Something completely different, or more of the same? TikTok and the "democratization" of cultural content production

Keywords: TikTok; Challenges; Digital content production; Participation inequalities.

Extended Abstract

Social media and digital production tools have provided a big impulse to the possibility of a democratization of cultural content production. However, enhanced viability does not imply an actual capacity to tap into the opportunity. Research on established social media such as Twitter, Facebook and Instagram show that content production, and especially the most popular contents, is mostly driven by a small elite of super-users, and an additional but still small share of content creators. In this paper, we wonder whether a more recent social media such as TikTok - whose design features are optimized to facilitate individual content production and to enhance visibility of content from unknown users - actually makes a difference in this regard. What we find is that the production and networking patterns observed do not show substantial departures from those observed for older social media.

We work on a dataset made publicly available by Jason Baumgartner (at https://files.pushshift.io/tiktok/) that contains information related to 25 millions videos posted on TikTok between October 2014 and December 2019 from 6,973,120 unique users, who posted 4,593,835 unique hashtags. The mean number of comments per video is 46.15, with a standard deviation of 605.86, while for the number of videos per user the figures are 3.62 and 12.58, respectively. The average video attracts a non-negligible amount of comments, and therefore caters some attention, but with a huge variability. On the other hand, the average number of videos posted by users is small, but again with a huge variability.

As already mentioned, TikTok has brought some relevant innovations compared to preexisting social media. These features allow a variety of possible networking schemes across the platform: homophilic communities built through direct friendship relations or interest in key social issues, aggregation around influencers, ephemeral communities shaped by social trends or specific memetic content, etcetera. Whereas influencer-centered communities reflect the widely described centralization of traditional social media such as Facebook and Instagram, homophilic and ephemeral communities are characterized by a fragmented structure closer to that of the communities that are typical of the vernacular web. Once again, it is difficult to predict a priori what kind of overall networking structure emerges from the conflation of different sub-communities with distinctive characteristics. What is then the social networking structure of TikTok communities, and are there any major differences between the classical modes of aggregation enabled by older social media and those enabled by the specific affordances of TikToK?

In order to study the characteristics of TikTok communities, we build two types of networks through some features offered by the platform: one based on direct interactions, and one based on indirect interactions. As to the former, we use mentions, that is, we connect a user posting a video with the other users mentioned in the text accompanying the post. In the latter, we use the one-mode projection typical of bipartite networks to select all users that have participated in the same challenge — i.e, a call to take some sort of action and record it. Typically, challenges originate from popular videos usually involving a song, dance move,

movie quote, etc —, and we connect all those users who have chosen the same song in their challenge-related TikTok post. Whereas the first kind of network is built around online influence or direct acquaintance, as the connection is established by direct tags to specific users, this second kind of network reflects the logic of an ephemeral community held together by the common momentary interest in, and orientation toward, a certain cultural content. We also associate a weight to each node, quantifying the number of challenges/songs that each user has participated in. Looking at their degree distributions, the distribution of the number of connections per node, we find that in all cases their decay is compatible with a power-law-like function, see Figure 1 upper row. We also look at how the weights of the nodes in the challenge and song networks are distributed, finding again that these are not peaked around a well-defined value, and display a slow but steady decay (Figure 1, bottom row). Finally, we look at the community structure. We use the Louvain algorithm, based on modularity maximization, to find groups of tightly connected users inside the network of mentions. The first thing to note, again, is that the distribution does not peak around a certain value and decays as a fat tail. Interestingly, it shows a non-monotonic behavior and an increase in probability at intermediate community size (i.e., 100-200 users). A hypothesis for this phenomenon could be a Dunbar number-like effect, that is, a socio-cognitive upper bound to the size of a viable community for optimum participation.
These results restate once more, yet from a different angle, what we have found before: namely, on the one hand the heterogeneity in the modes by which connections and communities are formed within TikTok, while on the other hand the general persistence of unequal participation that rewards a few relevant influencers and challenges while leaving the majority of users on the periphery of the platform.

References

Nielsen, J. (2006) The 90-9-1 rule for participation inequality in social media and online communities.
Boulianne, S. (2015) Social media use and participation: A meta-analysis of current research. Information, communication & society 18, 524–538 .

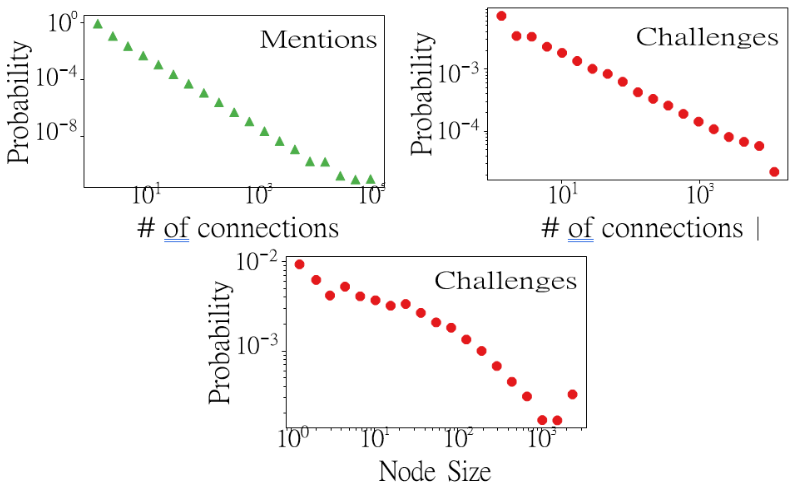

Figure 1. Topological characterization of the different networked representations discussed in the text. On the top row we display the normalized histogram of the number of connections per node, i.e., the degree distribution, for the mention (left) and challenge (right) networks. On the bottom row we show the normalized histogram of node size – number of users that participate in a challenge.