QUANTITATIVE POLITICAL ECONOMY RESEARCH GROUP
Department of Political Economy
King's College London

**Word-Of-Mouth Centrality and the Probabilities of
Node-To-Node Diffusion in Fixed Networks**

QPE Working Paper 2020-03

Maia King

March 17, 2022

Word-of-mouth centrality and the probabilities of node-to-node diffusion in fixed networks*

Maia King[†]

King's College London

March 17, 2022

Abstract

Calculating the true probability that a signal will be transmitted between any pair of nodes in a network is computationally hard. Diffusion centrality, which counts the expected number of times that a signal will be transmitted, is often used as a heuristic for this probability. But this approach can lead to distorted results, because it does not take account of the inclusion-exclusion principle in probability theory. This paper provides a new heuristic for the probabilities of node-to-node network diffusion, which uses De Morgan's laws to account for the inclusion-exclusion principle. It is quick to calculate and also provides a new centrality measure, *word-of-mouth centrality*, which is the average probability that a signal emitted by a node will be received by other nodes in the network by diffusion. The paper also defines *blocking centrality*, for the cases when some nodes block signals.

JEL classification: D83, D85, Z13

Keywords: Information, diffusion centrality, Katz-Bonacich centrality, networks

*I am grateful for helpful comments from Nizar Allouch, Sergio Currarini, Matthew Elliott, Arun Chandrasekhar, Yann Bramoullé, Garance Genicot, Nicholas Christakis, Andrew Oswald, Amrita Dhillon, Aniol Llorente-Saguer, Rachel Kranton, Francesco Nava, Matthew O. Jackson, Tomas Rodriguez-Barraquer, Amrish Patel, Alex Teytelboym, Nicole Tabasso, Chara Tzanetaki, Richard Stead, Andrew King, two anonymous reviewers, and to seminar participants at ESNIE, UECE, QMUL networks conference, Warwick PhD conference, CTN, Harvard Conference on Information Transmission in Networks, the French Symposium on Games, and a Nuffield College, Oxford seminar on Networks and Games. Elements of this work previously appeared in a working paper entitled “Whom can you trust? Reputation and cooperation in networks” (King, 2017). I declare no relevant or material financial interests that relate to the research described in this paper.

[†]Department of Political Economy, Aldwych, London WC2B 4BG. E-mail: maia.king@kcl.ac.uk

1 Introduction

The optimal seeding problem is a canonical question in network theory. It asks which nodes, if they are the first to be ‘seeded’ with new information or actions, will lead to the greatest impact across the network. Studies of this question in different contexts abound, for example in marketing (Richardson and Domingos, 2001; Iyengar et al., 2011), innovation (Rogers, 2010), insurance (Cai et al., 2015), health interventions (Kim et al., 2015; Shakya et al., 2017), internet adoption (DiMaggio and Garip, 2011), obesity (Christakis and Fowler, 2007), and the interaction between network influence and mass advertising (Rossman and Fisher, 2021). Kempe, Kleinberg and Tardos (2003) showed that the optimal seeding problem is NP hard, which has led to the use of heuristics (Akbarpour et al., 2018; Jackson, 2017; Rose, 2019), and methods to elicit centrality measures without full network data (Banerjee et al., 2019; Christakis and Fowler, 2010; Breza et al., 2020).

Within the broader topic of optimal seeding for network influence, a key question for both empirics and theory is how to calculate the probabilities of information transmission in fixed networks between each pair of nodes. These probabilities are important for analysing many phenomena, including the diffusion of microfinance opportunities (Banerjee, Chandrasekhar, Duflo and Jackson, 2013), political intermediation of clientelistic favours (Cruz, Labonne and Querubín, 2017; Fafchamps and Labonne, 2019) and the monitoring of savings behaviour within communities (Breza and Chandrasekhar, 2019). But finding the ‘true’ probabilities of diffusion is computationally hard (Bramoullé and Genicot, 2018), because tracking the probabilities of a signal travelling along multiple walks in a network is combinatorially complex. In each of the empirical papers cited above, the heuristic of diffusion centrality, defined by Banerjee et al. (2013), has been used to approximate the probabilities of information transmission.

Diffusion centrality has two key characteristics which make for computational ease. First is the assumption of *independence*: that the probability of a signal travelling along any walk in the network is independent of the probability that it travels along any other walk. Second, diffusion centrality uses the Katz-Bonacich (Bonacich, 1987) approach to combine these probabilities (of the signal travelling along different walks), by adding them together in summation. This formulation means that diffusion centrality calculates the *expected number of times that a signal will be transmitted between nodes* (under the independence assumption).

This formulation leads to two important consequences for researchers. First, due to the properties of Katz-Bonacich centrality, they are constrained to use a specific value of the primitive p in their calculations, which is the probability that a signal is passed between direct neighbours. With diffusion centrality, p is assumed to be equal to the inverse of the largest eigenvalue of the adjacency matrix (Banerjee et al., 2013).

Second, as Section 3 of the paper explores, if a summation of probabilities is used as a proxy for a probability, this can lead to distortions, because a summation does not take account of the inclusion-exclusion principle. As such, since the numbers returned by the diffusion centrality formula may be greater than one, it would require arbitrary scaling to convert them into numbers which can be used as probabilities.

Furthermore, a method which does not take account of the inclusion-exclusion principle could lead to an overestimation of the inequality of centrality measures. This is because we would expect this distortion to increase for those nodes which are more central, and who would be more likely to be connected to other nodes via a large number of different walks. If the overlapping nature of the probabilities of signals travelling along these multiple walks is not accounted for, centrality could be overestimated.

In Section 4 of this paper, I provide a new heuristic for the probabilities of node-to-node information diffusion in fixed networks, which uses De Morgan’s laws to account for the inclusion-exclusion principle. This gives a formula for *word-of-mouth probabilities*, which are bounded from above by 1, and are as simple to calculate as diffusion centrality. The word-of-mouth formula can use any value for edgewise communication p , in contrast to the constraint for diffusion centrality of using the inverse of the largest eigenvalue of the adjacency matrix. As such, the word-of-mouth formula could provide researchers with more reliable and flexible estimates for the probabilities of information transmission in networks.¹

¹This probabilistic flow of signals between network connections, means that it could also be used in theoretical models with Bayesian updating, in contrast to the behavioural approaches that are often used for information transmission in networks. For example, Degroot (1974); Golub and Jackson (2012); Mueller-Frank and Neri (2015); Levy and Razin (2014). In other settings, Hagenbach and Koessler (2010), Galeotti et al. (2013) and Acemoglu et al. (2014) also focus on the transmission of signals in networks, using the cheap talk framework of Crawford and Sobel (1982).

In Section 5, I use these probabilities to construct a new centrality measure: *word-of-mouth centrality*, and use examples to compare it to diffusion centrality as a measure of probabilistic information transmission in networks. This shows that diffusion centrality could overestimate inequality in information transmission in networks. It can also affect the relative ranking of nodes, as the number of rounds of information transmission increase.

Section 6 applies the word-of-mouth approach to the case when some nodes in a network ‘block’ signals. I find that the effect of this blocking depends on the structure of the network, in particular 2-connectedness and cycles of length four. We can also construct a centrality measure that depends on these blocked probabilities, showing how the communication possibilities of different nodes are affected by others nodes’ blocking behaviour. This is called *blocking centrality*.

The paper is organised as follows. Section 2 sets out the network model, and Section 3 discusses communication centrality and diffusion centrality. Section 4 sets out the word-of-mouth probabilities, and Section 5 compares word-of-mouth centrality to diffusion centrality. Section 6 looks at word-of-mouth probabilities when some nodes block signals, and Section 7 concludes.

2 The network

Let us consider a network with n nodes in $N = \{1, \dots, n\}$ who occupy the nodes of a fixed directed weighted network $\mathbf{g}$ such that $\{i, j\} \in \mathbf{g}$ if there is a directed edge from i to j . A walk of length a between two nodes i and j in network $\mathbf{g}$ is a sequence of nodes ($i = x_0, x_1, \dots, x_{a-1}, x_a = j$) such that for every $r \in \{1, 2, \dots, a\}$, we have that $\{x_{r-1}, x_r\} \in \mathbf{g}$. If the nodes are distinct, the sequence is a path, and if in addition $i = j$, it is a cycle.

Let $\mathbf{G} = [g_{ij}]$ be the adjacency matrix of the network $\mathbf{g}$, where $g_{ij} \in [0, 1]$ and $g_{ij} > 0$ indicates that $\{i, j\} \in \mathbf{g}$, and $g_{ij} = 0$ otherwise (and $g_{ii} = 0 \ \forall i \in N$ by convention). Let $\mathcal{N}_i = \{j : g_{ij} > 0\}$ be the set of node i 's neighbours and $|\mathcal{N}_i|$ be i 's degree. Let $d_{ij}(\mathbf{G})$ be the length of the shortest path between two nodes i and j in the network $\mathbf{G}$, which is known as the geodesic distance or *social distance*. If two nodes are not connected, their social distance is infinite. Let $D_{\mathbf{G}} = \max\{d_{ij}(\mathbf{G})\}$ be the diameter of the network $\mathbf{G}$: the length of the longest shortest path. Two nodes are *connected* if there exists a path of finite length between them, and a network is connected if all nodes are connected to each other. If $\mathbf{G}$ is not connected, its diameter is infinite.

Let $\mathbf{G}_{-k}$ be the adjacency matrix of the network when node k is disconnected: this is the $n \times n$ adjacency matrix created when all the entries in the k th row and k th column of $\mathbf{G}$ are set to zero; that is, k remains as a node in the network, but all its links are removed and it is isolated as a singleton. If $\mathbf{G}_{-k}$ is connected then the network is *2-connected with respect to k* . The network $\mathbf{G}$ is *2-connected* if it is 2-connected with respect to all nodes (Renault and Tomala, 1998; Wolitzky, 2015).

3 Diffusion centrality and communication centrality

Banerjee, Chandrasekhar, Duflo and Jackson (2013) develop an innovative approach to investigate the diffusion of microfinance information in Indian villages. They use an SIR model² within a fixed network structure, and define the *communication centrality* of a seeded household as “the fraction of households who would eventually participate [in the microfinance scheme] if this household were the only one initially informed”.

²The Susceptible, Infected, Recovered (SIR) model is a standard approach for examining epidemics (Kermack and McKendrick, 1927; Bailey, 1957).

Communication centrality is based on the following SIR-like structure. First, seeded nodes decide whether to participate in the microfinance scheme or not. Next, they pass on signals about the scheme to their neighbours, with a higher probability if the seeded node decided to participate. Some neighbours are now ‘informed’ by this signal, and then they decide whether or not to participate. The participation decision depends on the household characteristics and also potentially on the participation decisions of a household’s neighbours. In each further round, all previously ‘informed’ nodes — including the initial seeds — pass on the signal to their neighbours, with a probability which depends on their participation decision.³ Via simulation, Banerjee et al. (2013) find that communication centrality gives a good prediction of the proportion of village households who would participate in a microfinance scheme, for each initial seed node.

Banerjee et al. (2013) also define diffusion centrality as useful heuristic for communication centrality, given the latter’s computational difficulty. They state that “*the diffusion centrality of a node i in a network with an adjacency matrix $\mathbf{G}$, passing probability p , and iterations T , as the i th entry of the vector $\mathbf{DC}(\mathbf{G}, p, T) = \sum_{\tau=1}^T (p\mathbf{G})^\tau \mathbf{1}$. Consider T iterations of information passing from a single initially informed node i where at each iteration every informed node tells each neighbor with probability p . The diffusion centrality of node i then corresponds to the expected total number of times that all nodes taken together hear about the opportunity.*”

A key difference between communication centrality and diffusion centrality is that diffusion centrality dispenses with the participation decision, and also with different probabilities of information transmission depending on this. As such, diffusion centrality depends on only three elements: the network structure $\mathbf{G}$; the passing probability p which denotes how likely nodes are to meet and/or exchange information with their neighbours; and the number of rounds of information transmission, T . For example, if $p = 1$ and $T = 1$, information is passed with certainty only to a node’s direct neighbours. Banerjee et al. (2013) find a 0.86 correlation between the two centrality measures in their data.

³Banerjee et al. (2013) describe how this approach differs from the standard SIR model, because informed nodes who do not choose to participate in microfinance may also pass on information about the scheme, but with a lower probability than participating nodes. The other key difference from the SIR model is that the communication centrality model runs for finite time T .

Bramoullé and Genicot (2018) highlight an inconsistency in Banerjee et al. (2013)’s definition above, showing that “diffusion centrality does not emerge from a model where *‘at each iteration every informed node tells each neighbor with probability p ’*.” They point to two assumptions of the implied process of information transmission with diffusion centrality, which make it different from communication centrality. One is that information is a flow, so that nodes only pass on a signal that they received in the last period. If it was a stock, in the case of communication centrality, ‘informed’ nodes would continually emit the signal, even if they had not just received it. The other implicit assumption is that if nodes receive multiple signals, they also independently emit the same number of signals. This again contrasts with the behaviour of ‘informed’ nodes for communication centrality, where nodes pass on signals depending on their status as ‘informed’ nodes, not whether they have just received signals in the previous period.

In fact, these differences come down to the independence assumption, which is implicit in the formula for diffusion centrality. The independence assumption implies that the probability of a signal travelling any of the walks in a network is independent of the probability the signal travels along any of the other walks in the network. In terms of information transmission, one way to conceptualise this assumption of independence is as follows. The source node emits a signal in each round of information transmission. The other nodes pass on the signal to each other, but do not recall anything of these signals until the last round. Nodes only observe the signals that they receive in the *last* round of information transmission. As such, nodes do not learn anything until the final period, and only ‘know’ information that they receive at the end of the process.

We can also go further and say that the independence assumption effectively means that there are no such thing as ‘informed’ nodes. For example, if ex post at the end of the T rounds of information transmission, if we observed that node h in Figure 1 had learnt of a signal, with diffusion centrality, this would not imply that node g also had that knowledge. This is due to the independence assumption. This is quite different from the model of communication centrality and more generally, from the computationally complex information transmission processes modelled by Kempe et al. (2003) and Akbarpour et al. (2018), where one can, ex post, identify connected components that are linked by ‘live’ network edges.

We can therefore provide here the following revised definition of diffusion centrality. The example in Figure 1 shows in more detail the impacts of the differences in the information transmission protocol between communication centrality and diffusion centrality.

Definition 3.1 (Diffusion centrality). (*Banerjee et al., 2013*): *Node i 's diffusion centrality provides the expected total number of times that nodes in the network will receive a signal, when node i is the seed node. For a network $\mathbf{G}$, information transmission rounds T and passing probability p , diffusion centrality is given by the i th entry of the vector $\mathbf{DC}(\mathbf{G}, p, T) = \sum_{\tau=1}^T (p\mathbf{G})^\tau \mathbf{1}$. The diffusion process assumes that the probability of a signal flowing along any of the walks in the network is independent from the probability that it flows along any other walks in the network.*

Another key aspect of diffusion centrality is the way that its formula deals with the possibility that signals could travel along multiple paths in a network. By using the Katz-Bonacich summation approach, it adds together the (assumed independent) probabilities of a signal travelling along each different walk in the network. Banerjee et al. (2013) state that diffusion centrality measures the “*expected total number of times that all nodes taken together hear about the opportunity*”. As such, it is a sum of probabilities, and not a probability in itself, because using the summation formula does not take account of inclusion-exclusion principle in probability theory. This means that the probabilities of a node received a signal via multiple paths are counted multiple times, without accounting for the fact that a signal may already have been received by another path.⁴

⁴This issue is related to but distinct from correlation neglect, where nodes observe the same signal more than once through different walks in the network and erroneously treat each report as a distinct signal.

$T = 1$	$T = 2$	$T = 2$	$T = 2$	$T = 2$	$T = 3$
$CC_{ab} = p$	$CC_{ab} = 2p - p^2$	$CC_{ac} = p^2$	$CC_{ad} = 2p - 2p^3 + p^4$	$CC_{ag} = 2p^2 - p^4$	$CC_{ah} = 2p^3 - p^5$
$DC_{ab} = p$	$DC_{ab} = p$	$DC_{ac} = p^2$	$DC_{ad} = p + p^2$	$DC_{ag} = 2p^2$	$DC_{ah} = 2p^3$
$w_{ab} = p$	$w_{ab} = p$	$w_{ac} = p^2$	$w_{ad} = p + p^2 - p^3$	$w_{ag} = 2p^2 - p^4$	$w_{ah} = 2p^3 - p^6$

Figure 1: Differences between the approaches for calculating or approximating probabilities of information transmission in networks, illustrated in the above example and table.

- CC_{ij} is the probability that node j will be informed if node i is the seed node, under the information transmission protocol of communication centrality;
- DC_{ij} gives the (i, j) th entry in the diffusion centrality matrix $\sum_{\tau=1}^T (p\mathbf{G})^\tau$;
- w_{ij} gives the word-of-mouth probabilities, defined in this paper.

There are three key differences between these three approaches. Let a be the seed node. First, as shown in the second column of the table, with communication centrality, information is retained by a node if it is received at any time during the information transmission process. This is not the case for DC_{ab} and w_{ab} , where nodes only retain the information they receive in the final round of information transmission. Second, as shown in the fifth column of the table, for diffusion centrality with $T = 2$, we have that $DC_{ag} = 2p^2$. But for communication centrality, due to the inclusion-exclusion principle, the probability that a signal emitted by a is received by g would be $2p^2 - p^4$. This also appears for the new measure w_{ag} , calculated as $(1 - p^2)^2$. Third, as shown in the sixth column, there is difference between CC_{ah} and w_{ah} over longer transmission chains. This is due to the independence assumption of the diffusion process, discussed in the text. (The examples in all Figures in the paper assume an unweighted and undirected network, for clarity of exposition).

Following Banerjee et al. (2013), diffusion centrality has been used in empirical work as an approximation of the probability that information is transmitted between nodes. Example applications include models of savings monitors (Breza and Chandrasekhar, 2019) and political intermediation (Fafchamps and Labonne, 2019; Cruz et al., 2017). However, this issue of the inclusion-exclusion principle leads to two main concerns, if diffusion centrality is used as a proxy for network transmission probabilities. First is that the numbers returned by the diffusion centrality formula may sum to more than one, and need to be arbitrarily scaled in order to be used as probabilities. Second, neglecting to account for the overlapping probabilities of receiving a signal via multiple walks may lead to an overestimation of the inequality in the propensity of nodes to send and receive signals, since this distortion would be expected to be more pronounced for more central nodes which are connected to others via multiple walks.

Bramoullé and Genicot (2018) tackle the inclusion-exclusion issue with a new measure called ‘reachability’. This measure finds the ‘true’ probabilities, is computationally hard, and does not require the independence assumption. The new formula for word-of-mouth probabilities presented in this paper, in the next section, tackles the inclusion-exclusion issue while also retaining the independence assumption, and is thereby as simple to calculate as diffusion centrality. It is compared to diffusion centrality and communication centrality in Figure 1.

4 Word-of-mouth probabilities

This paper presents a new formula for approximating the probabilities of node-to-node information transmission by diffusion, using De Morgan’s laws to account for the inclusion-exclusion principle. To do so, it maintains the structure of the diffusion process given in Definition 3.1, including the independence assumption. As such, it is a new heuristic which allows us to quickly and easily approximate the *probability* that a signal is transmitted and received between the nodes of a network, under the independence assumption. We call this new measure *word-of-mouth centrality*, capturing the intuitive concept whereby information travels within a community via conversations between nodes and their connections (Ahn and Suominen, 2001; Lippert and Spagnolo, 2011).

Let Ω denote the information structure of the network such that $\Omega = \{p, T, \mathbf{G}\}$. Let

$s_i \in \{\{1\}, \emptyset\}$ be the signal emitted by node i , and $\rho_j \in \{\{1\}, \emptyset\}$ be the signal received by node j , and $w_{ij}(\mathbf{\Omega})$ be the *word-of-mouth probability* that a signal emitted by i will reach j by diffusion:

$$\Pr[\rho_j = 1 \mid s_i] = \begin{cases} w_{ij}(\mathbf{\Omega}) & \text{if } s_i = 1 \\ 0 & \text{if } s_i = \emptyset \end{cases} \quad \forall i, j \in N. \quad (4.1)$$

One of De Morgan's laws of duality states that for a family of sets $\mathbb{A} = \{A_i; i \in I\}$ in the universal set X , where I is some index set, we have that $\sim (\cup_{i \in I} A_i) = \cap_{i \in I} (\sim A_i)$.⁵ Applied to our case of network transmission of signals, this means that the complement of $w_{ij}(\mathbf{\Omega})$ is given by the probability that j does *not* hear a signal from i along *any* of the walks that connect i and j in $\mathbf{G}$. In other words, for a signal *not* to travel from i to j , we need it *not* to travel along every possible walk that connects i to j in $\mathbf{G}$, of length $\leq T$.

For the example in Figure 1 above, we can use De Morgan's laws to directly obtain w_{ag} from $1 - w_{ag} = (1 - p^2)(1 - p^2)$. More generally, for each walk of length τ , the probability that the signal does not travel along all the links in that walk by diffusion is $1 - p^\tau$. Based on this approach, next we provide the formula for word-of-mouth probabilities and examine how the three aspects of the information structure $\mathbf{\Omega}$ affect the word-of-mouth probabilities.

Definition 4.1. *Word-of-mouth probability* given by $w_{ij}(\mathbf{\Omega})$ is the probability that a signal passes from i to j by diffusion, given in Definition 3.1. For any $\mathbf{\Omega}$, we have that⁶

$$w_{ij}(\mathbf{\Omega}) = 1 - \prod_{\tau=0}^T [1 - p^\tau]^{l_{ij}(\tau, \mathbf{G})}.$$

Where $l_{ij}(\tau, \mathbf{G}) = [\mathbf{G}^\tau]_{ij}$ is the number of walks between i and j of length τ in $\mathbf{G}$.

Proposition 4.1. *The word-of-mouth probabilities are weakly increasing in p or T or in adding a link to the network $\mathbf{G}$. There are complementarities between the three aspects of the information structure $\mathbf{\Omega}$ for the diffusion probabilities. An increase in p or T or an additional link in $\mathbf{G}$ cannot lead to a decrease in any word-of-mouth probabilities.*

⁵For example, with two events, we have that $\neg(A \cup B) = \neg A \cap \neg B$

⁶Note that this formulation, particularly the stipulation that we begin at $\tau = 0$, implies that $w_{ii}(\mathbf{\Omega}) = 1$. This implicitly defines period 0 as the period in which information is seeded to the source node.

It is important to note that the independence assumption of diffusion centrality continues to apply for the word-of-mouth probabilities in Definition 4.1. For this reason, the word-of-mouth probabilities remain an approximation of the probabilities that would result from the model used for communication centrality. For example, let us return to w_{ah} in Figure 1, with $T = 3$. Node g could receive the signal from a through two walks, either via nodes e or f . With independence, the probability that he passes either of them on to h is independent, giving $w_{ah} = p^3 + p^3 - p^6$. But with communication centrality, we would have that $CC_{ah} = pw_{ag} = p(2p^2 - p^4) = 2p^3 - p^5$. This example shows that there is some loss in precision due to the independence assumption, which could perhaps be offset by its computational ease.

5 Word-of-mouth centrality and diffusion centrality

We can also construct a centrality measure using these word-of-mouth probabilities. We call this new measure *word-of-mouth centrality*.

Definition 5.1. *Word-of-mouth centrality is the average probability of information transmission by diffusion for each node in a network and is given by*

$$w_i(\Omega) = \frac{1}{n} \sum_j w_{ij}(\Omega) \quad \forall i \in N.$$

As well as diffusion centrality and communication centrality (Banerjee et al., 2013, 2019), there are several related measures of centrality, in particular: Bonacich centrality (Bonacich, 1987); information centrality (Stephenson and Zelen, 1989); random walk closeness centrality (Noh and Rieger, 2004); cascade centrality (Teytelboym et al., 2015); and percolation centrality (Moore and Newman, 2000; Piraveenan et al., 2013) in the epidemiological literature.

As described above, Bramoullé and Genicot (2018)’s computationally hard ‘reachability’ measure also provides true probabilities, and does not depend on the independence assumption for the probabilities of a signal travelling along each walk in the network.

It is also interesting to note that, in moving beyond diffusion centrality’s additively separable measure, word-of-mouth centrality responds to the call from Bloch et al. (2019): “despite the seeming abundance of centrality measures, there is still ample room for new measures that process information about nodes’ positions in novel ways - moving beyond the narrow class of additively separable measures”.

Word-of-mouth centrality ranks the nodes in a network according to their average probabilities of information transmission by diffusion. It could also be seen as providing the proportion of a population who would become infected, for each source node of initial infection. In contrast, diffusion centrality describes the expected number of times a signal or infection is received, for a given source node. Both diffusion centrality and word-of-mouth centrality have similar computational complexity.

Next we use some examples to investigate the relationship between the two measures. In their empirical work, Banerjee et al. (2013) set p equal to the inverse of the largest eigenvalue of the adjacency matrix, $\lambda_{\max}(\mathbf{G})$. They state that this “*is the critical value of p for which the entries of $[p\mathbf{G}]^T$ tend to 0 as T grows if $p < 1/\lambda_{\max}(\mathbf{G})$, and some entries diverge if $p > 1/\lambda_{\max}(\mathbf{G})$* ”. They also point out that, as T tends to infinity, diffusion centrality “*becomes proportional to either Bonacich centrality or eigenvector centrality, depending on whether p is smaller than the inverse of the first eigenvalue or exceeds it, respectively*”. The analysis in Figures 2 and 3 uses both this benchmark value of p , and a wider range of values to make these comparisons.

Figure 2 provides a cardinal comparison of the rankings of different nodes using diffusion centrality and word-of-mouth centrality. This analysis suggests that if there is any empirical reason to suspect that p differs from $1/\lambda_{\max}(\mathbf{G})$, diffusion centrality (or a linear transformation of it) may no longer be a good approximation of signal transmission probabilities. In this case, using word-of-mouth centrality may increase accuracy and reliability. Figure 3 provides an ordinal comparison of the two centralities, with an example where the word-of-mouth centrality rankings of the nodes changes as the parameter T increases, as the relative importance of shorter and longer walks changes. This change does not occur with diffusion centrality.

Figure 2: Cardinal comparison of diffusion centrality and word-of-mouth centrality, using data from one of the Indian villages studied by Banerjee et al. (2013) as an example (No. 1; $n = 182$). Centralities are calculated at the household level. The value of $p = 1/\lambda_{\max}(\mathbf{G}) = 0.059$ is used to compare the two centralities in Chart (a). At this value of p , a linear transformation of diffusion centrality appears to be a good approximation of word-of-mouth centrality. Chart (b) shows the profile of diffusion centralities with this value of p for all households in this example village. Chart (c) shows diffusion centralities for each node for a range of p between 0 and 0.5, showing an affine transformation which explodes to large numbers as p increases. Chart (d) shows the range of word-of-mouth centralities for the nodes of the example village for a range of p between 0 and 0.5. This shows an empirically meaningful variation in centralities, with changes in the shape of the centrality rankings as p increases (where p is the probability that direct neighbours in the network transmit information to each other). The word-of-mouth centralities converge to ≤ 1 at high p . In addition, it appears that the shape of the rankings in the diffusion centralities in charts (b) and (c) gives a more unequal profile than those for word-of-mouth centrality on chart (d), which might be expected because the distorting effect of multiple counting of signals would be greatest for the most central nodes.

Figure 3: Ordinal comparison of diffusion centrality and word-of-mouth centrality in an example network of 11 nodes, shown below. Charts (a)-(d) shows how the rankings of the word-of-mouth centralities for this network change as T increases. Charts (a) and (b) use $p = 1/\lambda_{\max}(\mathbf{G}) = 0.3216$ to compare the two centrality measures; Charts (c) and (d) use a slightly higher value of $p = 0.32$. We can observe that the numbers for word-of-mouth centrality stay below 1, whereas for diffusion centrality they increase above 1. There are also differences in the rankings, which change as T changes. In all 4 charts, the relative rankings of the two centralities are the same when T is low, but after this point the rankings in word-of-mouth centrality switch around. In Chart (b) node b remains the most central but c and d overtake a . In Chart (d) at a higher p , node d becomes the most central as T increases. This switching effect may be due to the fact that, at low T , short walks are more important, and node b is most central. When T is higher, longer walks are also feasible, and now node d overtakes the other nodes to have the highest centrality. This switching effect does not occur with diffusion centrality.

6 Blocked word-of-mouth probabilities

Another important issue in information transmission in networks is the potential blocking of signals, where some signals are not retransmitted by some nodes. One general reason for this blocking might be the presence of ‘antagonistic’ or negative links between nodes (Isakov et al., 2019; Everett and Borgatti, 2014). With antagonistic ties, nodes might not pass on signals that were sent from (or to) their enemies.

In a different example of blocking behaviour, in the savings monitor model of Breza and Chandrasekhar (2019), one might expect that a message about a node who did not meet their savings commitments, would not be retransmitted by that node itself. In the political favours model of Cruz et al. (2017), where voters send requests for favours to politicians, Bramoullé and Genicot (2018) point out that both the sender (voter) and receiver (politician) might not retransmit the message. They also note that, for the gossip model of Banerjee et al. (2019), where the information transmitted is “agent i first said [...]”, one would not expect this information to be retransmitted to (and hence by) i . Bramoullé and Genicot (2018) tackle this issue using ‘targeting centrality’, which finds the expected number of times that a message is transmitted, if the sender and receiver do not pass on the message, at intermediate steps in the chain.

In our framework, we can provide a general form of the word-of-mouth probabilities when signals are blocked by different nodes. In what follows, we define these *blocked word-of-mouth probabilities* when just one node blocks a signal, while Appendix A.2 shows how this approach can be generalised to more complex blocking patterns, where different nodes block the signal in different rounds of information transmission.

Let $X \subset N$ be the subset of nodes who block a signal in all rounds τ of information transmission. As before let $s_i \in \{\{1\}, \emptyset\}$ be the signal emitted by node i , and let $\rho_j \in \{\{1\}, \emptyset\}$ be the signal received by node j .

$$\Pr[\rho_j = 1 \mid s_i; X = \{k\}] = \begin{cases} p_{ij}(k, \Omega) & \text{if } s_i = 1 \\ 0 & \text{if } s_i = \emptyset \end{cases} \quad \forall i, j, k \in N \quad (6.1)$$

The *blocked word-of-mouth probability* $p_{ij}(k, \mathbf{\Omega})$ defines the probability a signal emitted by node i , will be received by node j , if it is blocked by node k . To find these probabilities, we can modify the word-of-mouth probabilities in Definition 4.1 to take account of blocking by node k . To do so, we can use the network $\mathbf{G}_{-k}$ to calculate the number of walks, which is the $n \times n$ matrix $\mathbf{G}$ with the k th row and column set to zero: that is, where k is present in the network but disconnected from all his neighbours.

Definition 6.1. *Blocked word-of-mouth probability* given by $p_{ij}(k, \mathbf{\Omega})$ is the probability that a signal passes from i to j by diffusion given in Definition 3.1, when the signal is blocked by k . For any $\mathbf{\Omega}$, we have that

$$p_{ij}(k, \mathbf{\Omega}) = 1 - \prod_{\tau=0}^T [1 - p^\tau]^{l_{ij}(\tau, \mathbf{G}_{-k})}.$$

Where $l_{ij}(\tau, \mathbf{G}_{-k}) = [\mathbf{G}_{-k}^\tau]_{ij}$ is the number of walks between i and j of length τ in $\mathbf{G}_{-k}$.

Proposition 4.1, which describes the effects of the information structure $\mathbf{\Omega}$ on word-of-mouth probabilities, also applies to the case with blocking, if we replace $\mathbf{G}$ with $\mathbf{G}_{-k}$. Since $\mathbf{G}$ has weakly more links than $\mathbf{G}_{-k}$, from Proposition 4.1, we have that $w_{ij}(\mathbf{\Omega}) \geq p_{ij}(k, \mathbf{\Omega}) \forall i, j, k$.

We can use the blocked word-of-mouth probabilities to explore how the structure of the network affects information transmission in the presence of blocking. First we define some key concepts. The concept of a ‘blocker’ relates to the effect that a node has on the ability of other nodes to communicate with each other.

Definition 6.2. Nodes i and j can **communicate** iff $w_{ij}(\mathbf{\Omega}) > 0$, or in the case of blocking, if $p_{ij}(k, \mathbf{\Omega}) > 0$. This requires $p > 0$ and, respectively $d_{ij}(\mathbf{G}) \leq T$ and $d_{ij}(\mathbf{G}_{-k}) \leq T$.

Definition 6.3. A network is **informative** iff every pair of nodes in the network can communicate, that is $w_{ij}(\mathbf{\Omega}) > 0 \forall i, j \in N$. This requires $p > 0$ and $D_{\mathbf{G}} \leq T$.

Definition 6.4. A node is a **blocker** if and only if its disconnection from the network means that two or more nodes who could previously communicate no longer can. That is, node k is a blocker iff $\exists i, j \in N$ such that $w_{ij}(\mathbf{\Omega}) > 0$ and $p_{ij}(k, \mathbf{\Omega}) = 0$. In particular, when $p > 0$, node k is not a blocker if for all $\{i, j\}$ with $d_{ij}(\mathbf{G}) \leq T$, we also have that $d_{ij}(\mathbf{G}_{-k}) \leq T$.

Figure 4: Node k cannot be a blocker in a cycle of length four (4a), but can be in a cycle of length five (4b). In both example networks above, $\mathbf{G}$ is made up of all the solid and dashed links, while $\mathbf{G}_{-k}$ includes only the solid links. In (4a), k is in a cycle of four in network $\mathbf{G}$, and $d_{lm}(\mathbf{G}) = 4$. If k is removed in network $\mathbf{G}_{-k}$, then l and m can still communicate because $d_{lm}(\mathbf{G}_{-k}) = 4$. This is because k is in a cycle of length 4: node o is connected to k 's neighbours i and j . This means that k cannot be a blocker, no matter what the value of T is. On the other hand, (4b) shows a case where k could be a blocker, depending on the value of T . If $T = 4$, nodes l and m could communicate in the network $\mathbf{G}$ via k , where $d_{lm}(\mathbf{G}) = 4$, but cannot when he is removed in $\mathbf{G}_{-k}$, because $d_{lm}(\mathbf{G}_{-k}) = 5$. This means that in network (4b), node k is a blocker when $T = 4$.

Remark 6.1. Assume that the network is informative. If k has only one neighbour, k is not a blocker. If k has more than one neighbour, then k is not a blocker if for each pair of k 's neighbours $i, j \in \mathcal{N}_k$, they share another common neighbour, or are connected to each other. This sufficient condition implies that for each pair of k 's neighbours $i, j \in \mathcal{N}_k$, there is a cycle of length ≤ 4 which includes the sequence (i, k, j) . Alternatively, it also implies that for all $l, m \in N$ who have the sequence (i, k, j) as part of the shortest path(s) between them, we have that $d_{lm}(\mathbf{G}_{-k}) = d_{lm}(\mathbf{G})$.

Figure 4 shows an example of this result. The length of the cycles in the two example networks determines whether or not k 's neighbours can still communicate, even when k is blocking signals. The importance of cycles of length four, our sufficient condition for a node not to be a blocker, recalls well-known results on the importance of network cycles of length three: Coleman's (1988) *closure*; and Jackson et al.'s (2012) *support*.

Definition 6.5. A node k is **visible** if and only if $p_{ij}(k, \Omega) > 0 \quad \forall i, j \in N \setminus k$: that is, if every pair of nodes can still communicate, even when node k is blocking the signal. This requires $p > 0$ and $D_{\mathbf{G}_{-k}} \leq T$.

Definition 6.6. *A network is **blocked informative** if and only if every node is visible. This requires $p > 0$ and $D_{\mathbf{G}_{-k}} \leq T \quad \forall k \in N$. If also $p = 1$, there is full information.*

The visibility of an individual node depends on the network structure when that node is disconnected: the network $\mathbf{G}_{-k}$. Of critical importance is the relative length of social distances in $\mathbf{G}_{-k}$, compared to the total number of steps that a signal can travel, T . Remark 6.2 describes the conditions for visibility, and also links visibility to the concept of 2-connectedness (Renault and Tomala, 1998; Wolitzky, 2015). 2-connectedness relates to the presence of cycles in the network, which can be important for successful outcomes in different settings, for example mutual credit systems (Iosifidis et al., 2018).

Remark 6.2. Assume that the network is informative. A necessary condition for k to be visible is that $\mathbf{G}_{-k}$ is connected: that is, that $\mathbf{G}$ is 2-connected with respect to k . This is because a visible node does not disconnect the network by his absence. A sufficient condition for k to be visible is that $D_{\mathbf{G}_{-k}} \leq T$.

For example, only the leaf nodes of a tree network could be visible, because every non-leaf node connects the network. The blocked word-of-mouth probabilities can also be used to construct another centrality measure, blocking centrality, which shows the extent to which information transmission in a network is reduced, if a given node blocks the signals sent by others. Figure 5 compares the three centrality measures that we have discussed: diffusion, word-of-mouth, and blocking centrality.

Definition 6.7. *Blocking centrality* $\Psi_k(\Omega)$ *is the proportional reduction in the average probability of network transmission by diffusion, if node k blocks signals sent by other nodes, given by:*

$$\Psi_k(\Omega) = \frac{w(\Omega) - P(k, \Omega)}{w(\Omega)}.$$

Where $w(\Omega) = \frac{1}{n} \sum_i w_i(\Omega)$ is the average probability of information diffusion in the network $\mathbf{G}$; and $P(k, \Omega) = \frac{1}{n^2} \sum_i \sum_j p_{ij}(k, \Omega)$ is the average probability of information diffusion in the network $\mathbf{G}$ when node k blocks all signals.

Figure 5: Comparing centrality measures in a line network. The figure uses a line network of 8 nodes, and the benchmark $p = 1/\lambda_{\max}(\mathbf{G}) = 0.53$, and $T = 7$. We can observe that diffusion centrality gives the highest numbers, exceeding 1 for the centre, and with the greatest difference between the centre and the ends. Word-of-mouth centrality is lower and with less inequality. For these parameters, seeding a signal with either of the central nodes would give around 60 per cent average probability that the signal would be received by the other nodes. Blocking centrality is also highest in the centre: if the central nodes block signals, the average probability of information transmission between nodes in the network is reduced by around 40 per cent.

7 Conclusion

This paper has provided a new heuristic for the probabilities of node-to-node diffusion in networks. The simple formula is quick to calculate and accounts for the inclusion-exclusion principle, which is omitted when diffusion centrality is used as a heuristic for these probabilities. This omission can lead to an overestimate of the inequality of diffusion capabilities of the nodes of a network, and can affect the ranking of nodes. The new measure also allows us to use any level of p , the link probability of diffusion, which may help in empirical applications. These probabilities can be used to generate a new centrality measure called word-of-mouth centrality, which provides the average probability that each node can inform the rest of the network, when the independence assumption holds.

The measure can also be used to examine the case of blocking, where certain nodes may not pass on certain signals. It provides two new concepts. First, visibility, where nodes cannot prevent others from communicating, which is linked to 2-connectedness. Second, a node is a blocker if it can prevent others from communicating, which is linked to cycles of length 4. The analysis provides another centrality measure — blocking centrality — which shows how information transmission in a network is affected if a given node blocks signals sent by other members of their community.

Word-of-mouth probabilities, and the centrality measures they provide, could be useful to both empiricists and theorists, as a simple new heuristic for the probabilities of node-to-node diffusion in fixed networks.

A Appendix

A.1 Proof of Proposition 4.1

1. $w_{ij}(\Omega)$ is increasing in p if and only if $\exists \tau \leq T$ such that $l_{ij}(\tau, \mathbf{G}) \geq 1$.

We have that $\frac{\partial w_{ij}(p, T, \mathbf{G})}{\partial p} \geq 0$ with a strict inequality if and only if $l_{ij}(\tau, \mathbf{G}) > 0$ for any $\tau \leq T$, that is, if and only if i and j are connected by one or more walks of length $\leq T$ in the network $\mathbf{G}$.

2. $w_{ij}(\Omega)$ increases as T increases to $T + 1$ if and only if $l_{ij}(T + 1, \mathbf{G}) \geq 1$. Let

$$F_{ij}(p, T, \mathbf{G}) = 1 - w_{ij}(\Omega) = \prod_{\tau=1}^T [1 - p^\tau]^{l_{ij}(\tau, \mathbf{G})}. \quad (\text{A.1})$$

If T increases to $T + 1$, we have that

$$F_{ij}(p, T + 1, \mathbf{G}) = F_{ij}(p, T, \mathbf{G})(1 - p^{(T+1)})^{l_{ij}(T+1, \mathbf{G})}. \quad (\text{A.2})$$

So $F_{ij}(p, T + 1, \mathbf{G}) < F_{ij}(p, T, \mathbf{G})$ if and only if $l_{ij}(T + 1, \mathbf{G}) \geq 1$; otherwise they are equal. The same argument holds for any further increases in T . Hence $F_{ij}(\Omega)$ is weakly monotonically decreasing in T and so $w_{ij}(\Omega)$ is weakly monotonically increasing in T .

3. $w_{ij}(\Omega)$ increases as a link is added to $\mathbf{G}$ if and only if the new link leads to an increase in $l_{ij}(\tau, \mathbf{G})$ for any $\tau \leq T$.

Suppose that we add a link to the network $\mathbf{G}$, creating the network $\mathbf{G}'$, which has an additional walk of length τ between i and j so that $l_{ij}(\tau, \mathbf{G}') = l_{ij}(\tau, \mathbf{G}) + 1$. Now we compare $F_{ij}(p, T, \mathbf{G}')$ and $F_{ij}(p, T, \mathbf{G})$:

$$F_{ij}(p, T, \mathbf{G}') = F_{ij}(p, T, \mathbf{G})(1 - p^\tau). \quad (\text{A.3})$$

We have that $F_{ij}(p, T, \mathbf{G}') < F_{ij}(p, T, \mathbf{G})$ as required. Meanwhile any change to the network that leaves $l_{ij}(\tau)$ unchanged $\forall \tau \leq T$ leaves the probabilities unchanged.

A.2 Obstruction by subsets of nodes in different rounds of information transmission

Now let $X_\tau \subset N$ be the subset of nodes who obstruct a signal in round τ of information transmission, and let $\mathbf{X} = \{X_\tau\}_{1 \leq \tau \leq T}$. Let $l_{ij}(\tau, \mathbf{G}, \mathbf{X})$ be the number of walks between i and j of length τ when the set of obstructing nodes is $\mathbf{X}$. So we have that, for example, $l_{ij}(1, \mathbf{G}, \mathbf{X}) = [\mathbf{G}_{-X_1}]_{ij}$ where $\mathbf{G}_{-X_1}$ is the network $\mathbf{G}$ with those nodes in X_1 disconnected. Then we have that $l_{ij}(2, \mathbf{G}, \mathbf{X}) = [\mathbf{G}_{-X_2} \mathbf{G}_{-X_1}]_{ij}$ and $l_{ij}(3, \mathbf{G}, \mathbf{X}) = [\mathbf{G}_{-X_3} \mathbf{G}_{-X_2} \mathbf{G}_{-X_1}]_{ij}$, and so on. In general, $l_{ij}(\tau, \mathbf{G}, \mathbf{X}) = [\prod_{\tau=1}^\tau \mathbf{G}_{-X_\tau}]_{ij}$, ensuring that the ordering of matrices $\mathbf{G}$ is preserved. We can use these walk lengths in the formula for obstructed diffusion probability in Definition 6.1 to generalise to any set of obstructing nodes.

References

- Acemoglu, Daron, Kostas Bimpikis, and Asuman Ozdaglar (2014) “Dynamics of information exchange in endogenous social networks,” *Theoretical Economics*, Vol. 9, pp. 41–97.
- Ahn, Illtae and Matti Suominen (2001) “Word-Of-Mouth Communication and Community Enforcement,” *International Economic Review*, Vol. 42, pp. 399–415.
- Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi (2018) “Just a few seeds more: value of network information for diffusion,” *Available at SSRN 3062830*.
- Bailey, Norman T. J. (1957) *Mathematical Theory of Epidemics*: Arnold.
- Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson (2013) “The Diffusion of Microfinance,” *Science*, Vol. 341, p. 1236498.

- (2019) “Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials,” *The Review of Economic Studies*, Vol. 86, pp. 2453–2490.
- Bloch, Francis, Matthew O. Jackson, and Pietro Tebaldi (2019) “Centrality Measures in Networks,” SSRN Scholarly Paper ID 2749124, Social Science Research Network, Rochester, NY.
- Bonacich, Phillip (1987) “Power and Centrality: A Family of Measures,” *American Journal of Sociology*, Vol. 92, pp. 1170–1182.
- Borgatti, Stephen P. (2005) “Centrality and network flow,” *Social Networks*, Vol. 27, pp. 55–71.
- Bramoullé, Yann and Garance Genicot (2018) “Diffusion Centrality: Foundations and Extensions.”
- Breza, Emily and Arun G. Chandrasekhar (2019) “Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment,” *Econometrica*, Vol. 87, pp. 175–216.
- Breza, Emily, Arun G. Chandrasekhar, Tyler H. McCormick, and Mengjie Pan (2020) “Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data,” *American Economic Review*, Vol. 110, pp. 2454–2484.
- Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet (2015) “Social Networks and the Decision to Insure,” *American Economic Journal: Applied Economics*, Vol. 7, pp. 81–108.
- Christakis, Nicholas A. and James H. Fowler (2007) “The Spread of Obesity in a Large Social Network over 32 Years,” *New England Journal of Medicine*, Vol. 357, pp. 370–379, Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMsa066082>.
- (2010) “Social Network Sensors for Early Detection of Contagious Outbreaks,” *PLOS ONE*, Vol. 5, p. e12948, Publisher: Public Library of Science.
- Coleman, James S (1988) “Social capital in the creation of human capital,” *American journal of sociology*, pp. S95–S120.

- Crawford, Vincent P. and Joel Sobel (1982) “Strategic Information Transmission,” *Econometrica*, Vol. 50, pp. 1431–1451.
- Cruz, Cesi, Julien Labonne, and Pablo Querubín (2017) “Politician Family Networks and Electoral Outcomes: Evidence from the Philippines,” *American Economic Review*, Vol. 107, pp. 3006–3037.
- Degroot, Morris H. (1974) “Reaching a Consensus,” *Journal of the American Statistical Association*, Vol. 69, pp. 118–121.
- DiMaggio, Paul and Filiz Garip (2011) “How Network Externalities Can Exacerbate Inter-group Inequality,” *American Journal of Sociology*, Vol. 116, pp. 1887–1933, Publisher: The University of Chicago Press.
- Everett, Martin G. and Stephen P. Borgatti (2014) “Networks containing negative ties,” *Social Networks*, Vol. 38, pp. 111–120.
- Fafchamps, Marcel and Julien Labonne (2019) “Family Networks and Distributive Politics,” *Journal of the European Economic Association*.
- Galeotti, Andrea, Christian Ghiglino, and Francesco Squintani (2013) “Strategic information transmission networks,” *Journal of Economic Theory*, Vol. 148, pp. 1751–1769.
- Golub, Benjamin and Matthew O. Jackson (2012) “How Homophily Affects the Speed of Learning and Best-Response Dynamics,” *The Quarterly Journal of Economics*, Vol. 127, pp. 1287–1338.
- Hagenbach, Jeanne and Frédéric Koessler (2010) “Strategic communication networks,” *The Review of Economic Studies*, Vol. 77, pp. 1072–1099.
- Iosifidis, George, Yanick Charette, Edoardo M. Airoidi, Giuseppe Littera, Leandros Tassioulas, and Nicholas A. Christakis (2018) “Cyclic motifs in the Sardex monetary network,” *Nature Human Behaviour*, Vol. 2, pp. 822–829, Number: 11 Publisher: Nature Publishing Group.
- Isakov, Alexander, James H. Fowler, Edoardo M. Airoidi, and Nicholas A. Christakis (2019) “The Structure of Negative Social Ties in Rural Village Networks,” *Sociological Science*, Vol. 6, pp. 197–218.

- Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011) “Opinion Leadership and Social Contagion in New Product Diffusion,” *Marketing Science*, Vol. 30, pp. 195–212, Publisher: INFORMS.
- Jackson, Matthew O. (2017) “A Typology of Social Capital and Associated Network Measures,” *arXiv:1711.09504 [physics]*, arXiv: 1711.09504.
- Jackson, Matthew O., Tomas Rodriguez-Barraquer, and Xu Tan (2012) “Social capital and social quilts: Network patterns of favor exchange,” *The American Economic Review*, Vol. 102, pp. 1857–1897.
- Kempe, David, Jon Kleinberg, and Éva Tardos (2003) “Maximizing the spread of influence through a social network,” in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146: ACM.
- Kermack, William and Anderson McKendrick (1927) “A contribution to the mathematical theory of epidemics,” *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, Vol. 115.
- Kim, David A, Alison R Hwong, Derek Stafford, D Alex Hughes, A James O’Malley, James H Fowler, and Nicholas A Christakis (2015) “Social network targeting to maximise population behaviour change: a cluster randomised controlled trial,” *The Lancet*, Vol. 386, pp. 145–153.
- King, Maia (2017) “Whom can you trust? Reputation and Cooperation in Networks,” *School of Economics and Finance, QMUL*.
- Levy, Gilat and Ronny Razin (2014) “A simple Bayesian heuristic for social learning and groupthink.”
- Lippert, Steffen and Giancarlo Spagnolo (2011) “Networks of relations and word-of-mouth communication,” *Games and Economic Behavior*, Vol. 72, pp. 202–217.
- Moore, Cristopher and Mark E.J. Newman (2000) “Epidemics and percolation in small-world networks,” *Physical Review E*, Vol. 61, p. 5678.
- Mueller-Frank, Manuel and Claudia Neri (2015) “Quasi-Bayesian updating in social networks,” *Available at SSRN*.

- Noh, Jae Dong and Heiko Rieger (2004) “Random Walks on Complex Networks,” *Physical Review Letters*, Vol. 92.
- Piraveenan, Mahendra, Mikhail Prokopenko, and Liaquat Hossain (2013) “Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks,” *PLoS ONE*, Vol. 8, p. e53095.
- Renault, Jerome and Tristan Tomala (1998) “Repeated proximity games,” *International Journal of Game Theory*, Vol. 27, pp. 539–559.
- Richardson, Matthew and Pedro Domingos (2001) “Mining Knowledge-Sharing Sites for Viral Marketing,” p. 10.
- Rogers, Everett M. (2010) *Diffusion of Innovations, 4th Edition*: Simon and Schuster, Google-Books-ID: v1ii4QsB7jIC.
- Rose, Christiern (2019) “Optimal injection points for information diffusion,” *Economics Letters*, Vol. 175, pp. 67–70.
- Rossmann, Gabriel and Jacob C. Fisher (2021) “Network hubs cease to be influential in the presence of low levels of advertising,” *Proceedings of the National Academy of Sciences*, Vol. 118, p. e2013391118, Publisher: Proceedings of the National Academy of Sciences.
- Shakya, Holly B., Derek Stafford, D. Alex Hughes, Thomas Keegan, Rennie Negron, Jai Broome, Mark McKnight, Liza Nicoll, Jennifer Nelson, Emma Iriarte, Maria Ordonez, Edo Airoidi, James H. Fowler, and Nicholas A. Christakis (2017) “Exploiting social influence to magnify population-level behaviour change in maternal and child health: study protocol for a randomised controlled trial of network targeting algorithms in rural Honduras,” *BMJ Open*, Vol. 7, p. e012996, Publisher: British Medical Journal Publishing Group Section: Global health.
- Stephenson, Karen and Marvin Zelen (1989) “Rethinking centrality: Methods and examples,” *Social Networks*, Vol. 11, pp. 1–37.
- Teytelboym, Alex, Yongwhan Lim, and Asu Ozdaglar (2015) “A simple model of cascades in networks.”
- Wolitzky, Alexander (2015) “Communication with tokens in repeated games on networks,” *Theoretical Economics*, Vol. 10, pp. 67–101.