

UNIVERSITY OF WAIKATO

**Hamilton
New Zealand**

Zipf's Law and #econtwitter

Michael P. Cameron

Working Paper in Economics 8/22

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

Abstract

This letter investigates the distribution of Twitter followers for the top 521 economists and for the top 5000 Twitter users, testing whether each distribution follows a power law. We find strong evidence for a power law, both for economists and for Twitter more generally. However, the inequality in the distribution of followers is greater among economists, which has potentially negative implications for the quality of economic debate on Twitter.

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.

1. Introduction

Zipf's Law states that the distribution of ranked data follows a power law. The original exposition of Zipf's Law described the frequency distribution of word usage (Zipf, 1932), but this empirical regularity has been observed across a range of domains in the natural and physical world, as well as in the human sphere. More recently, Zipf's Law has been shown to characterise a number of features of the internet, including the inter-domain topology (Faloutsos et al., 1999), the distribution of website links (Adamic and Huberman, 2002), and the distribution of email connections (Ebel et al., 2002). In social media, the distribution of connections (friends, followers, or subscribers) has been shown to follow a power law distribution for many of the most popular online social networks (Cameron, 2022).

In this letter, we focus on the distribution of connections (followers) for top economists on Twitter. The motivation for focusing on a subset of Twitter users is that economists on Twitter constitute an important online community, using the platform to share their ideas and discuss economic research and emerging economic issues. The Twitter hashtag #econtwitter has become a signal of membership in this community, and is also used by outsiders to attract the attention of the community of economists on Twitter. Economists on Twitter also represent a key bridge (Coleman, 2005) between the economics profession and the general public, who may not otherwise be directly exposed to the thoughts and views of economists. It is therefore important to understand how well-connected economists are on Twitter, and how equal or unequal the distribution of connections is. A power law distribution in economics Twitter might be an indicator that the 'voice' of economists is dominated by a few key figures, with many others having much less influence.

Previous studies have investigated the distribution of Twitter followers for subgroups of Twitter users, identifying power law relationships and consistency with Zipf's Law for Norwegian politicians (Enjolras, 2014), and for celebrities, politicians, and sportspeople (Rastogi, 2016). However, to date no study has looked specifically at economists. A power law relationship is not a given in all social media networks, however. Falck-Ytter and Overby (2012) found that a power law distribution is not a good description of the number of followers for the 10,020 most popular Twitter users, while Ribeiro (2009) showed that the distribution of MySpace friends exhibited a 'double-Pareto' shape. Nevertheless, more recent work has shown that the distribution of Twitter followers is consistent with a power law (Cameron, 2022). Consistent with this, we find strong evidence of power laws in the distribution of Twitter followers, both for economists and for Twitter users overall.

The remainder of this letter proceeds as follows. The next section briefly outlines the data and methods we use, and the following section presents our results. The final section concludes).

2. Data and Methods

We focus our analysis on Twitter, a ‘microblogging’ platform where users post ‘tweets’ – short text-based messages, or short audio or video clips. Users can choose to ‘follow’ other users, thereby ensuring that they always receive notification of tweets from the followed user. Thus, Twitter consists of a network of users, each of which has a number of followers. We investigate the distribution of the number of followers per user.

We use data on the number of Twitter followers for economists ranked by RePEc, collated on 27 September 2021.¹ RePEc indexes over 2000 economists on Twitter,² and records their numbers of followers. We limit our analysis to the top 521 economists by count of followers, being the top 25% of all economists on Twitter as indexed by RePEc. In our data, Paul Krugman has the most followers (over 4.6 million), followed by Xavier Sala-i-Martin (511,833), Alejandro Gaviria (382,971), and Joseph Stiglitz (365,915). For comparative purposes, we also use data on the top 5000 Twitter users overall, collated on 30 September 2021 by the social media analytics website socialblade.com. There is limited overlap between the two datasets – of the top 521 economists, only Paul Krugman (ranked 1347th) appears in the top 5000 Twitter users overall.

To investigate whether power law relationships are apparent in our data, we follow Gabaix and Landier (2008) by first illustrating the distribution of the number of social media connections on a plot of the natural log of the number of connections against the natural log of rank. To avoid small sample bias, we follow Gabaix and Ibragimov (2011) and use rank minus one half rather than the absolute value of rank. A straight line relationship in these plots would provide evidence of a power law relationship. To provide more quantitative evidence of power law relationships, we then regress the natural log of rank (minus one half) on the natural log of the number of social media connections. A statistically significant coefficient on the number of social media connections provides evidence for the power law relationship. Following Gabaix and Ibragimov (2011), we report revised standard errors alongside the usual OLS standard errors.

3. Results

Figure 1 illustrates the relationship between the natural log of rank (minus one half) and the natural log of social media connections, for economists on Twitter and for Twitter users overall. Both distributions exhibit a broadly linear shape for most of the ranking, but with a notable departure in the tail of the distribution, i.e. for Twitter users with the most followers. In addition, it is clear that in the distribution for economists, Paul Krugman (PK) is a substantial

¹ <https://ideas.repec.org/top/top.person.twitter.html>

² <https://ideas.repec.org/i/etwitter.html>

outlier. As noted in Cameron (2002), the slope of the distributions represents the degree of inequality in the number of social media connections. It is apparent from Figure 1 that there is greater inequality in the number of connections among top economists on Twitter than among top Twitter users more generally.

Figure 1. Distribution of number of Twitter followers, September 2021

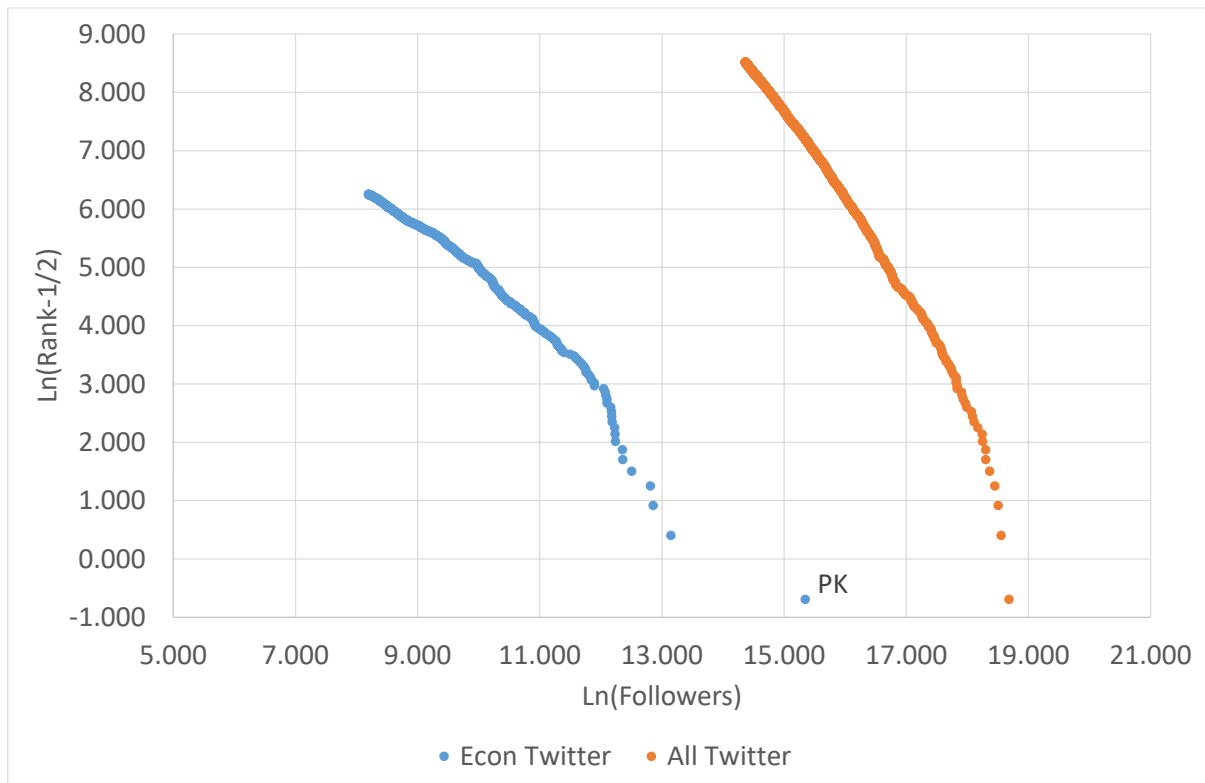


Table 1 further investigate these power law relationships, presenting the results of regression models of the natural log of the number of social media connections against the natural log of rank (minus one half). In each case, the results confirm those illustrated in Figure 1. The coefficients on the natural log of rank (minus one half) are highly statistically significant and, along with the high adjusted R-squared values, indicate a strong linear relationship between the natural log of rank (minus one half) and the natural log of the number of Twitter followers. This demonstrates that a power law relationship likely exists in these data. However, as Figure 1 indicates, the relationship breaks down at the upper tail of the distribution. In the Appendix, Table A1, we report results excluding the top 20 economists on Twitter, and excluding the top 20 Twitter users overall. The results are similar and, as expected, they demonstrate an even stronger linear relationship. Also in the Appendix, we show results for economists on Twitter, excluding Paul Krugman as an outlier. The results are robust to Krugman’s exclusion as well.

Table 1. Power law regression results

	Econ Twitter	All of Twitter
Ln(Rank-1/2)	-0.908*** (0.018) [0.018]	-1.506*** (0.007) [0.030]
Constant	13.88*** (0.163)	30.24*** (0.112)
Adjusted R ²	0.967	0.992
Sample Size	521	5000

N.B. Each column reports the result of a separate linear regression of $\ln(\text{connections})$ on $\ln(\text{Rank}-1/2)$ for the top 521 economists on Twitter, or the top 5000 Twitter users overall. Robust standard errors are reported in parentheses, with revised standard errors based on Gabaix and Ibragimov (2011) in square brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

4. Discussion

We find strong evidence for power law distributions in the number of Twitter followers, both for economists on Twitter and for Twitter users overall. This is an important finding, because it illustrates the inequality in Twitter connections between economists. When the interpretations of economics research and economic events are dominated by a small number of high profile economist ‘influencers’, valuable checks may be lost and the quality of economic debate potentially suffers. We don’t seek to call into question the quality of interpretation by the top economists on Twitter; instead, we note that there are advantages to the general public having access to a greater variety of perspectives on important issues. Fortunately though, Twitter is not the only source of these perspectives. However, mainstream media use of Twitter alongside economists could easily lead to the perpetuation of ‘echo chambers’ (Usher et al., 2018). We find that the inequality in the number of Twitter followers is greater for economists than it is for Twitter users generally (Figure 1 and Table 1), even when Paul Krugman is excluded from the analysis (Appendix, Table A1). This suggests a greater risk of these echo chambers for economics than for Twitter more generally (at least among the most popular Twitter users). However, future research should compare the inequality in Twitter followers for other subgroups of Twitter users, particularly politicians.

Another attractive opportunity for future research is to take a longitudinal perspective on the dynamics of the distribution of Twitter followers. While it appears that the RePEc list of top economists on Twitter is fairly stable over time, longitudinal data would allow researchers to uncover the mechanisms that drive the observed power law distribution (Cameron, 2022), and whether it corresponds to a ‘Yule process’ (Newman, 2005). Also, network analysis of links between economists on Twitter and other groups, such as politicians and the mainstream media, might yield valuable insights into the dissemination of economic views. We highlight these as extensions for future work.

References

- Adamic, L.A., and Huberman, B.A. (2002). Zipf's law and the Internet. *Glottometrics*, 3, 143-150.
- Cameron, M.P. (2022). *Zipf's Law across social media*, Working Paper in Economics 22/7. Hamilton: School of Accounting, Finance and Economics, University of Waikato.
- Coleman, S. (2005). New mediation and direct representation: reconceptualizing representation in the digital age. *New Media and Society*, 7(2), 177-198.
- 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.
- 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/>.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. *Computer Communication Review*, 29(4), 251-262.
- 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.
- Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323-351.
- Rastogi, T. (2014). A power law approach to estimating fake social network accounts. *arXiv: 1605.07984*.
- 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.
- Usher, N., Holcomb, J., and Littman, J. (2018). Twitter makes it worse: Political journalists, gendered echo chambers, and the amplification of gender bias. *International Journal of Press/Politics*, 23(3), 324-344.
- Zipf, G.K. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, Ma.: Harvard University Press.

Appendix

Table A1. Supplementary power law regression results

	Econ Twitter, excluding the top 20	Econ Twitter, excluding Paul Krugman	All of Twitter, excluding the top 20
Ln(Rank-1/2)	-0.818*** (0.006) [0.016]	-0.901*** (0.018) [0.018]	-1.478*** (0.003) [0.030]
Constant	13.06*** (0.052)	13.82*** (0.165)	29.82*** (0.043)
Adjusted R ²	0.988	0.966	0.997
Sample Size	501	520	4980

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