Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM 2022)

Network Analysis of the Information Consumption-Production Dichotomy in Mastodon User Behaviors

Lucio La Cava, Sergio Greco, Andrea Tagarelli*

Dept. Computer Engineering, Modeling, Electronics, and Systems Engineering (DIMES)
University of Calabria, 87036 Rende (CS), Italy
{lucio.lacava, greco, tagarelli}@dimes.unical.it

Abstract

Decentralized Online Social Networks (DOSNs) are today an established alternative to the popular centralized counterparts. In this work, we push forward research on user behaviors in a decentralized context, by exploring the dichotomy between information consumption and production. Using the Mastodon user network as a proxy for the Fediverse landscape, we address two main research questions: Do the consumers, resp. producers, identified in one instance exhibit the same behavior consistently while interacting with other instances? and, Are there users who behave as consumers in one instance and simultaneously as producers in other instances, or vice versa? In this respect, our results reveal interesting traits of Mastodon users, yet unveil the emergence for further studies that can embrace other services in the Fediverse.

Introduction

In the past few years, the online social network (OSN) landscape has witnessed an increased diffusion of a *decentralized* paradigm as opposed to the most popular *centralized* one. In a centralized context, platforms generally see sole players as owners (e.g., companies). Although this feature might positively reflect in terms of administration and development of the platform itself, it also determines a lack of "owner scalability", potentially causing concerns in terms of privacy and policies; for instance, OSN companies are commonly aimed at monetization through advertisements, and this choice is to be passively accepted by their users. On the contrary, Decentralized Online Social Networks (DOSNs) (Datta et al. 2010; Guidi et al. 2018) are built upon sound principles of privacy and transparency, and avoid artificial mechanisms that can influence relationships between users or capture their attention for marketing purposes. Decentralization is obtained through the availability of specific open-source software that allows anyone to create a new server, a.k.a. *instance*, thus building communities guided by spontaneous interest towards certain topics. Moreover, by exploiting such protocols as *ActivityPub*, DOSNs favor seamless interactions between their users, i.e., any user subscribed to a server, said *home* instance, can interact with users on another server, even of a different networking service, without registering on that as well. This unique trait has favored the development of the so-called *Fediverse*, the federated universe of a number of interconnected decentralized services, from micro-blogging to multimedia hosting.

The novelties introduced by DOSNs have attracted the attention of numerous research groups. For instance, Guidi et al. (2018) extensively focused on data management and availability, information diffusion, and privacy on DOSNs, highlighting the main limitations and issues that characterize the DOSNs. Mastodon, the decentralized alternative to Twitter, is recognized as the most relevant platform in the Fediverse in terms of notoriety and user base, and is indeed used as a proxy for the study of DOSNs (Cerisara et al. 2018), (Trienes et al. 2018), (Zignani et al. 2018, 2019), (Raman et al. 2019), (Zulli et al. 2020), (La Cava et al. 2021). Notably, Zignani et al. (2018) provided the first network-analysis work on Mastodon. La Cava et al. (2021) analyzed the network of Mastodon instances, providing insights into macroscopic and mesoscopic structural traits and into the platforms evolution over the years. In (La Cava and Tagarelli 2022), we have analyzed the Mastodon user network, also including relations that emerge from the dualism between information consumption and boundary spanning.

Contributions. Despite such a recent corpus of studies on Mastodon, several questions still remain open particularly about how users behave in the decentralized scenario. We believe one important direction concerns how the seamless interaction between users of different instances — which, in contrast to centralized platforms, does not require a user having multiple accounts or subscriptions — might impact on the users' contribution to the community life, and the role(s) they might take in their home instance as well as in any other instance where they are involved.

Within this view, in this work we aim to fill a gap in understanding the user behaviors in the Fediverse, through the lens of Mastodon. Our focus is on the dichotomy between information-consumption and information-production behaviors of users across Mastodon instances. In this respect, we want to answer the following research questions: (**RQ1**) *Do the consumers, resp. producers, identified in one instance exhibit the same behavior consistently while interacting with other instances?* (**RQ2**) *Are there users who behave as consumers in one instance and simultaneously as producers in other instances, or vice versa?*

*Corresponding author

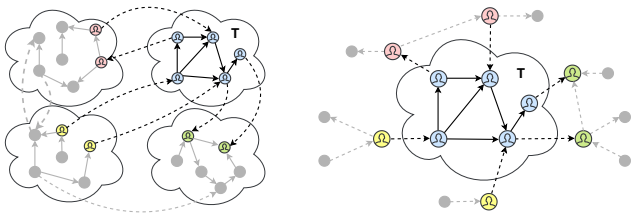

Figure 1: Illustration of our ego-network model applied to a target instance (T) in the Mastodon user network. Node colors indicate home instances. Solid black links are between users belonging to the target instance, whereas dashed black links are from/to other instances. Gray nodes/edges refer to existing entities not involved in the ego-network of T.

The underlying motivation for the above research questions stems from our interest in understanding whether DOSNs, and Mastodon in particular, may exhibit similar user behavioral patterns as those observed in some groups of centralized OSNs. For instance, as discussed in (Perna et al. 2018), it is often the case that a user can produce information and actively interact in certain platforms where s/he is subscribed, but it may also be the case that the same user can assume a *silent* behavior on other platforms.

To the best of our knowledge, information-consumption vs. information-production has not been studied so far in Mastodon. We also point out that, differently from (La Cava and Tagarelli 2022) where information consumption is studied isolatedly, the dichotomous coupling with information-production needs to be analyzed through an unprecedented modeling of the user relations in Mastodon.

Methodology

Data. We used the data provided by (La Cava et al. 2021), which is claimed as the most complete and up-to-date network dataset of Mastodon. This dataset contains more than 1.3M unique users and about 17M unique links between users. Also, according to (La Cava and Tagarelli 2022), it covers about 78% of the Mastodon user base to date, thus enabling a representative study of the Mastodon scenario.

Network model. Let us denote with $\mathcal{U}$ and $\mathcal{I}$ the set of users and instances, respectively, available in the Mastodon dataset. We define the Mastodon user network as a directed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, where the node set $\mathcal{V}$ contains user-instance pairs, i.e., $\mathcal{V} = \{(u,i) \mid u \in \mathcal{U},\ i \in \mathcal{I}\}$, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ where any $(x,y) \in \mathcal{E}$, with $x = (u,i)$ and $y = (v,j)$, means that user v in instance j *receives information produced by* user u in instance i. It should be noted that u and v may coincide only if $i \neq j$.

Given a specific instance $i \in \mathcal{I}$, we define the *extended ego-network* of i as the directed subgraph $\mathcal{G}_i = \langle \mathcal{V}_i, \mathcal{E}_i \rangle$, induced from $\mathcal{G}$, such that $\mathcal{V}_i \subseteq \mathcal{V}$ and $\mathcal{E}_i = \{(x,y) |\ x = (u,j),\ y = (v,k) \wedge (j = i \vee k = i)\} \subseteq \mathcal{V}_i \times \mathcal{V}_i$. Figure 1 reports an illustration of our ego-network model.

Identification of consumers and producers. To identify users that tend to over-consume information produced by others, we take a perspective that relies on the theory of *lurking behavior analysis* (Sun et al. 2014), (Edelmann 2013): the majority of OSN users does not actively contribute, rather it mostly remains hidden or "silent", gaining benefit from information produced by other users. Modeling and analyzing lurking behaviors has been formulated as an eigenvector-centrality-based node ranking problem, which is content-agnostic, and builds upon three key principles (Tagarelli and Interdonato 2014): (*i*) content over-consumption, (*ii*) the authoritativeness of the information received, (*iii*) the non-authoritativeness of the information produced. The first shapes the imbalance between the amount of information a user consumes w.r.t. the amount of information she/he produces, whereas the others refer to the importance as information producer of her/his in-neighbors, and the importance as information consumer of her/his out-neighbors, respectively.

Given a directed graph $\mathcal{G}$, here corresponding to the Mastodon user network or to any instance-specific ego-network, the LurkerRank score $LR(v)$ of any node v according to the *in-out-neighbors-driven lurker ranking* formulation (Tagarelli and Interdonato 2014) is defined as:

$$LR(v) = \alpha[LR_{in}(v)(1 + LR_{out}(v))] + (1-\alpha)p(v), \quad (1)$$

where LR_{in} (*in-neighbors-driven* lurking function) is:

$$LR_{in}(v) = \frac{1}{|\mathcal{N}_v^{out}|} \sum_{u \in \mathcal{N}_v^{in}} \frac{|\mathcal{N}_u^{out}|}{|\mathcal{N}_u^{in}|} LR(u), \quad (2)$$

and LR_{out} (*out-neighbors-driven* lurking function) is:

$$LR_{out}(v) = \frac{|\mathcal{N}_v^{in}|}{\sum_{u \in \mathcal{N}_v^{out}} |\mathcal{N}_u^{in}|} \sum_{u \in \mathcal{N}_v^{out}} \frac{|\mathcal{N}_u^{in}|}{|\mathcal{N}_u^{out}|} LR(u). \quad (3)$$

$\mathcal{N}_u^{in}, \mathcal{N}_u^{out}$ are the in-, out-neighbor sets of u, α is a damping factor in $[0,1]$ (by default 0.85), and $p(v)$ is the value of the PageRank-like *personalization vector* (by default $1/|V|$). To avoid zero or infinite ratios, the values of the in/out-neighborhood size of a node are Laplace add-one smoothed.

According to the LurkerRank, the higher the LR-score of a node, the stronger is the status of the node as consumer in the network. Conversely, as demonstrated in (Tagarelli and Interdonato 2014), (Perna et al. 2018), the bottom of a LR ranking can be used to identify the users that act as opposed to consumers, i.e., producers. We hence leverage an analysis of ranking *heads* and *tails* to model the dichotomy between consumers and producers in the ego-network model.

It should be noted that either social and interaction relations can be seen as proxy for information consumption by users; indeed, LurkerRank has been extensively evaluated on both followee-follower and comment/like/mention graphs (Tagarelli and Interdonato 2014, 2015). Since the available Mastodon user relations are of "following" type, in this work LurkerRank is applied to followship graphs.

Evaluation goals and assessment criteria. Our goal is to answer the previously stated **RQ1**-**RQ2** based on an evaluation of the LR ranking solutions obtained on each instance's ego-network. To this purpose, we first use the *Jaccard similarity* coefficient to measure the matching degree between

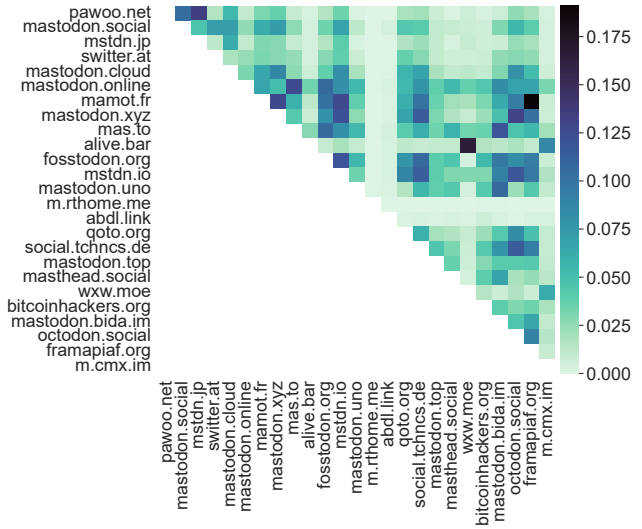

Figure 2: Jaccard similarity between the $tail@25$ from the LR rankings, i.e., top-producers, for each pair of instances in the top-25 instances context.

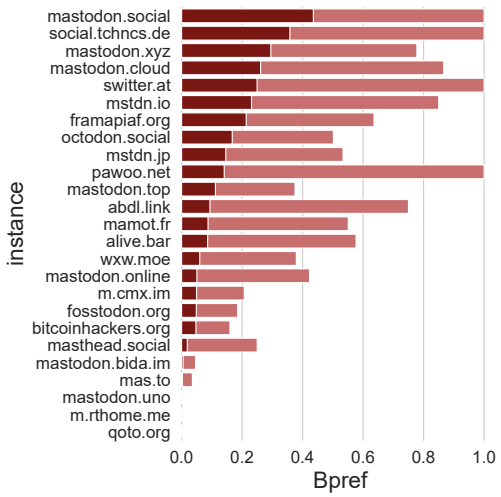

Figure 3: Bpref on the $head@25$ from the LR rankings, i.e., top-consumers, for each of the top-25 instances. The average and maximum values for each reference instance over the others are shown in red and light red, respectively. Instances are ordered by decreasing average Bpref values.

(portions of) LR rankings L_i, L_j of any two given ego-networks: $Jaccard(L_i, L_j) = (|L_i \cap L_j|)/(|L_i \cup L_j|)$.

To delve into the comparison of two rankings, we also resort to the *binary preference* criterion, which measures how often judged relevant items are retrieved in a list L before judged non-relevant ones (Buckley and Voorhees 2004):

$$Bpref(L) = \frac{1}{R}\sum_{r}\left(1 - \frac{min(\#n \text{ ranked above } r, R)}{min(N, R)}\right), \quad (4)$$

where r and n are *relevant* and *non-relevant* judged items, resp., on a total number R (resp. N) of relevant (resp. non-relevant) items. Note that *not-judged* items may be present in L, although their position is discarded. Bpref ranges within $[0, 1]$, whereby the closer to 1 the better the Bpref.

Given a pair of ego-networks $(\mathcal{G}_i, \mathcal{G}_j)$, and (portions of) their respective LR rankings L_i, L_j, for any choice of a *reference instance*, say L_i, we compute the Bpref of L_j w.r.t. L_i, such that: the users common to the instances i and j correspond to the relevant items, the users in $\mathcal{G}_j$ having j as home instance correspond to the not-judged items, and the users in $\mathcal{G}_j$ not having j as home instance correspond to the non-relevant items. Note that our definition of Bpref is asymmetric as it depends on the choice of the reference instance, i.e., Bpref of L_j w.r.t. L_i is not necessarily equal to Bpref of L_i w.r.t. L_j.

Settings. To ensure significance of our results yet representativeness of the currently active Mastodon landscape, we focused on the top-25 instances by user base according to the *instances.social* platform, which is widely recognized as the de-facto tracker for Mastodon. We organized such instances into three subsets called *contexts*, namely top-5, top-10, and top-25 instances, and for each instance in a given context, we induced its ego-network. Note that the same instance might have less/more external edges, hence a different ego-network, depending on the selected context.

Moreover, given an LR ranking solution, we considered the top-$k\%$, resp. bottom-$k\%$, users by score in LR, dubbed as $head@k$, resp. $tail@k$. We set by default $k = 25$.

Results

To answer our first research question (**RQ1**), we begin with a pairwise comparison of the ego-network LR heads, resp. tails. Considering the top-5 instances subset, there is some evidence of Jaccard similarity between the head, resp. tail, at $k = 25$, of their corresponding LR rankings, up to 0.04 and 0.14, respectively. Broadening the scope at the top-10 context, we observe analogous Jaccard coefficient values, with peaks of 0.04 resp. 0.14 for pairwise comparisons of $head@25$ resp. $tail@25$, and of 0.09 resp. 0.19, for pairwise comparisons of $head@25$ resp. $tail@25$ when we account for the largest context (i.e., top-25 instances). These results suggest the existence of a few users that, in at least two different instances, exhibit behavioral consistency, and this holds more for producers than consumers. Moreover, not only this finding is robust to the extent of the context of comparison of instance pairs, but also to the size of the ranking heads and tails (results with $k \in \{5, 10\}$ follow trends analogous to $k = 25$). Due to space limitations, in Fig. 2 we report results only for the largest ranking scope ($k = 25$) and context (top-25 instances) of top-producers.

Note that the above results are valid for pairwise comparisons. When extending to triplets of instances, however, we find Jaccard values close or equal to zero in all cases.

We then inspected the interesting cases observed for the pairwise scenario through our Bpref-based evaluation. Figures 3-4 show results for each instance in the top-25 context. Interestingly, the reference instance with the highest average Bpref turns out to be the first established instance in Mastodon, i.e., *mastodon.social*, for consumers resp. pro-

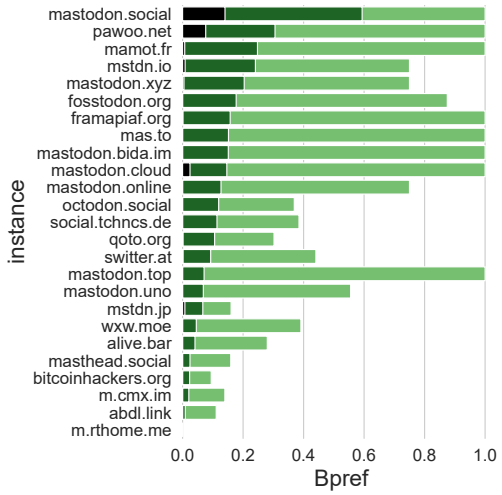

Figure 4: Bpref on the $tail@25$ from the LR rankings, i.e., top-producers, for each of the top-25 instances. The average, maximum, and minimum values for each reference instance over the others are shown in green, light green, and dark green, respectively.

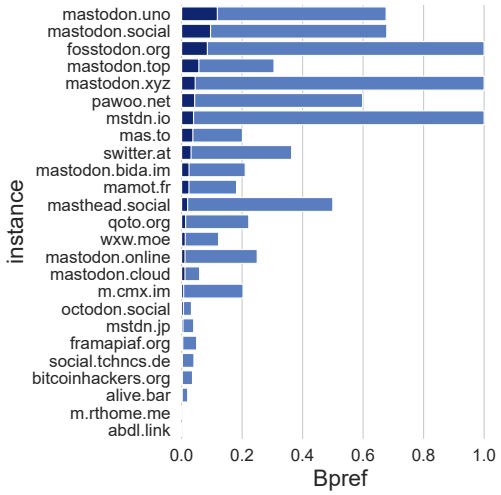

Figure 5: Bpref on the $head@25$ w.r.t. $tail@25$ from the LR rankings, and vice versa, for each of the top-25 instances. The average resp. maximum values for each reference instance over the others are shown in blue resp. light blue.

ducers, with average values up to 0.436 resp. 0.593, for $k = 25$ in the top-25 context. Moreover, 4 resp. 8 out of 25 instances reach maximum Bpref equal to 1 for the consumer resp. producer evaluation. It should be noted that the maximum Bpref scores relating to producers (Fig. 4) are on average greater than those of consumers (Fig. 3), as analogously observed for the Jaccard based evaluation.

As concerns our second research question (**RQ2**), we start again by considering a pairwise instance comparison, choosing this time the ranking head (i.e., the top consumers) from one instance and the tail (i.e., the top producers) from the other instance in each pair. Jaccard similarity results for $k = 25$ (not shown) reveal maximum values below 0.1 regardless of the instance context. Nonetheless, we also take a finer-grain perspective through Bpref, whereby, for each instance j and reference instance i in a top-N context (with $i \neq j$), we aggregated over both the Bpref of $(tail@k)_j$ w.r.t. $(head@k)_i$ and the Bpref of $(head@k)_j$ w.r.t. $(tail@k)_i$. Figure 5 shows results for the top-25 instances and $k = 25$. We observe that 7 out of 25 instances show maximum Bpref above 0.4, with 3 instances reaching Bpref at 1. On average, however, the scores appear to be quite lower than the previous evaluation concerning **RQ1**, with *mastodon.uno* as the instance with the highest average score (0.118).

Discussion

We investigated on the impact that the Fediverse decentralization might have on the user behaviors in terms of information production and consumption, either in a repeated or an alternate fashion across two or more instances in Mastodon.

Our analysis of the instance-specific ego-networks' LR ranking heads and tails to capture the consumption-production dichotomy unveiled a few interesting facts, which are summarized as follows.

- There exists a small number of largest Mastodon instances (less than 25) in which either pairwise behavioral consistency and alternation can be observed.
- The fraction of users that are regarded as information consumers, resp. producers, simultaneously in two instances is below 0.1, resp. 0.2, Jaccard similarity. This holds to a less extent when comparing top-consumers of one instance with top-producers of another instance, and vice versa.
- Although statistically not relevant in terms of absolute number of users, the behavioral consistency exhibited by certain consumers resp. producers (**RQ1**) appears to be relatively strong according to our Bpref evaluation of the rankings' heads resp. tails. This is more evident on average for producers, which might be explained by the ease of interaction and content dissemination favored by Mastodon, thus reducing the attitude of having a silent behavior.
- The behavioral alternation (**RQ2**) also turns out to be limited to low fractions of users per instance-pairs, and with lower Bpref strength than for the behavioral consistency.

Overall, **RQ1** and **RQ2** get moderately affirmative answers, and we suspect that this may generally hold in other platforms of the Fediverse. However, our findings need to be taken with a grain of salt, because of two main reasons: (1) our analysis context is content-agnostic and the consumption-production dichotomy is modeled through follower-followee relations, which are in principle a weaker proxy of the user activity in OSNs. (2) Mastodon instances are designed to be interrelated, rather than being perceived as independent and separate OSN platforms; in fact, differently from what happens when the same user has to subscribe to multiple centralized platforms, Mastodon users could feel a limited need of taking a silent behavior on some instances while being producers on other instances. This, however, might change when extending the analysis also to instances of other services in the Fediverse than Mastodon.

References

Buckley, C.; and Voorhees, E. M. 2004. Retrieval Evaluation with Incomplete Information. In *Proc. of the 27th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, SIGIR '04, 25–32.

Cerisara, C.; Jafaritazehjani, S.; Oluokun, A.; and Le, H. T. 2018. Multi-task dialog act and sentiment recognition on Mastodon. In *Proc. of the 27th Int. Conf. on Computational Linguistics*, 745–754. Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Datta, A.; Buchegger, S.; Vu, L.-H.; Strufe, T.; and Rzadca, K. 2010. *Decentralized Online Social Networks*, 349–378. Boston, MA: Springer US. ISBN 978-1-4419-7142-5.

Edelmann, N. 2013. Reviewing the definitions of "lurkers" and some implications for online research. *Cyberpsychology, Behavior, and Social Networking*, 16(9): 645–649.

Guidi, B.; Conti, M.; Passarella, A.; and Ricci, L. 2018. Managing social contents in Decentralized Online Social Networks: A survey. *Online Social Networks and Media*, 7: 12–29.

La Cava, L.; Greco, S.; and Tagarelli, A. 2021. Understanding the growth of the Fediverse through the lens of Mastodon. *Appl. Netw. Sci.*, 6(1): 64.

La Cava, L.; and Tagarelli, A. 2022. Information Consumption and Boundary Spanning in Decentralized Online Social Networks: the case of Mastodon Users. *CoRR*, abs/2203.15752.

Perna, D.; Interdonato, R.; and Tagarelli, A. 2018. Identifying Users With Alternate Behaviors of Lurking and Active Participation in Multilayer Social Networks. *IEEE Trans. Comput. Soc. Syst.*, 5(1): 46–63.

Raman, A.; Joglekar, S.; Cristofaro, E. D.; Sastry, N.; and Tyson, G. 2019. Challenges in the Decentralised Web: The Mastodon Case. In *Proc. of the Internet Measurement Conference*, 217–229. ACM.

Sun, N.; Rau, P. P.-L.; and Ma, L. 2014. Understanding lurkers in online communities: a literature review. *Computers in Human Behavior*, 38: 110–117.

Tagarelli, A.; and Interdonato, R. 2014. Lurking in social networks: topology-based analysis and ranking methods. *Soc. Netw. Anal. Min.*, 4(1): 230.

Tagarelli, A.; and Interdonato, R. 2015. Time-aware analysis and ranking of lurkers in social networks. *Soc. Netw. Anal. Min.*, 5(1): 46:1–46:23.

Trienes, J.; Cano, A. T.; and Hiemstra, D. 2018. Recommending Users: Whom to Follow on Federated Social Networks. *CoRR*, abs/1811.09292.

Zignani, M.; Gaito, S.; and Rossi, G. P. 2018. Follow the "Mastodon": Structure and Evolution of a Decentralized Online Social Network. In *Proc. of the Int. Conf. on Web and Social Media (ICWSM)*, 541–551. AAAI Press.

Zignani, M.; Quadri, C.; Gaito, S.; Cherifi, H.; and Rossi, G. P. 2019. The Footprints of a Mastodon: How a Decentralized Architecture Influences Online Social Relationships. In *Proc. of the IEEE Conf. on Computer Communications (INFOCOM) Workshops*, 472–477.

Zulli, D.; Liu, M.; and Gehl, R. 2020. Rethinking the social in social media: Insights into topology, abstraction, and scale on the Mastodon social network. *New Media & Society*, 22(7): 1188–1205.