When to target hubs? Strategic Diffusion in Complex Networks

Aamena Alshamsi$^{1,2,*}$, Flávio L. Pinheiro$^{1,*}$, and César A. Hidalgo$^1$

$^1$The MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA
$^2$Khalifa University of Science and Technology, Masdar Institute, Abu Dhabi, UAE

*These authors contributed equally to this work

Abstract

What is the most effective way to spread a behavior on a network? The recent literature on network diffusion has focused mostly on models of simple contagion[1, 2, 3, 4]—where contagion can result from contact with a single ”infected” individual—and complex contagion [5, 6, 7, 8]—where contagion requires contact with multiple ”infected” sources. While in the case of simple contagion, strategies focused on central nodes are known to be effective [9, 10, 11], the strategies that are most effective in the case of complex contagion are relatively unknown. Here, we study the strategies that optimize the diffusion of a behavior on a network in the case of complex contagion by comparing algorithms that choose which nodes to target at each step. We find that, contrary to the case of simple contagion, where targeting central nodes is an effective strategy, in the case of complex contagion minimizing the total diffusion time requires the use of dynamic strategies targeting less connected nodes in the beginning and hubs at a critical intermediate time. That is, the strategic question in the case of complex contagion is not who to target, but when to target hubs. We solve the model analytically for simple network structures and also use numerical simulations to show that these dynamic strategies outperform simpler strategies that could be hypothesized to be effective, like always choosing the node with the highest probability of infection. These findings shed light on the dynamic strategies that optimize network diffusion in the case of complex contagion.

The study of contagion is one of the central applications of network science [1, 2, 12, 13, 14]. Originally, the literature on contagion focused on problems of simple contagion. These are problems where contagion can happen in the presence of a single “infected” neighbor. Simple contagion applies mostly to the spread of infectious diseases [2, 1, 15], but also, it has been used to model processes of social influence [3, 4, 16]. In the last decade, however, network scientists came to the realization that simple contagion processes are an inadequate model for the spread of information, or social behaviors, since the adoption of ideas and behaviors in social networks often requires reinforcement by multiple sources [5, 6, 7, 17]. This realization begot a literature on complex contagion, which focuses on studying the adoption of ideas and behaviors when contagion requires multiple contacts.

But simple and complex contagion have different implications for the spread of diseases, information, and behaviors. In the case of simple contagion, the weak and long ties that connect distant parts of the network accelerate diffusion [9, 10]. In the case of complex contagion, long and weak ties can hinder diffusion, since thick bridges and communities are required to reinforce contagion and induce adoption [6, 5].

Here, we show that there is a second important difference between simple and complex contagion. This is a difference in the strategies needed to accelerate spreading. In the case of simple contagion, diffusion can be maximized by targeting central nodes, or hubs [18, 19, 20, 11, 1]. In the case of complex contagion, the question of when to target hubs becomes strategic, since hubs are less susceptible to be infected when they are not surrounded by enough early adopters. We test this hypothesis by solving complex contagion models in simple network structures analytically, and by numerically simulating contagion processes over heterogeneous networks.
Figure 1: Complex contagion model. a, we model complex contagion by assuming nodes in a network can be in three possible states: inactive (blue), potentially active (green), and active (red). Potentially active nodes are inactive nodes that are connected to active nodes, and can be activated with a probability that depends on their degree and the number of active neighbors (equation 1). b, Activation probability (equation 1) for different values of $\alpha$.

We find that the strategies needed to maximize adoption are non-trivial and dynamic, meaning that they require infecting nodes with different levels of connectivity at different times. The strategies nevertheless, go through three stages: first, they target highly susceptible low degree nodes to build critical mass around a hub, then they target hubs, and finally they return to targeting low degree nodes. These strategies spread faster than simple strategies that could be considered effective, such as always targeting the most susceptible node. Also, unlike the case of simple contagion, we find that relentlessly targeting hubs throughout the diffusion process is among the worst strategies for complex contagion problems (performs worst than random).

Finally, we discuss applications of complex contagion to the spread of information and behaviors in an organization, and the diversification of economies in the product space [21]. These findings shed light into the implications of complex contagion for the diffusion of information, diseases, and behaviors, in complex networks.

Strategies in Complex Contagion

Imagine you are working in an organization and just had a great idea. To put that idea to work, you need to convince several of your colleagues. So you ask yourself: who should you try to convince first? Your influential and skeptical boss, who is unlikely to adopt an idea unless he hears it from multiple people in the organization? Or the colleagues on your team, who can help you build the critical mass you need to convince your boss and others?

Now imagine you are running a government or private effort to promote innovation, research, or new exports. Your job involves disbursing funds through programs designed to support research, innovation, or economic ventures that could generate employment. But since funds are limited, and people from different sectors are applying to them, you need to disburse them strategically. Should you fund programs that target existing sectors? Or should you also support sectors that are related to the ones that already exist? How about unrelated sectors? Should you support initiatives that have low chance of succeeding, but that could open new opportunities in the future?

These two examples raise questions of strategic diffusion in the case of complex contagion in a network. In the first example, the key limiting factor is that people are conformist learners [22, 23, 24], meaning that they adopt behaviors when these are popular among their peers. In the second example, the key limiting factor is that the probability that a country, region, or organization will succeed in the development of a new industry [25], scientific area [26, 27], or the export of a product [21, 28], is proportional to the number of related activities present in that country, region, or organization. So the return to a “bet” is not simply its payoff times its probability of success, but also, the increase it generates in the probability of success of future bets.
Figure 2: Network models. (a) Wheel network, (b) generalized wheel network, (c) and scale-free network

To model problems of complex contagion strategically we consider a network where nodes can be in one of three states: active ($A$), potentially active, ($P$) or inactive ($I$). Active nodes can be interpreted, for instance, as individuals who have adopted ideas or as countries that are succeeding in the export of a product. Potentially active nodes are nodes who are connected to active nodes and could be activated. Inactive nodes, are nodes that do not share connections with active nodes. The key assumption of the model is a function governing the probability that a potentially active node will become active (see Figure 1 a).

Here we use a family of power-functions that allow us to interpolate through cases in which complex contagion is stronger or weaker, or even, make the spreading equivalent to that of simple contagion. This function is:

$$p_i = \left( \frac{B \sum_{j=1}^{k_i} a_{ij} M_j}{k_i} \right)^\alpha$$

Where $A_{ij}$ is the adjacency matrix indicates whether there is a link between node $i$ and node $j$ ($a_{ij} = 1$ if there is a link and 0 otherwise). $M_j$ is a memory vector indicating if a node is active ($M_j = 1$) or inactive ($M_j = 0$). $k_i$ is the degree of node $i$ ($k_i = \sum_j a_{ij}$). $B$ is a constant, and $\alpha$ is the key parameter that helps us interpolate the model between simple contagion and complex contagion. The simple contagion case is recovered when $\alpha = 0$ and $B$ is equal to the epidemic constant of the simple contagion process. When $\alpha$ is between 0 and 1, the probability of activation is a convex function of the fraction of active neighbors, meaning that it grows fast with the first active neighbors, and slow thereafter (Figure 1 b). When $\alpha = 1$ the probability of activation is linearly proportional to the fraction of active neighbors. When $\alpha > 1$ the probability of activation becomes concave, meaning that a large number of active neighbors are required to trigger activation.

Results

Now, we use our model to identify the strategies that optimize complex contagion in three cases: (i) a wheel network (a network with a central hub surrounded by a ring of nodes), (ii) a generalized wheel network (a network with multiple hubs connected to a large number of nodes in a random network), and (iii) a scale-free network constructed using the configuration model (Figure 2). Moreover, we compare the identified strategies with five simple diffusion strategies and their combinations: (1) random strategy (used as a benchmark) where the potentially active nodes targeted at each step are selected at random; (2) high-degree strategy where we target the potentially active node with the highest degree; (3) low-degree strategy where we target the potentially active node with the lowest degree; (4) greedy strategy where we target the node
with the highest probability of activation; and (5) conformist strategy where we target the potentially active node with the highest number of active neighbors. See Section 1 in the supplementary material for more details.

The Wheel Network

Consider a wheel network populated by \(Z\) nodes (Figure 2 a). In a wheel network we have a central hub, which is connected to all nodes, and a ring of nodes that are connected to two neighbors and the hub. The wheel network is particularly instructive because the problem of strategic diffusion reduces simply to the problem of choosing at which point you should target the hub. In the wheel network, the probability of activating peripheral nodes \((p_i = \left(\frac{1}{3}\right)^\alpha\) does not change unless the hub is active. After the hub is active, the probability of activating peripheral nodes grows to \((\frac{2}{3})^\alpha\).

We assume the system starts with a single active peripheral node. The naive dynamic strategy in this case, is to activate \(\frac{1}{3}\) of the peripheral nodes, and at that point target the hub (since at that moment the probability of activating the hub is identical to that of activating low degree nodes). This naive strategy would be equivalent to the greedy strategy mentioned above. The conformist strategy would target the hub after one peripheral node is infected, and would be almost identical to the high degree strategy. The low degree strategy would target the hub last.

We can obtain an optimal strategy, however, by leveraging the symmetry of the wheel network to write an equation for the average total time needed to activate the full network.

Let \(L\) be the time when we target the hub, measured as the number of peripheral nodes that have been activated. Also, note that the expected time \(t\) required to activate a node with activation probability \(p\) is \(t = \frac{1}{p}\). Then, the total time required to activate the entire network will be equal to:

\[
T(L, \alpha) = 3^\alpha (L - 1) + \left(\frac{Z - 1}{L}\right)^\alpha + \left(\frac{3}{2}\right)^\alpha (Z - 2 - L) + 1 \tag{2}
\]

where \(3^\alpha\) is the time required to activate each of the first \(L - 1\) nodes, \((Z - 1/L)^\alpha\) is the time required to activate the hub, and \((3/2)^\alpha\) is the time required to activate all remaining peripheral nodes after activating the hub, except for the last one which takes one unit of time.

Figure 3 a shows the total time needed to activate the entire network as a function of \(L\) and \(\alpha\) \((T(L, \alpha))\). We find that for \(\alpha > 1\) the time is minimized when between 5% and 20% of the peripheral nodes are active before targeting the hub (which is smaller than the naive \(\frac{1}{3}\) strategy (greedy strategy), but also, large enough to imply that targeting the hub should be avoided in the early stages of complex contagion). We can solve \(2\) to find the minimum time by setting \(dT/dL = 0\) and obtain the optimal time to activate a hub \(L^*\):

\[
L^*(\alpha) = (Z - 1)\left((\frac{3}{2})^\alpha + 3^\alpha (Z - 1)\right)^{-\frac{1}{\alpha+1}} \tag{3}
\]

This simple model has three important implications. First, the optimal fraction of active peripheral nodes \(L^*/Z\) is always lower than we would expect from the naive greedy strategy \(L_{\text{naive}}/Z = \frac{1}{3}\), which targets the hub when the probability of activating it is equal to that of activating a low degree node. Second, targeting hubs from the beginning is a strategy that performs poorly. Third, the difference between the optimal solution and the suboptimal solutions increases with \(\alpha\), meaning that the value of using a dynamic strategy is small in cases that are close to simple contagion \((\alpha << 1)\), and large for concave complex contagion \((\alpha > 1)\). Finally, we find that the values of the optimal solution increase with \(\alpha\), meaning that we need to target hubs later in the process in cases that depart more strongly from simple contagion.

Generalized Wheel Network

Next, we extend these results to a generalized wheel network, a network with \(m\) hubs connected to \(k_H\) low degree nodes that form a random network among themselves with average degree \(k_L << k_H\) (Figure 2 b).
In equation 4, the first sum accounts for the activation time of the initial \( L \) low degree nodes, the second term for the activation of the \( m \) hubs, and the third term accounts for the activation time of the remaining \( n - L \) low degree nodes. Since this expression doesn’t have a closed form solution, we explore it numerically.

Figure 4 compares the predictions of equation 4 with the results obtained for numerical simulations for the five strategies described above: random, high-degree, low-degree, greedy, and conformist, for networks with different levels of heterogeneity (with increasing \( k_H \)). We consider generalized wheel networks with 1,000 nodes and 2, 6, and 10 hubs, and vary the connectivity of these nodes from 100 to 1,470,000 (to simulate increasingly more heterogeneous networks).

In all cases we find that the differences between solutions is larger for more heterogeneous networks. This means that choosing the right dynamic strategy is more important in heterogeneous networks. In fact, for networks with low heterogeneity, Equation 4 fails to provide the optimal solution. For high levels of heterogeneity, however, we find the largest differences among all strategies. In general, we find that strategies focused on high-degree and low-degree nodes perform worse than random in all cases. Whereas the greedy strategy (targeting most susceptible nodes) and the conformist strategy (targeting nodes with the highest number of active neighbors), perform better than random. Yet, the greedy and conformist strategies are inferior to the solutions of equation 4 for highly heterogeneous networks. In fact, Equation 4 provides a strategy that activates the whole network in 80 percent of the time of the greedy and conformist strategy, and in 70 percent of the time required using the random strategy, for the most heterogeneous networks.
Figure 4: Total activation times for five static strategies and our dynamic strategy in generalized wheel networks with 1000 nodes and 2 (a), 6 (b), and 10 (c) hubs. The total activation times are presented as a function of the heterogeneity of the network, characterized by the degree of the hubs. As heterogeneity increases, the difference in the total activation between the dynamic and static strategies increases, meaning that the use of dynamic strategies is more important in more heterogenous networks.

Scale-Free network

Finally, we explore complex contagion in scale free networks [29] (Figure 2 c). These are heterogeneous networks characterized by a power-law degree distribution, meaning that the probability that a node will be connected to $k$ other nodes follows a distribution of the form $P(k) \approx k^{-\gamma}$. Scale free networks are relevant because they are common in social, economic, technological, and biological systems [30].

Figure 5 a compares the performance of the five aforementioned strategies (random, low-degree, high-degree, greedy, and conformist) with a mixed strategy that stochastically alternates between the greedy strategy (with probability $p$) and the conformist strategy (with probability $1-p$). The performance of strategies is measured in terms of the total time they take to activate the entire network.

In general, we find that the greedy and conformist strategy are always better than the random strategy, and that strategies focused on low and high degree nodes perform worst than random. The mixed strategy, however, outperforms all simple strategies.

Next, we explore whether the mix strategy is behaving similar to strategies identified for the wheel and generalized wheel network: are they targeting hubs at a critical intermediate time? To do this, first we identify the optimal mix of strategies for each $\alpha$ by varying $p$ (the probability of choosing the greedy strategy) (Figure 5 b). We find that more heterogeneous networks ($P \approx k^{\gamma}$ with $\gamma \approx 2$) require almost a 50-50 mix of the greedy and conformist strategy and that less heterogeneous networks ($P \approx k^{\gamma}$ with $\gamma \approx 4$) require a mix that is mostly based on the greedy strategy, but still need a tad of the conformist strategy.

Using the optimal mix we identify the time at which these strategies target hubs, and compare that with the time at which using a purely greedy or conformist strategy would target hubs. Figure 5 c shows these results by plotting the expected rank of the activation time of nodes of different degrees in a network with 1,000 nodes. For random strategies, that expected rank is 500 for nodes of all degrees, since random strategies do not have a preference for targeting high degree nodes early or late in the process. The mixed strategy, however, targets hubs at a time that is later than the conformist, but earlier than the greedy, showing that its ability to infect the network faster comes from its ability to target hubs at the right time.

Discussion

The study of contagion has a long tradition in the natural sciences, dating back to the seminal work of Kermack and McKendrick in the 1920’s [15]. Yet, only in the last two decades scholars began to understand how the networks underlying contagion phenomena affected contagion dynamics [1]. The introduction of networks into problems of contagion brought an important distinction, that between problems of simple
Methods

Networks

Generalized-wheel networks with $m$ hubs and $n$ leaves were generated by first creating a random network, following the Erdős-Rényi model, among the $n$ leaves so that they exhibit an average degree of $k$. Finally, each of the $m$ hubs is randomly connected with one of the $n$ leaves. Here leaves correspond to low-degree nodes and hubs to high-degree nodes.

Scale-free networks were generated using the configuration model \[^{21}\] which requires networks to have a degree distribution that follows a power-law distribution ($\approx k^{-\gamma}$). For each value of $\gamma$, we have generated 100 independent networks with 10,000 nodes. More information on the topological properties of the generated networks can be found in the supplementary material Section 2.

Computer simulations

Simulations were used to analyze the performance of different strategies in scale-free networks and to validate the analytical results in the wheel and generalized wheel networks. Each simulation starts with a single low degree node active and sequentially attempts to activate a potential node with probability given by equation 1. The target node is selected according to one of the strategies under analysis. In case of ambiguity, if more than one node is a target candidate, one is selected at random. In the case of mixed strategies, in which the strategies are selected with a given probability $p$ and $1-p$, respectively, the target is chosen at random from the set of potential targets.

Results

Our findings have implications for different domains. Our results contribute to the literature by showing that strategic diversification is not only about knowing what products to target next, but also about knowing when to target products that are unrelated. Our findings suggest that these efforts to target unrelated products should be optimal at an intermediate level of diversification.

Here, we show that the strategies needed to maximize spreading in the case of sequential complex contagion are drastically different from those that are optimal in simple contagion. In simple contagion, central nodes need to be targeted at the right time. In the former case, they teach us how to spread an idea in a social context to the largest possible extent. In the latter case, our results are relevant because it is known that the probability that a country or region will succeed in developing a new product or industry depends on the number of related industries present in that location. Our findings suggest that these efforts to target unrelated products should be optimal at an intermediate level of diversification.

Figure 5. Computer contagion in Scale-Free networks. a) total time needed to activate a scale-free network of 1000 nodes as a function of the exponent of the Scale-free distribution ($\gamma$). b) the expected rank of the activation time when $\gamma$ is 1.25, 1.50, 1.75, 2.00, 0.00, 0.25, 0.50, 0.75. c) the expected rank of the activation time when a strategy targets hubs too late, while the greedy strategy targets hubs too early. The mixed strategy, however, targets the hubs at the right intermediate time, after recruiting enough low-degree nodes.
A selection of a target node can be done through two or more actions, at the beginning of an activation step one of them is selected at random according to the designated weights.

Acknowledgement This work was funded by the Cooperative Agreement between the Masdar Institute of Science and Technology (Masdar Institute), Abu Dhabi, UAE and the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA - Reference 02/MI/MIT/CP/11/07633/GEN/G/00. Also, this work was supported by the Center for Complex Engineering Systems (CCES) at King Abdulaziz City for Science and Technology (KACST) and the Massachusetts Institute of Technology (MIT) and the MIT Media Lab consortia.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to César A. Hidalgo (email: hidalgo@mit.edu).

References

[1] Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. Physical review letters 86, 3200 (2001).

[2] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. Reviews of modern physics 87, 925 (2015).

[3] Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. New England journal of medicine 357, 370–379 (2007).

[4] Fowler, J. H. & Christakis, N. A. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. Bmj 337, a2338 (2008).

[5] Centola, D. & Macy, M. Complex contagions and the weakness of long ties1. American journal of Sociology 113, 702–734 (2007).

[6] Centola, D., Eguiluz, V. M. & Macy, M. W. Cascade dynamics of complex propagation. Physica A: Statistical Mechanics and its Applications 374, 449–456 (2007).

[7] Hill, A. L., Rand, D. G., Nowak, M. A. & Christakis, N. A. Emotions as infectious diseases in a large social network: the sisa model. Proceedings of the Royal Society of London B: Biological Sciences 277, 3827–3835 (2010).

[8] Centola, D. The spread of behavior in an online social network experiment. Science 329, 1194–1197 (2010).

[9] Granovetter, M. S. The strength of weak ties. American journal of sociology 78, 1360–1380 (1973).

[10] Onnela, J.-P. et al. Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences 104, 7332–7336 (2007).

[11] Dezso, Z. & Barabási, A. Can we stop the aids epidemic? Phys. Rev. E 65, 055103 (2001).

[12] Lerman, K. & Ghosh, R. Information contagion: An empirical study of the spread of news on digg and twitter social networks. ICWSM 10, 90–97 (2010).

[13] Elliott, M., Golub, B. & Jackson, M. O. Financial networks and contagion. The American economic review 104, 3115–3153 (2014).

[14] Dodds, P. S. & Watts, D. J. A generalized model of social and biological contagion. Journal of theoretical biology 232, 587–604 (2005).
[15] Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. In Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences, vol. 115, 700–721 (The Royal Society, 1927).

[16] Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. Reviews of modern physics 81, 591 (2009).

[17] Hill, A. L., Rand, D. G., Nowak, M. A. & Christakis, N. A. Infectious disease modeling of social contagion in networks. PLoS Comput Biol 6, e1000968 (2010).

[18] Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. Science 341, 1236498 (2013).

[19] Borgatti, S. P. Centrality and network flow. Social networks 27, 55–71 (2005).

[20] Christakis, N. A. & Fowler, J. H. Social network sensors for early detection of contagious outbreaks. PloS one 5, e12948 (2010).

[21] Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. Science 317, 482–487 (2007).

[22] Rendell, L. et al. Cognitive culture: theoretical and empirical insights into social learning strategies. Trends in cognitive sciences 15, 68–76 (2011).

[23] Laland, K. N. Social learning strategies. Animal Learning & Behavior 32, 4–14 (2004).

[24] Henrich, J. The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter (Princeton University Press, 2015).

[25] Neffke, F., Henning, M. & Boschma, R. How do regions diversify over time? industry relatedness and the development of new growth paths in regions. Economic Geography 87, 237–265 (2011).

[26] Estellés, E., Del Moral, E. & González, F. Social bookmarking tools as facilitators of learning and research collaborative processes: The diigo case. Interdisciplinary Journal of E-Learning and Learning Objects 6, 175–191 (2010).

[27] Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M. & Hidalgo, C. A. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. Scientometrics 109, 1695–1709 (2016).

[28] Hausmann, R. et al. The atlas of economic complexity: Mapping paths to prosperity (Mit Press, 2014).

[29] Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. science 286, 509–512 (1999).

[30] Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. Reviews of modern physics 74, 47 (2002).

[31] Newman, M. E. The structure and function of complex networks. SIAM review 45, 167–256 (2003).