Which friends are more popular than you? Contact strength and the friendship paradox in social networks

3 authors, including:

James P. Bagrow
University of Vermont
79 PUBLICATIONS **2,693** CITATIONS

Lewis Mitchell
University of Adelaide
79 PUBLICATIONS **799** CITATIONS

Some of the authors of this publication are also working on these related projects:

Analysing Social Networks to Predict Civil-Unrest View project

Which friends are more popular than you? Contact strength and the friendship paradox in social networks

James P. Bagrow[1], Christopher M. Danforth[1] and Lewis Mitchell[2]

Abstract— The friendship paradox states that in a social network, egos tend to have lower degree than their alters, or, "your friends have more friends than you do". Most research has focused on the friendship paradox and its implications for information transmission, but treating the network as static and unweighted. Yet, people can dedicate only a finite fraction of their attention budget to each social interaction: a high-degree individual may have less time to dedicate to individual social links, forcing them to modulate the quantities of contact made to their different social ties. Here we study the friendship paradox in the context of differing contact volumes between egos and alters, finding a connection between contact volume and the strength of the friendship paradox. The most frequently contacted alters exhibit a less pronounced friendship paradox compared with the ego, whereas less-frequently contacted alters are more likely to be high degree and give rise to the paradox. We argue therefore for a more nuanced version of the friendship paradox: "your closest friends have slightly more friends than you do", and in certain networks even: "your best friend has no more friends than you do". We demonstrate that this relationship is robust, holding in both a social media and a mobile phone dataset. These results have implications for information transfer and influence in social networks, which we explore using a simple dynamical model.

INTRODUCTION

The "friendship paradox" is the observation that the friends of individuals tend to have more friends on average than the individuals themselves. This notion was first articulated for social networks [1] but it plays a role on other settings as well [2], [3], [4]. The friendship paradox stands alongside other famous features of social networks: the strength of weak ties, that people tend to gain novel information not from close friends but from weak acquaintances [5], and Dunbar's number, which argues that human's have a cognitive upper limit on the effective size of an individual's social circle [6]. Researchers have studied and validated these properties in a number of modern datasets, including Twitter [7], [8], Facebook [9], [10] and mobile phones [11], [12], [13]. Network degree relates to many personality traits such as introversion [14], and perceptual biases also factor into the effects of the friendship paradox [15]. All of these phenomena in concert provide a deeper understanding of how social networks form under cognitive and attentional limits, how information can move between individuals via the network, and to what extent the individual's embedded

view corresponds to the actual network or information within it.

The friendship paradox along with the strength of weak ties play significant roles in how one is exposed to information socially. By increasing the size of his or her social circle, an individual can gain access to a potentially overwhelming array of information. But in the modern "attention economy", limited attention and bombardment of information may dull the effectiveness of social information spread by making it more difficult to exchange information with weak ties [16]. This can be due to reduced opportunities for communication and, if those weak ties are popular individuals, limited attention to spend on any one of their social ties. Therefore, it is crucial to understand the interplay between an individual and his or her social ties, not simply from the static features of those ties but also from the dynamics and relative frequencies of those interactions. Focussing as they do on the structural properties of static, unweighted social networks, previous analyses ignore valuable information on the nature and strength of social ties.

In this work, we study two social network datasets, one derived from Twitter and one from mobile phone activity. We move beyond the classic friendship paradox analysis to consider what effects the volume of contact between egos and alters may have on the paradox. We show that there is a significant relationship between these quantities, in particular that the more heavily or frequently contacted alters tend to have fewer social ties than the less frequently contacted alters. This implies that much of the driving force behind the friendship paradox is due to these weak ties. Expanding upon this analysis, we investigate the generality of our results by studying similarities and differences between the properties of these two very different datasets and determining how those differences track with the relationship between contact volume and the number of friends of alters. For example, the outdegree of Twitter users is more heavily right-skewed than mobile phone users, yet both sets of users display the same contact volume–outdegree relationship. We apply a null model to show the significance of this relationship while accounting for data sampling effects. The friendships paradox plays a role in information diffusion over a social network alongside other features such as the strength of weak ties. We use a dynamical model of a simple information diffusion process to show that the relationship plays a significant role in slowing down the spread of information. The work we present here moves beyond static network properties and strengthens the connection between the friendship paradox, the strength of weak ties, and Dunbar's number.

[1]Mathematics & Statistics, Vermont Complex Systems Center, University of Vermont, Burlington, VT, USA {`james.bagrow,` `chris.danforth`}`@uvm.edu`

[2]School of Mathematical Sciences, University of Adelaide, Adelaide 5005, Australia `lewis.mitchell@adelaide.edu.au`

DATASETS

We focused on two social network datasets, one taken from Twitter.com social media users and one from Mobile Phone billing records. We extracted thousands of egocentric networks from each dataset, gathering the outdegrees of egos and alters, and ranking alters by the volume of contact from the ego (rank-1 is the most frequently contacted alter).

Twitter

- Here the outdegree k_{out} was the number of followers of users and alters were ranked by the number of at-mentions from the ego.

Specifically, we collected the Twitter dataset by selecting a random sample of individuals for study from the 10% Gardenhose feed collected during the first week of April 2014. From this, we uniformly sampled individuals who had tweeted in English (as reported by Twitter in the metadata for each tweet) during this time period and, to control for confounding effects of the highly right-skewed distribution of Twitter users, had $50 \leq k_{\text{out}} \leq 500$ followers, as reported in the tweet metadata. We then collected their complete public tweet history (up to 3200 most recent messages by Twitter's Public REST API limit [17], excluding retweets. These individuals formed the egos of the dataset. For each ego, we determined through their messages their top-15 most at-mentioned alters, and then proceeded to gather the tweet histories and other metadata (number of followers) of these alters using the Public REST API in the same manner. We limited ourselves to the top-15 alters because gathering the complete tweet history from the API is time-intensive. When finished, this collection process gathered 3,300 individual egos and 38,650 alters, 32,980 of which are unique across ego networks.

Sampling— To be retained as part of the Twitter data we required that all egos and alters have posted at least 10k total words in their complete, collected public tweet history. It is possible that some of the top-15 alters for a given ego do not satisfy this criterion, leading to gaps in that ego's set of alters. However, we still use all at-mentioned alters for ranking purposes, so that if, e.g., an ego with 3 alters has the rank-2 alter excluded from calculations, we still know and retain the correct rank of 3 for the third alter.

Mobile Phone

- Here k_{out} was the number of unique recipients of phone calls or text messages and alters were ranked by the number of calls and/or texts they received from the ego.

This dataset is comprised of de-identified billing records of approximately 10M subscribers to a mobile phone service provider in a western European country [18], [19], [20]. We sampled approximately 90 thousand subscribers from a nine-month period of these billing records to form the egos of the dataset. These records contain the full call histories of each subscriber, from which we can extract three time-series: (i) timestamps when phone calls were made or text messages sent, to an hourly resolution; (ii) approximate locations from where calls/texts were made, as quantified by the cellular

tower transmitting the call/text; and (iii) the recipients of these calls and texts. To ensure a complete picture of the activities of the egos, we applied the sampling criteria of [19]: selected egos must place calls from at least two distinct cellular towers during the nine-month data window, and their average call frequency must exceed 0.5 hour^{-1}. No constraints were placed on the outdegrees. For each ego, the third time series was used to reconstruct their corresponding egocentric communication network, up to the 100 most-contacted alters, with call/text recipients forming the ego's set of alters.

Sampling— The mobile phone records are limited to subscribers of a single service provider, which means we only have call histories for subscribers. Yet, subscribers will often call non-subscribers. In these data, non-subscribers are consistently identified for us, but their activities are not captured except when they interact with subscribers. This means that, while all egos in our study have full data, not all alters do, and we cannot compute the outdegree of non-subscriber alters. However, because we rank alters based on the contact volume from the ego, we still rank all alters correctly. To account for this, we only include subscriber alters within our calculations, but using their correct ranks. For example, an ego with three alters where the rank-2 alter is not a subscriber will still accurately provide two data points (ego-alter dyads) for any calculations, the rank-1 alter and the rank-3 alter.

RESULTS

The friendship paradox states that the mean degree of alters is greater than the degree of the ego for most egos in a social network. Using the data and our choices for outdegree, we found that the friendship paradox holds in both datasets. For Twitter it is highly prevalent, holding for 99.7% of egos, comparable to that found by other researchers on Twitter datasets [8]. (It also holds for 99.3% of egos when comparing the median outdegree of alters to the ego.) The friendship paradox is prevalent in the mobile phone data as well, with 76.8% of egos having lower outdegree than the mean outdegree of their alters. This percentage also matches closely with the 74% identified for one of the datasets analyzed in the original friendship paradox work [1].

We compared the outdegrees of alters as a function of how frequently the ego contacted them, as quantified by the rank of the alter, with rank-1 being the most frequently contacted alter. Ranking alters allows us to make comparisons across egos of differing overall activity levels. The mobile phone data allows us to examine a larger number of alters per ego (up to 100) whereas the Twitter data collection API effectively limited us to at most 15 alters.

Figure 1 displays the relationship between alter k_{out} and alter rank. Despite their very different natures, usage patterns, and social roles fulfilled by Twitter and mobile phone communication activity, as well as the different scopes of the data we have collected, we observed the same qualitative trend: the more frequently contacted alters tend to be of lower degree than the less frequently contacted alters. Indeed,
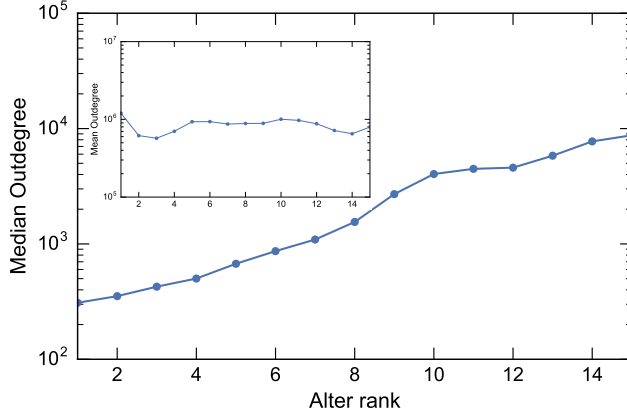
(a) Twitter
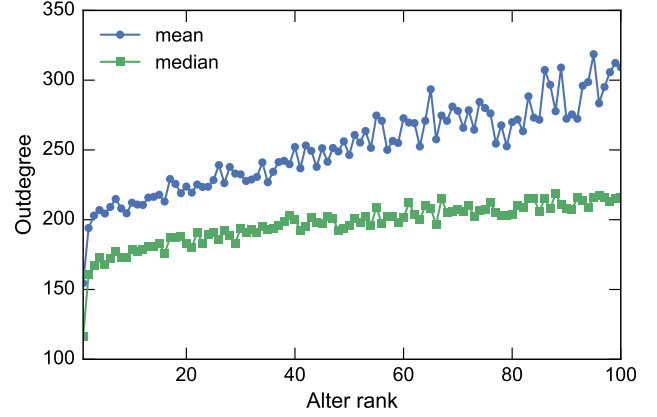
(b) Mobile Phone

Fig. 1: Your friends are more popular than you, but your most frequently contacted alters are the least popular. Rank 1 = most frequently contacted.

in the mobile phone dataset, the mean ego outdegree for the egos (166.3) is in fact slightly *greater* than that for the rank-1 alter (154.2), reversing the friendship paradox. Only 48% of egos have outdegree lower than their rank-1 alters, in stark contrast to Feld's original finding that 74% of egos have outdegree less than the mean outdegree for all of their alters. Similarly, for the mobile phones the median ego outdegree is 120, compared with 116 for the rank-1 alter. While small, this difference is significant ($p < 10^{-6}$, Wilcoxon signed-rank test), meaning in our mobile phone dataset we can reasonably claim "your best friend is no more popular than you". Therefore, we have shown that, at least in these datasets, the main drivers of the friendship paradox are not the "best friends", but the more distant contacts.

While the trend of increasing k_{out} with increasing rank holds for both Twitter and Mobile Phone, there are differences. In particular, the extent of the increase is greater for Twitter: the median k_{out} for rank-1 alters was an order of magnitude smaller than it was for rank-15 alters. In comparison, the change in k_{out} is far more gradual for mobile phone egos. (Note that Fig. 1a has a logarithmic vertical scale while Fig. 1b does not.) The main driver behind this difference is the relative skewness of the degree distributions of the two social networks (Fig. 2). While the two distributions look similar in shape overall—and both peak at $k_{out} \approx 100$, a value comparable with Dunbar's number of 150 [6]—the hubs on Twitter are far higher degree, nearly four orders of magnitude more than the mobile phones. This makes sense because of the "broadcast" nature of social media relative to a true social network: a mobile phone user simply cannot contact so many people, but online social media enables such large audiences. Because of this skew in k_{out} we focused on the median outdegree in Fig. 1a. Indeed, when considering the mean outdegree instead (Fig. 1a inset), the trend is swamped by outliers. Both mean and median measures show similar results for mobile phones (Fig. 1b).

Figure 3 shows the relationship between decreasing con-

tact volume and increasing alter outdegree directly, without ranking alters. Ranking is a useful technique here because it eliminates much of the *heterogeneity* across egos: very active egos and less heavily active egos will have very different overall contact volume, but will appear similar in "rank space". Despite this, it is important to also look at the direct relationship between contact volume and alter outdegree. We took one step in Fig. 3 to account for heterogeneity—we have binned the data by deciles of the outdegree degree of egos (which strongly correlates with the total contact volume of the ego). Due to the skew in outdegree of alters in Twitter (Fig. 2) we first log-transformed the outdegree for Twitter. We see for both datasets a significant decreasing trend. With the possible exception of the 10th decile of mobile phone egos, increases in contact volume overwhelmingly correlate with decreases in alter outdegree.

To further investigate the link between contact volume and alter outdegree, in particular to move away from some of the statistical averaging (mean or median k_{out}) applied so far, we defined the **hub alter** of an ego as the alter with the maximum k_{out} (excluding any alters whose outdegree was unavailable in the data). We investigated where the hub alter tends to fall. Is the hub alter more likely to occur at lower ranks, among the heavily contacted alters? Or at higher ranks, among the less contacted alters. To find out, we computed the proportion of ego-alter dyads for each rank r where the alter at rank r was the hub alter, as a function of r (Fig. 4) [1]. In both datasets there is a significant, increasing trend (Twitter: Spearman's $\rho = 0.8321$, $p < 0.001$; Mobile Phone: $\rho = 0.7208$, $p < 10^{-16}$), meaning that it is the less frequently contacted alters who tend to be the hub alter.

Permutation test— Is there a significant relationship between the hub alter and its rank (Fig. 4), or could it be

---

[1]For this calculation we only considered egos with at least 5 alters with an available k_{out}. Egos with few alters must necessarily have the hub alter at lower ranks, and this filter criterion prevents this dataset limitation from masking any trends.
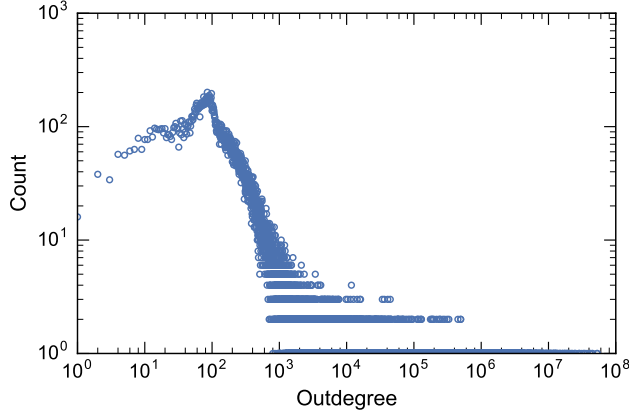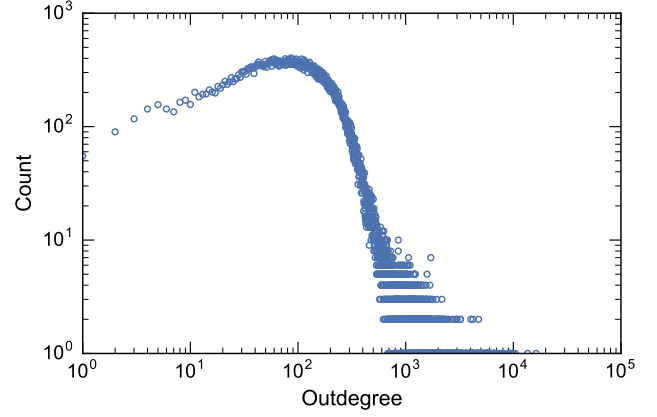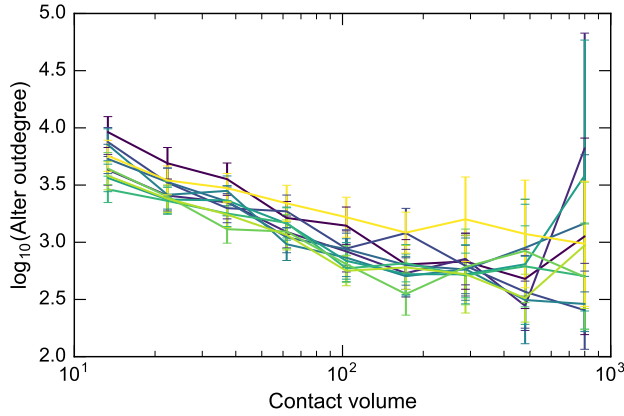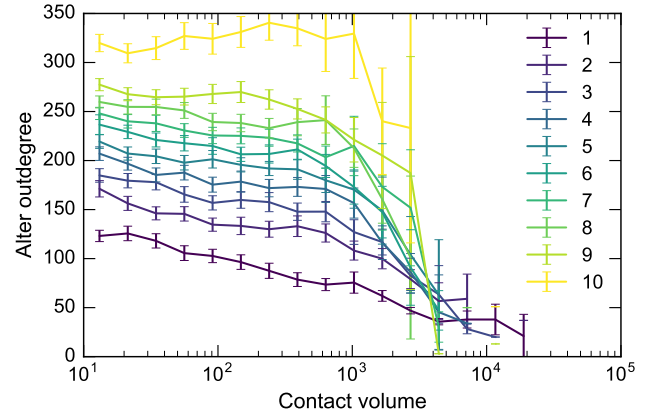
(a) Twitter   (b) Mobile Phone

Fig. 2: Degree distributions both peak near Dunbar's number of $k \approx 150$ while the Twitter degree distribution is more right-skewed than Mobile Phone.
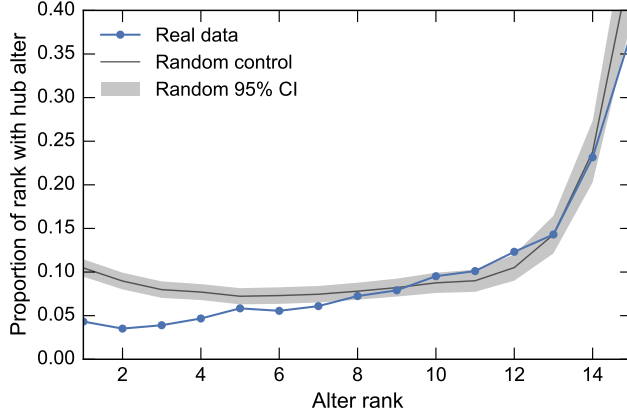

(a) Twitter   (b) Mobile Phone

Fig. 3: Direct (unranked) relationship between ego-to-alter contact volume and alter outdegree. Curves correspond to deciles of the ego outdegree distribution (colors the same in both panels).

explained by a random effect or bias in the data? To test the significance of what rank the hub alters fall at, we need an appropriate null model. We assume that, if there is no relationship, each possible rank is equally likely to be the hub alter. This is a Monte Carlo permutation model that is easy to compute (we did 1000 independent permutations for each dataset). However, there is an important caveat: some alters were excluded from our data either due to insufficient written text (Twitter) or those alters are not subscribers to the service provider (Mobile Phone); see Datasets for details. This means that we did not have their outdegrees even though for both data we ranked all the alters correctly. However, this issue is easy to bypass—since we did not use those alters we simply do not allow their ranks to be chosen randomly under the null. In other words, for an ego with n total alters, the null model samples a rank at random not from all ranks $1, \ldots, n$ but only from the $0 < n_{\text{available}} \leq n$ available

ranks for that ego. Crucially, accounting for "missing" alters makes sure that any results we observe cannot be explained by sampling effects. We simulated these null models and overlaid the results on Fig. 4 (shaded regions denote the middle 95% of random realizations not confidence intervals on the mean proportion). For both datasets, there was almost no relationship between alter rank and the proportion of hub alters under the null, except for high ranks in the Twitter data (which is explainable by sampling effects discussed in Datasets). In fact, otherwise, if anything, there was actually a weak decreasing trend in the null model for the first few (5–10) alters of both datasets, likely due to averaging over egos with different numbers of alters. In contrast, the real data displays a strong increasing trend. Further, for both datasets, the proportion of hub alters for the first few (approximately 7) ranks is significantly lower than expected from chance.

We also observed a Zipfian ranking relationship [21] be-

(a) Twitter
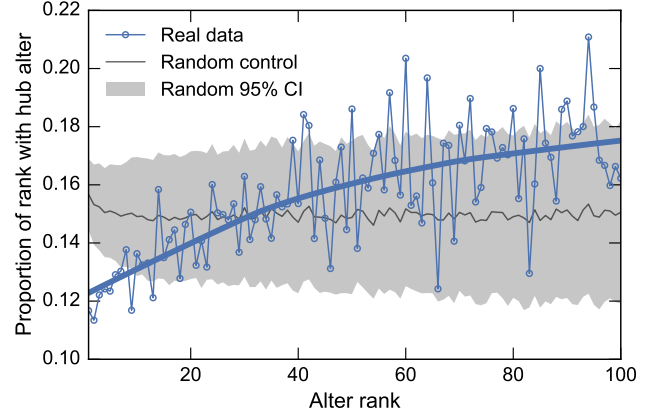
(b) Mobile Phone

Fig. 4: The proportion of ego-alter dyads where the alter is the ego's "hub alter". In both datasets, the likelihood of a hub alter increases with rank and the likelihood for a hub alter in the first ≈ 7 ranks is significantly lower than expected by chance. The guide line in (b) is a LOESS curve.

tween contact volume and alter rank. In Fig. 5 we show that both datasets obey a power law relationship with power law exponent ≈ 1.2 (although our Twitter data only provides a little over a decade of data). This scaling has previously been observed for mobile phones when ranking locations by the number of visits [18] and when ranking social ties by contact volume [20] as we did here. A similar scaling of ≈ 1.3 was also found by Frank et al. for the probability of tweeting from a location as a function of location rank [22]. Remarkably, despite their different natures, potential differences in how individuals use the respective communication services, and sampling differences in our collected data, the Twitter and Mobile Phone datasets demonstrate nearly identical scaling.

A key difference between the Twitter and Mobile Phone data is sampling heterogeneity across ego-alter dyads. To quantify this difference, we measured how many ego-alter pairs (dyads) were present in the data for each alter rank. This gives us the distribution of number of alters across egos (which may differ from ego k_{out}; see Datasets). As shown in Fig. 6, most Twitter egos have comparable numbers of available alters, at least up to around 8 alters. In contrast, the Mobile Phone egos, while potentially having up to 100 alters, tend to have fewer than ten alters and unlike Twitter this number drops off quickly. There are still at minimum hundreds (Twitter) or thousands (Mobile Phone) of dyads available at every alter rank. This points to distinct differences in the form of the social network on these two communication platforms and how the data is collected. Yet, despite these differences, we have already observed a number of similarities between Twitter and Mobile Phone in how egos relate to alters.

Information cascades

The friendship paradox, in concert with the strength of weak ties and Dunbar's number, plays a significant role in information diffusion on social networks. To explore how the

contact volume–outdegree relationship affects information transfer, we used a simple dynamical model of information transfer on the mobile phone network. We generated a network with 88,137 nodes and 8,774,126 edges from the degree distribution shown in Fig. 2b using the configuration model [23], [24] as implemented in the Python NetworkX package [25, ver. 1.11]. Starting from an initial infected node, we then ran outbreaks using a basic Susceptible-Infected model [26], [27], [28] under two different transmission regimes:

(i) *uniform transmission without contact volume dependence*: transmission probability $\beta = 0.01$ is held constant across all nodes;

(ii) *alter-dependent transmission with contact volume dependence*: each ego i has a per-alter transmission probability $\beta_i = C_i/r$ inversely proportional to alter rank r, similarly to Fig. 5.

To set equal the expected number of secondary infections from an infected node across experiments, we choose

$$C_i = \frac{n_i}{H(n_i)}\beta \quad (1)$$

where $H(n_i)$ is the harmonic number and n_i is the number of alters of ego i. Furthermore, inspired by the trends shown in Fig. 4, we defined alter rank in the model network by the outdegree of the alter, so that the highest-ranked alters were the least "famous". Recognizing that this represents a particularly strong version of the contact volume–alter rank relationship where high-ranked alters are *always* more famous, we also introduced a parameter p which controls the strength of this effect: For each potential infection during an outbreak, we infected a new node as per scenario (i) with probability $1 - p$ and otherwise as per scenario (ii) with probability p. This has the effect of adding some randomness into the contact volume-alter relationship, making the ranking less strictly enforced in the numerical experiments.
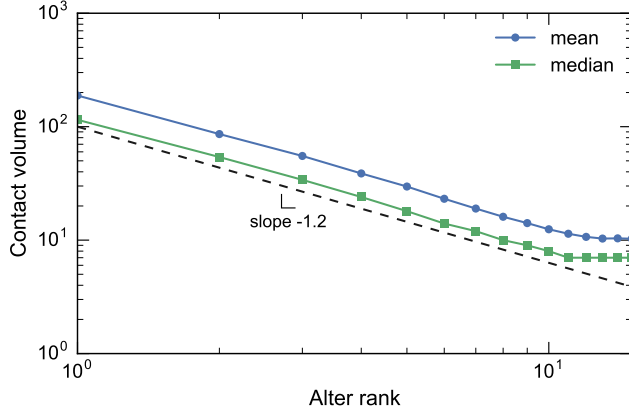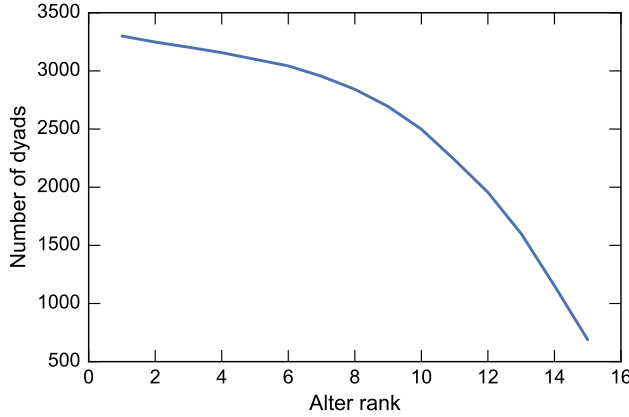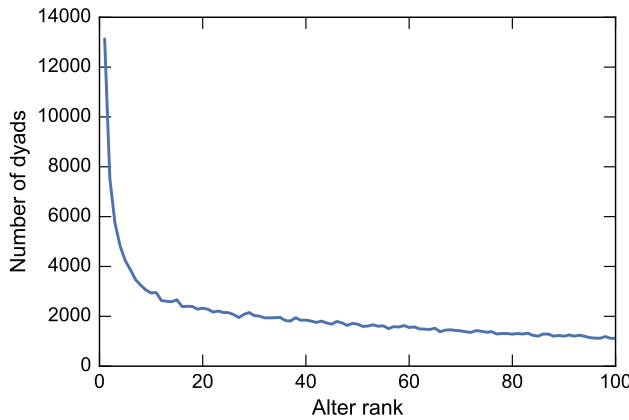
(a) Twitter
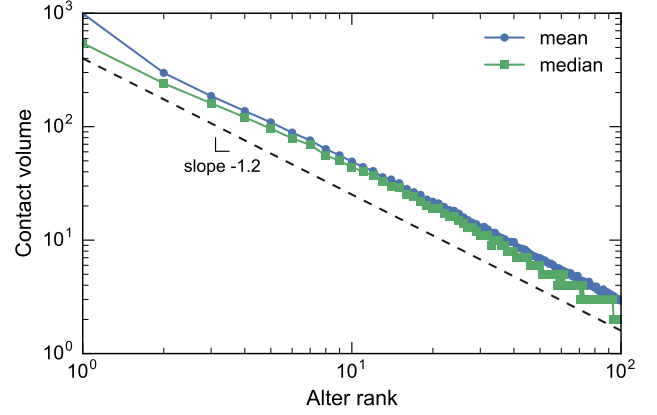
(b) Mobile Phone

Fig. 5: Contact volume (number of at-mentions, number of calls/text messages) versus alter rank. Note that the two plots have different aspect ratios and that the dashed guideline has the same slope for both: Contact volume $= C \times$ (Alter rank)$^{-1.2}$.

(a) Twitter

(b) Mobile Phone

Fig. 6: Number of ego-alter dyads across alter ranks.

Overall, scenario (i) provides a baseline and comparative control, while scenario (ii) captures rough approximations of the dynamical ingredients we have observed here.

We ran 100 independent simulations of 20 time steps of an outbreak using synchronous updating on the same network, starting from a single infected node. Figure 7 shows the results without contact volume $p = 0$ (red), with contact volume relationship $p = 1$ (blue), and for an intermediate value $p = 0.75$ (green). Error bars show 95% confidence intervals on the simulation averages. Despite the two different models having the same expected number of secondary infections from any infected node, the contact volume dependence dramatically reduces the rate of information propagation over the network, as high-degree nodes are much less likely to become infected by a low-degree node.

RELATED WORK

The key observation of the friendship paradox is that, because highly popular individuals are over-represented across the sets of alters of egos, the average degree of alters is right-skewed, leading to the colloquial "your friends have more friends than you" [1]. Perceptions can be biased because of the paradox [15] and and personality traits which correlate with network degree, such as extroversion [14] will be influenced by the paradox.

The original observations of the friendship paradox have led to a number of papers investigating it and similar paradoxes and how they affect the views of individuals within a social network. The original author showed that the variation in classroom sizes exhibits [2], while Ugander et al. showed that the friendship paradox exists at a very large scale through analyzing the anatomy of the Facebook social graph [10]. Generalizations of the friendship paradoxes have also been studied [29]. That work shows that any observable quantity which correlates with degree will be biased in the same way as degree itself. Interestingly, the friendship paradox has also been proposed as an exploitable feature of
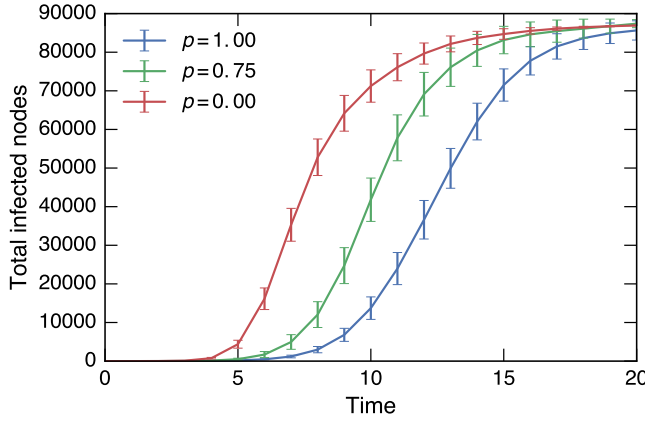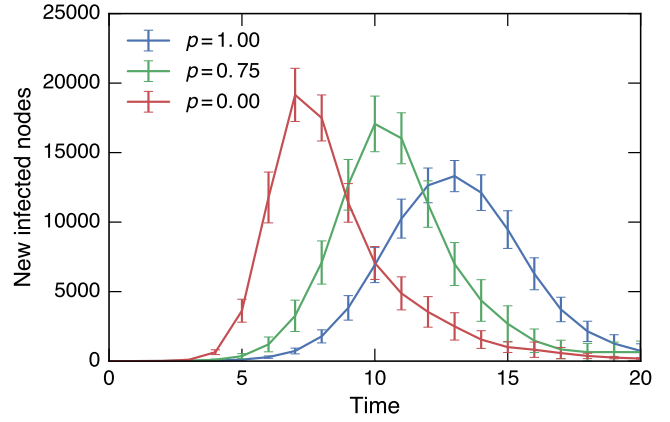
(a) Total infections   (b) New infections

Fig. 7: Epidemic curves for friendship paradox with (blue) and without (green) contact volume dependence.

networks. Christakis and Fowler used the friendship paradox to suggest a social monitoring system for early detection of influenza outbreaks [30]. They also investigated this in Twitter data for information propagation [31]. Dynamics of communication have been studied via telecommunications data by many researchers. For example, Karsai et al. and Borge-Holtheofere et al. studied bursty communication dynamics and information transfer in mobile phone and other datasets [32], [33], [34], including egocentric networks [35] but did not investigate the friendship paradox. To the best of our knowledge, research has not jointly investigated the relationships between contact volume or other individuals' contact behavior with the friendship paradox and its effects.

The work which is perhaps most similar to our own is the work of Hodas et al. [8]. They showed that, not only does the paradox hold in Twitter, it is very strong, affecting $\approx 99\%$ of egos, as we also found. In addition, they investigated several other related paradoxes, including that alters are more active (activity paradox) and that alters are exposed to more viral tweet content (virality paradox). This shows that both the structure of the Twitter follower graph and the dynamics of viral content are influenced by paradoxes such as the friendship paradox. Their work was limited to Twitter only and considered numbers of tweets and retweets, but unlike the work here did not investigate contact volume (at-mentions) between users.

Most work on the friendship paradox has been structural in nature, comparing the static degrees of egos and alters. Yet, social networks host dynamic activity patterns, and egos will maintain very different levels of contact among their social ties, due to differing degrees of social intimacy, cognitive limits on the size of a social network, and simply because of time constraints: there are only so many hours in the day. The work we present here moves beyond the static network and strengthens the connections between the friendship paradox, the strength of weak ties, and Dunbar's number.

CONCLUSIONS

We have investigated how the strength of the friendship paradox relates to contact volume between individuals in two real, social network datasets. Specifically, we have found that while "your friends are more famous than you" holds on average in the overwhelming majority of cases, this effect is in general far less pronounced for your "closest friends" or most frequently contacted social ties. Instead, more distant friends are more often the main drivers of the friendship paradox. In general, more frequently contacted alters are of lower outdegree, and conversely an ego's "hub alter", the highest outdegree alter, is more likely to be less frequently contacted by the ego. Using a conceptual Susceptible-Infected (SI) model of information transmission we demonstrated numerically that the contact volume-alter rank inverse relationship can reduce the rate of information transfer over the network. Future work will look in more detail at the content of the messages passing between egos and alters with a view towards characterizing social information flow.

The contact volume-outdegree relationship may be relevant in epidemiology, where "acquaintance immunization" [36] has been shown to be an effective strategy. If individuals actually end up immunizing their closest, or high-contact volume alters, then this immunization strategy may in fact turn out to be less effective than if they immunized more distant friends.

Online communication provides the means to vastly increase the apparent size of a user's social circle, leading to information overload from the social inputs of all the newly available alters. Yet, if egos preferentially contact the lowest degree alters as we have studied, then this information overload may be attenuated. Further investigation of information transfer under this relationship is warranted, in particular how it may relate to the formation of filter bubbles or echo chambers.

REFERENCES

[1] S. L. Feld, "Why your friends have more friends than you do," *American Journal of Sociology*, pp. 1464–1477, 1991.

[2] S. L. Feld and B. Grofman, "Variation in class size, the class size paradox, and some consequences for students," *Research in Higher Education*, vol. 6, no. 3, pp. 215–222, 1977.

[3] D. Hemenway, "Why your classes are larger than "average"," *Mathematics Magazine*, vol. 55, no. 3, pp. 162–164, 1982.

[4] I. J. Good, *Good thinking: The foundations of probability and its applications*. University of Minnesota Press, 1983.

[5] M. Granovetter, "The Strength of Weak Ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[6] R. I. Dunbar, "Coevolution of neocortical size, group size and language in humans," *Behavioral and brain sciences*, vol. 16, no. 04, pp. 681–694, 1993.

[7] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 591–600.

[8] N. O. Hodas, F. Kooti, and K. Lerman, "Friendship paradox redux: Your friends are more interesting than you," in *Proceedings of the 7th International AAAI Conference On Weblogs And Social Media (ICWSM)*, 2013.

[9] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook friends: social capital and college students use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, 2007.

[10] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *arXiv preprint arXiv:1111.4503*, 2011.

[11] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, p. 7332, 2007.

[12] J. P. Bagrow, D. Wang, and A.-L. Barabási, "Collective response of human populations to large-scale emergencies," *PLoS ONE*, vol. 6, no. 3, p. e17680, 2011.

[13] L. Gao, C. Song, Z. Gao, A.-L. Barabási, J. P. Bagrow, and D. Wang, "Quantifying information flow during emergencies," *Scientific Reports*, vol. 4, p. 3997, 2014.

[14] T. V. Pollet, S. G. Roberts, and R. I. Dunbar, "Extraverts have larger social network layers," *Journal of Individual Differences*, 2011.

[15] E. W. Zuckerman and J. T. Jost, "What makes you think you're so popular? self-evaluation maintenance and the subjective side of the" friendship paradox"," *Social Psychology Quarterly*, pp. 207–223, 2001.

[16] N. O. Hodas and K. Lerman, "How visibility and divided attention constrain social contagion," in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 2012, pp. 249–257.

[17] Twitter, "Twitter REST APIs," Available from: https://dev.twitter.com/rest/public, 2016, accessed: 2016-07-07.

[18] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[19] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[20] J. P. Bagrow and Y.-R. Lin, "Mesoscopic structure and social aspects of human mobility," *PloS one*, vol. 7, no. 5, p. e37676, 2012.

[21] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley, 1949.

[22] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Happiness and the patterns of life: A study of geolocated tweets," *Scientific Reports*, vol. 3, no. 2625, 2013.

[23] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random structures & algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.

[24] ——, "The size of the giant component of a random graph with a given degree sequence," *Combinatorics, probability and computing*, vol. 7, no. 03, pp. 295–305, 1998.

[25] D. A. Schult and P. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*, vol. 2008, 2008, pp. 11–16.

[26] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," in *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, vol. 115, no. 772. The Royal Society, 1927, pp. 700–721.

[27] R. M. Anderson and R. M. May, *Infectious diseases of humans: dynamics and control*. Wiley Online Library, 1992, vol. 28.

[28] M. E. J. Newman, *Networks: An introduction*. Oxford: Oxford University Press, 2010.

[29] Y.-H. Eom and H.-H. Jo, "Generalized friendship paradox in complex networks: The case of scientific collaboration," *Scientific Reports*, vol. 4, p. 4603, 2014.

[30] N. A. Christakis and J. H. Fowler, "Social network sensors for early detection of contagious outbreaks," *PLOS ONE*, vol. 5, no. 9, pp. 1–8, 09 2010.

[31] M. Garcia-Herranz, E. Moro, M. Cebrian, N. A. Christakis, and J. H. Fowler, "Using friends as sensors to detect global-scale contagious outbreaks," *PLOS ONE*, vol. 9, no. 4, pp. 1–7, 04 2014.

[32] M. Karsai, M. Kivelä, R. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, "Small but slow world: How network topology and burstiness slow down spreading," *Physical Review E*, vol. 83, no. 2, p. 025102, 2011.

[33] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, "Collective attention in the age of (mis) information," *Computers in Human Behavior*, vol. 51, pp. 1198–1204, 2015.

[34] J. Borge-Holthoefer, N. Perra, B. Gonçalves, S. González-Bailón, A. Arenas, Y. Moreno, and A. Vespignani, "The dynamics of information-driven coordination phenomena: A transfer entropy analysis," *Science Advances*, vol. 2, no. 4, 2016.

[35] M. Karsai, K. Kaski, and J. Kertsz, "Correlated dynamics in egocentric communication networks," *PLOS ONE*, vol. 7, no. 7, pp. 1–9, 07 2012.

[36] R. Cohen, S. Havlin, and D. ben Avraham, "Efficient immunization strategies for computer networks and populations," *Phys. Rev. Lett.*, vol. 91, p. 247901, Dec 2003.