Network Modularity Controls the Speed of Information Diffusion

Hao Peng,1 Azadeh Nematzadeh,2 Daniel M. Romero,1 and Emilio Ferrara3,*

1School of Information, University of Michigan
2SEP Global
3Information Sciences Institute, University of Southern California
(Dated: November 11, 2019)

The rapid diffusion of information and the adoption of ideas are of critical importance in situations as diverse as emergencies, collective actions, or advertising and marketing. Although the dynamics of large cascades have been extensively studied in various contexts, few have systematically examined the impact of network topology on the efficiency of global diffusion. Here, by employing the linear threshold model on networks with communities, we demonstrate that a prominent network feature—the modular structure—strongly affects the speed of diffusion at global scale. Our simulation results show that, when global cascades are enabled, there exists an optimal network modularity for the most efficient information spreading process. Beyond this critical value, either a stronger or a weaker modular structure actually hinders the speed of global cascades. These results are further confirmed by predictions using an analytical approach. Our findings have practical implications in disciplines from marketing to epidemics, from neuroscience to engineering, where the understanding of the structural design of complex systems focuses on the efficiency of information propagation.

The spread of information in complex networks controls or modulates fundamental processes that can have local effects on individual actors and groups thereof, and macroscopic effects on the whole system (e.g., global information cascades). Information diffusion has been studied by drawing analogies with epidemics. Many social behaviors, for example, act like infectious diseases: once triggered, they can spread to the entire population in a very short amount of time, generating a contagion process similar to an epidemic outbreak. Examples include collective actions such as voting and participation in social movements, the adoption of innovations such as vaccination and emerging technologies, the diffusion of viral memes in social media, and the spread of norms and cultural fads. The dynamics of these intriguing and complex phenomena have attracted research interest from a number of disciplines [2, 5, 6, 14, 15, 34].

There are two major models for the study of information diffusion: the independent cascade model and the linear threshold model. The former assumes that, similar to epidemic transmission, each exposure is independent from each other and a person only has one chance to “infect” their neighbors [14, 21]. The latter postulates that social reinforcement, or exposure to multiple sources, is needed in the contagion process and each person has a threshold to be met for successful adoption [16, 36]. The independent cascade model suits well with the simple contagion scenario, where the goal is to inform people rather than to convince them to take actions [15]. It thus has been adopted in the study of word-of-mouth spreading and viral marketing [14, 23]. However, some studies revealed that the threshold model is more applicable to the spread of risky or contentious social behaviors for which each additional exposure increases the likelihood of adoption [1, 4, 25, 31, 36]. This is sometimes referred to as complex contagion.

Social behavior spreads through social contacts, thus the structure of the underlying social network plays an important role in the process of information diffusion [2, 22, 29, 36]. Recent studies have examined the effects of different network properties on the dynamics of information diffusion [10, 14, 22, 36].

One prominent network feature is modular structure—the separation of a network into several subsets of nodes within which connections are dense, but between which connections are sparser [9, 27]. Note that we distinguish modular structure from another concept, clustering, which refers to the network transitivity and is quantified by the clustering coefficient [11, 37]. Networks with many “bridges” connecting nodes in different communities tend to have low network modularity [11, 28].

The strength of weak ties theory suggests that, networks with weak modular structure will promote both the scale and the speed of information diffusion since enough shortcuts, that tend to be weak ties, link relatively-separated groups and diffuse information across communities [15, 37]. In contrast, the weakness of long ties theory predicts that, in the case of complex contagion where the adoption requires multiple exposures, networks with strong modular structure, and thus an abundance of strong ties, can enhance the spread of certain social behaviors [3, 4]. The two competing hypotheses based on prior theoretical work manifest the interplay between social reinforcement and network modularity in most real social networks. Yet, empirical studies seem to reveal inconsistent results regarding the role played by community structure in complex contagion [3, 32, 38]. Recent findings reveal that network modularity plays two different roles in information diffusion, namely (i) enhancing intra-community spreading, and (ii) hindering inter-community spreading [26], providing an in-principle unifying explanation to the competing empirical evidence.
Overall, prior work on the relationship between network modularity and large cascades has mainly focused on one aspect of information diffusion—the size of global cascades, i.e., the total number of “infected” individuals in the steady state. Another important cascade feature—the efficiency of the diffusion process, i.e., the total time it takes to reach the steady state—has been underexplored [8, 17, 19, 40]. A better understanding of the speed of information diffusion can have many practical implications, such as informing the design of communication networks where the efficiency of information diffusion needs to be prioritized. For instance, a get out to vote campaign on election day may need to be optimized for adoption speed since the operation will be useless after the election is over.

The extant literature has also demonstrated how insights about the interplay between network modularity and information spread can provide a principled understanding of various complex system dynamics, from characterizing how brain connectomics facilitates information diffusion [24], to optimizing immunization strategies for public health and animal welfare [33, 39].

Here, by utilizing the linear threshold model [16, 36], we systematically examine the effects of network modularity on the speed of information diffusion. We show that, in complex networks, there exists an optimal amount of modularity for the most efficient information diffusion at global scale.

In the linear threshold model, a node can be in two states: either active or inactive. Each node $a$ is assigned a threshold $\theta_a$ uniformly at random from the interval $[0,1]$. Initially all nodes are inactive. At time step $t = 0$, a fraction $\rho_0$ of $N$ nodes (the seeds) are switched into active state. In the subsequent time steps, a node can become active if its fraction of active neighbors exceeds the threshold, and it stays active forever once being activated. Following these rules, we update a fraction $f$ of all nodes (selected randomly) at each step. In the synchronous updating scenario, where $f = 1$, the contagion process unfolds in a deterministic manner until the network reaches the steady state [21, 26, 36]. This model can be adapted to the case of asynchronous updating by setting $f < 1$. We assume that all nodes have the same threshold $\theta$ [26, 35].

We adopt the stochastic block model to generate networks with community structure [18]. The underlying network consists of $N$ nodes partitioned into $d$ communities $\{C_1, C_2, \ldots, C_d\}$. Each community $C_i$ has a specified degree distribution $p_k^{(i)}$ and a mean degree $z^{(i)} = \sum k p_k^{(i)}$. The edges in the network are randomly distributed according to a $d \times d$ mixing matrix $e$, with $e_{ij}$ defined as the fraction of edges that connect nodes in $C_i$ to nodes in $C_j$. Although studies have indicated that tie strength is an important factor in modeling information diffusion [14, 29], here we consider edges to be unweighted, due to the unclear relationship between tie strength and network topology—some studies argue that strong ties mostly reside within tightly knit clusters and weak ties tend to link together distant communities [4, 14, 15, 29], while other empirical work reveals the opposite conclusion in social and scientific collaboration networks [7, 20, 30].

We use numerical simulations to compare the speed of diffusion across an ensemble of networks with different strength of network modularity. For simplicity, here we consider the case of two communities: let $d = 2$ and $|C_1| = |C_2| = N/2$. The following method, however, can be generalized to networks with an arbitrary number of communities (see Supplementary Information). We assume $p_k^{(1)}$ and $p_k^{(2)}$ both follow a Poisson distribution, with $z^{(1)} = z^{(2)} = z$. The expected total number of edges is: $M = z N/2$. Let $\mu M$ edges be randomly distributed between $C_1$ and $C_2$, and the remaining $(1 - \mu M)$ edges be randomly placed between node pairs in the same community, thus $e = \frac{1}{2} \left[ 1 - \frac{\mu}{\mu + 1} \right]$. Here $\mu$ controls the strength of network modularity which turns out to be $Q = 1/2 - \mu$.

A larger $\mu$ gives a network with weaker network modularity since there are more edges running between two communities. Initially, the seeds are randomly selected from $C_1$ ($p_0^{(1)} = 2 \rho_0, \rho_0^{(2)} = 0$). For each $\mu$, we run 100 simulations, with each assuming a different realization of the network and the seeds. We measure the total time $t_s$ it takes to reach steady state and the total fraction $\rho_s$ of active nodes across the network at $t_s$.

We also study the dynamics in our system analytically. The topology of such a network can be well approximated by a tree structure with infinite depth and a single node at the top, a.k.a. a tree-like approximation [12, 13]. The top node is connected to $k_a$ neighbors at the next lower level, while any other node at level $n$ is connected to $k_a - 1$ neighbors at level $n - 1$, where $k_a$ is the degree of node $a$. We update the network from the bottom level ($n = 0$) to the top level ($n \to \infty$), one level at a time, according to the linear threshold diffusion model. Given an inactive node in $C_i$ at level $n+1$, the probability of reaching an active child at level $n$ by following an edge is

$$q_n^{(i)} = \frac{n+1}{n} \sum_{j} e_{ij} q_n^{(j)} - \frac{1}{n} \sum_{j} e_{ij} q_n^{(j)},$$

(1)

with $q_n^{(i)}$ being the probability that a node in $C_i$ at level $n$ is active, conditioning on its parent at level $n+1$ being inactive. Note that $q_n^{(i)}$ applies to nodes at any level except the top node since it has no parent in the tree. Here $q_0^{(i)} = \rho_0^{(i)}$, and the updating equation for $q_n^{(i)}$ is

$$q_n^{(i)} = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_k p_k^{(i)} \sum_{m=|mk|}^{k-1} \frac{1}{m} \left( \begin{array}{c} k-1 \cr m \end{array} \right) (q_m^{(i)})^m \times (1 - q_n^{(i)})^{k-1-m} \equiv g^{(i)}(q_n^{(i)}),$$

(2)
where $\tilde{p}_k^{(i)}$ is the probability that a node in $C_i$ reached by following an edge from its inactive parent has degree $k$, thus $\tilde{p}_k^{(i)} = k p_k^{(i)}/c^{(i)}$ [27].

Eq. 2 is the sum of two scenarios: (i) the probability that the node is among the seeds ($\rho_0^{(i)}$), and (ii) the probability that the node is not among the seeds ($1 - \rho_0^{(i)}$) but is connected to at least $[bk]$ active children (the second summation, note that this node connects to $k-1$ children), summed over all possible degrees $k$ of that node (the first summation).

Similar to $q_m^{(i)}$, the fraction of active nodes in $C_i$ at level $n$ can be approximated as the probability that the top node is active, assuming that it resides at level $n$ and is in $C_i$ (note that the top node connects to $k$ children):

$$p_n^{(i)} = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_k p_k^{(i)} \sum_{m=\lfloor bk \rfloor}^k \left( \frac{k}{m} \right) (\tilde{q}_n^{(i)})^m$$

The total fraction of active nodes in the whole network at level $n$ is the probability that the top node is active and can be calculated as: $\rho_n = \sum_i p_n^{(i)} \rho^i / N$.

In asynchronous updating, the tree level $n$ used in the above derivation can be directly mapped to the time step $t$ used in simulations [12], thus the speed of diffusion in $C_i$ at time $t$ can be approximated as: $v_t^{(i)} = dp_t^{(i)}/dt = [P_t^{(i)} - P_0^{(i)}/1]$, where the notation $[1]$ stands for max(0, -). The diffusion speed $v_t$ in the whole network at time $t$, the total diffusion time $t_s$, and the average speed $\bar{v}$ are

$$v_t = \sum_i \frac{N^{(i)} }{N} v_t^{(i)}, \quad t_s = t | v_t = 0, \quad \bar{v} = \frac{\rho_{ts} - \rho_0}{t_s}. \quad (4)$$

These equations can be adapted for asynchronous updating, provided that the fraction $f$ of nodes updated at each time step is sufficiently small such that they may be considered to be independent of each other [12]. We introduce the following notation: $\tilde{q}(t)$, $q(t)$ and $\rho(t)$. The evolution equations for asynchronous updating are

$$\tilde{q}^{(i)}(t+1) = \frac{1}{d} \sum_{j} e_{ij} q^{(i)}(t), \quad (5)$$

$$d\tilde{q}^{(i)}/dt = f[q^{(i)}(\tilde{q}^{(i)}(t+1)) - q^{(i)}(t)]^+, \quad (6)$$

$$v^{(i)}(t) = dp^{(i)}(t)/dt = f[h^{(i)}(\tilde{q}^{(i)}(t+1)) - \rho^{(i)}(t)]^+, \quad (7)$$

with $q^{(i)}(0) = \rho^{(i)}(0) = \rho_0^{(i)}$. The speed is calculated as,

$$v(t) = \sum_i \frac{N^{(i)} }{N} v^{(i)}(t), \quad \bar{v} = \frac{\rho_{ts} - \rho(0)}{t_s}. \quad (8)$$

Fig. 1 displays an interval of network modularity that can trigger global cascades, which concurs with the findings in [26]. Intuitively, one would imagine that a stronger modularity (smaller $\mu$) increases diffusion speed in $C_1$ since nodes in $C_1$ are exposed to more seeds, while a weaker modularity (larger $\mu$) increases diffusion speed in $C_2$ because more bridges connect nodes in $C_2$ to the seeds. This observation, however, raises the following question: is there an ideal network modularity at which the global cascade reaches the highest diffusion speed?

Let us first analyze the behavior of our system when only local diffusion is possible. Fig. 1 indicates that, when the network modularity is too strong (very small $\mu$), information only spreads among nodes in $C_1$ due to the lack of bridges between two communities, thus decreasing modularity (increasing $\mu$) decreases the average diffusion speed because it takes longer for spreading in $C_1$ and the cascade size stays the same.

When a global cascade is achieved, however, there is a quadratic relationship between diffusion speed and network modularity: decreasing modularity first increases the average diffusion speed, but only up to a critical point, after which a further reduction in modularity slows down the overall diffusion dynamics. The global cascade thus reaches its highest average speed at the optimal network modularity ($\mu = 0.17$). The analytical predictions show excellent agreement with the simulations (Fig. 1).

Next, we analyze the cascade dynamics in more detail to understand this phenomenon. Fig. 2 shows the diffusion speed per time step in each community, for three different levels of network modularity. The time lags of
spreading in two communities can help us to explain the influence of network modularity on the average diffusion speed of global cascades.

At $\mu = 0.13$, we reach the lower bound of the window for global cascades. However, the time difference between $C_1$ and $C_2$ is the longest: the spreading in $C_2$ merely gets started after $C_1$ reaches steady state (Fig. 2a). Thus the relatively long diffusion time in $C_2$ is the bottleneck for the average diffusion speed at global scale.

One may, therefore, predict that the highest average diffusion speed can be achieved when the time lag between the two communities is reduced as much as possible. For instance, since the time difference to finish spreading at $\mu = 0.21$ (Fig. 2c) is shorter than that at $\mu = 0.17$ (Fig. 2b), the average diffusion speed would be predicted to be faster in the former case (Fig. 2c). However, such an inference is incorrect, as the diffusion at $\mu = 0.21$ takes longer time than the scenario when $\mu = 0.17$, for which the global cascade finishes in the shortest amount of time.

Comparing the optimal network modularity (Fig. 2b) to the first scenario (Fig. 2a), it takes slightly more time to finish spreading in $C_1$, due to the decreasing number of edges in $C_1$. But the increasing connections between the two communities reduces the diffusion time in $C_2$. The time lag between $C_1$ and $C_2$ is much shorter, but not close to zero. Fig. 2 indicates that, at this optimal network modularity, neither $C_1$ nor $C_2$ achieves its highest diffusion speed, but both are pretty close to it, resulting in the most efficient global cascade.

Figure 2. Cross sections of three different $\mu$ values in Figure 1 that enable global cascades. (a-c) The diffusion speed $\bar{v}(t)$ in $C_1$ and $C_2$ as a function of time step $t$. (d-f) Same as (a-c), but for the cumulative cascade size $\bar{p}(t)$. The theoretical predictions of Eqs. 7 (curves) show excellent agreement with the numerical simulations (symbols), averaged over 100 runs. The optimal $\mu = 0.17$ achieves the shortest total diffusion time, thus the highest average diffusion speed.

However, at $\mu = 0.21$, the further reduction of the number of edges in $C_1$ slows down the speed of local spreading in $C_1$, and this becomes the bottleneck of the average speed in global scale. Although, under this condition, $C_1$ and $C_2$ reach the steady state almost concurrently (with a time lag close to zero), it cannot counteract the increase in diffusion time of both communities (Fig. 2c).

Fig. 3 shows two phase diagrams of the average diffusion speed $\bar{v}$ as a function of the seed size $\rho_0$ and the threshold $\theta$, respectively. It shows that, in the region of global cascades, there always exists an optimal network modularity for the most efficient information diffusion, and this critical value of $\mu$ depends on both $\rho_0$ and $\theta$.

Notice that, in Fig. 3a, a minimal seed size is needed to trigger global cascades, and once above this threshold, when $\rho_0$ is not too large (e.g., $\rho_0 = 0.1$), the average speed of global diffusion first increases and then decreases as one reduces the community strength (increasing $\mu$), resulting in an intermediate value of $\mu$ as the optimal modularity. However, when $\rho_0$ is sufficiently large (e.g., $\rho_0 = 0.2$), the average speed of global cascades always increases as one increases the number of cross-community links, making the network with no-community structure ($\mu = 0.5$) the ideal case for the most efficient spreading process. This can be explained by the fact that, when increasing $\mu$ never blocks local spreading in $C_1$ as a result of the presence of enough seeds in $C_1$, more external links are always going to make the diffusion faster in $C_2$. Similar phases can be obtained for the threshold $\theta$ when the seed size is fixed (Fig. 3b).

Although the results presented here are based on random networks with two communities and a fixed number of nodes and a specific average degree, our findings generalize to scale-free networks with arbitrary number of communities, different network sizes, and variable average degrees, based on both analytical predictions and...
simulations. We also examined conditions with different seed placements and cases when the seeds are randomly selected across the whole network, and obtained consistent results (see Supplementary Information).

In conclusion, we investigated the effect of community structure, as measured by network modularity, on the speed of information cascade. Through simulations and analytical approximations, we demonstrate that, when global cascades are enabled, there always exists an optimal strength of network modularity—under which information diffuses at the highest speed globally. We demonstrate that such an efficient spreading behavior is achieved by making the right compromise between internal connectivity and cross-community bridges for synchronized spreading in different communities.

This framework can allow the study of many naturally-occurring complex systems in biological, animal, and human networks, and enable the understanding of evolutionary dynamics in complex networks exhibiting a certain level of network modularity that facilitates or hinders information diffusion speed. For example, network modularity has already been used to study brain connectomics and to explain global cascades [24], but the role of information diffusion speed in this context is unexplored.

From a practical standpoint, our framework has many potential applications. For example, it may help to design better organizational structure for institutions where the efficiency of information diffusion is important. Drawing on the communication network of employees in a company (e.g., from email or instant messaging), managers could make office assignments such that the optimal amount of modularity in the communication network is achieved, and thus making the information flow more efficient. Analogously with transportation networks, urban planners may account for the modular structure of the system to increase (decrease) modularity in order to obtain more efficient traffic flows.

From a methodological standpoint, by incorporating the effect of network modularity on the speed of information diffusion, machine learning algorithms can utilize network modularity to better predict the speed of global cascades. Last but not least, this study has implications for online advertising. For instance, a common social media marketing strategy is to promote a product by holding a raffle among users who shared the post within a fixed time window. Our study suggests that advertisers may try to target social networks with a good level of modular structure to achieve the fastest spreading process and therefore to maximize the number of informed audiences.

Future work of this study can focus on the empirical validation of the relationship between network modularity and the efficiency of information diffusion, and to examine its variations by considering different diffusion mechanisms on networks with more complex structure such as the hierarchical organization of communities.

Acknowledgements We thank Ashok Deb and Yong-Yeol Ahn for helpful discussions and suggestions. This work is partly supported by DARPA (W911NF-17-C-0094) and by the Air Force Office of Scientific Research under award number FA9550-19-1-0029.

Author contributions H.P., A.N., D.M.R. and E.F. collaboratively conceived and designed the study. H.P. carried out the experiments and performed the analyses. H.P., D.M.R. and E.F. drafted and revised the final manuscript.

* Correspondence should be addressed to E.F. emiliof@usc.edu

[1] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. In KDD, pages 44–54, 2006.
[2] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In WWW, pages 519–528, 2012.
[3] Damon Centola. The spread of behavior in an online social network experiment. Science, 329(5996):1194–1197, 2010.
[4] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. American Journal of Sociology, 113(3):702–734, 2007.
[5] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In WWW, pages 925–936. ACM, 2014.
[6] Justin Cheng, Lada A Adamic, Jon M Kleinberg, and Jure Leskovec. Do cascades recur? In WWW, pages 671–681, 2016.
[7] Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. On facebook, most ties are weak. Communications of the ACM, 57(11):78–84, 2014.
[8] Jean-Charles Delvenne, Renaud Lambiotte, and Luis EC Rocha. Diffusion on networked systems is a question of time or structure. Nature Communications, 6:7366, 2015.
[9] Santo Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75–174, 2010.
[10] Aram Galstyan and Paul Cohen. Cascading dynamics in modular networks. Physical Review E, 75(3):036109, 2007.
[11] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. PNAS, 99(12):7821–7826, 2002.
[12] James P Gleeson. Cascades on correlated and modular random networks. Physical Review E, 77(4):046117, 2008.
[13] James P Gleeson and Diarmuid J Cahalane. Seed size strongly affects cascades on random networks. Physical Review E, 75(5):056103, 2007.
[14] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters, 12(3):211–223, 2001.
[15] Mark S Granovetter. The strength of weak ties. American Journal of Sociology, 78(6):1360–1380, 1973.
[16] Mark S Granovetter. Threshold models of collective behavior. American Journal of Sociology, 83(6):1420–1443, 1978.
[17] Jose Luis Iribarren and Esteban Moro. Impact of human
activity patterns on the dynamics of information diffusion. Physical Review Letters, 103(3):038702, 2009.

[18] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. Physical Review E, 83(1):016107, 2011.

[19] Márton Karsai, Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, A-L Barabási, and Jari Saramäki. Small but slow world: How network topology and burstiness slow down spreading. Physical Review E, 83(2):025102, 2011.

[20] Qing Ke and Yong-Yeol Ahn. Tie strength distribution in scientific collaboration networks. Physical Review E, 90(3):032804, 2014.

[21] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In KDD, pages 137–146, 2003.

[22] Kristina Lerman and Rumi Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. In ICWSM, 2010.

[23] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. The dynamics of viral marketing. ACM Transactions on the Web (TWEB), 1(1):5, 2007.

[24] Bratislav Mišić, Richard F Betzel, Azadeh Nematzadeh, Joaquín Goni, Alessandra Griffa, Patric Hagmann, Alessandro Flammini, Yong-Yeol Ahn, and Olaf Sporns. Cooperative and competitive spreading dynamics on the human connectome. Neuron, 86(6):1518–1529, 2015.

[25] Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. Evidence of complex contagion of information in social media: An experiment using twitter bots. Plos One, 12(9):e0184148, 2017.

[26] Azadeh Nematzadeh, Emilio Ferrara, Alessandro Flammini, and Yong-Yeol Ahn. Optimal network modularity for information diffusion. Physical Review Letters, 113(8):088701, 2014.

[27] Mark Newman. Networks: An Introduction. Oxford University Press, 2010.

[28] Mark EJ Newman. Modularity and community structure in networks. PNAS, 103(23):8577–8582, 2006.

[29] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. Structure and tie strengths in mobile communication networks. PNAS, 104(18):7332–7336, 2007.

[30] Alexander Michael Petersen. Quantifying the impact of weak, strong, and super ties in scientific careers. PNAS, 112(34):E4671–E4680, 2015.

[31] Daniel M Romero, Brendan Meeder, and Jon Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In WWW, pages 695–704, 2011.

[32] Daniel M. Romero, Chenhao Tan, and Johan Ugander. On the interplay between social and topical structure. In ICWSM, 2013.

[33] Pratha Sah, Stephan T Leu, Paul C Cross, Peter J Hudson, and Shweta Bansal. Unraveling the disease consequences and mechanisms of modular structure in animal social networks. PNAS, 114(16):4165–4170, 2017.

[34] Samuel V Scarpino, Antoine Allard, and Laurent Hébert-Dufresne. The effect of a prudent adaptive behaviour on disease transmission. Nature Physics, 12(11):1042, 2016.

[35] Pramesh Singh, Sameet Sreenivasan, Boleslaw K Szymanski, and Gyorgy Korniss. Threshold-limited spreading in social networks with multiple initiators. Scientific Reports, 3:2330, 2013.

[36] Duncan J Watts. A simple model of global cascades on random networks. PNAS, 99(9):5766–5771, 2002.

[37] Duncan J Watts and Steven H Strogatz. Collective dynamics of ’small-world’ networks. Nature, 393(6684):440, 1998.

[38] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Virality prediction and community structure in social networks. Scientific Reports, 3:2522, 2013.

[39] Shu Yan, Shaoting Tang, Wenyi Fang, Sen Pei, and Zhiming Zheng. Global and local targeted immunization in networks with community structure. Journal of Statistical Mechanics: Theory and Experiment, 2015(8):P08010, 2015.

[40] Jiang Yang and Scott Counts. Predicting the speed, scale, and range of information diffusion in twitter. In ICWSM, 2010.
Supplementary Information

Hao Peng,1 Azadeh Nematzadeh,2 Daniel M. Romero,1 and Emilio Ferrara3

1School of Information, University of Michigan
2S&P Global
3Information Sciences Institute, University of Southern California

(Dated: November 11, 2019)

I. INTRODUCTION

The network representation of social relationships between people is a core ingredient in modeling the dynamics of information diffusion since the adoption of ideas or social behaviors are often influenced by one’s social neighbors. Therefore the structure of the underlying social network strongly affects the process of information diffusion. In this paper, we study how a salient network property—the modular structure—influences the speed of information diffusion by using the linear threshold diffusion model on networks with varying degree of network modularity. Through both numerical simulations and an analytical approach, we demonstrate that there exists an optimal network modularity for the most efficient information diffusion at global scale.

In this supplementary document, we provide further evidence to support our findings by exploring the behavior of our diffusion model under more general conditions with a wide range of parameters. We investigate the speed of information diffusion on modular networks with varying (i) average degree $z$, (ii) anti-modular structure $u$, (iii) network size $N$, and (iv) number of communities $d$. Additional, we report results based on networks with power-law distributions for both the node degrees and the community sizes, and we further generalize to cases where the seed nodes are randomly selected from the whole network instead of from a single community.

II. RESULTS

A. Average degree

Figure S1 shows the approximated average diffusion speed as a function of the seed size and the network modularity, on networks with different average degrees. Figure S2 shows the simulation results based on networks with $z = 100$. The results indicate that, as one increases the average degree, the minimal number of inter-community edges required for global cascades also increases, so does the minimal number of seeds. This is expected since more active neighbors are needed to achieve the same adoption threshold when the neighbor size increases.

However, the optimal network modularity for the highest speed of information diffusion always exists when global cascades are enabled. And the optimal value depends on the seed size, which agrees with our findings in the main text. In other words, the average degree does not change the behavior of our system qualitatively.
B. Anti-modular network and seed arrangement

Although our focus is on networks with community structure, it is intriguing to examine the diffusion dynamics on anti-modular networks. Figure S3 shows the average speed of information diffusion in the whole range of $\mu$, where the network shifts from exhibiting a modular structure to displaying a bipartite structure. Interesting patterns emerge: different from the dynamics on modular networks, where information spreads from the originating community to the other, the diffusion process on anti-modular networks temporally alternates between the two communities.

In such a scenario, global cascades still require a minimal number of seeds, but unlike modular networks, when $\rho_0$ is not too large (e.g., $\rho_0 = 0.2$), strong anti-modular structure (large $\mu$) always promotes the diffusion speed, making the strict bipartite networks the ideal conditions for global cascades. However, when $\rho_0$ is sufficiently large (e.g., $\rho_0 = 0.4$), the most efficient global diffusion happens at an intermediate strength of anti-modular structure (Figure S3).
Figure S3. Phase diagrams of the average speed of information diffusion in the whole range of $\mu$. The results are based on both analytical predictions (left) and numerical simulations (right), averaged over 100 runs. Model parameters are: $N = 1 \times 10^5, z = 10, \theta = 0.35, f = 0.01$. There are three regions: $\mu < 0.5$ (assortative and modular); $\mu = 0.5$ (random); and $\mu > 0.5$ (disassortative and anti-modular). The anti-modular networks behave quite differently from the modular networks.

We also examine the diffusion dynamics in our system under conditions where the seeds are not entirely placed in a single community. Figure S4 shows that, at any given seed distribution in the network (draw a horizontal slice), when global cascades are possible, there is a window of network modularity for information diffusion at global scale. For example, when all seeds are placed in $C_2$, the global cascade window of $\mu$ is $[0.13, 0.24]$, and the highest diffusion speed happens at a middle level of modularity ($\mu = 0.17$), which is exactly what we see in the first figure in the main text. The same pattern holds for other slices on the $y$-axis in Figure S4. In other words, our finding of an intermediate strength of network modularity being the ideal condition for efficient global cascades can be generalized to all other seed arrangements in the two communities, when the seed size $\rho_0 = 0.1$.

Figure S4. Phase diagrams of the average information diffusion speed in the whole range of $\mu$, as a function of seed arrangement in two communities. The results are based on analytical predictions. The $y$-axis represents the fraction of seeds placed in $C_1$. The seed size $\rho_0$ is 0.1. Other model parameters are: $N = 1 \times 10^5, z = 10, \theta = 0.35, f = 0.01$

C. Network size

Figure S5 and Figure S6 present results based on networks with different number of nodes, derived through the analytical approach and the numerical simulation, respectively. It shows that the network size does not change our findings of the most efficient spreading behavior with respect to the network modularity.
Figure S5. Phase diagrams of the average diffusion speed based on networks with different number of nodes $N$. The results are derived from the analytical approach. Other model parameters are: $z = 10, \theta = 0.35, f = 0.01$.

Figure S6. Phase diagrams of the average diffusion speed based on different network sizes $N$, derived from numerical simulations (averaged over 100 runs). Other model parameters are: $z = 10, \theta = 0.35, f = 0.01$.

D. Number of communities

So far, all our experiments are limited to conditions where the network consists of two communities. Here, we examine the diffusion dynamics based on networks with different number of communities. As a first step, we assume that all communities still have the same number of nodes and links are randomly placed according to the parameter $\mu$, as is the case in the main text. The mixing matrix is:

$$e = \frac{1}{d} \begin{bmatrix} 1 - \mu & \frac{\mu}{d-1} & \ldots \\ \vdots & \ddots & \vdots \\ \frac{\mu}{d-1} & & 1 - \mu \end{bmatrix}.$$  

(1)

where $d$ is the number of communities, and $e$ is $d \times d$. The diagonal entries of $e$ are $1 - \mu$ and the off-diagonal entries are $\frac{\mu}{d(d-1)}$ [2]. The network modularity can be calculated as: $Q = 1 - \mu - \frac{1}{d}$, which means that, in order to generate modular networks, $\mu$ can be larger than $\frac{1}{2}$ when $d$ is large than 2.

Figure S7 shows the analytical results of diffusion speed in cases of different number of communities. Please note that, at any given $\mu$, the number of bridges running between a pair of communities decreases as the number of communities $d$ increases. Thus networks with more communities require smaller adoption threshold $\theta$ in order to achieve global cascades. Figure S7 indicates that our findings of the optimal network modularity for the most efficient information diffusion can be generalized to networks with multiple communities.
However, it is still a bit unrealistic to model information diffusion on networks with equal-sized communities and random edge placement, and with seeds concentrated in one community. To further test the robustness of our results, we simulate information diffusion on the widely used LFR benchmark graphs [1], which generate networks with community structure similar to that observed in real networks by accounting for the heterogeneity in the distributions of node degrees and of community sizes.

We also randomly distribute the initial seeds across the whole network, instead of placing them in a single community. Figure S8 shows the simulation results based on the LFR networks. An optimal network modularity for the most efficient global information diffusion still emerges as in all previous conditions.

Figure S7. Phase diagrams of the average diffusion speed on networks with different number of communities $d$, derived from the analytical approach. Other model parameters are: $N = 1\times10^5, z = 10, f = 0.01$.

**Figure S8.** The phase diagram of the speed of information diffusion based on the LFR benchmark networks with $N = 25000, z = 10, \theta = 0.3, f = 0.01, \gamma = 2.5, \beta = 1.5, k_{\text{max}} = 30$. The $\mu$ (x-axis) controls the strength of community structure in the network. The results are derived from numerical simulations (averaged over 500 runs), and the seed nodes are randomly selected across the whole network. The node degrees and the community sizes in the LFR benchmark networks are distributed according to a power law with different exponents, $\gamma$ and $\beta$, respectively.

---

[1] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4):046110, 2008.

[2] Mark EJ Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.