Interaction Communities in Blockchain Online Social Media

Barbara Guidi
Department of Computer Science
University of Pisa
Pisa, Italy
Email: guidi@di.unipi.it

Andrea Michienzi
Department of Computer Science
University of Pisa
Pisa, Italy
Email: andrea.michienzi@di.unipi.it

***Abstract*—Surfing Online Social Media (OSM) websites have become a daily activity for a large number of people worldwide. People use OSMs to satisfy their innate need to socialise, but also as a source of information or to share personal facts. Thanks to the massive success of cryptocurrencies, the blockchain technology gained popularity among researchers, giving birth to a new generation of social media. Steemit is the most well-known blockchain-based social media, and it is based on the public blockchain Steem. Steemit employs Steem as data storage, and to implement a rewarding mechanism that grants cryptocurrency to pieces of content that are considered relevant by the users. Steem represents the first experiment that integrates OSMs and an economic rewarding system on the same platform, and in this paper, we inspect the interactions among the users from a community perspective. We apply two community detection algorithms on five graphs that model just as many facets of the Steem blockchain and test the detected structure against three measures for community structure evaluation. Findings show that communities tend to be very large, index of how much users are encouraged to interact as much as possible, and in particular, in the monetary graph, we detect a large number of the block producers of Steem.**

I. Introduction

Online Social Networks (OSNs) have become the most important channels through which people can socially interact with each other. Creating posts, reading, commenting and evaluating other user's posts have become one of the daily activity of people to satisfy their need for social interaction. OSNs, such as Facebook, and other social media sharing platforms, such as Instagram or TikTok, are usually implemented exploiting a fully centralized architecture, meaning that all data is stored and maintained by a single entity. Using a centralized architecture is a flaw if we consider problems like scalability, dependence on a provider, privacy [1], and poor value redistribution [2]. To overcome the weaknesses of centralized architectures, researchers and developers have focused their attention to the possibility of implementing an OSN platform on top of decentralized architectures.

Primarily fuelled by the revelations of the Cambridge Analytica scandal [3], many Decentralized Online Social Networks (DOSNs), implemented on P2P architectures, have been proposed [4], [5], [6]. However, since the revamp started in the early 2010s thanks to the success of Bitcoin [7], the blockchain technology, has started to gain importance. Blockchain quickly evolved since then to support not only the storage of economic transactions, but also the storage of more general data, up to the point where the blockchain can support the execution of code, as in Ethereum. One of the outcomes of this enormous interest towards the blockchain technology is its integration in numerous scenarios, including the one of social networks. Thanks to the support of blockchain, a new generation of decentralized Social Media, referred to as Blockchain Online Social Media (BOSM), have been proposed [8]. BOSMs had a relevant impact on the OSN scenario, and there are many BOSMs platforms which rely on blockchain technology for the storage and rewarding systems to encourage their users to create and share high quality contents. One of the most well-known BOSM platform is Steemit[1], implemented on top of the blockchain Steem. The blockchain Steem natively supports the development of social applications through a rich set of 38 transaction types. The transactions are used to store social, economic and management information on the blockchain, and thanks to the rewarding system users are able to gain based on their social activity. The rewarding system is based on the attention economy, and is such that the more users vote a piece of content, the higher the reward will be. The economic incentive is granted as cryptocurrency, and is such to encourage the production of better contents that can generate more votes. While the main motivation behind the rewarding system is to fight the problems of creation of low quality content, fake news, and censorship [8], the real effect on the social behaviour of users is still unclear. Indeed, it was observed that BOSMs encourages polarisation [9] or the distortion of the social interactions. Our main concern is that economic incentives may fuel numerous interactions which are not strictly social, but also have economic gain as main goal. This leads to situations in which platforms have unusually large interactions communities.

In this paper, we conduct an investigation concerning the community structure detected on a set of graphs induced by the Steem blockchain in order to find details concerning the combination of both the social and economic side of this platform. We evaluate 5 different graphs, focusing on just as many aspects of the social blockchain. In detail, we

[1]https://steemit.com/

consider the Interactions graph, containing all the transactions, the Social Interactions Graph, which considers only social transactions, the Social No Bot Interactions Graph, which is obtained by removing a small set of identified bots from the Social Interactions Graph, the Follower-Following Graph, which considers only the *follow* transactions, and finally the Monetary Interactions Graph, which considers only the economic transactions. We apply two community detection algorithms, namely the Louvain Method and the Label Propagation. This choice was mainly driven by the fact that the five graphs considered contain 1.2 million nodes and up to almost 200 millions of edges. Indeed, due to the size of the graphs, algorithms that cannot be parallelized or have high time and space complexity cannot be easily used. We check the community structure returned by the two approaches on all the five graphs, and assess the quality of the structures using three measures: modularity, intra-cluster density, and conductance. Results show that the structure returned by Label Propagation is too simplistic and is usually made of one large community which contains almost all the nodes of the considered graphs. On the other hand, Louvain Method returns a much more relevant community structure, as shown by the values of modularity, intra-clsuter density and conductance.

The rest of the paper is structured as follows. In Section 2 we review the relevant literature concerning BOSM, and the most important concepts of community detection. In Section 3 we describe how the transactions contained in the blockchain can be interpreted as interactions. In Section 4 we describe the relevant features of the dataset used. In Section 5 we show the experimental results, focusing on the community structures identified on the graphs. Section 6 draws conclusions and points to possible future works.

II. Literature review

The most used OSN platforms are implemented through centralized servers storing all the information of the users and interfaces which take the form of web or mobile applications. Adopting a set of centralized servers has several drawbacks including privacy, dependence on a provider, and scalability [1]. In recent years, thanks to the fact that OSNs have become the most popular web services, they have become one of the main channels through which privacy is violated but at the same time they allow and extremely rapid and efficient spread of information. Alongside the technical challenges that need to be faced when implementing massive worldwide, real time systems, an ever-increasing interest in privacy preservation has led researchers and developers to rethink social networking platforms considering a decentralized architecture.

The decentralization presents some benefits, but also introduces a number of new challenges. The main ones are: **Data availability and persistence**, which addresses the problem of where to store and how to make available users' data over the service; **Information diffusion**, which addresses the problem of how information is spread on the platform; **Privacy**, which addresses the problem of ensuring that users data are protected and not accessible without consent; **Mobile users & Physical locality**, addresses a wide range of problems and possibilities connected to the fact that most users access OSN services through mobile devices; **Trust**, addresses the problem of how to define and use the concept of trust between users and devices.

The first generation of Decentralized Online Social Networks (DOSNs) did not have a major impact, however, thanks to the momentum gained by the blockchain technology, new solutions are being proposed. A DOSN implemented with the support of a blockchain is called a Blockchain Online Social Media (BOSM) [8]. The main goal of BOSMs is to overcome the issues connected to fake news diffusion and censorship, by exploiting the blockchain for the implementation of a rewarding strategy.

A. Blockchain for Online Social Media

Despite its innovative nature, there are already numerous active BOSMs platforms. The most famous is **Steemit** [2], [10], which has surpassed 1.5 million registered users, and represents an important and successful alternative to centralized OSNs. It was the first of its kind and it introduced the idea of rewarding user activity to oppose the spread of fake news and low quality content. It is implemented on its own blockchain, called Steem.

Peepeth[2] is a BOSM that runs over Ethereum, and has features similar to Twitter. It includes: Peepeth, an open-source smart contract running on the Ethereum blockchain, which acts as data storage; and Peepeth.com, the front-end for the interaction with the smart contract. Data is saved in the Ethereum blockchain, and anyone can monitor Peepeth's public data. Users can evaluate contents on Peepeth through a feature called Ensō. Users can give only one Ensō per day, and it should be given only to contents which are of very high quality. This feature should encourage the creation of "dignified, beautiful, and timeless content" [3], as stated by the developers. Since the number of Ensō is limited to one per day per user, there is a very high competition among content creators to create contents relevant to the platform. Additionally, an Ensō cannot be revoked. Lastly, Peepeth is moderated transparently, because of the open nature of the blockchain.

Sapien[4] provides a platform for users to create, publish, access, view, and evaluate content. Contrarily to other social media platforms, which are usually specialised in only one type of content (e.g. Youtube is primarily thought for videos), Sapien provides a common platform for articles, images, videos, and other. A piece of content can be public or private, implying that the platform guarantees a level of visibility of social data. The social services offered are: create a personal profile, add friends, form tribes (groups), share and comment on posts. Content creators are rewarded on Sapien directly through Peer-to-Peer (P2P) transactions.

[2]https://peepeth.com/welcome
[3]https://peepeth.com/about
[4]https://www.sapien.network/

Minds[5] is an open-source BOSM based on the contribution of its users. Contributions can take many forms: account setup and verification, referring new users, generating high quality content, maintaining an active channel, code improvement or bug fixing, and more. Each contribution is rewarded in tokens on a daily basis. The amount of tokens granted as reward is directly proportional to the contribution of the user and inversely proportional to the activity of all the users. The tokens earned can then be used for several purposes, such as tipping other users or channels, or they can be exchanged for more views on specific pieces of content. To ensure that the platform is always available, its core is hosted on Amazon Web Services, but the project is open source. The the economic system is implemented on Ethereum, through a smart contract, and users need an Ethereum identity to be able to receive rewards.

Hive Blog is one of the many interfaces of the Hive blockchain, a new blockchain that originated as a fork of Steem after a controversial series of events involving a large sum of cryptocurrency. Although it is a new platform, its core mechanisms are similar to Steemit.

Appics is a social media application that revolves around image sharing. It runs on top of the Telos blockchain [6], but its tokens, called APX, can be exchanged with tokens on the EOS blockchain. It was initially implemented with the same blockchain technology of Steem.

In terms of academic proposals, **BCOSN** [11] focuses primarily on privacy issues. The blockchain is used as a trusted server to provide central control services.

B. Community Detection

A research field relevant to the scenario of OSNs is graph theory and in particular the problem of community detection, whose aim is to find relevant sets of nodes in a given network. Contrarily to other graph problems, there is no unique or widely-accepted definition of what a community is, but, intuitively, a community can be seen as a set of "well-knit" nodes of the whole graph. The absence of a formal definition let to the proliferation of different possible definitions, each of which is based on different ideas and concepts. The most well-known definitions rely on a pure topological aspects, such as graph structure, visits and distances. For instance, some methods aim at building partitions of the network that minimize or maximize an objective function or find good approximations with strict time/space bounds. One such example is the Louvain Method [12], which aims at maximising the modularity based on the contribution of single nodes. Other methods for community detection aim at finding particular structures, such as cliques, others still rely on more complex procedures. For instance, Label propagation is based on assigning nodes to communities based on labels propagated through edges [13]. A detailed survey of community detection techniques is presented in [14].

[5]https://www.minds.com/

[6]https://telos.net/

Given a community structure of a network, it is possible to assess the quality of the structure through some measures. Among the most popular and important we find: Modularity, Intra-cluster density, and Conductance. The most used measure in literature is the *Modularity* [15]. The modularity of a community structure is based on the density of the edges of the given communities compared to the expected density of random defined communities. Modularity can be written as $Q = \frac{1}{2m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})\delta(i,j)$, where m is the number of edges of the graph, i and j are two vertices, A is the adjacency matrix and $A_{ij} = 1 \iff i$ and j are connected by an edge, k_i is the degree of node i, and δ is a function that returns 1 if the two nodes passed as argument belong to the same community, and 0 otherwise. Although this is a very commonly used measure, it was also shown that in some cases community structures with the highest modularity value are affected by the resolution limit [16].

The *intra-cluster density* of a community is the density of the subgraph consisting of the nodes inside the community. To assess the quality of the whole community structure of the whole graph, the intra-cluster density values of all communities are aggregated. Among the possible aggregation functions we find the minimum and the average.

Lastly, we cite the *Conductance*, which measures how a community is well separated from the other communities. Considering that also for the conductance we have a list of values, one for each community, we need to consider possible aggregation functions, such as the average or the maximum.

C. Community detection in Online Social Media

The study of the community structure can be beneficial for multiple problems in the scenario of OSNs. In [17] the authors propose to tackle the problem of sentiment analysis at a community level, rather than a global level, to have a more detailed and granular evaluation of the sentiment. In particular, the study shows how the results obtained from the community detection task and the sentiment analysis task can be mutually beneficial to increase the quality of each result. In [18] the authors propose a study on Twitter specifically targeting users discussing about human papillomavirus vaccines. They propose a parallel study of topic detection on the tweets and community detection on the follower-following networks, showing that there is some alignment between the results. Community detection was also used for the role identification task. In [19] the authors show that including the direction of the edges can play a crucial role towards the quality of the results. Thanks to their analyses, the community structure uncovers that some users have more important roles in their network. The authors of [20] propose a stance classification method based on community detection. Starting from a handful known profiles, through the execution of community detection on multi layered networks, they are able to understand the stance of each user towards a certain topic. In [21] the authors study the community structure in content sharing social networks. They show that their community-detection analysis framework is beneficial for link prediction in follower-following networks.

D. Community detection in blockchain

The task of community detection was also studied specifically in networks derived from blockchain scenarios. In [22] the authors propose a technique for de-anonymisation of Bitcoin addresses which is guided by the application of a community detection algorithm. The authors firstly build a hint network, which is derived from the transactions appearing in the Bitcoin network. In the hint network, nodes represent users and edges are added whether the two users are suspected to be the same user. A community detection algorithm is applied to find dense parts of this graph or, in other words, to find sets of users for which there are many hints suggesting that they are the same user. A similar approach is presented in [23], where a graph is created according to the transactions, but the nodes are also annotated with some attributes (such as maximum amount of bitcoin sent in a single transaction). Once the graph is created, a community detection algorithm based on attribute propagation to detect users with a similar activity.

III. Interaction in Blockchain Online Social Media

A BOSM is an Online Social Media platform implemented with the support of the blockchain technology. Steemit is the first successful BOSM and is implemented on Steem, a blockchain specifically designed to support the development of socioeconomic applications. Indeed, all activity on Steemit is stored in the form of (unencrypted) transactions on the Steem blockchain [2]. Transactions are then validated, packed into blocks, and put in the blockchain by the witnesses. A witness is a Steem user which runs a witness node, and has many tasks for the management of the blockchain. Witnesses are voted by the Steem users, but only the top 20 witnesses by votes are entitled to create blocks for the blockchain. Users on Steemit, similarly to other social platforms can create, comment, evaluate, and share contents. Additionally, the platform incorporates an economic side which is tightly connected to the social side [10]. The blockchain Steem provides support in terms of storage, indeed all user activity, including creating a post, voting or transferring cryptocurrency are all actions that are stored in the blockchain as a transaction.

Steem provides 38 transaction types to support the storage of social and economic interactions among the users and the development of social applications. We classify the transaction types according to the same categorisation proposed in [2], which can be summarized as following:

- **Social.** Social transactions are the transactions used to model social actions, such as creating posts or voting pieces of content.
- **Monetary.** Monetary transactions are the transactions used to model economic actions, such as cryptocurrency transfers or escrow transactions.
- **Management.** Management transactions includes transactions for account management, and transactions that can be used only by witnesses for blockchain management.

The decentralization introduced by the blockchain technology had an impact on multiple aspects of the platform. The most important aspect introduced by BOSMs platform is the possibility to grant economic rewards to the users, based on their social activity. Indeed, as explained in [10] each post or comment on the platform can be evaluated (curated) by the users and depending on the evaluation received, a sum of cryptocurrency is assigned as reward to the piece of content. The reward is then granted to both the creator of the piece of content and its curators. At least half of the reward is granted to the creator and the remaining is granted to the curators but the first curators get more.

This rewarding strategy highly encourages phenomena like automatic content curation or evaluations buying and selling (buy the votes of the most influential users to increase the reward assigned to a post) powered by bots [24]. These types of phenomena fuel a new kind of interaction, one that is not strictly social, but has potentially other goals. Indeed, while in the most famous OSNs interactions are made to establish social ties or to increase one's social capital, in BOSMs and Steemit in particular interactions can have an economic implication [25]. Clearly, the introduction of the economic aspect in the scenario of OSNs may have a set of repercussions and can force the social and non-social interactions among users that would have ignored each other otherwise.

With the aim to show that the social interactions are highly influenced on BOSMs, we extract the interactions graph, taking Steem as case study. Interactions graph are made such that the nodes represent the users on Steem, and the edges represent the interactions among them. Interactions in the Steem blockchain are stored in the form of transactions, therefore we need to parse the relevant transactions to extract how people interact with each other. We extracted 5 different graphs, explained in detail in Section IV. To prove our hypothesis, we applied two state-of-art algorithms for community detection. The choice fell on Label Propagation [13], [26] and Louvain Method [12], [27] because they are largely known and applied. The choice was also partially driven by the fact that the graphs we are dealing with contain millions of nodes and hundreds of millions of edges, making unfeasible to apply algorithms with high time and space complexity. The application of the community detection algorithms will return a community structure, one for each algorithm used. Considering the nature of the graph, the community structures can be seen as sets of users which represent meaningful interaction groups.

IV. Dataset

The dataset considered for this analysis contains the blocks included in the Steem blockchain starting from the block at height 10.000.000, produced on the 8th of March 2017 at 17:34:21, up to block at height 29.059.999, produced on the 1st of January 2019 at 00:40:27. Data is organised in JSON objects for ease of use. From the blocks we extracted the interactions

Graph	#nodes	#edges
Interactions Graph	1.244.889	191.205.168
Monetary Interactions Graph	1.074.722	4.204.473
Social Interactions Graph	1.230.568	187.953.267
Social no bot Interactions Graph	1.230.554	182.884.418
Follower-Following Graph	1.223.473	98.092.374

TABLE I
NUMBER OF NODES AND EDGES OF THE 5 GRAPHS CONSIDERED.

	LabProp	Louvain
#communities	2	44
min community size	2	1
max community size	$1.24489\ e^{06}$	775,319
avg community size	622,444	28,292.9
modularity	$3.491\ e^{-08}$	$3.749\ e^{-01}$
density (graph)	$2.217\ e^{-04}$	
intra-cluster density	$2.217\ e^{-04}$	$3.089\ e^{-04}$
conductance	$2.140\ e^{07}$	$8.363\ e^{-01}$

TABLE II
COMMUNITY STRUCTURE ON THE INTERACTIONS GRAPH

among the users with which we were able to build 5 graphs. The size of the graphs in terms of nodes and edges is shown in Table I.

In the graphs node represent users and edges represent interactions among them in the form of transactions present in the blockchain. In detail, the graphs are built according to the following rationale:

- The **Interactions Graph** is built using all the transactions available in the dataset.
- The **Monetary Interactions Graph** is built using only the transactions that belong to the Monetary category.
- The **Social Interactions Graph** is built using only the transactions that belong to the Social category.
- The **Social no bot Interactions Graph** is a subgraph of the Social Interactions Graph where the most influential bots detected through in-degree centrality are removed.
- The **Follower-Following Graph** is built using only transactions that express the intention to "follow" another user (similarly to Twitter or Instagram).

As shown in Table I, the graphs have a similar number of nodes, meaning that almost all are involved in at least a social and a monetary transaction. It is however interesting to notice how the Follower-Following graph contains approximately half the edges of the Interactions Graph, sign that a lot of interactions happen also outside this kind of expressed relationships. Additionally, the Monetary Interaction Graph contains only 4 million edges, which hints that the economic side of the platform is highly underdeveloped and underused with respect to the social counterpart. This is a somewhat unexpected effect, considering that the economic side of Steemit is what makes it unique with respect to other well-known OSN platforms.

V. EXPERIMENTAL RESULTS

In this Section, we summarise the most relevant findings detected by the application of the Label Propagation and Louvain Method for community detection on the five graphs introduced in Section IV. For the analyses we made use of Networkit, a powerful tool for managing large networks [28]. In particular, we applied the parallel implementations of Label Propagation [26] and Louvain Method [27].

A. Interaction Graph

Table II describes the community structure detected by Label Propagation and Louvain Method in the Interactions Graph. The community structure detected by Label Propagation is made of solely 2 communities, of which one of them contains just two nodes, and the other contains all the remaining nodes. The detected community structure hardly has any significance because almost all the nodes belong to the same community, and this lack of community structure is also identified by the values of modularity, intra-cluster density and conductance. On the other hand, the community structure detected by the Louvain Method is able to partition the graph in a much more meaningful structure. Indeed, there are 44 identified communities, where the smallest one contains just one node and the largest contains more than half of the nodes. The intra-cluster density, although still low, is higher with respect to the graph density. The modularity score of 0.37 is higher with respect to the one obtained by Label Propagation, meaning that the community structure has a modular structure. The conductance score of 0.83 shows that the communities are very well separated from each other.

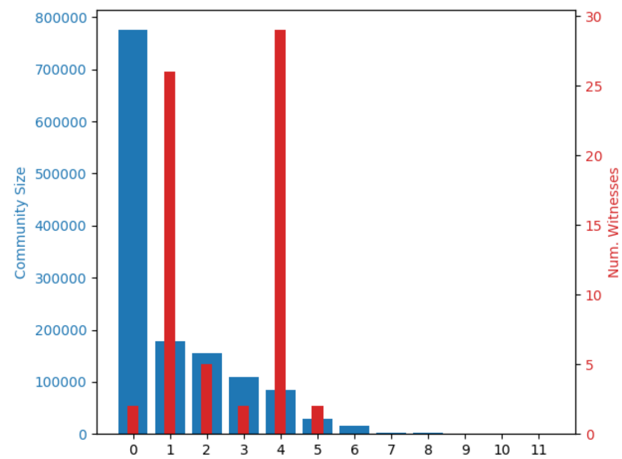

Fig. 1. Witness recap in the 12 largest communities (Louvain) of the Interactions Graph

Considering the high modularity and the very low conductance scores obtained by the community structure, we decide to delve deeper in the community structure detected by Louvain Method. In particular, we decided to inspect the identity of the nodes inside the communities and check how many witness nodes are present in each community. The number of witnesses in each community is shown in Figure 1. We only show the top 12 communities sorted by the

	LabProp	Louvain
#communities	51	189
min community size	2	1
max community size	1.03981 e^{06}	756,312
avg community size	21,073	5,686.36
modularity	4.196 e^{-02}	4.648 e^{-01}
density	6.595 e^{-06}	
intra-cluster density	6.985 e^{-06}	8.964 e^{-06}
conductance	1.469	7.152 e^{-01}

TABLE III
COMMUNITY STRUCTURE ON THE MONETARY INTERACTIONS GRAPH

number of nodes in the community. The Figure shows that two communities, namely the second largest community and the fifth largest community, have a much higher concentration of witness users in them, if compared to the other communities. We find that just these two communities host 55 witnesses, which are more than half, which may hint the presence of two "factions" of witnesses which ultimately brought to the Hive hard fork in early 2020. The Figure additionally shows that the witnesses are unevenly distributed among the communities, indeed in the largest community, only a few of them are present.

B. Monetary Interactions Graph

Table III describes the community structure detected by Label Propagation and Louvain Method in the Monetary Interactions Graph. With respect to the Interactions graph, the two community detection algorithms are able to detect a much more articulated structure in this graph, despite the fact that the graph contains less nodes and much less edges. Label Propagation detected a set of 51 communities, with sizes ranging from 2 nodes for the smallest community and over 1 million nodes for the largest community. The conductance score of the community structure detected by Label Propagation is relatively low, which underlines the fact that communities are well separated. This result is partly confirmed by the modularity score, which is not as close to 0 as in the case of the Interactions Graph. Overall, the modularity is quite low, hinting to the fact that communities are well separated but poorly knitted. This fact is confirmed by the value of inter-cluster density which is only marginally higher with respect to the whole graph. The community structure detected by the Louvain method is made up of 189 communities with sizes ranging from 1 and 750,000 nodes. Also in this case the chosen measures for community quality evaluations are better for the community structure detected by the Louvain Method. While the intra-cluster density is only marginally higher, the conductance is less than half, which suggests that the community are better separated each other and more dense. The modularity score, which is higher by an order of magnitude, confirms an overall better community structure, made of several independent modules.

Considering their important role, and their dedication to the economic side of the platform, we inspect the presence of witnesses in the largest communities of the Monetary Interactions Graph. We only focus on the community structure detected by the Louvain Method because the community structure returned by this algorithm was proven to be better according to modularity, intra-cluster density and conductance. The number of witness accounts per community in the 12 largest communities is shown in Figure 2. Contrarily to what we observed in the Interactions Graph, in the Monetary Interactions Graph, over half of the witnesses gathered in a single community: the second in terms of number of nodes. One of the possible motivations why we do not observe the same phenomenon in this case is clearly related to the fact that the economic side of the platform is still quite underdeveloped and only few accounts have a relevant economic activity on the platform. Witness nodes are more active in this aspect due to the fact that they have access to block producing rewards, as part of their dedication to the network.

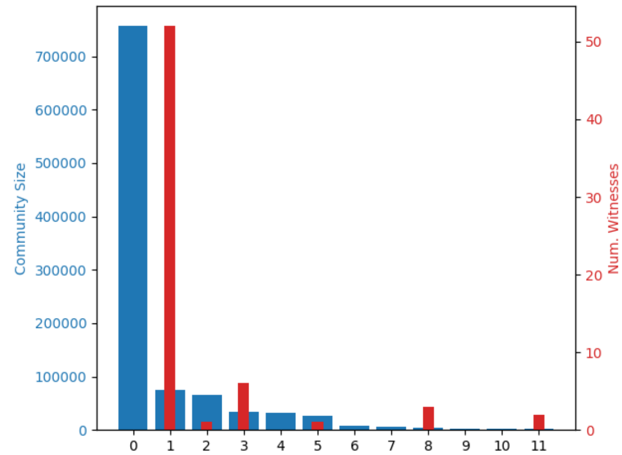

Fig. 2. Witness recap in the 12 largest communities (Louvain) of the Monetary Interactions Graph

	LabProp	Louvain
#communities	3	31
min community size	2	2
max community size	1.230 e^{06}	560,373
avg community size	410,189	39,695.7
modularity	9.453 e^{-06}	3.614 e^{-01}
density	2.235 e^{-04}	
intra-cluster density	2.235 e^{-04}	4.809 e^{-04}
conductance	5.357 e^{06}	7.152 e^{-01}

TABLE IV
COMMUNITY STRUCTURE ON THE SOCIAL INTERACTIONS GRAPH

C. The social graphs

Table IV describes the community structure detected by Label Propagation and Louvain Method in the Social Interactions Graph. The community structure identified by the two algorithms is more similar to the one detected in the Interactions Graph with respect to the one detected in the Monetary Interactions Graph. Label Propagation detected only three communities, one of which contains the vast majority of the nodes of the graph. The poorness of this community

	LabProp	Louvain
#communities	17,697	17,715
min community size	2	1
max community size	1.212 e^{06}	492,924
avg community size	69.5	69.4
modularity	1.130 e^{-06}	3.667 e^{-01}
density	2.173 e^{-04}	
intra-cluster density	2.237 e^{-04}	4.448 e^{-04}
conductance	9.916 e^{01}	1.179 e^{-03}

TABLE V
COMMUNITY STRUCTURE ON THE SOCIAL NO BOT INTERACTIONS GRAPH

	LabProp	Louvain
#communities	69,970	69,957
min community size	2	1
max community size	1.153 e^{06}	707,999
avg community size	17.485	17.488
modularity	1.659 e^{-05}	3.739 e^{-01}
density	1.310 e^{-04}	
intra-cluster density	1.474 e^{-04}	2.033 e^{-04}
conductance	1.274	4.145 e^{-04}

TABLE VI
COMMUNITY STRUCTURE ON THE FOLLOWER-FOLLOWING GRAPH

structure is also confirmed by the scores achieved in the quality measures considered. Indeed, modularity is very low, intra-cluster density is equal to the whole graph density, and conductance is extremely high. The Louvain Method detects a community structure made of 31 communities, and the communities size range from 2 to 560.000 nodes. Concerning the quality of the community structure detected, we acknowledge that the modularity score, although slightly lower than the community structure detected in the Interaction Graph, is still sign of a good community structure. Additionally, we see that the intra-cluster density is more than double of the density of the who graph, while the conductance is lower than 1.

Table V describes the community structure detected by Label Propagation and Louvain Method in the Social no bot Interactions Graph. In this graph, both algorithms are able to find a very complex community structure with more than 17,000 communities each, however the structures detected are very different. Indeed, once again Label Propagation groups the largest part of the nodes in the same community and, even though it is not a partition algorithm, the remaining nodes form very small communities. The scores of the community structure evaluation measures are better with respect to the Social Interactions Graph, but in particular the low modularity and the fact that the intra-cluster density is very similar to the graph density, denote a poor community quality overall. While at a first glance the community structure obtain via the Louvain Method may seem the same, it is indeed very different. Indeed, the communities are made of up to 492,924 nodes, which is the lowest maximum size for a community in our analyses. The community structure evaluation measures tell us that that the structure is slightly worse than the one detected by the same algorithm in the Social Interactions Graph, save for the conductance which is instead two orders of magnitude less. The big differences concerning the community structures that we see in Tables IV and V show us the importance of bots an in particular how much they are invasive in the platform. Indeed, by just removing a small set of the most influential bots of the platform, the number of communities exploded, the structure became more fragmented. To confirm this fact, we observe that the conductance and intra-cluster values for both community structures dropped sensibly, meaning that communities are much better separated thanks to the removal of few nodes.

Lastly Table VI describes the community structure detected by Label Propagation and Louvain Method in the Follower-Following Graph. The community structure in this case is even more complex as in both cases we see that the number of communities is more than 69.000, the highest in all graphs. Despite the very large number of communities, for both algorithms the largest community contains the majority of the nodes of the graph. In the case of Label Propagation, the modularity score indicates that the community structure is poor, and the result is confirmed by a low intra-cluster density. However, the conductance of this community structure is the lowest among the community structures detected by Label Propagation in all the graphs. The structure detected by Louvain Method in this graph is modular as confirmed by the modularity score. The intra-cluster density score is only marginally higher with respect to the graph density, but the low conductance suggests that the community are very well separated from each other.

In these three social graphs, we can observe the overall impact of the rewarding system on the social interactions among the users. The very small number of communities, especially in the Social Interactions Graph, and the fact that in most cases the largest community contains all the nodes of the graphs, is a clear hint that people are highly encouraged to interact socially with other users because of the potential rewards. This effect is most emphasised in the Social Interaction Graph, where the community structure has the lowest scores and bots are included in the graph, possibly hinting the impact of the bots on the platform.

VI. CONCLUSION

In this paper we studied the community structure obtained from five graphs obtained by the interactions of users on the Steem blockchain. Steem is the main blockchain used for social applications, with more than 1.5 million users at the time of writing this paper. Our paper considered 5 different aspects of the platform, which let to the creation of just as many graphs. Thanks to the toolkit provided by Networkit, we were able to extract two community structures for each graph and evaluate them using three measures to assess their quality.

Results highlight that the community structure detected by Label Propagation often tends to return one community which contains almost all the nodes of the graph, and many smaller communities. On the other hand, as demonstrated by the community structure evaluation measures adopted in our methodology, Louvain Method is able to find a more meaningful community structure. We also detected than witnesses, so are called the block producers, are mostly in the same communities, clearly indicating, especially from an economic point of view, that there is a "caste" of rich users on Steem. From a social point of view, we detect that people tend to create large aggregated communities, mainly because of the activity of the bots.

As future works we plan to deepen our understanding of the economic aspect of the platform, in particular understanding if the phenomenon of the Pareto principle (rich get richer) holds on Steem as well. Furthermore, we plan to enrich the dataset, by collecting the fist months of Steem, and the latest ones, in order to understand how the community structure is evolved over time. Additionally, we will shed light on other social phenomena that have been observed in other social networks, such as the presence of the so-called Dunbar Circles, and investigate the communication among users through temporal motifs.

ACKNOWLEDGMENT

This work was partially funded by the European Commission under contract number H2020-825585 HELIOS.

REFERENCES

[1] A. Datta, S. Buchegger, L.-H. Vu, T. Strufe, and K. Rzadca, "Decentralized online social networks," in *Handbook of social network technologies and applications*. Springer, 2010, pp. 349–378.

[2] B. Guidi, A. Michienzi, and L. Ricci, "Steem blockchain: Mining the inner structure of the graph," *IEEE Access*, vol. 8, pp. 210 251–210 266, 2020.

[3] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach," *The guardian*, vol. 17, p. 22, 2018.

[4] S. Buchegger, D. Schiöberg, L.-H. Vu, and A. Datta, "Peerson: P2p social networking: early experiences and insights," in *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*. ACM, 2009, pp. 46–52.

[5] L. A. Cutillo, R. Molva, and T. Strufe, "Safebook: A privacy-preserving online social network leveraging on real-life trust," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 94–101, 2009.

[6] B. Guidi, T. Amft, A. De Salve, K. Graffi, and L. Ricci, "Didusonet: A p2p architecture for distributed dunbar-based social networks," *Peer-to-Peer Networking and Applications*, pp. 1–18, 2015.

[7] S. Nakamoto *et al.* (2008) Bitcoin: A peer-to-peer electronic cash system.

[8] B. Guidi, "When blockchain meets online social networks," *Pervasive and Mobile Computing*, vol. 62, p. 101131, 2020.

[9] K. Kapanova, B. Guidi, A. Michienzi, and K. Koidl, "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*, 2020, pp. 163–168.

[10] 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.

[11] L. Jiang and X. Zhang, "Bcosn: A blockchain-based decentralized online social network," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1454–1466, 2019.

[12] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[13] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical review E*, vol. 76, no. 3, p. 036106, 2007.

[14] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[15] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical review E*, vol. 74, no. 3, p. 036104, 2006.

[16] J. M. Kumpula, J. Saramäki, K. Kaski, and J. Kertész, "Limited resolution in complex network community detection with potts model approach," *The European Physical Journal B*, vol. 56, no. 1, pp. 41–45, 2007.

[17] W. Deitrick and W. Hu, "Mutually enhancing community detection and sentiment analysis on twitter networks," *Journal of Data Analysis and Information Processing*, vol. 1, no. 3, pp. 19–29, 2013.

[18] D. Surian, D. Q. Nguyen, G. Kennedy, M. Johnson, E. Coiera, and A. G. Dunn, "Characterizing twitter discussions about hpv vaccines using topic modeling and community detection," *Journal of medical Internet research*, vol. 18, no. 8, p. e6045, 2016.

[19] B. R. Amor, S. I. Vuik, R. Callahan, A. Darzi, S. N. Yaliraki, and M. Barahona, "Community detection and role identification in directed networks: understanding the twitter network of the care. data debate," in *Dynamic networks and cyber-security*. World Scientific, 2016, pp. 111–136.

[20] O. Fraisier, G. Cabanac, Y. Pitarch, R. Besançon, and M. Boughanem, "Stance classification through proximity-based community detection," in *Proceedings of the 29th on Hypertext and Social Media*, 2018, pp. 220–228.

[21] N. Natarajan, P. Sen, and V. Chaoji, "Community detection in content-sharing social networks," in *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining*, 2013, pp. 82–89.

[22] C. Remy, B. Rym, and L. Matthieu, "Tracking bitcoin users activity using community detection on a network of weak signals," in *International conference on complex networks and their applications*. Springer, 2017, pp. 166–177.

[23] X. Xueshuo, W. Jiming, Y. Junyi, F. Yaozheng, L. Ye, L. Tao, and W. Guiling, "Awap: Adaptive weighted attribute propagation enhanced community detection model for bitcoin de-anonymization," *Applied Soft Computing*, vol. 109, p. 107507, 2021.

[24] B. Guidi and A. Michienzi, "Users and bots behaviour analysis in blockchain social media," in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2020, pp. 1–8.

[25] N. Lin, "Building a network theory of social capital," *Social capital*, pp. 3–28, 2017.

[26] M. Ovelgönne and A. Geyer-Schulz, "An ensemble learning strategy for graph clustering." *Graph Partitioning and Graph Clustering*, vol. 588, p. 187, 2012.

[27] C. L. Staudt and H. Meyerhenke, "Engineering parallel algorithms for community detection in massive networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 1, pp. 171–184, 2015.

[28] C. L. Staudt, A. Sazonovs, and H. Meyerhenke, "Networkit: A tool suite for large-scale complex network analysis," *Network Science*, vol. 4, no. 4, pp. 508–530, 2016.