The Footprints of a "Mastodon": How a Decentralized Architecture Influences Online Social Relationships

Matteo Zignani*, Christian Quadri*, Sabrina Gaito*, Hocine Cherifi† and Gian Paolo Rossi*
*Department of Computer Science
University of Milan, Milan, Italy
Email: firstname.lastname@unimi.it
†LE2I UMR CNRS University of Burgundy, Dijon, France
Email: hocine.cherifi@iut-dijon.u-bourgogne.fr

***Abstract*—Decentralized online social networks (DOSNs) have recently emerged as a viable solution to preserve the users' privacy and ensure higher users' control over the contents they publish. However, little is known about the backlashes that the decentralized organization and management of these platforms may have on the overlaid social network.**

This paper fills the gap. Specifically, we investigate how a decentralized architecture based on distributed servers impacts the structure of the users' neighborhood and their ego-networks. Our analysis relies on social data gathered from the decentralized micro-blogging platform Mastodon, the newest and fastest-growing decentralized alternative to Twitter. Our findings highlight that the social network supported by each server, namely instance, has a specific footprint in terms of degree distribution and clustered structure of the ego-networks of its members. Further, how users connect to people hosted in other instances is heavily bound by the server they are in. Moreover, users who tend to establish relationships in outer instances prefer to use a bunch of servers. Finally, we show that the ego-networks of the users are more clustered within the instance boundary, i.e. triangles are more likely to form among members of the same instance. All these findings suggest that the decentralization drives the social network to a structure that can be potentially very different from the usual one typical of centralized online social networks. Thus, the architecture of a DOSN is a factor developers and researchers should take into account when designing this kind of social platforms.

***Index Terms*—decentralized online social networks, node's neighborhood, clustering coefficient, online social networks**

I. INTRODUCTION

In the last few years, techno-activists, open-source software developers and researchers have proposed various forms of online social networks with the goal of both preserving user privacy and putting users' communications and contents back to the focus of these platforms. Decentralized online social networks (DOSNs) [1] represent one of the most promising and widely accepted solutions, as they weaken or even remove the dependency on a centralized provider, thus giving more control back to the users over their own data.

While user privacy and privacy management [2], data ownership [3] and data portability [4] are the main topics where research on DOSNs is focusing, much remains to uncover about the structure of the social relationships they maintain. This latter aspect is as important as the former ones to allow DOSNs to follow the growth and wide adoption we observed for their centralized counterpart. In fact, there is a strict relationship between the social network, and some of its properties, and the success of many centralized online social networks [5], [6]. From this perspective, one of the key points to understand and, somehow, control the evolution and growth of DOSNs lays on the understanding of the interplay between the system design and the network of social relationships the system supports. An interplay quite evident since users have to choose in advance which server will host their data and they have a limited access to the social network - in terms of navigability and friend searchability - imposed by the connections among the servers.

The above interplay between the system and the overlying social network is exactly the subject of this paper. To this purpose, we rely on Mastodon, a new and fast emerging decentralized microblogging platform, designed as a *federated* architecture, in which an overlay of networked servers acts as proxy to support the social features of the system. The Mastodon platform exhibits some features well suited to the purpose of our work: (i) each server, namely a *instance* in Mastodon jargon, supports a community orientated towards specific topics and interests; (ii) data about the hardware capacity, the position and the publication policy of the instances are publicly available through APIs; and (iii) users can use all the functions which ease social interactions, such a "follow" button, hashtags, post publication, mentions and replies. The resulting set of data, i.e. the Mastodon social network and the meta-data about the instances, allows us to investigate how a federated architecture impacts the neighborhood and the ego-network of the users, and whether each instance has a specific social footprint. The main findings on the above aspects may be summarized as follows:

- The impact of the instances on the creation of links are weaker than those imposed by the nationality of the user, i.e. users are more bound by their geographical, cultural and linguistic background than by the architecture of the platform. It is, however, suggestive that people alternate phases in which they form links within the same instance with people whom they share an interest with, and phases in which they explore other communities to look for social relationships.

- The exploring phase, i.e. searching for users outside the hosting server, is targeted on a few instances. In fact, users who are likely to search and establish social relationships outside their home-instance do not equally spread their social links among the other instances, rather they prefer a bunch of instances, which likely fit their interests. From the user's viewpoint not all the servers in the network overlay are equally important.
- The architecture based on independent instances has a stronger impact on how users' ego-networks cluster. Indeed, the instance-based design influences the triadic closure process, sometimes limiting the formation of closed triads within the instance boundary, other times promoting the clustering of nodes' neighborhood outside the instance they belong to.
- Finally, each instance has a peculiar footprint which reflects how its members establish "follower"/"followee", mutual relationships and close triads. We suppose that the underlying mechanism which drives the formation of the nodes' neighborhood might vary from instance to instance.

The paper is organized as follows. In Section II we briefly describe the main feature of the decentralized microblogging platform Mastodon. In Section III we describe which kind of data we collect from Mastodon – the instance meta-data and the structure of the social network – and how we gathered these pieces of information. In Sections IV and V we discuss the main findings about the impact instances exert on their members when they have to establish new relationships and the footprint of the instances. Here we focus our analysis on two microscopical aspects of the social network: the node's neighborhood and its ego-network structure.

II. An overview of Mastodon

Mastodon is a DOSN with microblogging features, where each server runs open source software. The basic aim of the project, which dates to 2016, is to restore control of the content distribution channels to the people by avoiding the insertion of sponsored users or posts in the feeds.

From an architectural viewpoint, the platform follows a federated architecture organized into two layers implementing the ActivityPub protocol,[1] as shown in Fig.1. The ActivityPub protocol allows it both to manage the communications (black links) among the servers – *instances* – comprising the federation and to offer a client-to-server interface which enables interactions (blue links) among the users having their accounts on the instances. In the server-to-server layer the instances form a network, and each of them administrates its

[1] https://www.w3.org/TR/activitypub/

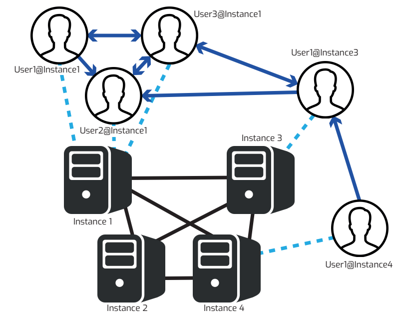

Fig. 1. The decentralized architecture of Mastodon. We distinguish between two layers: the server-to-server layer, made up of interconnected (black links) instances (*InstanceN*); and the social network layer, formed by the 'follow' relationships (blue links) between users (*UserK@InstanceN*) hosted (dotted cyan links) by the different instances.

own rules, account privileges, and whether or not to share messages coming to and from other instances. Each server hosts individual user accounts, the content they produce, and the content they subscribe to.

From a user experience viewpoint, Mastodon releases the major features of a microblogging platform:

- users can follow one another, whether or not they are hosted on the same instance;
- users can post short messages consisting of up to 500 text characters, called 'toots', for others to read.
- toots are aggregated in local and federated timelines. The former show messages from users hosted on a specific instance, while the latter aggregate the messages across all participating Mastodon instances.
- users have control over the visibility of their posts. They can choose whether the post can be displayed on the federated timeline, on the timelines of the user's followers only, or on the timeline of the users mentioned in the toot.

Mastodon differs from other commercial microblogging platforms w.r.t. two key points. First, it is oriented towards small communities and community-based services. In fact, each instance may support and favor specific topics. So prior to registration, a user is encouraged to choose the instance better suited to her own tastes. Second, the Mastodon platform does not provide any algorithm for recommending new friends or promoted contents. So, the only way to establish a connection or consume a content is by searching an already known account through the search functions or by exploring the feeds of the instances in search of users with similar interests or interesting posts. These two characteristics are fundamental for the study of the interplay between the physical layer of a decentralized social network and its overlaid social network, since: (i) the coupling server/topic affects how people are distributed on the servers set and how they interact among themselves, this is true because it is well-known that common interests shape the structure of social networks [7], [8]; (ii) the lack of recommendation systems removes different external factors – often hidden by recommendation algorithms – from the mechanisms driving the formation of new links [9], [10]. The individual goes back to being the focus of the link creation process and the decentralized organization of the instance might be one of the few factors affecting the choice of who to "follow".

III. Dataset

The dataset has been completely gathered by combing queries to the Mastodon API and a custom web crawler. Here, we describe its two main components: the instance meta-data and the Mastodon social network [11].

A. Instance meta-data

One of the signature elements in the Mastodon platform is the idea of instance, the cornerstone of the server-to-server layer in Fig.1. Within a decentralized system, the collection of information about the instances may raise some issues due to the search for the servers. Indeed, each server is independent and the ActivityPub protocol does not provide specifications for polling the other servers. To overcome this issue, Mastodon developers have introduced an API to query different kinds of information about the instances. [2] APIs provide a full description of the instance, the list of topics the instance supports, the number of registered users, the number of posts and the IP address of the instance. By continuously monitoring the instance meta-data, we are able to estimate the size of the server-to-server layer and to track its growth over a six-month period. The number of instances on January 21, 2018 (the date we got the first snapshot of the Mastodon social network) is 1733, an increase of about 450 servers w.r.t. the start of the monitoring.

We also enriched the instance meta-data by adding the geographical position of the instances at a country-granularity. To this end, we exploited the global IP database provided by *ipstack.com* for assigning to each server the country it is in. The localization of the instances allows us to *i)* understand where the instance and the users are distributed worldwide; and to *ii)* quantify the strength of the interplay between the geographical position of the servers and the overlaid social network. One third of the instances is located in Japan, while the remaining are distributed in North America, in Europe (most notably France) and in other countries such as China, Australia, Brazil and India. Similarly, distribution of the users follows the instance one; indeed we do not find countries with few instances and many users.

B. The Mastodon social network

To gather the social relationships among Mastodon users, we developed a crawler targeted to the web pages of the platform. From each profile page we extract both the followers and the followees, i.e. the in-going and out-going relationships of a user. The opportunity to follow both directions represents an advantage in building the network, since the crawl of a directed network using out-going links only may not result in the entire weakly connected component [12]. We also highlight that the information in following/follower web pages are also available to visitors who are not logged on. To build a seed set as large as possible and reach as many users as possible, we exploit the list of the instances and their global timelines, since they report all the statuses with public visibility (see Section II) in chronological order. Specifically, from each public timeline we extract the users who posted at least one status – Mastodon APIs provide a specific end-point to get posts in a public timeline – and put them into the seed set. The resulting seed set contains more than 62K users. Finally, in our crawler we implement a breadth-first search (BFS) strategy which traverses both out-going and in-going links, where the latter are traversed in the opposite direction. In the crawler we also add a filter which discards links towards profiles hosted in other social platforms supporting ActivityPub or OStatus protocols, which allow users to interact with users on other decentralized platforms, i.e. the "fediverse". After the end of the crawling process we obtained a network made up of 479,425 nodes and 5,649,762 directed links, covering 46% of users in Mastodon (the total amount of users can be obtained by the instances metadata). Specifically, the six biggest instances are covered on average to an extent of 52%.

In the following analysis we model the followee and follower relationships by defining two networks: (i) the *directed network*, which reflects the asymmetric nature of the social relationships in Mastodon; and (ii) the *mutual network*, formed by the reciprocated links in the directed network.The latter allows us to compare our findings on the Mastodon networks with previous studies on centralized online social networks [13].

IV. The impact of the instances

In this section we present our findings about the impact of instance-based decentralized architecture on how people establish new relationships with the other members of Mastodon. Specifically, we investigate this aspect from a microscopic viewpoint by analyzing the characteristics of the neighborhood and the ego-network of Mastodon users.

1) Are users bound by instances?: In the Mastodon context, it should be easier to find friends within the same instance than in outer instances, since users within an instance are supposed to share a common interest and each instance has a dedicated timeline showing their users' activities only.[3] So, here we are wondering to what extent users are bound by instances when they establish relationships. To evaluate how likely users establish connections outside their own instances, we compute the *border-index BI*. Namely, the border-index of a user u is defined as the fraction of u's neighbours who are hosted in different instances:

$$BI(u) = \frac{|\{v \in \Gamma(u)^{(\cdot)} | i(u) \neq i(v)\}|}{|\Gamma(u)^{(\cdot)}|} \quad (1)$$

where $\Gamma(u)^{(\cdot)}$ indicates the set of u's successors ($\Gamma(u)^{(+)}$)/ predecessors ($\Gamma(u)^{(-)}$) in the directed network, or the set of u's neighbors ($\Gamma(u)^{(mut)}$) in the mutual network; while

[2] https://instances.social/api/doc/

[3] Finding interesting users or posts in the federated timeline is much harder since users have to scroll many more elements coming from most of the Mastodon instances.

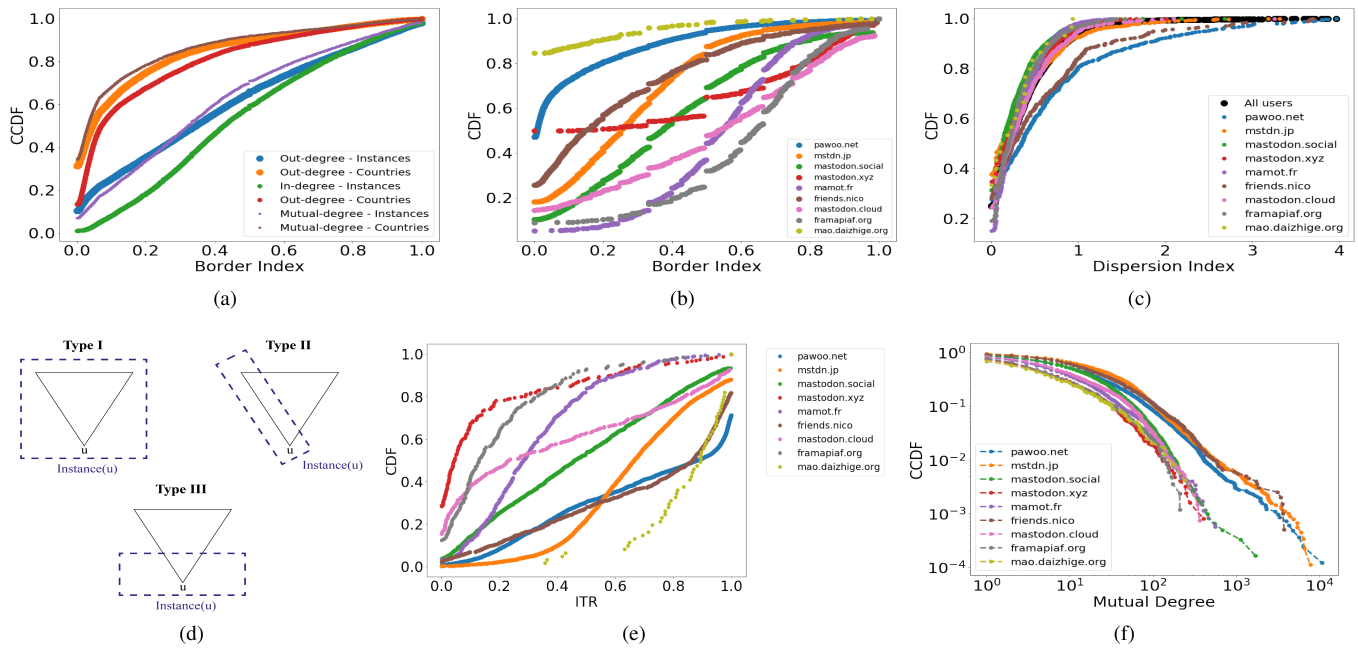

Fig. 2. In (a) the border-index distribution (CDF) taking into account different types of degree (out-degree, in-degree and mutual degree) and different kinds of node attribute (instance or country). In (b) the distribution of the border-index based on the mutual degree and the instance attribute. The figure reports the distribution for the ten biggest instances. In (c) the distribution (CDF) of the dispersion index measured on the whole network (black dots) and on the ten biggest instances (coloured dots). In (d) the three types of triangle according to the position of the nodes in the instances. In (e) the distribution (CDF) of the intra-triangle index measured on users having a mutual degree greater than 10, for the ten biggest instances. In (f) the distribution (CCDF) of the mutual degree in the 10 biggest instances of Mastodon.

$i(u)$ denotes the instance the user u is on. A border-index close to 1 indicates that a user in an instance is likely to establish relationships with users outside the instance. In this case the user is scarcely limited by the instance. In Fig. 2a we report the cumulative distribution function (CDF) of the border-index measured on both the directed (blue and green lines) and mutual (purple line) networks. We observe that in all the cases about 35-40% of the users establish or receive connections to/from people in other instances, i.e. they are more likely to search for friendships outside their instance or to receive friendship requests from users who lie outside their home-instance. In the same figure, we also display the CDF of the border-index where we substitute the instance of u (see equation 1), with its country. This way, we can compare the impact of the geographical, cultural and linguistic limits given by the country, with the influence of an architectural aspect, given by the instances. As we can see in Fig. 2a, users are more bound by the country than by the instance they belong to. In fact, only 10% of the Mastodon users have more than a half of their neighbors lying in instances located in other countries. Thus, from the analysis of the border-index we can assert that *the constraints imposed by the instances on the creation of links are weaker than those imposed by the nationality of the user, i.e. users are more bound by their geographical, cultural and linguistic backgrounds than by the architecture of the platform.* This fact is also supported by the analysis of the assortativity for the attributes instance and country. In fact, the former is 0.56 and the latter is 0.75 in the mutual network, meaning that country-homophily is stronger than instance-homophily.

The above finding holds for a generic user in Mastodon, since the distribution of the border-index has been computed on the whole Mastodon network. However, it may be that users in some instances tend to establish intra-instance relationships more than others. To cope with this aspect, we analyze the distribution of the border-index in each single instance, as shown in Fig. 2b. Here we reported the CDF of the border-index for the ten largest instances in Mastodon. We observe that the above general behavior is a mixture of traits of the instances very different from one another. On the one hand, there are instances, e.g. "mastodon.cloud", "mamot.fr" and "framapiaf.org", whose users are likely to create connections outside their home-instance; on the other hand, there are instances whose users are bound by the instance they belong to ("pawoo.net" or "friends.nico"). We should note that this phenomenon scarcely depends on the size of the instance, in fact "mao.daizhige.org" and "pawoo.net" show similar behaviors, but the former is ten times smaller than the latter. In general, the *tendency of users to establish relationships with members of outer instances depends on the instance.*

2) Do users connect to specific outer instances?: A further aspect which may shape the neighborhood is how each node's links are distributed among the outer instances. In fact, people may equally distribute their relationships among the instances or address their efforts on specific instances; where , the latter behavior is more likely, since people focus on instances which are inclined to their interests. We investigate how users distribute their relationships among the instances by exploiting a statistical measure of dispersion. Specifically, for a generic user u, we compute the Kullback–Leibler (KL) divergence between the distribution of the instances of the u's neighbors – discarding u's neighbors lying on the u's instance – and a uniform distribution on the same instance set. Since KL divergence represents the distance between these distributions, the closer the KL divergence to 0 the more similar the distributions. In Fig. 2c we report the distribution (CDF) of the KL divergence measured on users having a border-index greater than 0.5 and a degree greater than 10, grouped by instance, along with the KL divergence distribution measured on the aggregation of the instances (black line). We reasonably define as "similar to the uniform" those distributions which get a KL divergence less than 0.3, meaning that users spread almost uniformly their social relationships among the instances. From the figure a clear trait emerges as in all the instances nearly 75% of the members are not "similar to the uniform"; i.e., to a lesser or greater extent they prefer a few instances when they have to create a social relationship. This behavior is even more evident in "pawoo.net" and "friends.nico", where more than 20% of their members prefer one or two outer instances. In general, *in all the instances, users who are likely to seek out and establish social relationships outside their home-instance do not equally spread their connections among the other instances; rather, they prefer a a few instances, which likely fit their interests.*

TABLE I
α AND x_{min} PARAMETERS ESTIMATED BY THE FITTING. THE FIRST TWO COLUMNS REFER TO THE OUT-DEGREE, THE THIRD AND FOURTH REFER TO THE IN-DEGREE, THE LAST TWO COLUMNS TO THE MUTUAL DEGREE.

Instance	α^{+}	x^{+}_{min}	α^{-}	x^{-}_{min}	α^{m}	x^{m}_{min}
pawoo.net	2.4	205	1.6	25	2.2	65
mstdn.jp	2.2	71	2.4	25	2.4	63
mastodon.social	2.9	130	2.3	155	3.3	100
mastodon.xyz	2.8	69	2.1	65	2.5	30
mamot.fr	3.1	130	2.4	167	3.0	79
friends.nico	1.9	50	2.3	182	2.2	75
mastodon.cloud	2.8	95	2.4	7.2	3.0	66
framapiaf.org	2.4	47	1.8	19	2.3	22
mao.daizhige.org	2.1	28	1.9	18	2.1	14

3) The impact of instances on triangles.: Moving from neighborhood viewpoint to an ego-network perspective we seek out to what extent instances impact the formation of triangles a node belongs to. To this aim, we start by getting the ego-network of each node, which consists of the subgraph induced over the neighborhood of the node in the network. Then, we focus on the triangles involving the node and its two neighbours. The organization into instances introduces three types of triangle shown in Fig. 2d. Here we deal with the first type, i.e. a triangle whose elements are restricted to the same instance. Based on it, for each node u we compute the *intra-triangle index ITR*, defined as the ratio between the number of triangles of type I and the overall number of triangles involving the node u. An intra-triangle index close to 1 indicates that the user is almost always in triangles limited to the same instance, i.e. her ego-network clusters within the same instance. In Fig. 2e we report, for the ten largest instances, the distribution (CDF) of the intra-triangle index measured on users having a mutual degree greater than 10. From the figure two opposite behaviors emerge: (i) in some instances, such as 'pawoo.net" or "friends.nico", we observe a tendency of the neighborhood's nodes of being clustered within the instance boundary (ii) in other instances, most of nodes are involved in triangles whose members lay in outer instances, i.e. their neighborhood tends to cluster outside the instance they belong to. In general we confirm that the architecture based on independent instances has a stronger impact on the how users' ego-networks cluster; sometimes instances act as a bound on how the neighborhood of their members clusters, other times instances promote an external clustering.

V. THE FOOTPRINT OF THE INSTANCES

In this section we present our findings about the diversity of the instances in terms of the social network they support, i.e. their footprint.

The properties of the node neighborhood, and consequently the node degree, are fundamental in defining how people in an online social network behave. In particular, in the context of DOSNs, the distribution of the degree might be a footprint of the instances since it captures the propensity of the instance members to follow or be followed by other Mastodon users, and how this propensity is distributed among the members of the instance.

1) Degree distribution footprint: Since aggregating all the nodes' degree does not allow us to highlight the contribution of each instance, we compute and compare the distribution of the in-degree (k^{-}), out-degree (k^{+}) and mutual degree (k^{mut}) for the instances with more than one thousand members, i.e. the ten biggest instances in Mastodon. For reasons of readability and space, in Fig. 2f we report the complementary cumulative distribution function (CCDF) of the mutual degree. But our findings can be extended to the out-degree and in-degree cases. From the figure it is evident that two different traits emerge: a group of instances ("pawoo.net", "mstdn.jp" and "friends.nico") hosting users who are more likely to establish mutual relationships with other Mastodon users; and a second group, made up by the remaining instances, containing fewer connected people. With respect to [11], we examine the latter aspect in greater depth, starting from the observation that all the distributions follow a heavy-tail trait, so we adopt a widely used framework for quantifying and fitting heavy-tail and power-law behaviors in empirical data [14]. For all the instances, the power-law distribution was the best candidate.

In TABLE I we report the estimated parameters $\alpha^{(\cdot)}$ – the exponent of the power-law – and $x^{(\cdot)}_{min}$ – the minimal value from which the scaling behavior of the power-law begin – for the out-degree, in-degree and mutual degree distributions of the ten largest instances. In all the cases the α exponents differ from one another, but can be roughly merged into two or three groups. For instance, in the case of mutual degree, "mastodon.social", "mastodon.cloud" and "mamot.fr" have an exponent greater than or equal to 3, while the remaining

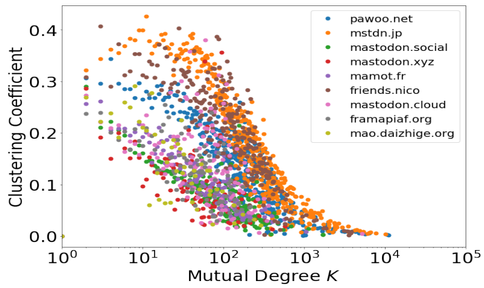

Fig. 3. The local clustering coefficient as a function of the mutual degree in the ten largest instances.

instances lie in the range $[2.1 - 2.5]$. Similar results hold for the out-degree and the in-degree.

In general, by combining the previous observations on the degree distributions and the outcomes of the fitting procedure, we can assert that *each instance has a peculiar footprint which reflects on how its members establish "follower"/"followee" or mutual relationships*. Moreover, we suppose that the underlying mechanism which drives the formation of the nodes' neighborhood may vary among the instance, since the variability of the slope parameter (α) is strictly related to different network growth models.

2) Clustering coefficient footprint: In the previous section we found that instances condition how triangles form, meaning that also the clustering coefficient of the nodes may be a footprint of the instance. So, we analyze the differences in the local clustering between the main Mastodon instances, as shown in Fig. 3. In the figure we report the average local clustering coefficient versus the mutual degree in the ten biggest instances. Instances are very different from one another: *i)* "mstdn.jp", the second largest instance, has a higher average clustering coefficient (0.35) than the other instances (e.g. "pawoo.net" - 0.26, "mastodon.social" - 0.13 and "mastodon.xyz" - 0.08), and it is also higher than the clustering coefficient of the entire network; *ii)* in the "mstdn.jp" subnetwork the clustering coefficient increases up to a peak ($cc = 0.46$) at degree around 30, then slows down. That indicates the presence of clustered regions around nodes with a small-medium connectivity. The same behavior, at a different magnitude order, has been observed in the Twitter Japanese subgraph [13], where there are quasi-clique subgraphs centered around high degree nodes. The above results highlight a further footprint of the instances, i.e. the trend of the local clustering coefficient as a function of the degree.

VI. Conclusion

Decentralized online social networks have recently emerged as a novel paradigm able to better preserve the user's privacy and to ensure higher users' control over the contents they publish. The design of such a decentralized architecture has been mainly accomplished without giving sufficient consideration to the overlaid social network, despite the fact that it played a significant role in the successful design of their centralized counterpart, in terms of both new services and efficiency of data management. This paper investigates the interplay between the system design and the network of social relationships the system supports. Our analysis relies on a novel large dataset about the decentralized microblogging platform Mastodon, and highlights to what extent an instance/community-based infrastructure conditions the way people connect to each other over the platform. Our findings show that the impact instances exert on how their members establish social relationships is instance-dependent, however to a lesser or greater extent people cross over the instance boundaries to search for new friendships. The underlying factors driving this behavior are actually unknown, however it might be different across the instances, since each of them has a specific footprint in terms of degree distribution and clustering coefficient.

As a future work we wonder which role instances and privacy settings play on the diffusion of contents on the Mastodon social network.

References

[1] S. R. Chowdhury, A. R. Roy, M. Shaikh, and K. Daudjee, "A taxonomy of decentralized online social networks," *Peer-to-Peer Networking and Applications*, vol. 8, no. 3, pp. 367–383, 2015.

[2] A. D. Salve, P. Mori, and L. Ricci, "A survey on privacy in decentralized online social networks," *Computer Science Review*, vol. 27, pp. 154 – 176, 2018.

[3] L. Bahri, B. Carminati, and E. Ferrari, "Decentralized privacy preserving services for online social networks," *Online Social Networks and Media*, vol. 6, pp. 18 – 25, 2018.

[4] G. Zyskind, O. Nathan *et al.*, "Decentralizing privacy: Using blockchain to protect personal data," in *Security and Privacy Workshops (SPW), 2015 IEEE*. IEEE, 2015, pp. 180–184.

[5] L. Lorincz, J. Koltai, A. F. Gyor, and K. Takacs, "Collapse of an online social network: Burning social capital to create it?" *Social Networks*, vol. 57, pp. 43 – 53, 2019.

[6] A. Patil, J. Liu, and J. Gao, "Predicting group stability in online social networks," in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW '13, 2013.

[7] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[8] J. Yang and J. Leskovec, "Community-affiliation graph model for overlapping network community detection," in *Proceedings of the IEEE 12th International Conference on Data Mining*, ser. ICDM '12. IEEE, 2012, pp. 1170–1175.

[9] J. Su, A. Sharma, and S. Goel, "The effect of recommendations on network structure," in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW '16, 2016.

[10] M. Zignani, S. Gaito, G. P. Rossi, X. Zhao, H. Zheng, and B. Y. Zhao, "Link and triadic closure delay: Temporal metrics for social network dynamics," in *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, ser. ICWSM'14, 2014.

[11] M. Zignani, S. Gaito, and G. P. Rossi, "Follow the "mastodon": Structure and evolution of a decentralized online social network," in *Proceedings of the 12th International AAAI Conference on Weblogs and Social Media*, ser. ICWSM'18, 2018, pp. 541–551.

[12] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Growth of the flickr social network," in *Proceedings of the First Workshop on Online Social Networks*, ser. WOSN '08. ACM, 2008, pp. 25–30.

[13] S. A. Myers, A. Sharma, P. Gupta, and J. Lin, "Information network or social network?: the structure of the twitter follow graph," in *Proceedings of the 23rd International Conference on World Wide Web*, ser. WWW '14. ACM, 2014, pp. 493–498.

[14] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.