Information Diffusion in Heterogeneous Groups

Jennifer M. Larson

Abstract Standard approaches to the study of information diffusion draw on analogies to the transmission of diseases or computer viruses, and find that adding more random ties to a network increases the speed of information propagation through it. However, a person sharing information in a social network differs from a computer transmitting a virus in two important respects: a person may not have the *opportunity* to pass the information to every tie, and may be *unwilling* to pass the information to certain ties even when presented with the opportunity. Accounting for these two features reveals that, while additional random ties allow information to jump to distant regions of a network, they also change the composition of network neighborhoods. When the latter increases the proportion of neighbors to whom people are less willing to pass information, the result can be a net decrease in diffusion. I show that this is the case in heterogeneous, homophilous networks: the addition of random ties strictly impedes information dissemination, and the impediment is increasing in both original homophily and the number of new ties.

1 Introduction

The study of information diffusion in social systems applies insights from epidemiology to the spread of ideas, innovations, or behavior from node to node in a social network [24, 20, 11, 23, 26, 1]. The basic logic holds that nodes "infected" with an idea or behavior are "contagious"; network neighbors of the infected are exposed and hence susceptible to the infection, with variants accounting for the consequences of exposure to multiple sources [4, 5], variation in motivation [10, 7], the cumulative effect of repeated exposures [8, 9], and homophily with respect to susceptibility [6].

Jennifer M. Larson
Department of Politics, New York University, 19 w. 4th St. New York, NY 10012, e-mail: jenn.larson@nyu.edu

The analogy to disease spread has generated important findings about the relationship between network structure and information diffusion. Increasing the proportion of random ties in a regular network dramatically increases the propagation rate of cascades [19, 12], the presence of particularly well-connected nodes is beneficial for diffusion [25, 21, 20, 16], and random rewiring in small world networks accelerates diffusion [19, 12]. In general, adding random ties to a network will improve diffusion.

While the epidemiological approach has offered valuable insights, ties in a social network function quite differently for the spread of information than ties in a contact network function for the spread of a disease. In the case of a contact network, a tie by definition makes an alter susceptible to the disease of the ego. In the case of a social network, a tie does not *by definition* spread information to an alter. A tie indicates a social relationship. Whether or not this social relationship results in an ego passing information to an alter depends on a variety of factors: whether the two happen to encounter each other while the information is salient, whether they are together for long enough for the information to come up, whether the ego thinks the information is relevant to the alter, whether the ego is willing to share with the particular alter, and so on.

In fact, for the type of information that is often the subject of diffusion studies, an ego may have good reason to prefer to share it with some social ties over others. In the case of collective action, the information may be a person's dissatisfaction with a regime or her willingness to participate in a protest [7, 5]. Given the sensitivity of this information, especially in oppressive regimes, a person may only be willing to pass it to her most trusted social ties. In the case of technology adoption, especially in the developing world, relevant information may be news of a development organization offering startup loans or handing out new technology like fertilizer [2]. A person may judge the opportunity to be finite or selectively beneficial and prefer to share information of it with her social ties that are kin or members of her salient ingroup like her tribe [14]. In social networks, a person can choose whether to share information or whether to withhold it on a tie-by-tie basis.

In this conceptualization of information diffusion, a person in a social network will only spread information to a particular network neighbor if (1) she is presented with an opportunity to do so, and (2) is willing to share the information with that neighbor.

I account for these two features in a model in which a person has a finite number of opportunities to spread information with network neighbors. Individuals in the network have a type, which could represent ethnicity, tribe, political party, or any other salient division correlated with willingness to share new information. Given an opportunity, a person always shares information with a same-type neighbor but occasionally withholds information from a different-type neighbor.

When only one type is present in the network, the results reproduce those of earlier work: the addition of random ties allows information to jump to distant regions of the network, increasing the speed of diffusion. When multiple types are present, however, random ties introduce a second effect: they change the composition of network neighborhoods, possibly increasing the chances that the limited number of encounters will be with different-type neighbors. I show that in heterogeneous networks with type-homophily, the addition of random ties can result in the second effect dominating. In heterogeneous networks, the addition of random ties can strictly reduce the speed of information diffusion. The reduction is increasing in the original homophily, the number of types in the network, and the number of added ties.

Since homophilous communities within a network would facilitate information spread, these results are consistent with others' findings that network modularity can improve information dissemination via social reinforcement [3, 18]. However, the result here is even stronger: not only would rearranging links to reduce modularity impede information spread, but adding *new* links to the network at random can strictly impede information spread as well.

These findings refine those of earlier work, showing that the benefit of additional random ties hinges on plentiful opportunities to share information with all network neighbors and perfect willingness to share the information at every opportunity. In the more realistic case of limited opportunities and differential willingness to share, the addition of random ties may be counterproductive. In heterogeneous groups, the greater the type-homophily, the more damaging random ties are to the wide reach of information.

2 An Opportunity Model of Information Diffusion

Suppose a network g is comprised of a finite number of nodes that each have one of n types $\tau \in \{\tau_1, \ldots, \tau_n\}$. A type is a descriptive feature of a node and is used to separate an in-group from out-groups, like membership in a certain tribe or political party. Call a network *homogeneous* if $n = 1$; that is, if all nodes have the same type. A network is *heterogeneous* if $n > 1$.

Consider a simple model of information diffusion over time in which individuals may pass along new information to some network neighbors when presented with the opportunity. Call i's neighbors in g $N_i(g)$. In the model, an individual's willingness to share information depends on type: she is more willing to share information with same-type nodes than with different-type nodes. Specifically, the diffusion process proceeds as follows:

- t = 0 One node i is randomly selected and endowed with information.
- t = 1 Seed i randomly encounters x of her network neighbors, $N_i(g)$. In each encounter, she passes information to the neighbor with probability p_{same} if she and the neighbor are both the same type, and probability $p_{dif} < p_{same}$ if they are different types.
- t = 2 All j who learned information in $t = 1$ randomly encounter x of their neighbors, $N_j(g)$, passing information with probabilities p_{same} and p_{dif}.
- ... Repeats for all who learned information in the previous period until the information has reached everyone in the network or the spread halts.

2.1 Consequences of randomly added links

Randomly added or rewired ties have been found to improve information diffusion in homogeneous networks because random ties allow information to "jump" to distant network locations [19, 12]. However, the diffusion process specified in section 2 introduces a second, potentially-competing effect in heterogeneous networks. Randomly added ties can change the composition of nodes' neighborhoods. If neighborhoods are comprised of more ties to other-type nodes, the expected number of neighbors who receive the information declines.

Dual Effects of Random Ties in Heterogeneous Networks

Jump effect: random ties allow information to jump across distant network locations, improving information dissemination.

Composition effect: random ties change the composition of a node's neighborhood, potentially impeding information dissemination.

In a heterogeneous network, which effect dominates– the jump effect which improves dissemination or the composition effect which hinders dissemination– depends on the relationship between homophily and the distribution of types in the network.

Node i's network neighborhood $N_i(g)$ can be decomposed into $N_i^{same}(g)$, the subset of his network neighbors that are the same type as i, and $N_i^{dif}(g)$, the subset that are different. The expected number of nodes who receive information from i can then be written

$$\frac{x}{\#N_i(g)}\left(\#N_i^{same}(g)p^{same}+\#N_i^{dif}(g)p^{dif}\right), \tag{1}$$

where # indicates the cardinality of a set.

The consequences of an additional tie added at random will depend on the proportion of the nodes in g that are the same type as i. Call q^{τ_k} the proportion of nodes in g that are type τ_k. For simplicity, from any node i's perspective, call q_i^{same} the proportion of nodes of i's type in g. Now a random link added to $N_i(g)$ will reduce the value of (1) whenever

$$\frac{\#N_i^{same}(g)}{\#N_i(g)} - q_i^{same} > 0. \tag{2}$$

That is, when the network is homophilous with respect to type so that a larger proportion of a node's neighbors are his same type relative to the frequency of his type in the overall network, the addition of random ties will strictly reduce the expected number of people that that node informs.

The extent to which the expected number of nodes who receive information from i declines depends on the magnitude of the left hand side of (2). The greater the type homophily, the bigger impact random ties will have on reducing the expected number of people that a node informs.

When this relationship is prevalent enough throughout a network, network-wide information dissemination can be strictly impeded by the addition of random ties. The next section demonstrates the aggregate results using a simulated information diffusion process.

3 Simulated Information Spread

In this section I simulate the information diffusion process from Section 2 on simple networks generated with varying levels of homophily, heterogeneity, and random tie additions.

3.1 The Downside to Density

I begin by generating four heterogeneous networks, each with two types of nodes. The networks have 234 nodes, half of which are each type, and 864 links. Each network is generated by randomly adding links according to a specified probability of attaching to a same-type node. One network is generated for each same-type node probability $\{.5, .65, .8, .95\}$. Let the difference between the proportion of same-type links present and the proportion of same-type links that would be observed by uniformly random link formation be called the network's "homophily." With two groups of equal size, the expected proportion of random same-type links is .5, yielding networks with homophily values $\{0, .15, .3, .45\}$.

I consider the consequences of increasing density for information diffusion by randomly adding links to the network. For each value of homophily, I add links such that the total number of links increases by a factor of 1, 2, 3, and 10.

Table 1 summarizes the interpretation of the model parameters and the values to which they are set in the simulations reported below.

Figure 1 shows the results of the simulated information diffusion process on each of these networks, grouped by homophily value. In each quadrant, the curves plot the average proportion of the network that is informed by the timestep on the horizontal axis over a set of 500 simulations for a particular value of density increase. Since the population is finite, $p_{same} > 0$, and $p_{dif} > 0$, diffusion follows the characteristic s-shape. The lower the curve, the slower the diffusion.[1]

When the network exhibits no homophily (top left), randomly adding links can improve information dissemination. In this case, since the composition of the pop-

[1] This represents an impediment to diffusion in the sense that information reaches people more slowly, and also in the sense that by any given point in time, fewer people are informed.

Table 1 Model Parameters

Parameter	Definition	Set to
x	Number of network neighbors a newly-informed node encounters in a period	2
p_{same}	Probability pass news to an encountered neighbor if neighbor is same type	1
p_{dif}	Probability pass news to an encountered neighbor if neighbor is different type	.5
$\tau = \{\tau_1, \ldots, \tau_n\}$	Set of types	$\{\tau_1, \tau_2\}$, $\{\tau_1, \tau_2, \tau_3\}$, $\{\tau_1, \tau_2, \tau_3, \tau_4\}$
q^{τ_k}	Proportion of type $\tau_k \in \tau = \{\tau_1, \ldots, \tau_n\}$ present in the network	$1/n$
Homophily	Proportion same-type ties in network minus proportion same-type ties expected under random tie formation	$\{0, .15, .3, .45\}$
Diversity	Number of types, or "groups", present in the network	$\{2, 3, 4\}$
Density Inc.	Factor by which number of links is increased; e.g. 2 adds 200% of original links as new links	$\{0, 1, 2, 3, 10\}$

ulation matches the composition of neighborhoods on average, randomly adding links has no composition effect. The jump effect dominates, improving information dissemination on net.

When network neighborhoods contain more same-type links than would be expected based on the overall network composition (exhibit positive homophily), the composition effect is present alongside the jump effect. In the cases of positive homophily shown in Figure 1, the composition effect dominates: an increase in density actually impedes information diffusion. The greater the number of links added, the worse the diffusion.

Note that the number of randomly-added ties is large in these simulations, in some cases increasing the number of links in the network many-fold. Under standard epidemiological models of information diffusion, the improvement in diffusion would be vast. Here, these large additions actually *reduce* the spread of information. Moreover, these simulations assume that individuals share with other-types half of the time ($p_{dif} = .5$). When people are more hesitant to share with other types so that p_{dif} is smaller, the reduction in information spread is even greater.

3.2 The Role of Diversity

Figure 2 holds the probability of same-type links constant and increases the number of equal-sized groups in the network (the network's "diversity"). The vertical bars display the proportion of the network that has been informed on average by the tenth timestep of the simulations for each network when it has ten times the number of original links added at random minus this value for the original network. In other

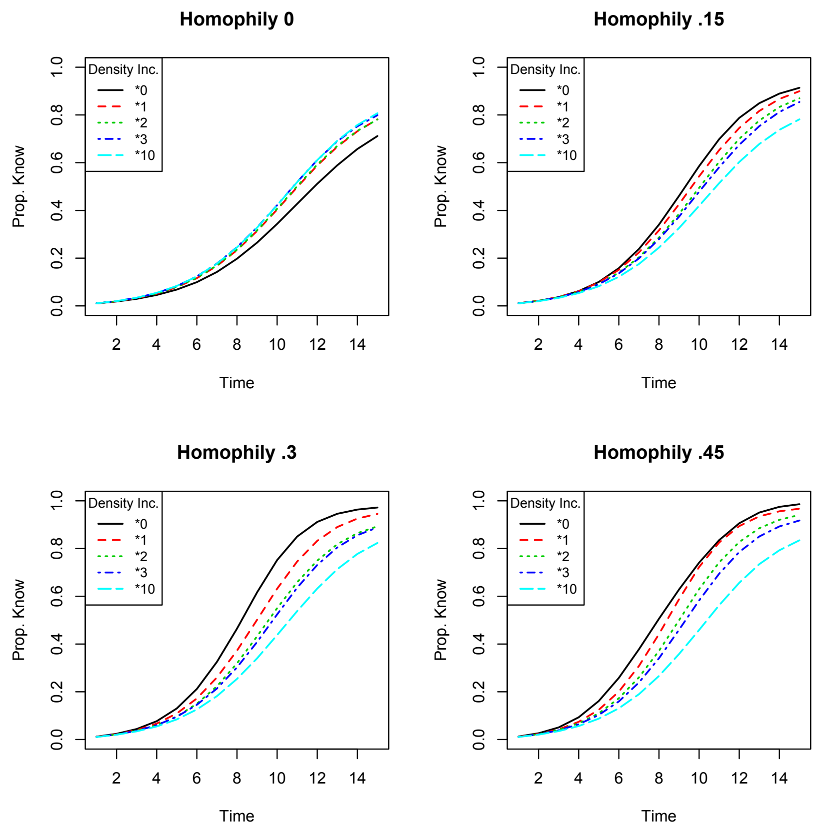

Fig. 1 Proportion of network informed by each timestep in simulated information spread on a network with $\tau = \{\tau_1, \tau_2\}$, and $q^{\tau_1} = q^{\tau_2} = \frac{1}{2}$. Simulation parameters set to $x = 2$, $p_{same} = 1$, and $p_{dif} = .5$. When homophily = 0, random ties will not change neighborhood compositions on average, so the jump effect dominants and increasing density strictly improves information diffusion. At greater values of homophily, increasing density does change neighborhood compositions and strictly impedes information diffusion.

words, this displays the gain or loss from increasing the density of each network given a certain number of groups present in the network.

The cluster of bars on the left translates the information from Figure 1 in which there are two groups present in the network. These show that when the probability of sharing a link with a same-type is greater than .5, greater density reduces the proportion of the network that has been informed by the tenth time step. The next two sets of clusters show the same from the case where there are three and four types of equal size present in the network, respectively. Comparing across clusters shows that the impediment to diffusion is greater when diversity is higher.

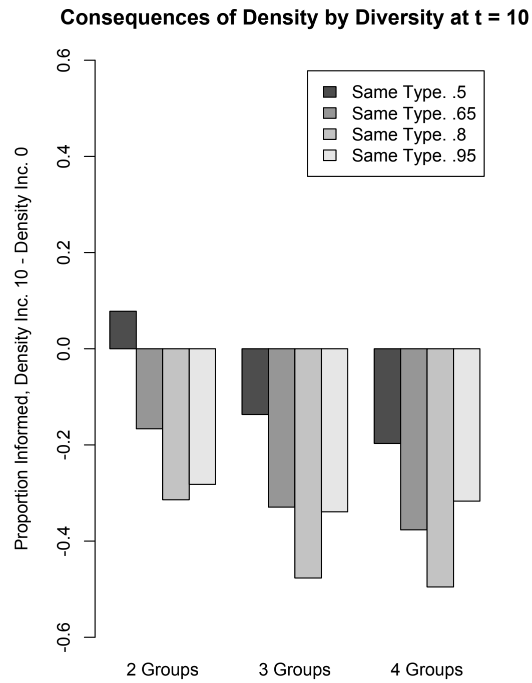

Fig. 2 Difference in proportion of network informed by timestep 10 when the density is increased by a factor of 10 compared to the proportion informed by timestep 10 given the original density. Shown for 2 groups ($\tau = \{\tau_1, \tau_2\}$ with $q^{\tau_1} = q^{\tau_2} = \frac{1}{2}$), 3 groups ($\tau = \{\tau_1, \tau_2, \tau_3\}$ with $q^{\tau_1} = q^{\tau_2} = q^{\tau_3} = \frac{1}{3}$), and 4 groups ($\tau = \{\tau_1, \tau_2, \tau_3, \tau_4\}$ with $q^{\tau_1} = q^{\tau_2} = q^{\tau_3} = q^{\tau_4} = \frac{1}{4}$). Simulation parameters set to $x = 2$, $p_{same} = 1$, and $p_{dif} = .5$. The downside to greater density is more pronounced in more diverse networks.

The negative consequences of adding random links to a network are even more acute in the presence of greater diversity.

4 Conclusion

Previous studies have found that the addition of random ties unambiguously improves information dissemination. Additional random ties generate a "jump effect," allowing information to jump from region to region within networks, speeding the spread of information. However, the present work suggests that there is an additional, at times competing effect that is masked when important features of information-sharing in social networks are unaccounted for.

Ties in social networks represent potential opportunities for the spread of information rather than certain conduits of information. People may be limited in the number of encounters that would permit information-sharing, and people can decide whether or not to share information with any candidate recipient when given the opportunity. This paper builds these two features into a model of information diffusion by assuming a uniform number of encounters per person and the presence of types such that people are more willing to share information with a same-type than a different-type neighbor.

Accounting for these features reveals that a “composition effect” can result in random ties impeding the spread of information. When random ties reduce the proportion of same-type nodes in nodes’ neighborhoods, opportunities to share information are more likely to arise with people of a different type. Since people are more hesitant to share with different types, random ties can impede overall information dissemination.

Note that the two effects can be on net negative, even when people are still willing to share information with different type ties *some of the time.* In heterogeneous groups, especially ones with high homophily, greater density can actually strictly reduce the speed with which information spreads throughout a network.

In addition to revealing a potentially negative consequence of network density in diverse groups, these results also help make sense of recent empirical findings in the social sciences showing that group composition is directly related to both trust [22] and the reach of novel information [14]. Areas that are heterogeneous in salient types– for instance those that are ethnically diverse– fare poorly in outcomes that require information to spread to coordinate outcomes like providing public goods [17], keeping aspiring rebel groups’ secrets from the government[15], and enforcing behavior through peer sanctions [13]. Heterogeneous groups may face difficulties due to problems with information dissemination that homogeneous groups are able to avoid.

References

1. Banerjee, A., Chandrasekhar, A.G., Duflo, E., Jackson, M.O.: The diffusion of microfinance. Science **341**(6144) (2013)
2. Banerjee, A.V., Chandrasekhar, A., Duflo, E., Jackson, M.O.: Gossip: Identifying central individuals in a social network. Available at SSRN 2425379 (2014)
3. Centola, D.: The spread of behavior in an online social network experiment. science **329**(5996), 1194–1197 (2010)
4. Centola, D., Macy, M.: Complex contagions and the weakness of long ties. American journal of Sociology **113**(3), 702–734 (2007)
5. Centola, D.M.: Homophily, networks, and critical mass: Solving the start-up problem in large group collective action. Rationality and society **25**(1), 3–40 (2013)
6. Chiang, Y.S.: Birds of moderately different feathers: Bandwagon dynamics and the threshold heterogeneity of network neighbors. Journal of Mathematical Sociology **31**(1), 47–69 (2007)
7. Chwe, M.: Communication and Coordination in Social Networks. Review of Economic Studies **67**(1), 1–16 (2000)
8. Dodds, P.S., Watts, D.J.: Universal behavior in a generalized model of contagion. Physical review letters **92**(21), 218,701 (2004)
9. Dodds, P.S., Watts, D.J.: A generalized model of social and biological contagion. Journal of theoretical biology **232**(4), 587–604 (2005)
10. Granovetter, M.: Threshold models of collective behavior. American journal of sociology pp. 1420–1443 (1978)
11. Jackson, M.O., Rogers, B.W.: Relating network structure to diffusion properties through stochastic dominance. The BE Journal of Theoretical Economics **7**(1) (2007)

12. Kleinberg, J.: Small-world phenomena and the dynamics of information. Advances in neural information processing systems **1**, 431–438 (2002)
13. Larson, J.M.: Networks and interethnic cooperation. Journal of Politics **Forthcoming** (2017)
14. Larson, J.M., Lewis, J.I.: Ethnic networks. American Journal of Political Science **Forthcoming** (2017)
15. Larson, J.M., Lewis, J.I.: Rumors, kinship networks, and rebel group formation. Working Paper (2016)
16. López-Pintado, D.: Diffusion in complex social networks. Games and Economic Behavior **62**(2), 573–590 (2008)
17. Miguel, E., Gugerty, M.K.: Ethnic diversity, social sanctions, and public goods in kenya. Journal of Public Economics **89**(11-12), 2325–2368 (2005)
18. Nematzadeh, A., Ferrara, E., Flammini, A., Ahn, Y.Y.: Optimal network modularity for information diffusion. Physical review letters **113**(8), 088,701 (2014)
19. Newman, M.E.: Models of the small world. Journal of Statistical Physics **101**(3-4), 819–841 (2000)
20. Newman, M.E.: Spread of epidemic disease on networks. Physical review E **66**(1), 016,128 (2002)
21. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Physical review letters **86**(14), 3200 (2001)
22. Robinson, A.L.: Nationalism and ethnic-based trust evidence from an african border region. *Comparative Political Studies* Advanced Online Publication. doi: 0010414016628269 (2016)
23. Siegel, D.A.: Social networks and collective action. American Journal of Political Science **53**(1), 122–138 (2009)
24. Valente, T.W.: Social network thresholds in the diffusion of innovations. Social networks **18**(1), 69–89 (1996)
25. Valente, T.W., Davis, R.L.: Accelerating the diffusion of innovations using opinion leaders. The Annals of the American Academy of Political and Social Science **566**(1), 55–67 (1999)
26. Young, H.P.: Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. The American economic review **99**(5), 1899–1924 (2009)