3], a protocol for data validation performed by a specific subset of elected users, called *witnesses*. Witness users can play an essential role in the network evolution, because they are allowed to do management operations that are restricted to regular users. In this scenario, we can extend this work by defining as hubs the witnesses who performed more management actions. The number of management actions a user performs is correlated with the number of times it is elected as witness. In the same scenario, we can extend the influence study to witness users, with the intent of observing the social pressure they exert [4] through financial and management operations.

References

[1] B. Guidi, A. Michienzi, and L. Ricci, "A graph-based socioeconomic analysis of steemit," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 2, pp. 365–376, 2021.

[2] ——, "Steem blockchain: Mining the inner structure of the graph," *IEEE Access*, vol. 8, pp. 210 251–210 266, 2020.

[3] "Steemit Whitepaper." [Online]. Available: https://steem.com/steem-whitepaper.pdf

[4] J. Sun and J. Tang, "A survey of models and algorithms for social influence analysis," in *Social network data analytics*. Springer, 2011, pp. 177–214.