# Gossip Spread in Social Network Models

*Tobias Johansson, Kristianstad University*

*Tobias.Johansson@hkr.se*

## Abstract

Gossip almost inevitably arises in real social networks. In this article we investigate the relationship between the number of friends of a person and limits on how far gossip about that person can spread in the network. How far gossip travels in a network depends on two sets of factors: a) factors determining gossip transmission from one person to the next and b) factors determining network topology. For a simple model where gossip is spread among people who know the victim it is known that a standard scale-free network model produces a non-monotonic relationship between number of friends and expected relative spread of gossip, a pattern that is also observed in real networks [1]. Here, we study gossip spread in two social network models [2,3] by exploring the parameter space of both models and fitting them to a real Facebook data set. Both models can produce the non-monotonic relationship of real networks more accurately than a standard scale-free model while also exhibiting more realistic variability in gossip spread. Of the two models, the one given in [3] best captures both the expected values and variability of gossip spread.

**Keywords**: Social Network; Gossip; Spread; Variability.

## 1. Introduction

Gossip is a pervasive feature of the human condition. Defined broadly as talk about social activities it accounts for about two-thirds of speaking time [4] and has been proposed to serve many functions, including cultural learning [5]; indirect aggression [6]; and social group bonding [4]. From a psychological perspective, gossip has been studied developmentally [7]; in terms of its effect on group members [8]; and as an individual differences variable [9]. However, gossip is an inherently social phenomenon [10] and thereby also depends intimately on the structure of relationships between people.

The structure of a social network and the way gossip spreads from one person to the next both imply constraints on how far gossip can spread in a social network. In this context, the relationship between the number of friends of a person and how far gossip about that person is expected to spread is not entirely obvious. Intuitively, one might expect that more friends should be associated with greater expected spread of gossip. However, based on a simple model of gossip transmission, Lind et al. [1] found expected relative spread of gossip to be a non-monotonic function of number of friends, both for a real social network dataset and for the commonly studied Barabási-Albert (BA) model [11]. For people with very few friends, the expected proportion of friends reached by gossip is high, decreasing up to a certain number of friends, from which point the expected proportion of friends reached by gossip increases. Hence,

in a sense, the optimal number of friends to reduce gossip spread is neither the minimal nor maximal number of friends, but somewhere in between.

The work of Lind et al. [1] was not conducted with the explicit purpose of modelling social networks. Indeed, the BA model is not viable as a social network model in general. Social networks typically exhibit greater levels of clustering, assortativity and community structure compared to the BA model. In the current work we set out to investigate spread of gossip in models explicitly designed to capture these distinguishing features of social networks. To this end, we simulate gossip spread in two different social network models [2,3] and in a real social network Facebook data set. We explore the parameter space of both models and we also fit the models to the Facebook data. The basic question is to what extent the social network models capture the relationship between gossip spread and number of friends. In this context, we consider not only the expected spread of gossip, but also variability in gossip spread within a network. An overall simple summary of our results is that the investigated social network models capture these features of gossip spread well and beyond that of the BA model.

A network consists of *nodes* and *edges* connecting the nodes. In the current simulations the nodes represent people and the edges represent friendship relations among people in a social network. Friendship is here treated as a symmetric relation: if A is a friend of B, then B is a friend of A. The network representing friendships is then undirected in that an edge connecting two nodes represent a symmetric relation. All friendship relations are given the same weight so the network is also unweighted. The degree $k$ of a node is given by the number of edges connected to it and the degree distribution $P(k)$ is given by the relative frequency of nodes with degree $k$ in a network. The extent of gossip spread for a person can be evaluated by the *spread factor f* [1]. Suppose there is some gossip information about a person being spread throughout the network. The quantity $f$ is then defined as the number of people reached by the gossip divided by the maximum number of people that could theoretically be reached by it. Thus, the spread factor is a relative measure as it designates a proportion.

An example of gossip spread may be helpful. Consider the network in Figure 1. At time $t = 0$ we have a person (green circle) possessing some gossip information about a target (red triangle). One of the gossip models investigated by [1] consists in gossip spreading from the initial gossiper (green circle) to friends common to the gossiper and the target. Thus, at time $t = 1$ in Figure 1, the information has spread to two additional persons. This spreading continues in the same way from the persons now possessing the gossip information, until no new persons can be informed by the gossip. At this point we ask what proportion of the target's friends have come to know the gossip and we have the spread factor for the target node starting from one particular gossiper, which turns out to be 1 at $t = 2$ in Figure 1. In order to get the overall spread factor $f$ for a target we compute spread factors starting from each friend of the target as the initial gossiper and take their average. This gives the expected proportion of friends of the target being reached by gossip starting from a random friend of the target. In the next section we evaluate gossip spread in an empirical Facebook data set and, for the sake of comparison, in the BA model.
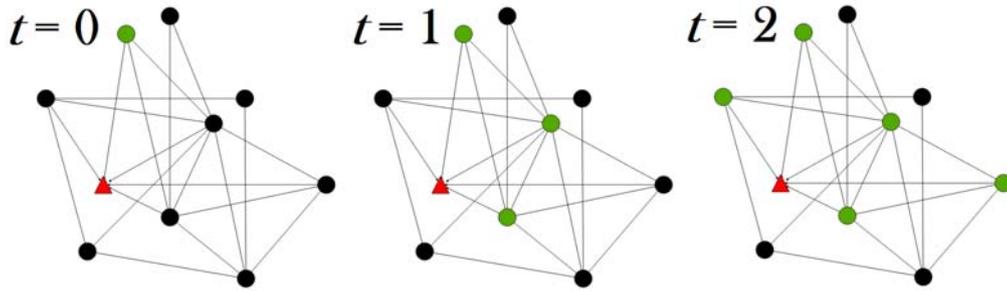
Figure 1. Spread of gossip through a network. At time $t = 0$ the initial gossiper (green circle) has gossip information about the target (red triangle). At $t = 1$ the gossip spreads to common friends of the gossiper and target. The gossip continues spreading in the same way until no new friends can be reached ($t = 2$). The fraction of target friends reached by the gossip is the spread factor for the target, given a specific initial gossiper.

## 2. Gossip Spread in a Real Network and the BA model

The real social network data set we use consists of a publicly available anonymized Facebook network data set, which was obtained by [12] by crawling the New Orleans regional Facebook network on two occasions in 2008 and 2009. The network is an undirected unweighted friendship network and we use the largest connected component amounting to $N = 60687$ nodes and $E = 690071$ edges. We simulated gossip spreading through this network the same way as described in Figure 1, so that gossip is constrained to spread only among common friends of target and gossipers. Although this constraint is not likely to be strictly true for real networks, it has been observed that sharing of friends is associated with a greater tendency for negative gossip [13].

Figure 2A shows the expected spread factor $f$ as a function of degree $k$ and the resulting non-monotonic relationship where $f$ first decreases and then increases. This relationship is very similar to that observed by Lind et al. [1] for a different data set. It can also be noted that the spread factor $f$ for individual nodes (grey circles) scatters over large ranges for different $k$, with tighter ranges as $k$ gets very small (which is trivial) or very large. The degree distribution in Figure 2B has a mean degree $\overline{k} = 23$, a median degree $\tilde{k} = 9$ and is heavily skewed but not strictly power-law.
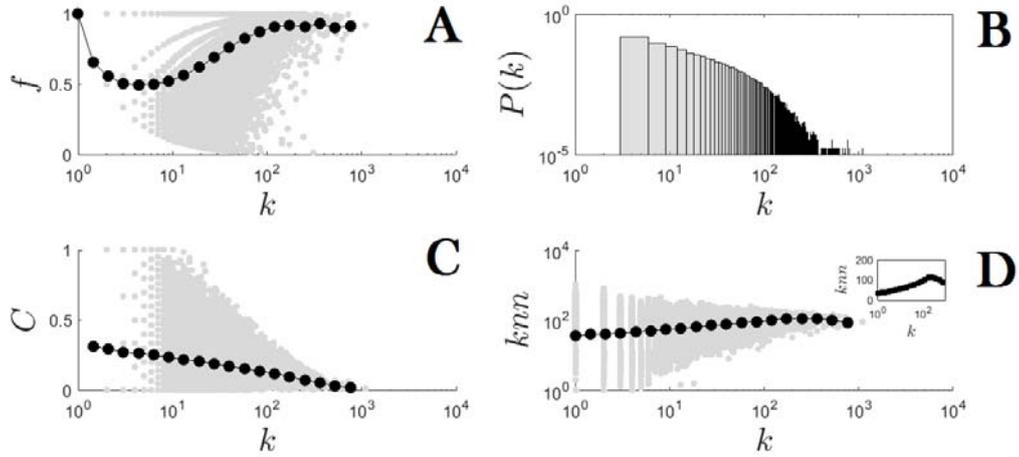
Figure 2. Results for Facebook data. A: Expected spread factor $f$ (black circles) as a function of degree $k$. B: Degree distribution. C: Expected local clustering coefficient $C$ (black circles) as a function of degree $k$. D: Expected average degree of neighboring nodes $knn$ (black circles) as a function of degree $k$. The inset in D shows semi-logarithmic plot for better resolution. Grey circles show individual nodes. Plots A, C and D use logarithmic binning for the expected values.
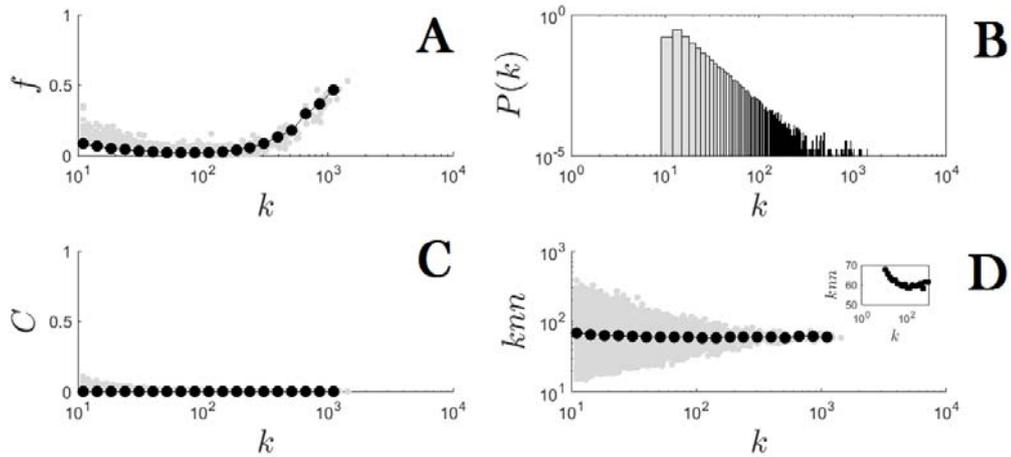


Figure 3. Results for BA model, $m = 11$. Plots are the same type as in Figure 2.

Figure 2C shows the expected local *clustering* coefficient $C$ as a function of degree $k$. The local clustering coefficient $C$ of a node ranges from 0 to 1 and quantifies the extent to which nodes connected to it are connected to each other [14]. Thus, $C$ is measure of the extent of connectivity among neighboring nodes. Clustering for the entire network can be measured by $\bar{C}$, the average of $C$ for all nodes. For this network, expected $C$ clearly decreases logarithmically with $k$ and the network as a whole shows appreciable clustering with $\bar{C} = .23$.

Figure 2D plots degree $k$ against average degree of nearest neighbors $knn$. This provides a visualization of the extent of *assortativity* with respect to degree. Assortativity refers to similarity between nodes and the nodes they are connected to, so that people with many friends are more likely to have friends with many friends. Hence, $knn$ should increase with $k$, which is the case in Figure 2D. A related quantity is the assortativity coefficient $r$ which is simply the Pearson correlation coefficient between all connected node pairs with respect to some property (Newman, 2002), in this case degree. For this network $r = .15$ revealing positive assortativity.

Lind et al. [1] explored gossip spread in the BA model and we include one such network simulation here for the sake of explicit comparison. In the BA model the network grows up to $N$ nodes, adding one new node and $m$ connections from it to $m$ existing nodes each time step. The $m$ existing nodes are selected randomly with a probability proportional to their degree $k$, a process called preferential attachment, because new nodes preferentially attach to high degree nodes. The degree distribution for the BA model is power-law $P(k) \sim k^3$ [11], expected assortativity is $r = 0$ [15] and the clustering coefficient $C$ depends only very weakly on degree $k$ [16]. These insights are visualized through simulation of a single BA network in Figure 3 along with the non-monotonic relationship between the spread factor $f$ and degree $k$. The results in Figure 3 are visualized just as in Figure 2 but for the BA model with $N = 60867$ nodes and $m = 11$. The $m$ parameter was set to approximate the average degree $\bar{k}$ of the Facebook data set.

The resulting simulated BA network has $E = 669471$ edges, a mean degree $\bar{k} = 22$, a median degree $\tilde{k} = 15$, an assortativity coefficient $r = -.01$, and a mean clustering coefficient $\bar{C} = .00$. The relationship between $f$ and $k$ (Figure 3, panel A) is qualitatively of the same non-monotonic nature as for the Facebook data, but there are clear differences as well. For the BA model, overall $f$ is low, the minimum $f$ occurs for relatively large $k$, and there is little variability in $f$ around the expected values. In the next section we consider two simple social network models that may be able to account for the data more accurately.

## 3. Two Social Network Models: $T$ and $Vaz$

The two social network models considered are introduced in [2] and [3]. We refer to these as the $T$ model and $Vaz$ model respectively after the first authors (in [3] the $Vaz$ model is denoted the connecting-nearest-neighbor model). Social networks typically display highly skewed degree distributions, relatively high clustering, positive assortativity and short average path lengths (the shortest distance between any two nodes). These features can be reproduced by both the $T$ and the $Vaz$ model.

The algorithm behind the $T$ model for generating a network with $N$ nodes is:

### Algorithm $T$

Start with an initial chain of $N_0$ nodes.

**for** $i = N_0+1, N_0+2, ..., N$

1) Randomly select a set $I = \{v_1, v_2, ..., v_{n_i}\}$ of $n_i$ existing nodes where $n_i$ $\sim F$ with expected value $\geq 1$.
2) Randomly select a set $S = \{v_1, v_2, ..., v_{m_i \cdot n_i}\}$ of $m_i * n_i$ nodes, consisting of $m_i$ neighbors to each member of $I$ where $m_i \sim G$ with expected value $\geq 0$.
3) Add a new node $v_i$.
4) Connect $v_i$ to the nodes in $I$ and $S$.

**end**

The $T$ model depends on the choice of distributions $F$ and $G$. In the current simulations we use the same distributions as in [2], as detailed further ahead. The sets $I$ and $S$ denote initial and secondary contacts respectively. Figure 4 ($T$) illustrates the algorithm behind the $T$ model in operation. At a given point in time an initial contact is selected (red). Then, a number of neighbors to the initial contact, so called secondary contacts, are selected (yellow). Finally, a new node (enclosed) is connected to the initial and secondary contacts. Attaching to secondary contacts resembles the process of getting to know friends of friends. This gives rise to implicit preferential attachment [2]: the higher the degree of a node, the more likely that node is to be a neighbor of a randomly selected initial contact. The implicit preferential attachment process is what gives rise to the high levels of clustering and also adds to the positive assortativity in the $T$ model. In terms of parameters the $Vaz$ model is simpler than the $T$ model, as the $Vaz$ model relies on a single probability parameter $u$, aside from the network size $N$. The algorithm behind the $Vaz$ model is:

### Algorithm $Vaz$

Start with a single isolated node (current network size = $n$), $n = 1$.

**while** $n < N$

*With probability 1-u*

1) Add a new node $v_{n+1}$ and connect to a randomly selected existing node $v_i$.
2) Add *potential* edges between $v_{n+1}$ and the neighbors of $v_i$ to the set P.
3) Update current network size: $n = n+1$.

*With probability u*

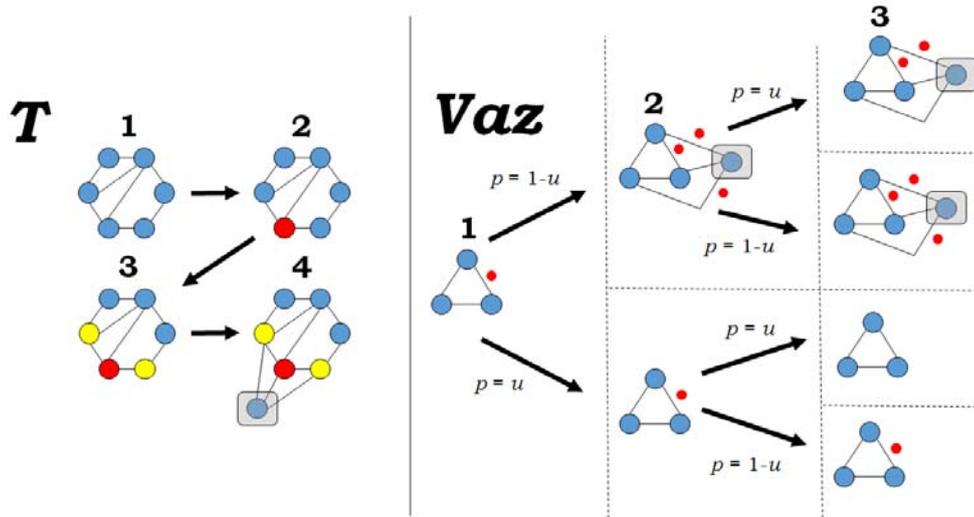4) Convert an edge in P to an *actual* edge.

**end**

**Figure 4.** Illustration of the growth processes in the $T$ (left panel) and $Vaz$ (right panel) models.

Figure 4 ($Vaz$) illustrates the algorithm behind the $Vaz$ model in operation as the network evolves along different possible trajectories. Edges with a red dot signify potential edges, i.e. edges that are not currently included in the network but could be converted to form actual edges. Step 1 is the network at some point in time. In step 2, with probability 1-$u$, a new node is added (enclosed) and connected to a randomly selected node $v$. Potential edges are then formed between the neighbors of $v$ and the new node. In step 3, with probability $u$, one potential edge is converted to an actual edge. These processes take place repeatedly until the network has grown to size $N$. Like the $T$ model, the $Vaz$ model incorporates a form of implicit preferential attachment as well: the higher the degree of a node, the more likely that node is to be at the end of a potential edge. This leads to both high levels of clustering and positive assortativity in the $Vaz$ model.

Although the $T$ and $Vaz$ models incorporate similar mechanisms, they are not identical. The statistical features of the networks they generate are often comparable, but clear differences can be found as well. For example, in a comparison of several different kinds of network models, the $Vaz$ model generated far too many large $k$-cliques (fully connected subgraphs of size $k$) compared to real data, while the $T$ model generated too few [17]. We now turn to how gossip spreads in these two models.

## 4. Gossip Spread in the $T$ and $Vaz$ Models

In order to evaluate gossip spread in the $T$ and $Vaz$ models we first report the results of simulations exploring the parameter space of both models. The gossip transmission model is the same as before, so that gossip spreads among common friends of victim and gossiper. For efficiency reasons the network size parameter $N$ is fixed to 1000 in all simulations. By all indications, larger $N$ gives very similar functional relationships

for the quantities assessed here. In Section 5 we fit the models to data and arrive at model estimates for the same $N$ as the Facebook data ($N$ = 60867).

For the $T$ model we use the same distributions for the number of initial and secondary contacts as in [2]. For the initial contacts, 1 contact is selected with probability $p$ and 2 contacts with probability 1-$p$. The number of secondary contacts follows a uniform distribution $Unif$ (0, $lim$) where $lim$ is the upper limit. In the simulations we explore the 2-dimensional parameter space of $p$ and $lim$, with $p$ = [.05, .5, .95] and $lim$ = [1, 3, 5]. This is a coarse exploration but still provides enough variability in the model parameters to assess model flexibility qualitatively. The parameter space also roughly encompasses that of previous simulations using the $T$ model [17]. The size $N_0$ of the seed network chain was set to 20.

When fitting the $Vaz$ model as specified in [3] to the Facebook data (see Section 5) we observed a poor combined fit to mean degree $\bar{k}$ and mean clustering $\bar{C}$. We therefore implement the $Vaz$ model so that 1) $m$ randomly selected existing nodes are connected to the same new node with probability 1-$u$ and 2) $m$ potential edges are converted to actual edges with probability $u$ each iteration. We explore the resulting 2-dimensional parameter space of $u$ and $m$ coarsely, with $u$ = [.2, .5, .8] and $m$ = [1, 3, 5]. Each parameter combination involves 100 simulated random networks.

Figure 5 shows the results for the $T$ and $Vaz$ models. Note that for the $T$ model, increasing $lim$ and decreasing $p$ produces less randomly connected and more clustered networks. The same occurs for the $Vaz$ model when decreasing $m$ and increasing $u$. The way these two models are parameterized they have the same of number of free parameters ($p$ and $lim$ in the $T$ model vs. $u$ and $m$ in the $Vaz$ model). Nevertheless, the parameters govern the two models in different ways so that the $T$ model is more flexible in its predictions. Comparing visually against the corresponding results for the Facebook data in Figure 2A, both models are able to capture the basic non-monotonic relationship between $f$ and $k$, but the $Vaz$ model does so more accurately and consistently across different parameter values. Both models show considerable variability in the spread factor $f$ around the expected values, with the $T$ model producing most variability and the $Vaz$ model producing quite sharp decreases in variability as $k$ increases along with $f$, at least for high values of $u$.
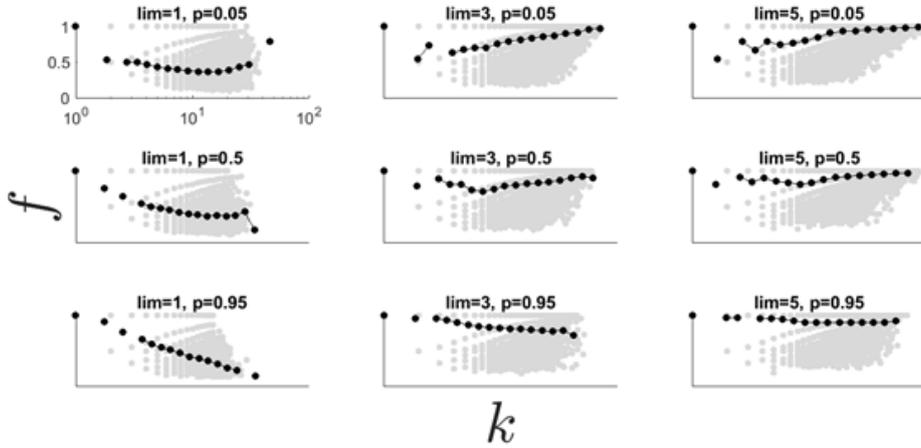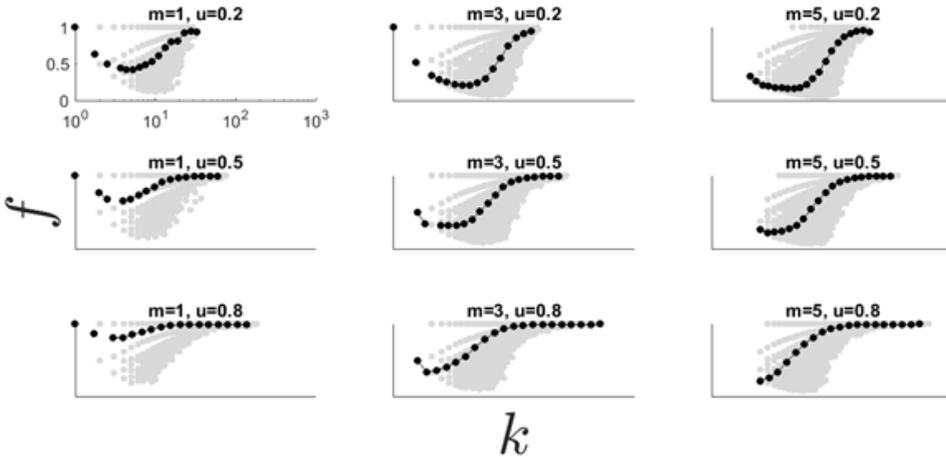
Figure 5. Gossip spread in the $T$ (upper) and $Vaz$ (lower) models: Expected spread factor $f$ (black circles) as a function of degree $k$ for different parameter values of $p$ and $lim$ for the $T$ model and of $u$ and $m$ for the $Vaz$ model. Each plot is based on 100 simulations. The axes are scaled the same within each model.

## 5. Fitting the Models to Data

In order to fit the models to the Facebook data set we use a similar type of fitting procedure as described in [17]. For the $T$ model, with parameters $p$ and $lim$, we minimize the norm of the weighted error function $f(p) = [w_k \varepsilon_k, w_c \varepsilon_c]$, in order to find the optimal value of $p$ separately for each of the integer values of $lim$ from 1 to 15. The error $\varepsilon_{q(p)}$ with respect to quantity $q$ and parameter $p$ is given by $\varepsilon_{q(p)} = |(q(p) - q_{target}) / q_{target}|$. That is, for each of the integer values of $lim$ from 1 to 15, we minimize the norm of the error function with respect to both mean degree $\overline{k}$ and mean clustering $\overline{C}$

to find the optimal value of $p$. We then select the ($p$, $lim$) pair with smallest norm as the optimal parameter pair for the $T$ model.

For the $Vaz$ model we use the same optimization routine to find the values of $u$ and $m$ that best match the mean degree $\bar{k}$ and mean clustering $\bar{C}$ in the data. The only difference is that we minimize the norm of the weighted error function $f(u)$ to find the optimal value of $u$ separately for each of the integer values of $m$ from 1 to 5. In [17] the optimal value of $u$ in the $Vaz$ model was obtained analytically. Our current implementation of the $Vaz$ model with the $m$ parameter differs from the original model in [3]. Hence, the analytical results in [3] do not apply to the current model and we use an optimization routine instead.

We set the weight parameters $[w_k, w_{\bar{c}}]$ manually through trial and error. As in [17] we put most weight on matching mean degree $\bar{k}$, ultimately settling for weight parameters $w_k = 4$ and $w_{\bar{c}} = 1$. The exact values turn out to have close to no impact on model parameter estimates, as long as $w_k > 2 w_{\bar{c}}$, and most definitely no relevant impact on network statistics nor the spread factor, as long as $w_k > w_{\bar{c}}$.

The optimization was implemented with the $fminbnd$ function in Matlab [18]. Optimizing model parameters with network size $N = 60867$, which is the size of the Facebook data, is very computationally inefficient given our model algorithms. Instead, we start by optimizing the parameters for $N = 1000$ and then manually refine these estimates for $N = 60867$.

**Table 1. Statistical Quantities and Optimal Model Parameter Values.**

|  | $\bar{k}$ | $\bar{C}$ | $r$ | $n_{edges}10^5$ | $p$, $lim$ | $u$, $m$ |
|---|---|---|---|---|---|---|
| *Facebook* | 22.67 | .23 | .15 | 6.90 | - | - |
| *T* model | 22.73 | .26 | .10 | 6.92 | .14, 11 | - |
| *Vaz* model | 22.52 | .18 | -.03 | 6.85 | - | .83, 2 |

Table 1 shows the estimated optimal model parameter values along with relevant observed quantities. The fitted models show decent agreement with the data, the only exception being that the *Vaz* model produced an assortativity coefficient $r$ close to zero. We now turn to the question of how gossip spreads through the networks implied by these models with the model parameters in Table 1. We know from Figure 5 that both models *can* produce the typical relationship between spread factor and degree. Now we want to know if the model parameters that best fit the mean degree and mean clustering of the Facebook data produce the same relationship. These results are shown in Figure 6, in which the red points represent the Facebook data. It is quite clear from the plots in Figure 6 that the *Vaz* model better captures the expected values of $f$ and $C$ in relation to $k$, whereas the $T$ model better captures the expected values of $knn$ in relation to $k$. Both models exhibit the typical non-monotonic relationship between $f$ and $k$, but the *Vaz* model clearly better captures the trend in the Facebook data.
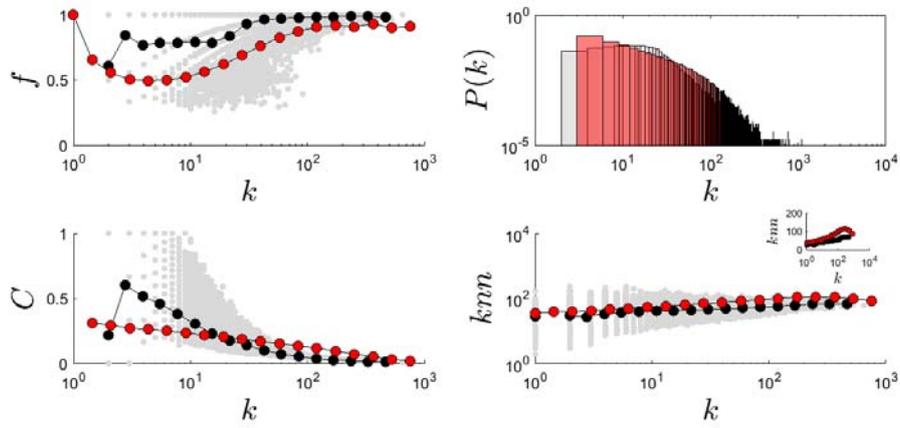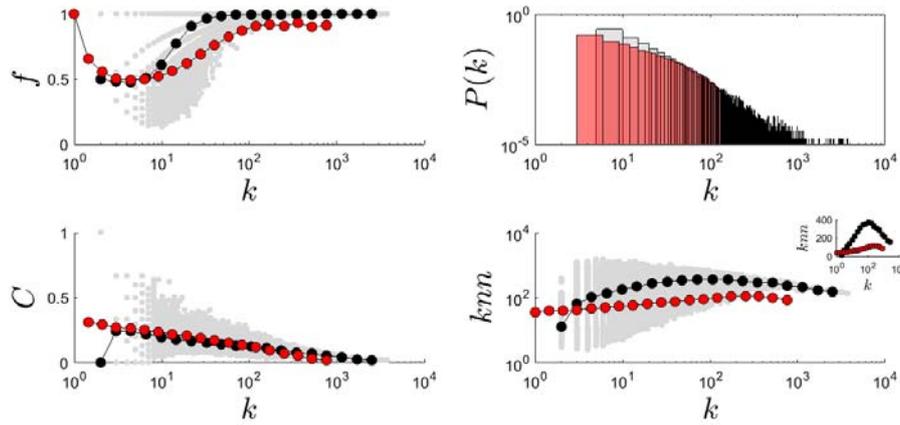
## $T$ model



## $Vaz$ model



Figure 6. Results for the fitted $T$ and $Vaz$ models based on the model parameters in Table 1. Plots are the same type as in Figure 2.
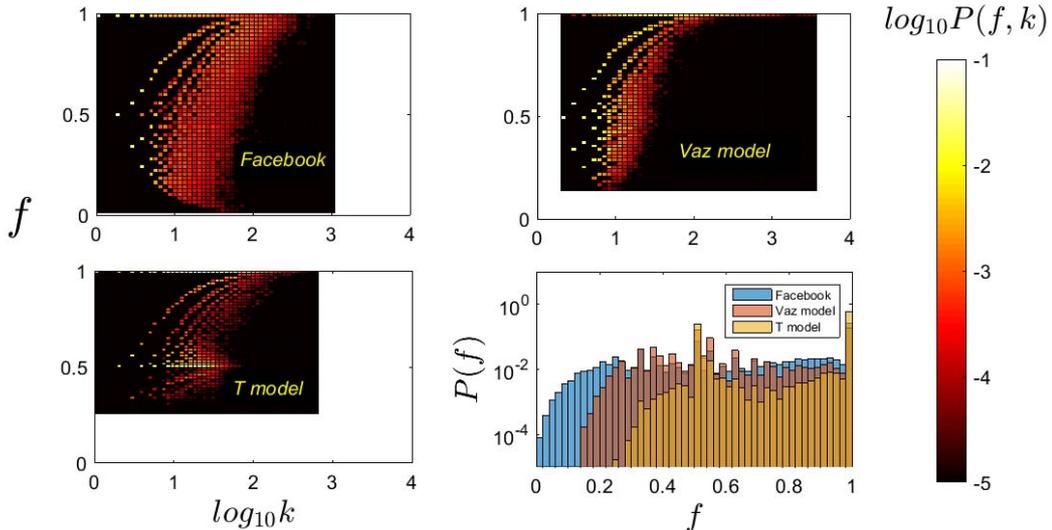
Figure 7. Lower right panel: Data and Model distributions of spread factor $f$. Lower left and upper panels: Color coded joint distributions of $f$ and $k$. The data stretch in each dimension as indicated by the black boundaries. This area is divided into a $50^*50$ matrix and the joint probability of $f$ and $k$ is colored according to the color bar.

As indicated in Figure 2, the Facebook data exhibit considerable spread factor variability. The probability distributions of the Facebook and model spread factors are shown in the lower right panel of Figure 7. Both models exhibit spread factor variability, but the *Vaz* model is clearly closer to the Facebook data. The other panels in Figure 7 provide visualizations of the joint probability distributions of $f$ and $k$, i.e. $P(f,k)$, where the joint probability is color coded. Both models show variability along the lines of the Facebook data, but the *Vaz* model is relatively compressed along the $k$-dimension in the two-dimensional probability space while the $T$ model is compressed along the $f$-dimension. This can be contrasted to the BA model which, rather unsurprisingly of course, shows almost no variability in this space (Figure 3A).

In order to compare the similarity between each of the joint model distributions (Figure 7, lower left and upper right) and the joint Facebook distribution (Figure 7, upper left) we computed the Jensen-Shannon divergence for these distributions. The Jensen-Shannon divergence $JSD(F||M)$ between distributions $F$ (in this case Facebook) and $M$ (in this case one of the models) is a symmetric smoothed version of the Kullback-Leibler distance, bounded between 0 and 1, where smaller values indicate less divergence. In order to avoid zero probabilities we added the smallest positive floating point number in Matlab to the zero entries. We used unbinned data along the $k$-dimension and linear binning in steps of .01 in the $f$-dimension. The results turn out very similar with other choices as well, e.g. with logarithmic binning of $k$. Of the two models, $T$ and *Vaz*, the *Vaz* model is closer to the Facebook data, $JSD(F||Vaz) = .35$ and $JSD(F||T) = .60$.

# 6. Discussion

Our aim in this article was to investigate the degree dependence of gossip spread in two social network models, the $T$ and $Vaz$ models. Both of these models were able to capture the basic non-monotonic relationship between degree and gossip spread when fitted to real data, although the $Vaz$ model was closer to the real data and also predicted the basic non-monotonic pattern more consistently across the model parameter space. The two models captured different aspects of the variability in the joint probability space of the spread factor and degree, but the $Vaz$ model captured more variability in total, both for the joint probability of spread factor and degree and for the spread factor per se.

Although the $Vaz$ model performed better overall than the $T$ model in this study, one should not make too much of this fact. In the current simulations we aimed to stay true to the original formulations of these two models as far as possible. However, the parameterization of these models is not obvious. For example, the number of secondary contacts per node is uniformly distributed with a lower bound of 0 in the $T$ model, but it is unclear why the number of secondary contacts has to be constrained exactly this way. Furthermore, we could make the $Vaz$ model more similar to the $T$ model by replacing the $m$ parameter in the $Vaz$ model with a uniform distribution. Thus, the extent to which differences in model behavior stem from differences in mechanisms vs. differences in distributional assumptions is not obvious.

Even though the social network models investigated here capture many of the distinguishing features of social networks, they are still limited. For example, real network components do not remain static once formed, but can change. Models of changing networks exist [19,20] and could be investigated with respect to gossip spread. Indeed, one of the factors affecting the evolution of such networks may be gossip itself [21]. Gossip may also be modelled as a probabilistic variable, as in [1], where each individual had the same fixed probability of gossiping. Such probabilities may themselves be distributed and time-dependent. For example, the average probability of gossiping could be modelled as decreasing at some rate with each time step of gossip spreading through the neighborhood of a gossiper.

Exact or approximate expressions relating the spread factor to other network quantities are difficult to obtain [22]. Clearly and by definition, the spread factor depends on neighborhood network connectivity of victim and gossiper. However, as noted in [1] the spread factor is different from the clustering coefficient, as the former depends on *how* the neighborhood is connected, not just how dense it is. Assuming that gossip spreads only among common friends, a high spread factor can come about in quite different ways, depending on network topology.

For example, in a network with a uniform degree distribution and very strong community structure friends will have common friends to a large extent. In the extreme case, we can picture such a network as consisting of equally sized communities where everyone knows each other within a community and no one knows anyone across communities, so that the spread factor is $f = 1$. On the other hand, a high spread factor can also come about in a network with close to no community structure if the network contains hubs, i.e. rare nodes of extremely high degree. In the extreme case,

consider a network where everyone except person P has two friends of which one is random and one is person P. In such a network the spread factor $f = 1$, because all gossip would go to or through P who can reach or be reached by anyone. The centrality of nodes in a network with respect to spreading is classical territory in network research. Typically, such scenarios and analyses involve epidemic spreading or information transmission to all directly connected nodes, where centrality is quantified in terms of, for example, betweenness-centrality [23], degree [24], or $k$-core decomposition [25]. Future work may want to identify central nodes in the context of gossip spread involving various constraints on gossip spread. Such constraints may include the restriction of gossip to spread only among common friends.

# Acknowledgements

# References

1 Lind, P.G. *et al.* (2007) Spreading gossip in social networks. *Phys. Rev. E* 76,

2 Toivonen, R. *et al.* (2006) A model for social networks. *Phys. Stat. Mech. Its Appl.* 371, 851–860

3 Vázquez, A. (2003) Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E* 67,

4 Dunbar, R.I.M. (2004) Gossip in evolutionary perspective. *Rev. Gen. Psychol.* 8, 100–110

5 Baumeister, R.F. *et al.* (2004) Gossip as cultural learning. *Rev. Gen. Psychol.* 8, 111–121

6 Archer, J. and Coyne, S.M. (2005) An integrated review of indirect, relational, and social aggression. *Personal. Soc. Psychol. Rev.* 9, 212–230

7 Björkqvist, K. *et al.* (1992) Do girls manipulate and boys fight? developmental trends in regard to direct and indirect aggression. *Aggress. Behav.* 18, 117–127

8 Feinberg, M. *et al.* (2014) Gossip and Ostracism Promote Cooperation in Groups. *Psychol. Sci.* 25, 656–664

9 Lyons, M.T. and Hughes, S. (2015) Malicious mouths? The Dark Triad and motivations for gossip. *Personal. Individ. Differ.* 78, 1–4

10 Rosnow, R.L. (2001) Rumor and gossip in interpersonal interaction and beyond: A social exchange perspective. In *Behaving badly: Aversive behaviours in interpersonal relationships.* (Kowalski, R. M., ed), pp. 203–232, American Psychological Association.

11 Barabási, A.-L. and Albert, R. (1999) Emergence of Scaling in Random Networks. *Science* 286, 509–512

12 Viswanath, B. *et al.* (2009) , On the evolution of user interaction in Facebook. , presented at the Proceedings of the 2nd ACM workshop on Online social networks, pp. 37–42

13   Grosser, T.J. *et al.* (2010) A Social Network Analysis of Positive and Negative Gossip in Organizational Life. *Group Organ. Manag.* 35, 177–212

14   Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of "small-world" networks. *Nature* 393, 440–442

15   Newman, M.E.J. (2002) Assortative Mixing in Networks. *Phys. Rev. Lett.* 89,

16   Fronczak, A. *et al.* (2003) Mean-field theory for clustering coefficients in Barabási-Albert networks. *Phys. Rev. E* 68,

17   Toivonen, R. *et al.* (2009) A comparative study of social network models: Network evolution models and nodal attribute models. *Soc. Netw.* 31, 240–254

18   (2015) *Matlab*, The Mathworks, Inc.

19   Deijfen, M. and Lindholm, M. (2009) Growing networks with preferential deletion and addition of edges. *Phys. Stat. Mech. Its Appl.* 388, 4297–4303

20   Jin, E.M. *et al.* (2001) Structure of growing social networks. *Phys. Rev. E* 64,

21   Shaw, A.K. *et al.* (2011) The effect of gossip on social networks. *Complexity* 16, 39–47

22   Lind, P.G. and Herrmann, H.J. (2007) New approaches to model and study social networks. *New J. Phys.* 9, 228–228

23   Freeman, L.C. (1978) Centrality in social networks conceptual clarification. *Soc. Netw.* 1, 215–239

24   Albert, R. *et al.* (2000) Error and attack tolerance of complex networks. *Nature* 406, 378–382

25   Kitsak, M. *et al.* (2010) Identification of influential spreaders in complex networks. *Nat. Phys.* 6, 888–893