

**UNIVERSITY OF WAIKATO**

**Hamilton  
New Zealand**

**Zipf's Law across social media**

Michael P. Cameron

**Working Paper in Economics 7/22**

March 2022

**Michael P. Cameron**

School of Accounting, Finance and Economics  
*and*

Te Ngira – Institute for Population Research

University of Waikato

Private Bag 3105

Hamilton 3240

NEW ZEALAND

Email: [mcam@waikato.ac.nz](mailto:mcam@waikato.ac.nz)

## **Abstract**

Zipf's Law describes an empirical regularity that appears across many human and physical domains, and states that ranked data exhibits a power law distribution. Although there are various extant studies illustrating power law relationships using social media data, we significantly extend these previous studies by looking at eight popular online social media networks: (1) Twitter; (2) YouTube; (3) Instagram; (4) Twitch; (5) DLive; (6) TikTok; (7) Daily Motion; and (8) Facebook. Specifically, we test whether the distribution of connections (followers, subscribers, or likes) follows a power law distribution for the top 5000 members of each social network. We find strong evidence that a power law relationship exists for every one of the social networks that we study, although this relationship breaks down for users at the top of the connections distribution. Despite the finding of a power law relationship for all of these social networks, the degree of inequality in social media connections differs substantially across the different networks, with the highest degree of inequality in DLive, and the lowest degree in TikTok and YouTube.

## **Keywords**

Social media

Zipf's Law

Power law

Pareto distribution

## **JEL Classification**

D85; L86

## **Acknowledgements**

We are grateful to Liam Bullen for excellent research assistance, to Sarah Cameron for advice on the descriptions of the social media networks, and to Gemma Piercy-Cameron for comments on an earlier draft.

## 1. Introduction

Zipf's Law describes an empirical regularity that has been observed across multiple domains, including the frequency distribution of word usage (Zipf, 1932), the distribution of city sizes (Zipf, 1949), the frequency distribution of family names (Zanette and Manrubia, 2001), the distribution of commercial success in the music industry (Cox et al., 1995), the market capitalisation of large firms (Gabaix and Landier, 2008) and the number of battle deaths in wars (Roberts and Turcotte, 1998), among many others. The idea underlying Zipf's Law is that, once elements in a set are ordered from largest to smallest, the resulting distribution of ranked data follows a power law (Newman, 2005). This empirical insight has implications for the measurement of various real-world phenomena (Clauset et al., 2009), and has been observed for a number of internet features, including the inter-domain topology (Faloutsos et al., 1999), the popularity of files requested (Cunha et al., 1995), the numbers of unique website visitors, website pages, and website links (Adamic and Huberman, 2002), and the distribution of the number of email connections (Ebel et al., 2002).

Despite a number of examples of networks, including online networks, exhibiting power law behaviour, power laws in modern social media applications have to date received little attention. In this paper, we investigate whether a number of different online social networks, of differing sizes, vintages, and growth trajectories, exhibit power law behaviour in relation to the number of social network connections. If the number of connections (friends, contacts, followers, or subscribers) follows a power law, then this provides an empirical regularity to the observed inequality in social media connections. This in turn may have implications for the development and maintenance of social capital (Cook, 2014), the diffusion of information (Xiong et al., 2012), the propagation of positive and negative messages (Tsugawa and Ohsaki, 2015; Subramanian, 2017), and trends in populism and other social movements (Kidd and McIntosh, 2016; Gerbaudo, 2018; Postill, 2018).

Specifically, we focus our attention on eight popular international social media networks: (1) Twitter; (2) YouTube; (3) Instagram; (4) Twitch; (5) DLive; (6) TikTok; (7) Daily Motion; and (8) Facebook. In the case of the first seven social networks, we have cross-sectional data on the number of social media connections (friends, contacts, followers, or subscribers) of the top 5000 members (or content creators) ranked by the number of connections. For Facebook, we have cross-sectional data on the number of 'likes' of the top 5000 Facebook pages ranked by the number of 'likes'. We use this data to identify and describe the distribution of connections, , to determine the extent to which the number of connections follows a power law distribution. We find strong evidence that all eight social networks exhibit power law distributions in the number of social media connections, but that this relationship breaks down in the upper tail of each distribution. We also find substantial differences in the degree of inequality in social media connections across the different networks, with the highest degree of inequality in DLive, and the lowest degree in TikTok and YouTube.

This paper builds on the extant literature on power law distributions in social networks. Ribeiro et al. (2009; 2010) demonstrated that the distribution of MySpace friends exhibited a ‘double-Pareto’ shape, with a clear break at around 100 friends, and different power law relationships above and below this threshold. Falck-Ytter and Overby (2012) found that a power law distribution is a good description of the number of views of the 160 most popular YouTube videos, but not for the number of followers for the 10,020 most popular Twitter users. This contrasts with Hurlimann (2015), who found that a close fit to a power law for the number of Twitter followers, as did Oshop and Foss (2015) for the Twitter followers of celebrities. Rastogi (2016) showed that a power law distribution approximately holds for the top 12 ranked Twitter users, the top 31 celebrities, the top nine politicians, and the top nine sportspeople for total tweets, average retweets, and total followers. Enjolras (2014) shows a similar result for the number of Twitter followers for 84 Norwegian politicians. We extend those earlier analyses to the top 5000 content creators across eight different online social networks.

The remainder of this paper proceeds as follows. The next section briefly describes the eight social networks that we have data for. We then briefly outline the data and methods we use, before presenting and discussing our results. The final section concludes.

## 2. Eight social media networks

A social network is a structure of relationships (or ‘links’) and interactions (or ‘communications’) between people, organisations, and other entities (or ‘nodes’) (Yu and Kak, 2014). Modern social media networks facilitate this structure using the Internet, providing a platform for people to share their ideas and experiences with many others locally or globally.

We analyse data on eight online social media networks, each of which has distinctive features. The first social media network we consider is Twitter, a ‘microblogging’ platform that began operation in 2006, and currently has over 330 million registered users.<sup>1</sup> On Twitter, users post ‘tweets’, which are short (less than 280 characters) text-based messages, or short audio or video clips (less than 140 seconds). The key networking aspect of Twitter is that users can choose to ‘follow’ other users, thereby ensuring that they always receive notification of tweets from the followed user. Users can also forward tweets (‘retweet’) with or without additional comment, or ‘like’ tweets, expressing enjoyment or approval of the tweet.

YouTube is an online video sharing platform owned by Alphabet (formerly Google). YouTube began operation as an independent platform in 2005 (and was acquired by Google in 2006), has about 2 billion active users, and is the second most visited website globally.<sup>2</sup> On YouTube, users upload videos that are either publicly viewable or may be ‘unlisted’ and viewable only by those with access to the particular web link – as of 2019, over 500 hours of video content was being uploaded per minute.<sup>3</sup> Content creators can generate channels for other users to follow. For our purposes, the key networking aspect of YouTube is that users can

---

<sup>1</sup> <https://en.wikipedia.org/wiki/Twitter>.

<sup>2</sup> <https://en.wikipedia.org/wiki/YouTube>.

<sup>3</sup> <https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/>.

choose to subscribe to YouTube channels, receiving notifications of new video content. Users can also comment on videos, as well as ‘like’ or ‘dislike’ videos.<sup>4</sup>

Instagram is an online video and photo sharing platform owned by Meta Platforms (formerly Facebook Inc.). Instagram began operation as an independent platform in 2010 (and was acquired by Facebook in 2012), and has over 1.3 billion users.<sup>5</sup> On Instagram, users post content that can be publicly shared or shared only with pre-approved followers. For our purposes, the key networking aspect of Instagram is that users can choose to follow other Instagram users, receiving notifications of new content. Users can also comment on and ‘like’ content.

Twitch is a video live streaming platform that is commonly used for streaming live video gaming. Twitch began operation in 2011 (as a spinoff of the general interest video streaming platform Justin.tv), and was acquired by Amazon in 2014. It currently has over 15 million daily active users.<sup>6</sup> On Twitch, content creators can live stream video gaming sessions or other content, which can be viewed by other users either live or on-demand. For our purposes, the key networking aspect of Twitch is that users can choose to follow content creators. Users can also participate in live chat during live streams or comment on on-demand videos. They can also ‘subscribe’ to particular content creators, which provides a regular payment to the creator from the user, or can ‘cheer’, which involves donating a small amount of money to them through microtransactions.

DLive is another video live streaming platform. DLive began operation in 2017, and is commonly used as an alternative for Twitch, with many similar features, but only 5 million daily active users as of 2019.<sup>7</sup> Similar to Twitch, users can choose to follow content creators, comment on streams and videos, and donate to content creators.

TikTok is a video-focused platform that is the international version of the Chinese platform Douyin, and began operation in 2017. TikTok has over 800 million monthly active users.<sup>8</sup> On TikTok, users post short videos (of less than three minutes), which often feature music, dancing, pranks, jokes or short clips. For our purposes, the key networking aspect of TikTok is that users can choose to follow content creators. They can also ‘like’ content, post ‘reaction’ or ‘duet’ videos, and can give donations or small gifts to content creators.

Daily Motion is an online video-sharing platform, similar to YouTube. Daily Motion began operation in 2005, and has about 300 million monthly users.<sup>9</sup> Similar to YouTube, users can follow content creators, like and comment on content.

Finally, Facebook is an online social media and social networking platform, owned by Meta Platforms (formerly Facebook Inc.). Facebook began operation in 2004, and has over 2.85

---

<sup>4</sup> Counts of ‘dislikes’ for each video were no longer displayed from November 2021.

<sup>5</sup> <https://www.businessofapps.com/data/instagram-statistics/>.

<sup>6</sup> [https://en.wikipedia.org/wiki/Twitch\\_\(service\)](https://en.wikipedia.org/wiki/Twitch_(service)).

<sup>7</sup> <https://www.globenewswire.com/news-release/2019/07/17/1883829/0/en/DLive-Daily-Active-Users-Grow-Six-Fold-in-New-Report.html>.

<sup>8</sup> <https://www.businessofapps.com/data/tik-tok-statistics/>.

<sup>9</sup> <https://en.wikipedia.org/wiki/Dailymotion>.

billion monthly active users.<sup>10</sup> Facebook offers a variety of options for sharing and commenting on content. Users create networks of friends. However, we focus on Facebook ‘pages’, which do not have friend networks. Instead, users can choose to follow Facebook pages, and like and comment on content from those pages.

These eight social networks differ substantially in vintage, network size, type of content, and the types of interactions between users. However, they all share a common feature in that the underlying network structure is comprised of connections between individual users, or between users and content creators. The structure of connections naturally leads to different content creators and users having different number of connections. We analyse the distribution of the number of connections for each social network, in order to determine whether each distribution follows a power law.

### **3. Data and Methods**

We use data on the number of social media connections (followers for Twitter, Instagram, Twitch, DLive, TikTok, and Daily Motion; subscribers for YouTube; and page likes for Facebook pages). These data were collated by the social media analytics website socialblade.com, and is limited to the top 5000 content creators or users on each platform. The data we use was collated on 30 September 2021. The raw number of social media connections in our dataset varies substantially between the social networks. The average number of Instagram followers for content creators in our dataset is 9.813 million, while the average number of Daily Motion followers is just 2867. We therefore analyse each social network separately, and then compare results.

To illustrate the power law relationships, we follow Gabaix and Landier (2008) by first illustrating the distribution of the number of social media connections by plotting the natural log of the number of connections against the natural log of rank. However, rather than using the absolute value of rank, we instead use rank minus one half, following Gabaix and Ibragimov (2011), in order to avoid small sample bias in the estimated power law relationship. A power law relationship in the social media connections distribution would appear in these plots as a straight line relationship.

To further quantify the power law relationship for each platform, we then regress the natural log of rank (minus one half) on the natural log of the number of social media connections. A power law relationship would be found if the coefficient on the number of social media connections is statistically significant. Gabaix and Ibragimov (2011) note that OLS standard errors are incorrect for this specification, so in addition to the OLS standard errors, we also report revised standard errors based on the formula in Gabaix and Ibragimov (2011).

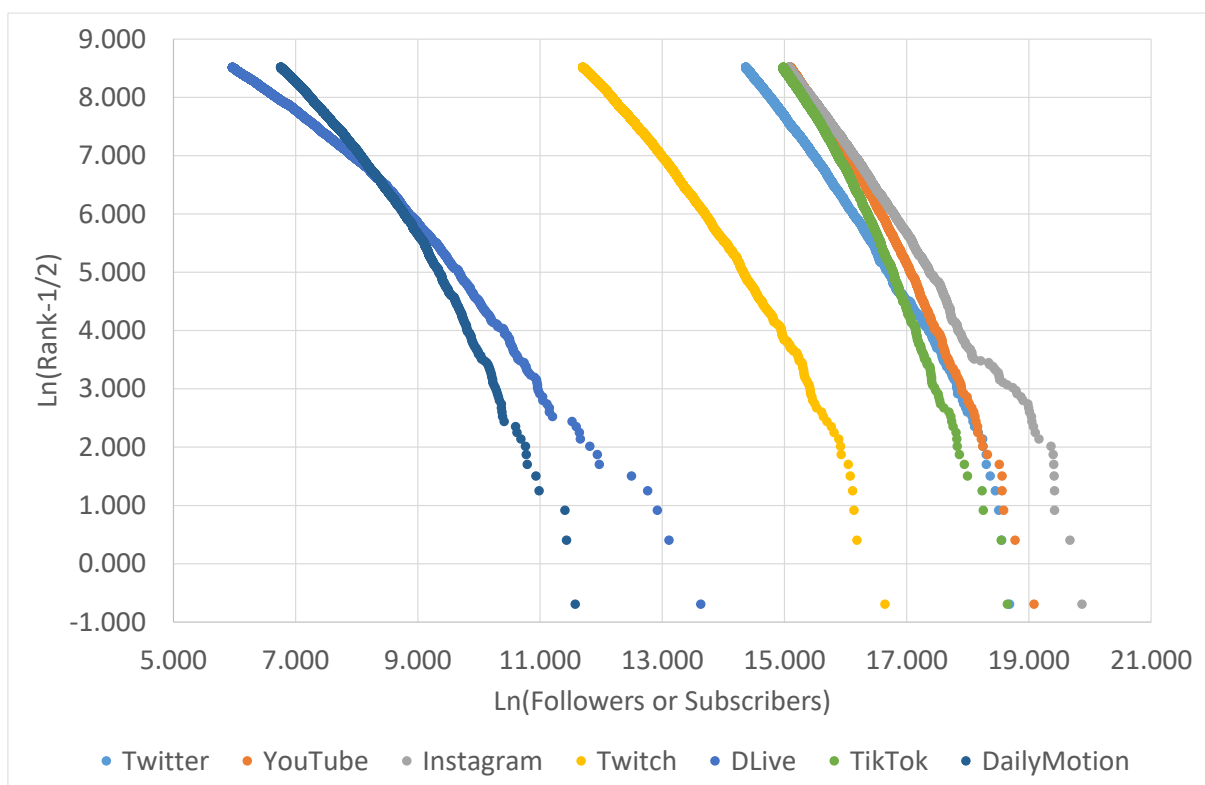
---

<sup>10</sup> <https://en.wikipedia.org/wiki/Facebook>.

#### 4. Results

Figure 1 illustrates the relationship between the natural log of rank (minus one half) and the natural log of social media connections, for the first seven social media networks (excluding Facebook page likes, which are included in a separate figure in the Appendix, Figure A1). The pattern in the relationship is similar for each platform, with a broadly linear shape for most of the ranking, becoming noticeably less linear in the tail, i.e. at the top of the distribution of social media connections. However, despite the consistent finding of linearity, the distributions clearly differ across the social media platforms, in two ways. First, the relative size or popularity of each social media network is evident in the clear ranking from left to right across Figure 1. DLive and Daily Motion have the smallest numbers of active users, while Instagram and YouTube have the most. Second, there are clear differences in the slopes in Figure 1. These slopes represent the degree of inequality in the number of social media connections within the top 5000 content creators or users in our sample. A flatter slope represents greater inequality, with the highest ranked content creators or users having a much higher relative number of connections than lower ranked content creators or users within the top 5000. Thus, DLive exhibits greater inequality in the number of social median connections than Daily Motion, although content creators for those two social networks in our sample have similar average numbers of connections, while Twitter and Instagram exhibit greater inequality than YouTube, which in turn exhibits greater inequality than TikTok.

**Figure 1. Distribution of number of social media network connections, September 2021**



The power law relationships are further illustrated in Table 1, which presents the results of regression models of the natural log of the number of social media connections against the natural log of rank (minus one half). In each case, the relationship is highly statistically significant, and the results confirm those illustrated in Figure 1. The adjusted R-squared values and the coefficients on the natural log of rank (minus one half) are uniformly high, demonstrating a strong linear relationship between the natural log of rank (minus one half) and the natural log of the number of social media connections. This demonstrates that the distribution may be adequately represented as a power law distribution in each case.

**Table 1. Power law regression results**

<b>Social media network</b>	<b>Ln(Rank-1/2)</b>	<b>Constant</b>	<b>Adjusted R<sup>2</sup></b>
Twitter	-1.506*** (0.007) [0.030]	30.24*** (0.112)	0.992
YouTube	-1.801*** (0.009) [0.036]	35.81*** (0.142)	0.990
Instagram	-1.522*** (0.006) [0.030]	31.52*** (0.087)	0.996
Twitch	-1.354*** (0.008) [0.027]	24.49*** (0.095)	0.988
DLive	-0.939*** (0.006) [0.019]	14.28*** (0.112)	0.976
TikTok	-1.958*** (0.011) [0.039]	37.97*** (0.171)	0.984
Daily Motion	-1.364*** (0.009) [0.027]	17.88*** (0.066)	0.981
Facebook page likes	-1.616*** (0.011) [0.032]	33.26*** (0.170)	0.981

N.B. Each row reports the result of a separate linear regression of  $\ln(\text{connections})$  on  $\ln(\text{Rank}-1/2)$  for the top 5000 content creators or users on that network. Robust standard errors are reported in parentheses, with revised standard errors based on Gabaix and Ibragimov (2011) in square brackets. \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ .

Comparing across the different regression models, the difference in slopes that was illustrated in Figure 1 is clear, with TikTok having the largest slope coefficient (representing the least inequality within the distribution of connections of the top 5000 content creators or users), and DLive having the smallest coefficient (and the greatest inequality in connections). Along with the other seven social networks, we find similar results (strongly statistically significant and linear) for the number of Facebook page likes.



However, referring back to Figure 1, it is clear that the linear relationship breaks down at the very tail of the distribution for some of the platforms. To further test this relationship, we re-estimated each regression model, omitting the top 20 content creators or users for each platform. The results are reported in the Appendix, Table A1. These results are similar and, as expected, demonstrate an even stronger linear relationship between the natural log of rank (minus one half) and the natural log of the number of social media connections, when the users at the upper tail of the connections distribution are excluded.

## 5. Discussion

In our analysis of eight online social media networks, we found strong evidence for power law distributions in the number of social media connections. However, while we contend that our results are consistent with power law distributions, we cannot unequivocally rule out other distributions, such as the lognormal, Pareto, or double Pareto (Fang et al., 2012). In part, that is because our research is based on only the distribution of the number of connections for the top 5000 members of each social network, and analysis of a larger sample might reveal that the distribution has different characteristics for users with a smaller number of connections. Ribeiro et al. (2009; 2010) demonstrate that the distribution of MySpace friends exhibited a ‘double-Pareto’ shape, with a clear break at around 100 friends, and different power law relationships above and below this threshold. Our results are based on users with substantially more followers or subscribers than the data observed by Ribeiro et al. (2009; 2010), and therefore could be consistent with a double-Pareto distribution. Our results also demonstrate that the power law breaks down for those with the greatest number of connections. As Newman (2005) notes, few real-world distributions exhibit a power law relationship over their entire range, with most relationships breaking down below some minimum threshold. We observe this breakdown of the relationship in the upper tail of the connections distribution for all of the online social networks that we studied.

In this paper, we have not sought to explain the specific mechanism that drives the power law distributions observed in our data. That is an exercise for future research, and is all the more important given that we have demonstrated the power law relationship holds across a multitude of social media networks. However, we note that a power law distribution in a social media network could be explained by a ‘Yule process’ (Yule, 1925, Newman 2005). Consider a social network with a number of nodes (members or content creators), with each node having a number of connections to other nodes (their followers or subscribers). As new members join the network, they follow or subscribe to existing members, creating new connections. Now assume that these new connections are distributed in proportion to the number existing followers or subscribers that content creators already have – a process known as ‘preferential attachment’ (Huberman and Adamic, 1999; Adamic and Huberman, 2000). Preferential attachment can be justified as an assumption because it is likely that the content creators who were previously most popular would continue to attract a disproportionate number of followers or subscribers. This multiplicative stochastic process yields a lognormal distribution of the

number of followers over time for each content creator (Adamic and Huberman, 2002). Taking an exponentially weighted mixture (as would be observed during the growth phase of a social media network) of lognormal distributions generates a power law distribution. Beyond the growth phase of the social media network, this structure persists into the future, although with the addition of new, stochastically generated, shocks in the form of newly-popular content creators. This perpetuates the power law distribution of connections into the future. Essentially, each social network exhibits a ‘rich get richer’ process, otherwise known as the Gibrat principle (Simon, 1955), or the Matthew effect (Merton, 1968). Newman (2005) provides a mathematical exposition of the Yule process and how it relates to power law distributions across a range of possible applications (see also Reed, 2001).

Our results also highlight differences in the degree of inequality in social media connections across the different networks, with the highest degree of inequality observed for DLive, and the lowest degree for TikTok and YouTube. These differences may reflect underlying differences in the egalitarian nature of social networks (Peters et al., 2013). Egalitarianism might be promoted by the video platforms’ practice of immediately serving users with algorithmically-determined content following the conclusion of a particular video. This content need not be from a content creator that the user already follows or is subscribed to, but can stochastically generate new subscriber or follower connections, reducing inequality in the number of connections. Future research should explore these differences in connection inequality in greater detail, including how algorithmic content recommendations may reduce connection inequality.

The power law distribution of connections in social networks may lead to a number of consequences. Ebel et al. (2002) observed that the structure of email networks facilitates the distribution of email viruses. In a communications network with a power law (or ‘scale free’) distribution of interactions, a virus can persist within the network without dying out, regardless of how infectious or otherwise it is (Pastor-Satorras and Vespignani, 2001b), because the distribution of viruses through the network is strongly affected by a small number of nodes with a large number of connections (Pastor-Satorras and Vespignani, 2001a). A power law distribution in the number of connections within social networks may likewise facilitate the distribution of disinformation and misinformation (Wardle and Derakshan, 2017). In an online social network, a small number of nodes with a large number of connections can become key distributors of disinformation and misinformation. One example of this is the recent ‘infodemic’ related to coronavirus misinformation (Gabarron et al., 2021). However, the power law nature of social networks also suggests countermeasures to the spread of disinformation and misinformation, including targeting the flows through nodes with a large number of connections. This contrasts with the current approach of surveillance, and improving health literacy and knowledge translation (e.g. Eysenbach, 2020; Zarocostas, 2020). The power law distribution of social networks can also explain the persistence and marketing effectiveness of social media ‘influencers’ – content creators that have a large number of followers or subscribers (Francalanci et al., 2015; Toscani et al., 2018). The distribution of information, including marketing material and recommendations, is most effective through content creators

with a large number of followers. Online social media networks, with power law distributions of connections between content creators (including influencers) and other users, can facilitate successful marketing and information campaigns.

The analysis in this paper has a number of limitations that should be noted. First, it is based on a cross-sectional snapshot of the distribution of social media connections at one point in time. While there is no reason to believe that the observed relationship would differ significantly for snapshots taken at different points in time, it would be useful to undertake a longitudinal analysis. A longitudinal approach would also enable greater understanding of the dynamics of the distribution of social media connections, as well as greater exploration of whether the newer social media networks such as DLive, which had the highest degree of connections inequality in our analysis, are converging to a similar level of inequality to the more established networks. Second, our dataset was limited to the top 5000 users of each social media network. Extending the analysis to a larger proportion of the user base of each network would help in determining whether the distribution of connections exhibits double Pareto behaviour, as found by Ribeiro et al. (2009; 2010) for MySpace. Moreover, investigating the connections behaviour of new users to each social media platform would allow the ‘Yule process’ mechanism to be tested. Third, our analysis was limited to eight social media networks, and similar analysis could easily be extended to other networks, where similar data are available.

Despite these limitations, the analysis in this paper covers a wide cross-section of social media networks, from long-established large networks such as Facebook and YouTube, to newer and smaller networks such as DLive. That a power law distribution is observed for the connections in all of these networks suggests that this result is robust, and that further analysis of the implications of power law distributions across social media is warranted.

## References

- Adamic, L.A., and Huberman, B.A. (2000). Power-law distribution of the world wide web. *Science*, 287(5461), 2115.
- Adamic, L.A., and Huberman, B.A. (2002). Zipf’s law and the Internet. *Glottometrics*, 3, 143-150.
- Clauset, A., Shalizi, C.R., and Newman, M.E.J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4), 661-703.
- Cook, K.S. (2014). Social capital and inequality: the significance of social connections. In MacLeod, J.D., Lawler, E.J., and Schwalbe, M. (Eds.). *Handbook of the Social Psychology of Inequality*, pp. 207-227. Dordrecht: Springer.

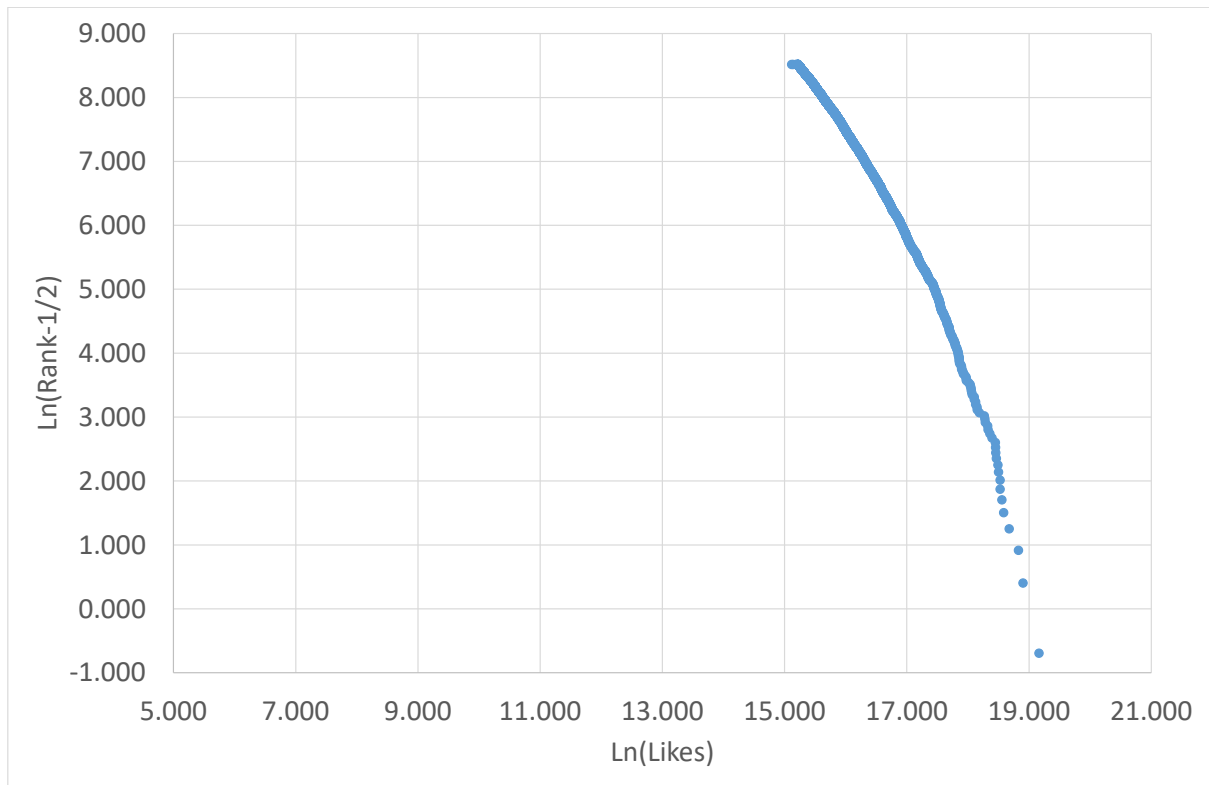
- Cox, R.A., Felton, J M., and Chung, K.H. (1995). The concentration of commercial success in popular music: an analysis of the distribution of gold records. *Journal of Cultural Economics*, 19(4), 333-340.
- Cunha, C.R., Bestavros A., and Crovella, M.E. (1995). Characteristics of WWW client-based traces, *Technical Report TR-95-010*. Boston: Computer Science Department, Boston University.
- Ebel, H., Mielsch, L.I., and Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Physical Review E*, 66(3), 035103.
- Enjolras, B. (2014). *How politicians use Twitter and does it matter? The case of Norwegian national politicians*, unpublished manuscript, Institute for Social Research, Oslo, 37 pgs.
- Eysenbach, G. (2020). How to fight an infodemic: the four pillars of infodemic management. *Journal of Medical Internet Research*, 22(6), e21820.
- Falck-Ytter, M., and Øverby, H. (2012). *An Empirical Study of Valuation and User Behavior in Social Networking Services*. Paper presented at the World Telecommunications Congress (WTC), 5-6 March. Available at IEEE: <https://ieeexplore.ieee.org/abstract/document/6170450/>, 6 pgs.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. *Computer Communication Review*, 29(4), 251-262.
- Fang, Z., Wang, J., Liu, B., and Gong, W. (2012). Double Pareto lognormal distributions in complex networks. In: Thai, M.T., and Pardalos, P.M. (Eds.). *Handbook of optimization in complex networks: Theory and application*, pp. 55–80. New York: Springer.
- Francalanci, C., Hussain, A., and Merlo, F. (2015). Representing social influencers and influence using power-law graphs. *Applied Mathematics and Information Sciences*, 9(5), 2453.
- Gabaix, X., and Ibragimov, R. (2011). Rank – 1 / 2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business and Economic Statistics*, 29(1), 24-39.
- Gabaix, X., and Landier, A. (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics*, 123(1), 49-100.
- Gabarron, E., Oyeyemi, S.O., and Wynn, R. (2021). COVID-19-related misinformation on social media: a systematic review. *Bulletin of the World Health Organization*, 99(6), 455.
- Gerbaudo, P. (2018). Social media and populism: an elective affinity? *Media, Culture and Society*, 40(5), 745-753.
- Huberman, B.A., and Adamic, L.A. (1999). Growth dynamics of the world-wide web. *Nature*, 401(6749), 131-131.

- Hurlimann, W. (2015). Benford's Law in scientific research. *International Journal of Scientific and Engineering Research*, 6(7), 143-148.
- Kidd, D., and McIntosh, K. (2016). Social media and social movements. *Sociology Compass*, 10(9), 785-794.
- Merton, R.K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*, 159(3810), 56-63.
- Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323-351.
- Oshop, R., and Foss, A. (2015). Twitter followers biased to astrological charts of celebrities. *Journal of Scientific Exploration*, 29(1), 9-34.
- Pastor-Satorras, R., and Vespignani, A. (2001a). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6), 066117.
- Pastor-Satorras, R., and Vespignani, A. (2001b). Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14), 3200.
- Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., and Pauwels, K. (2013). Social media metrics – A framework and guidelines for managing social media. *Journal of Interactive Marketing*, 27(4), 281-298.
- Postill, J. (2018). Populism and social media: a global perspective. *Media, Culture and Society*, 40(5), 754-765.
- Rastogi, T. (2014). A power law approach to estimating fake social network accounts. *arXiv: 1605.07984*, 16 pgs.
- Reed, W.J. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, 74, 15-19.
- Ribeiro, B., Gauvin, W., Liu, B., and Towsley, D. (2009). On the random nature of MySpace friendships, *UMass Technical Report UM-CS-2009-021*. Amherst, Ma.: University of Massachusetts, 8 pgs.
- Ribeiro, B., Gauvin, W., Liu, B., and Towsley, D. (2010). On MySpace account spans and double Pareto-like distribution of friends, *UMass Technical Report UM-CS-2010-001*. Amherst, Ma.: University of Massachusetts, 8 pgs.
- Roberts, D.C., and Turcotte, D.L. (1998). Fractality and self-organized criticality of wars. *Fractals*, 6(04), 351-357.
- Simon, H.A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4), 425-440.
- Subramanian, K.R. (2017). Influence of social media in interpersonal communication. *International Journal of Scientific Progress and Research*, 38(2), 70-75.
- Toscani, G., Tosin, A., and Zanella, M. (2018). Opinion modeling on social media and marketing aspects. *Physical Review E*, 98(2), 022315.

- Tsugawa, S., and Ohsaki, H. (2015). Negative messages spread rapidly and widely on social media. In *Proceedings of the 2015 ACM on conference on online social networks*, pp. 151-160.
- Wardle, C., and Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Strasbourg: Council of Europe.
- Xiong, F., Liu, Y., Zhang, Z. J., Zhu, J., and Zhang, Y. (2012). An information diffusion model based on retweeting mechanism for online social media. *Physics Letters A*, 376(30-31), 2103-2108.
- Yu, S.. and Kak, S. (2014). Social network dynamics: An attention economics perspective. In Pedrycz, W., and Chen, S. (Eds.). *Social Networks: A Framework of Computational Intelligence*, pp. 225-258. Cham, Switzerland: Springer.
- Yule, G.U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S. *Philosophical transactions of the Royal Society of London. Series B*, 213(402-410), 21-87.
- Zanette, D.H., and Manrubia, S.C. (2001). Vertical transmission of culture and the distribution of family names. *Physica A: Statistical Mechanics and its Applications*, 295(1-2), 1-8.
- Zarocostas, J. (2020). How to fight an infodemic. *The Lancet*, 395(10225), 676.
- Zipf, G.K. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, Ma.: Harvard University Press.
- Zipf, G.K. (1949). *Human Behaviour and the Principle of Least Effort*. Reading, Ma.: Addison-Wesley.

## Appendix

Figure A1. Distribution of number of Facebook likes, September 2021



**Table A1. Power law regression results, excluding the top 20 content creators**

Social media network	Ln(Rank-1/2)	Constant	Adjusted R <sup>2</sup>
Twitter	-1.478*** (0.003) [0.030]	29.82*** (0.043)	0.997
YouTube	-1.767*** (0.005) [0.035]	35.27*** (0.081)	0.993
Instagram	-1.511*** (0.003) [0.030]	31.33*** (0.049)	0.997
Twitch	-1.324*** (0.004) [0.026]	24.12*** (0.046)	0.994
DLive	-0.917*** (0.004) [0.018]	14.12*** (0.030)	0.981
TikTok	-1.917*** (0.007) [0.038]	37.34*** (0.112)	0.988
Daily Motion	-1.330*** (0.005) [0.027]	17.62*** (0.040)	0.988
Facebook page likes	-1.576*** (0.006) [0.032]	32.62*** (0.095)	0.989

N.B. Each row reports the result of a separate linear regression of  $\ln(\text{connections})$  on  $\ln(\text{Rank}-1/2)$  for the top 5000 content creators or users on that network, excluding the top 20 content creators. Robust standard errors are reported in parentheses, with revised standard errors based on Gabaix and Ibragimov (2011) in square brackets.  
\*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ .