

The spread of gossip in American schools

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2007 EPL 78 68005

(<http://iopscience.iop.org/0295-5075/78/6/68005>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 144.173.6.37

This content was downloaded on 16/08/2015 at 13:25

Please note that [terms and conditions apply](#).

The spread of gossip in American schools

P. G. LIND^{1,2}, L. R. DA SILVA³, J. S. ANDRADE jr.⁴ and H. J. HERRMANN^{4,5}

¹ *Institute for Computer Physics, University of Stuttgart - 70569 Stuttgart, Germany*

² *Centro de Física Teórica e Computacional - Avenida Professor Gama Pinto 2, 1649-003 Lisbon, Portugal*

³ *Departamento de Física, Universidade Federal do Rio Grande do Norte - 59072-970 Natal, Rio Grande do Norte, Brazil*

⁴ *Departamento de Física, Universidade Federal do Ceará - 60451-970 Fortaleza, Ceará, Brazil*

⁵ *IfB, HIF E12, ETH Höggerberg - CH-8093 Zürich, Switzerland*

received 5 December 2006; accepted in final form 8 May 2007

published online 4 June 2007

PACS 89.75.Hc – Networks and genealogical trees

PACS 89.65.Ef – Social organizations; anthropology

PACS 87.23.Ge – Dynamics of social systems

Abstract – Gossip is defined as a rumor which specifically targets one individual and essentially only propagates within its friendship connections. How fast and how far a gossip can spread is for the first time assessed quantitatively in this study. For that purpose we introduce the “spread factor” and study it on empirical networks of school friendships as well as on various models for social connections. We discover that there exists an ideal number of friendship connections an individual should have to minimize the danger of gossip propagation.

Copyright © EPLA, 2007

Introduction. – Gossip is so inherent to human nature that it was already worshiped in the Greek mythology through the many-tongued PHEME. Its impact in history and in sociology is so large that it has been studied from many points of view [1,2]. However, recent insights into the mathematical properties of social networks [3,4] and particularly those involving friendships [5], open up a new way of understanding how the propagation of gossip depends on connections of friendships. We will in particular consider gossip within the society of American school students.

In the last years many social network models have been studied serving to describe phenomena ranging from the Internet, epidemics and rumor spreading to scientific collaborations, movie actors and sexual intercourse [3–8]. Many properties have been identified to characterize these networks like the degree distribution, the shortest path, the cliquishness, the inbetweenness, etc. The study of gossip propagation, however, requires still another up to now never considered analysis, giving rise to what we call the “spread factor” and the “spreading time” which we will introduce in this letter and discuss their properties and applications. Using these properties we will present a striking finding: in real social systems there is the possibility to minimize the risk of being gossiped, by choosing an optimal number of friendship connections.

A simple model for gossip propagation. – As opposed to rumors, a gossip targets the behavior or private life of a specific person, *i.e.*, a target node (victim) in the network. To model the social network we consider that individuals are vertices connected by bonds representing their friendship connections as illustrated in fig. 1. At time $t = 0$ a gossip (be it truth or falsehood) is created about the victim by the “originator” (grey face). In the normal case the gossip is only of interest to those who know the victim personally and we therefore consider that it only spreads at each time step from the vertices that know the gossip to all vertices that are connected to the victim and do not yet know the gossip, *i.e.* over the arrows in fig. 1. Our dynamics is therefore similar to an epidemics or a burning algorithm [9], starting at the originator but limited to sites that are neighbors of the victim. The gossip spreads until attaining all reachable connections of the victim, *i.e.* after three time steps in fig. 1.

To measure how effectively the gossip attains the friends of the victim, we calculate the total number n_f of friends (nearest neighbors) who eventually hear the gossip in a network with N vertices (individuals), and the minimum time τ it takes to attain the accessible friends, which we call the “spreading time”. We define the “spread factor” as the fraction of attained nearest

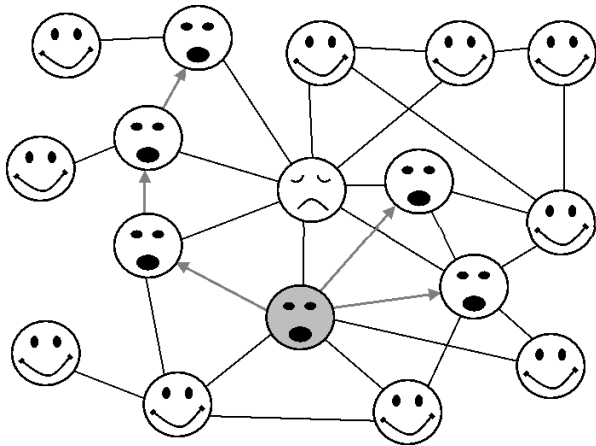


Fig. 1: Spreading of a gossip about a victim shown having an unhappy face. If the gossip starts from the gray face, in three time steps six neighbors (faces with open mouth) will know it, giving $f = 3/4$ and $\tau = 3$. The gossip spreads over the arrows. The clustering coefficient of the victim is $C = 3/14$ illustrating the significant difference between C and the spreading factor f (see text).

neighbors, $f = n_f/k$, where k is the degree of the victim. The range of this quantity f goes from $1/k$ (when only the originator knows the gossip and cannot spread it further giving $\tau = 0$), to 1, when all friends of the victim get to know the gossip. It is important to note that f and the clustering coefficient C defined in refs. [10,11] are similar but different because the latter only measures the number of bonds between neighbors and contains no information about the way they are distributed. In particular, a large spread factor indicates the existence of long paths connecting the neighbors of the victim, as discussed below.

The quantities τ and f are averages over all neighbors of the victim, *i.e.* over all victim-originator pairs. Since gossip is very relevant among children or students at school we will focus on studying these quantities on an empirical friendship network obtained by interviewing 90118 pupils from 84 schools [12]. In this network friendship connections were precisely defined by the evaluation of questionnaires [5,12], where students named the individuals inside their school with whom they had some personal contact, *e.g.* having lunch, talking on the phone or studying together. In this way, the friendship connections considered in these networks imply a somehow deep knowledge between linked students, different from the more general concept of “acquaintance” which could be interpreted as connections to someone in a more superficial manner. After addressing such empirical networks, we then compare these real network data to those obtained in scale-free networks, namely in the Barabási-Albert network [13] and in the Apollonian network (APL) [14], where the same features are also observed.

Comparative study of gossip propagation in real social networks. – In fig. 2a we plot the results for

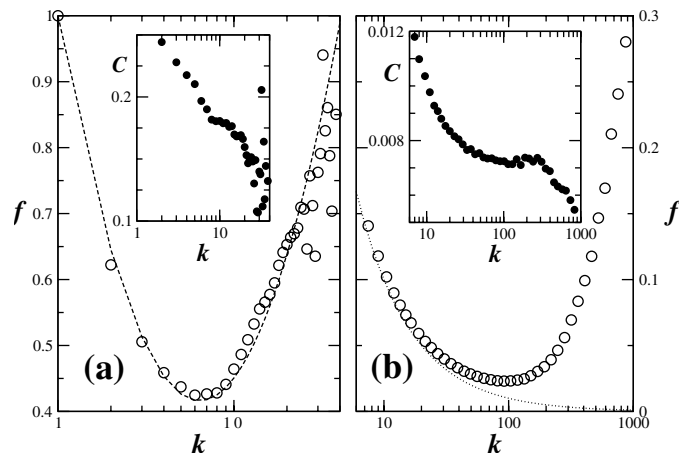


Fig. 2: (a) The spread factor f as a function of degree k for gossip propagation on the school friendship network of ref. [12]. The dashed line is the best fit to the data of the function $f(k) = 1 + a \ln k + b(\ln k)^2$, with parameters $a = -0.63$ and $b = 0.17$. In the inset we see the clustering coefficient $C(k)$, showing that both quantities are indeed very different (see text). (b) Results of a simulation with the BA network for $m = 7$ and $N = 10^4$, where m is the number of edges of a new site, and averaged over 100 realizations. The dotted line indicates $f = 1/k$ and the inset shows a different clustering coefficient spectrum from that of the empirical networks (see text).

the spread factor f of gossip in an empirical set of networks extracted from survey data [12] in U. S. schools. It is very surprising and unexpected that the curve shows a minimum, *i.e.* that we find a characteristic degree k_0 for which f and therefore the gossip spreading is the smallest. In fig. 2b we see the same quantity calculated for the Barabási-Albert (BA) network for $m = 7$ and also see a minimum at a characteristic degree k_0 which is larger because we could produce much larger BA networks on the computer thus having a larger average degree $\langle k \rangle$. The consequence of this finding is that the relative damage due to gossip can be minimized by choosing a specific connectivity k_0 which depends on the size of the network.

Since the increase of f reflects the increase of the maximal length of all paths connecting nearest neighbors of the victim, as explained above, the spectra $f(k)$ in fig. 2 show that for a larger number k of neighbors the length of such maximal path increases faster than k , a feature which cannot be ascertain from the clustering coefficient $C(k)$ shown in the insets. In fact, while for the schools the clustering coefficient decreases logarithmically with k , in BA networks the decrease is typically slower than $\log k$ [15].

For BA networks, one also is able to study the dependence of k_0 on network size N and average degree $\langle k \rangle = 2m$. We find that k_0 increases (respectively, decreases) logarithmically with N (respectively, m). Details will be reported elsewhere [16]. Therefore, this result is only valid for specific BA networks and already

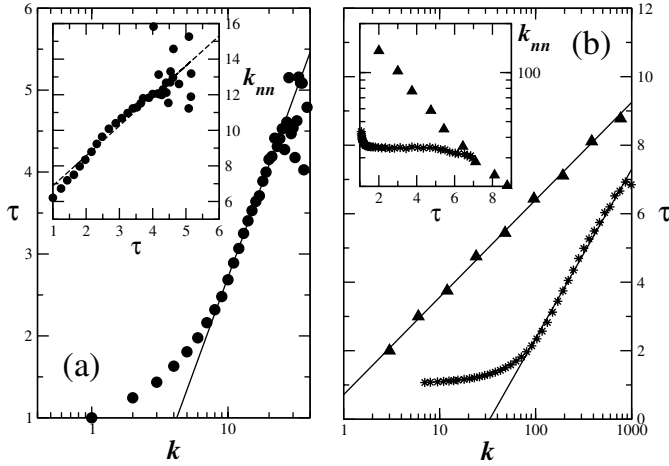


Fig. 3: Semi-logarithmic plot of the spreading time τ as a function of the degree k for (a) the real friendship network of American students [12] averaged over 84 schools and (b) the BA network with $N = 10^4$ nodes for 7 (stars), where m is the number of edges of a new site, and averaged over 100 realizations, and the APL network with $n = 9$ generations (triangles). Fitting eq. (1) for large k to these data (solid lines) we have $B = 4.5$ for the schools, $B = 2.28$ for BA and $B = 1.28$ for the APL case. In the insets, we see the average degree k_{nn} of neighbors of nodes as a function of τ , showing the non-trivial relationship between the logarithmic spreading time and the degree correlations (see text).

reducing m in the BA will, besides reducing globally f , also make the minimum less pronounced.

For an APL network all neighbors of a given victim site can be reached by a gossip since they are all connected in a closed path surrounding the victim site. This means that $f = 1$, a result that clearly shows the difference between the spread factor f and the clustering coefficient, which in this case is $C = 0.828$ [14].

In fig. 3 we see the dependence of the spreading time τ on the degree of the victim in a semi-logarithmic plot. The case of gossip in the empirical set of networks extracted from survey data [12] in U. S. schools can be seen in fig. 3a. In fig. 3b we see the data for two scale-free networks, namely the BA and the APL ones. In all cases τ clearly grows logarithmically,

$$\tau = A + B \log k, \quad (1)$$

for large k . In the case of the APL network, one can even derive this behavior analytically. In order to communicate between two vertices of the n -th generation, one needs up to n steps, which leads to $\tau \propto n$. Since [14], $k = 3 \times 2^{n-1}$, one easily obtains that $\tau \propto \log k$.

Another important point to address is to study the degree correlation [17], measured by the average number k_{nn} of neighbors of a node with k neighbors. Here, contrary to what one would expect from a first approach, the relation between degree correlations and the logarithmic behavior of the spreading time is not straightforward.

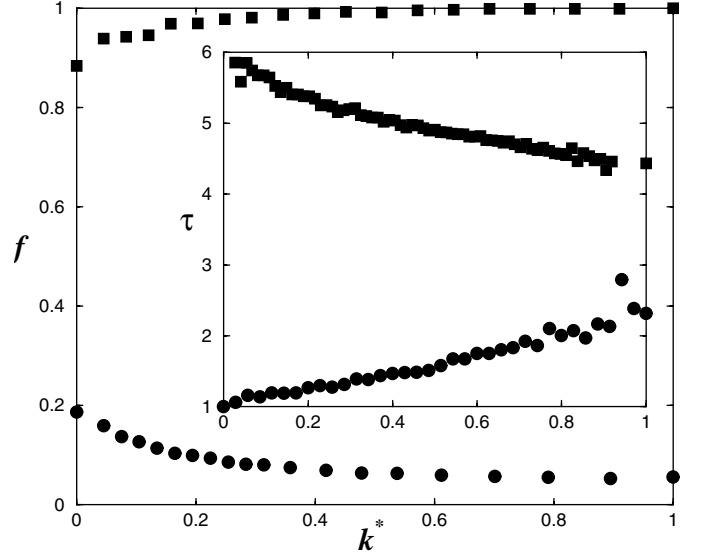


Fig. 4: Dependence of the spread factor f on the normalized degree $k^* = (k - k_{min}) / (k_{max} - k_{min})$, for the random graph with $N = 10^3$ sites and $p = 0.02$ (circles) and 0.08 (squares). In the inset: the spreading time τ of the random networks for the same parameter values.

In the empirical networks we found the same distribution for both k_{nn} and τ , yielding an approximate linear relationship between them, as illustrated in the inset of fig. 3a. However, in BA and APL networks, k_{nn} decreases with k [15] (disassortative networks). Consequently, as shown in the inset of fig. 3b, the degree correlations decrease with τ . In BA networks such decrease is at most logarithmic [15], while in APL it is exponential.

The typical behavior observed for the networks above is not observed in uncorrelated random networks, as shown in fig. 4. Here, one sees that, in random networks, the spread factor and spreading time do not depend so strongly on the average degree, i.e. on the connecting probability p . Namely, for low values of p (below the percolation threshold [3]) $f \sim 1/k$ and $\tau \sim 1$, while for larger values, more and more connections are introduced, in particular among the neighbors of each node, enabling that more nearest neighbors know about the gossip (larger f). We have also studied gossip spreading on other networks, like the Strogatz-Watts networks [10] and the configuration model [17] as will be presented in a forthcoming longer paper [16].

The large fluctuations in spreading for large τ are due to the small statistic for those values, as can be clearly seen from the distribution $P(\tau)$ in fig. 5. In fig. 5b we see that for the APL network this distribution decays exponentially. This behavior can be understood if we consider that $P(\tau)d\tau = P(k)dk$ and use eq. (1) together with the degree distribution, $P(k) \propto k^{-\gamma}$, to obtain

$$P(\tau) \propto e^{\tau(1-\gamma)/B}, \quad (2)$$

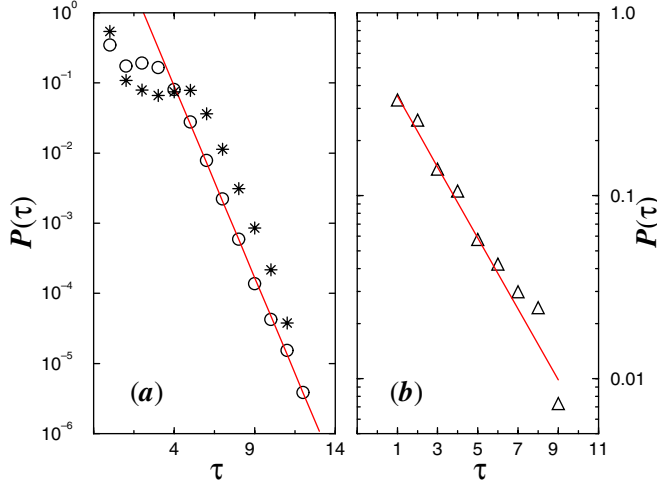


Fig. 5: Distribution $P(\tau)$ of spreading times τ for (a) the real school network (circles) and the BA network with $m=7$ and $N=1000$ (solid line), and (b) the APL network of 8 generations. The solid lines are the best fit to the data for large τ values of eq. (2), with parameters $(1-\gamma)/B = -1.26$ and -0.45 for the real school network in (a) and the APL network in (b), respectively.

for large k . The slope in fig. 5b is precisely $(1-\gamma)/B = -0.45$ using B from fig. 3b and $\gamma = 1.58$ from ref. [14]. Interestingly, $P(\tau)$ of the school network also follows an exponential decay for large τ , but with a 2.8 times smaller characteristic decay time, and has a local maximum for small τ , as seen in fig. 5a (circles). There we also plot $P(\tau)$ for the BA network for $m=7$ (solid line), which has a very similar shape but is slightly shifted to the right, due to the larger minimal number of connections. Here the distribution $P(\tau)$ counts the values of τ for every single originator, while in fig. 3a the value of τ is an average of all neighbors of a node with k neighbors. Therefore in fig. 5a a few larger values of τ appear.

Not everybody likes to gossip, so that it seems realistic to consider also the case in which the transfer from one person to the other happens with a probability q . Assuming that the person to which a gossip did not spread at the first attempt, will never get it, yields a regime similar to conditional percolation to the neighborhood of the victim. When the spreading probability q decreases, the minimum in f first shifts to larger k and finally disappears. The asymptotic logarithmic law of τ for large k seems preserved for all probabilities q . As before, the BA network for $m=7$ has a similar behavior as the school friendships.

If at each time-step the neighbors which already know the gossip repeatedly try to spread it to the common friends, one observes the same value of f measured for $q=1$, and the spreading time scales as $\tau' \sim \tau/q$, where τ is measured for $q=1$. In this way, from eq. (1) one can write

$$\tau \sim \frac{1}{q^\alpha} \log k \quad (3)$$

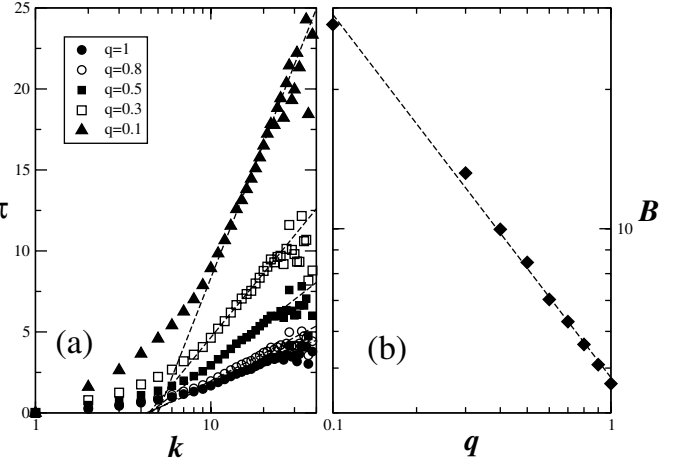


Fig. 6: Information or gossip propagation among first neighbors with probability q , when each node which knows the gossip tries to spread it (see text) on a real friendship network of American students [12] averaged over 84 schools. In (a) we show the spreading time τ for $q=1$ (\circ), $q=0.8$ (\bullet), $q=0.5$ (\square), $q=0.3$ (\blacksquare) and $q=0.1$ (\triangle). The dashed lines obey eq. (1) and have the slopes B indicated in (b).

with $\alpha \sim 1$. Figure 6 shows the result of such information propagation regime for the school networks. The spreading time is shown in fig. 6a for several values of $0 \leq q \leq 1$. In each case the dashed lines indicate the logarithmic dependence in eq. (1) with proper constants A and B . In fig. 6b we plot the slope B in eq. (1) for different values of q , yielding $\alpha = 0.8$ in eq. (3).

Finally, one can also imagine the case of a movie star for whom the gossip spreads also over strangers, *i.e.* over vertices which are not directly connected to him. We considered the cases, where nearest and next-nearest neighbors of the victim are involved in gossip spreading, and also the cases where all nodes are involved. In both cases the spreading time τ on school networks and also on the BA networks become independent of k and f loses its minimum. For the case of gossip spreading over the entire lattice, an additional quantity is also measured, namely the maximal time for the gossip to reach all students in the network. As expected such time decreases with k , since the empirical networks used in this work are assortative, as illustrated in fig. 3a.

Conclusion. – We have established a novel property of networks related to the spreading of gossip. Measuring this property in school friendship networks uncovers the striking result that there exists an ideal number of friends for which the spreading is minimized. Moreover, we also found that the spreading time grows logarithmically with the number of friends, related to degree correlations among friends, and has an exponential distribution, similarly to some scale-free networks. These results were analytically derived for the Apollonian network. When gossip also spreads among strangers, the non-trivial dependency on

the degree of connectivity disappears. We have investigated many more details on the issues presented here that will be published elsewhere [16]. In fact, due to its particular features and assumptions, our concepts to measure gossip are suited to other situations. For instance, in the case of the Internet, some Trojan horses need to connect to a specific host to download some data in order to become effective. For them the spread factor should be a good measure to assess the vulnerability to the spreading of this virus attack.

We profited from discussions with Constantino Tsallis and C. L. Nunes de Oliveira and thank the Max Planck prize, CAPES, CNPq and FUNCAP for financial support.

REFERENCES

- [1] ALLPORT G. W. and POSTMAN L., *The Psychology of Rumor* (Henry Holt & Co, New York) 1947.
- [2] BERGMANN J. R. and BEDNARZ J. jr., *Discreet Indiscretions: The Social Organization of Gossip* (Aldine, London) 1993.
- [3] BARABÁSI A. L., *Linked: The New Science of Networks* (Perseus Books Group, Cambridge MA) 2002.
- [4] WATTS D. J., *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press) 2003.
- [5] GONZÁLEZ M. C., LIND P. G. and HERRMANN H. J., *Phys. Rev. Lett.*, **96** (2006) 088702.
- [6] COHEN R., EREZ K., BEN-AVRAHAM D. and HAVLIN S., *Phys. Rev. Lett.*, **85** (2000) 4626.
- [7] ZANETTE D. H., *Phys. Rev. E*, **64** (2001) R050901.
- [8] ZANETTE D. H., *Phys. Rev. E*, **65** (2002) 041908.
- [9] HERRMANN H. J., HONG D. C. and STANLEY H. E., *J. Phys. A*, **17** (1984) L261.
- [10] WATTS D. J. and STROGATZ S. H., *Nature*, **393** (1998) 440.
- [11] AMARAL L. A. N., SCALA A., BARTHÉLEMY M. and STANLEY H. E., *Proc. Nat. Acad. Sci. USA*, **97** (2000) 11149.
- [12] Add Health program designed by J. R. Udry, P. S. Bearman and K. M. Harris funded by National Institute of Child and Human Development (PO1-HD31921).
- [13] BARABÁSI A. L. and ALBERT R., *Science*, **286** (1999) 509.
- [14] ANDRADE J. S. jr., HERRMANN H. J., ANDRADE R. F. S. and DA SILVA L. R., *Phys. Rev. Lett.*, **94** (2005) 018702.
- [15] DOROGTSEV S. and MENDES J. F. F., *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press) 2003.
- [16] LIND P. G., DA SILVA L. R., ANDRADE J. S. jr. and HERRMANN H. J., *Spreading gossip in social networks*, submitted (2007).
- [17] CATANZARO M., BOGUNÑA M. and PASTOR-SANTORRAS R., *Phys. Rev. E*, **71** (2003) 027103.