



**HAL**  
open science

## Diffusion and targeting centrality

Yann Bramoullé, Garance Genicot

► **To cite this version:**

Yann Bramoullé, Garance Genicot. Diffusion and targeting centrality. *Journal of Economic Theory*, 2024, 222, pp.106687. 10.1016/j.jet.2024.105920 . hal-04718273

**HAL Id: hal-04718273**

**<https://hal.science/hal-04718273v1>**

Submitted on 2 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Diffusion and targeting centrality <sup>☆</sup>

Yann Bramoullé <sup>a,\*</sup>, Garance Genicot <sup>b</sup>

<sup>a</sup> Aix-Marseille University, CNRS, Aix-Marseille School of Economics, France

<sup>b</sup> Department of Economics, Georgetown University, United States of America

## ARTICLE INFO

*JEL classification:*  
D83

*Keywords:*  
Diffusion  
Centrality  
Political intermediation  
Targeting

## ABSTRACT

This paper studies the dynamics of information diffusion within networks, encompassing both general and targeted dissemination. We first characterize the theoretical foundations of diffusion centrality. Next, we introduce two extensions of diffusion centrality: *targeting centrality* and *reachability*, that we believe to better capture situations involving targeted requests. We derive general explicit formulas for the computation of these novel centrality measures.

## 1. Introduction

Diffusion in networks plays a critical role in our interconnected societies, underlying the propagation of not only germs and technologies, but also ideas, misinformation, and data. In an important study, Banerjee et al. (2013) introduce a new metric termed *diffusion centrality*. This metric aims to quantify the extent to which a given piece of information disseminates among agents within a network over time. Specifically, diffusion centrality captures the expected frequency with which all agents become exposed to the given information within a stylized model of diffusion. Diffusion centrality admits a simple expression involving discounted sums of powers of the network's adjacency matrix, and nests eigenvector and Katz-Bonacich centralities. Banerjee et al. (2013) show that diffusion centrality performs well, empirically, in explaining take-up rates of a microfinance loan program in rural India. This notion has become a cornerstone in the research on diffusion.<sup>1</sup>

The first contribution of this paper is to characterize the precise theoretical foundations of diffusion centrality. Despite the success of this metric, the exact model of information diffusion that allows the interpretation of diffusion centrality as the expected frequency with which all agents become exposed to the specific information is only partially described in Banerjee et al. (2013). This can be a source of confusion (see the discussion in Appendix A). A *network diffusion process* must characterize how individuals transmit information to their neighbors, contingent upon the historical record of received messages. Section 3 characterizes the unique anonymous diffusion process that underpins the concept of diffusion centrality. This is a diffusion process in which individuals transmit indepen-

<sup>☆</sup> For financial support, Yann Bramoullé thanks the European Research Council (Consolidator Grant n. 616442), the French National Research Agency (grant ANR-17-EURE-0020) and the Excellence Initiative of Aix-Marseille University, A\*MIDEX. Garance Genicot acknowledges NSF grant SES 2242288 and is grateful to the Aix-Marseille School of Economics for their warm hospitality during the writing of this paper. We thank participants in seminars and conferences, as well as the editor and referees for their comments.

\* Corresponding author.

E-mail address: [yann.bramouille@univ-amu.fr](mailto:yann.bramouille@univ-amu.fr) (Y. Bramoullé).

<sup>1</sup> Banerjee et al. (2019) show how diffusion centrality correlates with one's effectiveness at propagating information. Duarte et al. (2019) use the same process of information diffusion to explain the information that vote brokers have on voters in Paraguay. Drago et al. (2020) investigate the diffusion of compliance behavior in local networks in Austria, probing the relationship between spillovers onto neighboring households and the diffusion centrality of the treated household. Additionally, Rose (2019) delves into optimal seeding strategies to maximize a function of the nodes' diffusion centralities.

<https://doi.org/10.1016/j.jet.2024.105920>

Received 8 September 2023; Received in revised form 18 July 2024; Accepted 22 September 2024

Available online 1 October 2024

0022-0531/© 2024 The Author(s).

Published by Elsevier Inc.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

dently all the messages received in the previous period to each of their neighbors. Messages received before the previous period are ignored, and messages received multiple times in the previous period are transmitted the same number of times.

Note that this diffusion process assumes that all agents, including the initial sender and the target, play a similar role in the diffusion process. This assumption can appear incongruous within scenarios where the conveyed information explicitly reveals the identities of these parties. Consider the example of bullying. Would it be reasonable to assume that the target of bullying spreads hurtful rumors about themselves? Identifying the individuals who are the most powerful bullies or the most susceptible to victimization is a recurring concern in academic and policy debates (see among others Atay et al. 2022<sup>2</sup> and Kaufman et al. 2021). Who we expect to spread gossip will impact predictions regarding the central individuals. In another context, envision a specific individual seeking employment making a request. It is improbable that others would echo this individual's job request back to them. Another scenario emerges when contemplating citizens appealing for favors from politicians, a concept explored in Cruz et al. (2017)'s model of political intermediation.<sup>3</sup> In this framework, the shared information entails statements like "agent  $i$  requires a favor from politician  $j$ ". In this context, presuming no retransmission to the sender and by the target might align more naturally with the dynamics at play.

Our second contribution is therefore to introduce and analyze centrality measures when the sender and/or the target do not retransmit information during diffusion. In Section 4, we define the *targeting centrality* of an agent as the expected number of times they receive others' requests for favors under no retransmission by either sender or target. We derive analytical formulas for targeting centrality in infinite and finite time, and for centrality measures under no retransmission by the sender only, by the target only, or both. Our computations involve elementary matrix operations over powers of the adjacency matrix of the network and can be programmed easily.

Observe that removing the retransmission of information by either the sender or the target introduces an inherent asymmetry. In an undirected network, governed by the information diffusion process that underlies diffusion centrality, an individual's centrality in receiving messages corresponds to their capacity to disseminate messages throughout the network. However, the removal of transmission by the target alone, as seen in the case of bullying, disrupts this symmetry. Those individuals with the highest centrality as potential targets of bullying will generally differ from those most central in terms of their capability to target others.

We then explore the relationship between diffusion and targeting centrality in infinite time. When the probability of information transmission is low, all that matters for ranking individuals in terms of centrality is their degree. In this case, removing retransmission by the sender and/or the target will make little difference in the ranking of individuals in terms of centrality. However, this dynamic shifts when the probability of information transmission is high. In Section 5.2, we show that rankings of nodes according to their targeting and diffusion centrality diverge when the target does not resend, yet converge when the sender does not transmit.

The impact of no retransmission by both the sender and the target amalgamates these two influences, and therefore depends on the particular network considered. Using numerical simulations on Erdős-Renyi random graphs, we find that the correlation and rank correlation between diffusion centrality and targeting centrality are typically low, and can even be negative when the probability of information transmission is high. This phenomenon holds for instance in the presence of bridges, since the absence of retransmission by bridge agents has a first-order impact on diffusion.

Lastly, we turn our attention to an alternative centrality metric: reachability. Diffusion and targeting centrality gauge the anticipated frequency with which information instigated by the sender reaches a target. However, when a sender requests a favor from the target, the favor's fulfillment often occurs only once. This scenario arises, for example, when a voter seeks employment or a politician strives to persuade a voter's decision. The question is then whether the likelihood of granting the favor depends on whether the information is received or on the number of times it is received. Our ultimate contribution, detailed in Section 6, encompasses a comprehensive formula for computing an agent's *reachability* - the sum of the probabilities that targeted messages will reach her - based on the inclusion-exclusion principle. Furthermore, we establish a formal connection between reachability and targeting centrality. Unsurprisingly, however, the formula for reachability is combinatorially complex.

## 2. Literature

Our analysis contributes to the literature on diffusion in networks. The structure of interactions among agents has a first-order impact on diffusion in many contexts, such as information, requests, ideas, new technologies, and infectious diseases, see Lamberson (2016), Jackson et al. (2017). One branch of the literature analyzes how long-run diffusion outcomes depend on macro features of the network such as density and degree distribution, see e.g. Jackson and Rogers (2007), López-Pintado (2008). Our analysis belongs to a second branch, which aims at understanding the relative importance of different nodes in diffusion.<sup>4</sup> This understanding is key to interventions aimed at expanding diffusion, such as optimal seeding, or reducing it, as with targeted immunization. In a benchmark model where one node initially receives some information, diffusion centrality captures the expected number of times all agents eventually hear about it. Banerjee et al. (2013) show that in the context of the adoption of a microfinance program in India, diffusion centrality is highly correlated with communication centrality, a different measure capturing the expected number of eventual adopters in a more complex model, and also performs well - empirically - in explaining actual take-up rates. Banerjee et al.

<sup>2</sup> Atay et al. (2022) models bullying as a crime activity with social conformity as in Ballester et al. (2006) and Ballester and Zenou (2012).

<sup>3</sup> Their model predicts that candidates with higher diffusion centrality should receive more votes. They show empirically that the eigenvector centrality of a politician in the family network is indeed associated with a higher vote share in data from the Philippines.

<sup>4</sup> Another branch studies how the impact of the network depends on properties of the diffusion process, such as whether contagion is "simple" or "complex", see Beaman et al. (2021), Guilbeault and Centola (2021).

(2013) justify diffusion centrality as an easy-to-compute approximation to communication centrality.<sup>5</sup> Diffusion centrality may also matter on its own, however, for instance, if agents are more likely to adopt a new technology when they repeatedly hear about it. In any case, diffusion centrality has become a reference measure in diffusion contexts, including those with targeted requests.

Our contribution is threefold. First, we characterize the precise theoretical foundations behind diffusion centrality. Second, we introduce targeting centrality, a new measure extending diffusion centrality to contexts with no retransmission by the sender and/or by the recipient. We provide explicit formulas for targeting centrality. Third, we introduce reachability to capture the expected number of eventually informed agents, and also provide an explicit formula for it.

Our analysis also contributes to the literature on centrality measures, an important part of network science. The centrality of a node reflects the relative influence of the node in a given network context. In economics, Ballester et al. (2006) first uncovered a deep connection between Katz-Bonacich centrality, from Katz (1953) and Bonacich (1987), and equilibrium action in network games. In general, researchers strive to understand which centrality measure matters in which context. This has given rise to a large, growing, and varied set of measures. Researchers have tried to discipline the analysis of centrality in several ways, one of which is to propose unifying axiomatic approaches. Dequiedt and Zenou (2017) develop an axiomatic approach for measures where the centrality of an agent is recursively related to the centralities of their neighbors. Bloch et al. (2023) propose a taxonomy of prominent centrality measures built as weighted averages of nodal statistics. Sadler (2022) develops an ordinal axiomatic approach based on recursive monotonicity.

An alternative has been to develop new measures extending classical indices. Banerjee et al. (2013) introduced diffusion centrality to capture the extent to which a piece of information given to an agent eventually diffuses in a network, and diffusion centrality nests eigenvector and Katz-Bonacich centralities. We build on this approach and propose two new measures extending diffusion centrality, targeting centrality and reachability. Targeting centrality generalizes diffusion centrality to contexts where the sender and/or the recipient do not retransmit the information during the diffusion process. Reachability captures the probability that the information reaches the recipient, rather than the number of times the recipient hears about it. We obtain explicit formulas to compute both measures and show that diffusion and targeting centrality may yield very different rankings over nodes.

### 3. Information diffusion in networks

As previously mentioned, diffusion centrality, as introduced by Banerjee et al. (2013), is a prominent centrality measure within the context of information diffusion. However, a certain level of ambiguity revolves around the underlying model of information diffusion, discussed in Appendix A. This section aims to characterize the exact theoretical foundations of diffusion centrality.

#### 3.1. Setup

A finite number  $n$  of agents are embedded in a fixed social network.  $G$  denotes the adjacency matrix, where an element  $g_{ij} = 1$  if  $i$  is linked with  $j \neq i$  and  $g_{ij} = 0$  otherwise. The network may be directed and, by convention,  $g_{ii} = 0$ .

At period 0, a designated agent, referred to as the *sender*, becomes the recipient of pertinent information. Subsequently, this information propagates *across the network through discrete iterations*. Within each iteration, individuals possessing the information can dispatch messages containing it to their immediate neighbors. The transmission of messages follows a stochastic process characterized by independent and identically distributed (iid) failures: each message sent is transmitted with probability  $\alpha$ , with  $0 < \alpha < 1$ .

Take a sender  $i$  and a possible *recipient* for the information  $j$ . For  $t \geq 1$ , denote by  $x_{ij}(t, G)$  and  $n_{ij}(t, G)$  the expected number of times the recipient  $j$  hears about the information at exactly time  $t$  and by time  $t$  respectively. That is,

$$n_{ij}(t, G) = \sum_{s=1}^t x_{ij}(s, G), \tag{1}$$

and taking  $t$  to the limit,  $n_{ij}(G) = \lim_{t \rightarrow \infty} n_{ij}(t, G)$ . Naturally, what these expressions correspond to depends on the exact information diffusion process.

#### 3.2. Diffusion processes in networks

Let  $T_{jk}^t$  be the number of messages sent to  $k$  by  $j$  at time  $t$ , and  $R_{jk}^t$  be the number of messages actually received by  $k$  from  $j$  at time  $t$ . To represent the initial seeding of information, we set  $R_{ii}^0 = 1$  where  $i$  is the sender and  $R_{jk}^0 = 0$  for all  $jk \neq ii$ . Since each message sent is received with probability  $\alpha$ , we have that

$$\mathbb{E}(R_{jk}^t | \{T_{jk}^s\}_{s \leq t}) = \alpha T_{jk}^t. \tag{2}$$

To fully specify the process of information diffusion, we must make assumptions on *who* transmits the information and *how often*.

A *network diffusion process* characterizes the way in which individuals transmit information to their neighbors, contingent upon the historical record of received messages.

<sup>5</sup> King (2024) proposes a new approximation for the probabilities of node-to-node network diffusion, based on the inclusion-exclusion principle.

**Definition.** A network diffusion process is deemed anonymous when the same information is sent deterministically to all neighbors and the information sent is unaffected by the identities and positions of individuals within the network.

In anonymous diffusion processes, individuals disseminate identical information to all of their neighbors, and the way information propagates does not depend on the sources of the information. In expectation, all resulting differences in the information received by different agents from a given sender stem purely from the architecture of the network. Note that this form of anonymity applies to both the *sender* and *recipient*: they participate in the retransmission of information just like any other agent. We will relax this particular assumption in the subsequent sections.

Denote by  $R_k^t = \sum_j R_{jk}^t$  the total number of messages received by  $k$  at time  $t$  and  $\mathbf{R}^t$  the associated vector. Let  $\mathcal{H}^t$  be the histories of the number of messages received by individuals in each period up to  $t$ :

$$\mathcal{H}^t := (\mathcal{H}^{t-1}, \mathbf{R}^t) \text{ and } \mathcal{H}_k^t := (\mathcal{H}_k^{t-1}, R_k^t).$$

An *anonymous diffusion process* is a series of functions  $\{f^t(\mathcal{H}_k^{t-1})\}_t$  such that the number of messages sent by  $k$  to each agent  $l$  at time  $t$  is

$$T_{kl}^t = g_{kl} f^t(\mathcal{H}_k^{t-1}). \tag{3}$$

The function  $f^t$  captures two key dimensions of the transmission process: (1) the extent to which the historical record of received messages matters, and (2) the number of transmitted messages. While anonymity restricts the spectrum of considered diffusion processes, it nevertheless permits a diverse array of potential scenarios. Information transmission can encompass intricate dependencies on the historical progression of the message influx across different periods.

A natural particular case is when history matters through an accumulated stock depreciating at a rate  $\delta$ . The number of messages sent at time  $t$  by agent  $k$  to each neighbor is then equal to  $f(\sum_{s=1}^{t-1} R_k^s \delta^{t-1-s})$  for some function  $f$ . The cases of  $\delta = 1$  and  $\delta = 0$  capture two benchmark assumptions regarding historical dependence. When  $\delta = 1$ , there is no depreciation and  $\sum_{s=1}^{t-1} R_k^s \delta^{t-1-s} = \sum_{s=1}^{t-1} R_k^s$ . All messages received in the past carry the same weight independently of when they were received, a form of perfect recall. When  $\delta = 0$ ,  $\sum_{s=1}^{t-1} R_k^s \delta^{t-1-s} = R_k^{t-1}$ . Messages transmitted at  $t$  only depend on messages received at  $t - 1$ , equivalent to one-period recall. Intermediate cases where  $0 < \delta < 1$  capture situations where messages received recently have more weight in current transmission than messages received in the distant past. Anonymous diffusion processes cover these three scenarios, as well as more complex ones such as transmitting information received in a specific time interval.

### 3.3. Diffusion centrality

Banerjee et al. (2013) define *diffusion centrality* as  $DC_i(t, G) := \sum_j dc_{ij}(t, G)$  where:

$$dc_{ij}(t, G) := \sum_{s=1}^t \alpha^s [G^s]_{ij} \quad \forall t \geq 1, \tag{4}$$

and interpret  $dc_{ij}(t, G)$  as the expected number of times  $j$  receives by time  $t$  the information originating in sender  $i$ .

We would like to make three concise remarks regarding diffusion centrality.

First, in undirected graphs, diffusion centrality exhibits symmetry:

$$\sum_j dc_{ij}(t, G) = \sum_j dc_{ji}(t, G).$$

This means that the number of messages reaching every node when node  $i$  is the sender is equal to the number of messages reaching node  $i$  when all nodes act as senders and node  $i$  is the recipient.

Second, when the number of periods extends to infinity, diffusion centrality is well-defined only for values of  $\alpha$  that are smaller than the inverse of the largest eigenvalue of the network's adjacency matrix, denoted by  $\lambda_1(G)$ . When  $\alpha < \alpha_{\max} \equiv 1/\lambda_1(G)$ , we have:

$$dc_{ij}(G) = \sum_{s=1}^{\infty} \alpha^s [G^s]_{ij} = [\alpha G(I - \alpha G)^{-1}]_{ij}. \tag{5}$$

Third, the elements of  $G^s$ , denoted as  $[G^s]_{ij}$ , represent the count of walks of length  $s$  between nodes  $i$  and  $j$ , as explained in Duncan (2004).<sup>6</sup> Let  $W_{ij}(s, G)$  be the set of walks of length  $s$  connecting  $i$  to  $j$  in  $G$ ,  $W_{ij}(G)$  be the set of all walks connecting  $i$  to  $j$ , and  $l(w)$  be the length of a walk  $w$ . Consequently, we can express:

$$dc_{ij}(t, G) = \sum_{s=1}^t \sum_{w \in W_{ij}(s, G)} \alpha^{l(w)} \quad \& \quad dc_{ij}(G) = \sum_{w \in W_{ij}(G)} \alpha^{l(w)}. \tag{6}$$

<sup>6</sup> A walk of length  $t$  in  $G$  connecting  $i$  to  $j$  is a set of  $t + 1$  agents  $i_1 = i, i_2, \dots, i_{t+1} = j$  such that  $\forall s \leq t, g_{i_s, i_{s+1}} = 1$ .

In our main result, we describe a diffusion process that justifies the interpretation of  $dc_{ij}(t, G)$  as the expected number of times  $j$  receives by time  $t$  the information originating in sender  $i$ , and we show that it is unique within the class of anonymous network diffusion processes. This result therefore provides an exact characterization of the diffusion process underlying diffusion centrality.

**Theorem 1.**  $dc_{ij}(t, G)$  from the diffusion centrality equals  $n_{ij}(t, G)$ , the expected number of times  $j$  receives the information originating in sender  $i$  by time  $t$ , for all networks and time  $t$  under an anonymous network diffusion process if and only if the network diffusion process satisfies  $\{f^t(\mathcal{H}_k^{t-1})\}_t = \{f^t, DC(\mathcal{H}_k^{t-1})\}_t$  where

$$f^t, DC(\mathcal{H}_k^{t-1}) := R_k^{t-1}. \tag{7}$$

Individuals transmit to each of their neighbors as many messages as they received in the previous period.

**STEP 1.** Recall that  $x_{ij}(t, G)$  is the expected number of times that  $j$  receives the information originating from sender  $i$  at exactly time  $t$ , that is

$$x_{ij}(t, G) = \mathbb{E} \left( \sum_k R_{kj}^t \right).$$

Before proving sufficiency, it is useful to prove the following Lemma that provides an inductive characterization of  $x_{ij}(t, G)$ .

**Lemma 1.**

$$x_{ij}(t, G) = \alpha^t [G^t]_{ij} \forall t \iff x_{ij}(t, G) = \begin{cases} 1 & \text{if } j = i \text{ \& } t = 0, \\ 0 & \text{if } j \neq i \text{ \& } t = 0, \\ \sum_k \alpha g_{kj} x_{ik}(t-1, G) & \text{if } t > 0. \end{cases}$$

**Proof.** Let us show sufficiency first. At  $t = 0$ ,  $x_{ij}(0, G) = [I]_{ij}$ . If  $t > 0$ , we can write:

$$[G^t]_{ij} = [G^{t-1}G]_{ij} = \sum_k [G^{t-1}]_{ik} g_{kj}$$

where the last equality relies on the definition of the product of matrices. Substituting this into  $x_{ij}(t, G) = \alpha^t [G^t]_{ij}$  gives us

$$\begin{aligned} x_{ij}(t, G) &= \alpha \sum_k \alpha^{t-1} (G^{t-1})_{ik} g_{kj} \\ &= \alpha \sum_k x_{ik}(t-1, G) g_{kj}. \end{aligned}$$

To show necessity, observe that the formula on the right-hand-side defines  $x_{ij}(t, G)$  unambiguously. There is a unique  $x_{ij}(t, G)$  that satisfies it. By sufficiency,  $x_{ij}(t, G) = \alpha^t [G^t]_{ij}$ . ■

Now we can proceed with our proof. Using (2), the expected number of times a recipient  $j$  receives the information at time  $t$  is

$$x_{ij}(t, G) = \sum_k \alpha \mathbb{E} \left( T_{kj}^t \right).$$

Under (7),

$$\begin{aligned} x_{ij}(t, G) &= \sum_k \alpha g_{kj} \mathbb{E} \left( \sum_l R_{lk}^{t-1} \right) \\ &= \sum_k \alpha g_{kj} x_{ik}(t-1, G) \\ &= \alpha^t [G^t]_{ij} \end{aligned}$$

where the last equality holds by Lemma 1.

[STEP 2] Next, we prove the necessity. That is if  $n_{ij}(t, G) = dc_{ij}(t, G)$  then the transmission process must be guided by (7).

We prove this by induction.

At time 1,  $n_{ij}(1, G) = x_{ij}(1, G) = \alpha \sum_k g_{kj} f^1(\mathcal{H}_k^0)$ . The history at time 0 is 1 for  $i$  and 0 for the others:  $\mathcal{H}_i^0 = R_i^0 = 1$  and  $\mathcal{H}_k^0 = R_k^0 = 0$  for  $k \neq i$ . Hence,  $n_{ij}(1) = x_{ij}(1, G) = \alpha g_{ij} f^1(1) + \alpha \sum_{k \neq i} g_{kj} f^1(0)$ . Under the diffusion centrality formula,  $dc_{ij}(1, G) = \alpha g_{ij}$  for all  $G$ . So in any network  $G$  where  $ij$  are not connected  $n_{ij}(1, G) = 0$ . Hence,  $\alpha \sum_k g_{kj} f^1(0) = 0$  tells us that  $f^1(0) = 0$ . Now add a link between  $i$  and  $j$ , under the diffusion centrality formula  $n_{ij}(1, G) = \alpha$  so that  $f^1(1) = 1$ .

At time 2,  $n_{ij}(2, G) = x_{ij}(1, G) + x_{ij}(2, G)$  and  $x_{ij}(2, G) = \alpha \mathbb{E}_{\mathcal{H}_k^1} \sum_k g_{kj} f^2(\mathcal{H}_k^1)$ . The history at time 1 is  $(1, 0)$  for  $i$ , it is  $(0, 1)$  with probability  $\alpha$  and  $(0, 0)$  otherwise for any  $k$  such that  $g_{ik} = 1$  and is  $(0, 0)$  for all others. Hence,

$$x_{ij}(2, G) = \alpha \left( g_{ij} f^2(1, 0) + \sum_{k \neq i} g_{kj} \left[ p_k^{1,G}(\mathcal{H}_k^1 = 1) f^2(0, 1) + p_k^{1,G}(\mathcal{H}_k^1 = 0) f^2(0, 0) \right] \right)$$

where  $p_l^{1,G}(\mathcal{H}_l^1 = 1)$  is the probability of a given history of messages received for agent  $l$  at time 1 given the network and process.

Under the diffusion centrality formula,  $x_{ij}(2, G) = \alpha^2 \sum_k g_{ik} g_{kj}$  for all  $G$ . Since in any network when  $i$  is isolated  $x_{ij}(2) = 0$ , we know that  $f^2(0, 0) = 0$ . When  $G$  is formed of the unique pair  $ij$ ,  $x_{ij}(2, G) = 0$  leading to  $f^2(1, 0) = 0$ . Now consider a network  $G$  formed of the pairs  $ik$  and  $kj$  and no other link. Then,  $x_{ij}(2) = \alpha^2$ . Given  $f^1(\cdot)$ ,  $k$ 's history is  $(0, 1)$  with a probability  $\alpha$  and it is  $(0, 0)$  with a probability  $(1 - \alpha)$ . This yields:

$$x_{ij}(2, G) = \alpha \left[ \alpha f^2(0, 1) + (1 - \alpha) f^2(0, 0) \right] = \alpha^2.$$

This gives us  $f^2(0, 1) = 1$ .

For our induction, assume that we have proved the claim for all networks at any time  $s \leq t - 1$ . That is, assume that we have shown that if for all time  $s \leq t - 1$

$$x_{ij}(s, G) = \alpha^s [G^s]_{ij}$$

for all  $ij$  and all networks  $G$  then

$$f^s(\mathcal{H}_k^{s-1}) := R_k^{s-1} \forall k \ \& \ \mathcal{H}^{s-1}.$$

Our claim is that if  $x_{ij}(s, G) = \alpha^s [G^s]_{ij}$  for all  $s \leq t$ , all  $ij$  and all networks  $G$  then  $f^s(\mathcal{H}_k^{s-1}) := R_k^{s-1}$  for all  $s \leq t$ ,  $k$  and  $\mathcal{H}^{s-1}$ .

To prove this claim, we will use a second induction and iterate over the total number of messages received up to time  $t$ .

Let  $\mathcal{H}^t[r]$  be the set of feasible history profile  $(R^0, R^1, \dots, R^{t-1})$  such that  $R^0 + R^1 + \dots + R^{t-1} \leq r$ . We want to show that  $f^t(R^0, R^1, \dots, R^{t-1}) = R^{t-1}$ .

Consider  $\mathcal{H}^t[0]$ . The only feasible history such that  $R^0 + R^1 + \dots + R^{t-1} \leq 0$  is  $R^0 = R^1 = \dots = R^{t-1} = 0$ . Consider the following network:  $i$  isolated, and then a line of length  $t - 2$  arriving at  $k$ , and then  $k$  connected to  $j$ . Under the diffusion centrality formula  $x_{ij}(t, G) = 0$ , and therefore

$$x_{ij}(t, G) = 0 = \alpha f(0, 0, \dots, 0)$$

tells us that  $f(0, 0, \dots, 0) = 0$ .

Now assume that we have shown that  $f^t(R^0, R^1, \dots, R^{t-1}) = R^{t-1}$  for all  $(R^0, R^1, \dots, R^{t-1}) \in H[r]$  for  $r \geq 0$ . We want to prove that  $f^t(R^0, R^1, \dots, R^{t-1}) = R^{t-1}$  for all  $(R^0, R^1, \dots, R^{t-1}) \in H[r + 1]$ .

Pick a feasible history profile  $(R^0, R^1, \dots, R^{t-1})$  such that  $R^0 + R^1 + \dots + R^{t-1} = r + 1$ . (If this is lower than  $r + 1$ , the equality holds by  $H[r]$ ). Consider the network in which  $j$  has only one neighbor  $k$ ,  $g_{kj} = 1$ , and the network is formed of the union of  $R^{t-1}$  disjoint directed lines of length  $t - 1$  from  $i$  to  $k$  not including  $j$ ,  $R_{t-2}$  disjoint directed lines of length  $t - 2$  from  $i$  to  $k$  not including  $j, \dots$ , and a direct directed link from  $i$  to  $k$  if  $R^1 = 1$ .

Considering all possible history  $(r^0, r^1, \dots, r^{t-1})$  for  $k$  and the ex-ante probability that this history happens under the DC diffusion process  $p(r^0, r^1, \dots, r^{t-1})$ , we get:

$$x_{ij}^{DC}(t, G) = \sum_{(r^0, r^1, \dots, r^{t-1})} p(r^0, r^1, \dots, r^{t-1}) \alpha r^{t-1}.$$

In the possible histories, there is only one where  $r^0 + r^1 + \dots + r^{t-1} = r + 1$ , and this is  $(R^0, R^1, \dots, R^{t-1})$ . This history is feasible: it happens when the signal goes through ALL the disjoint lines between  $i$  and  $k$ . Any other possible history is such that  $r^0 + r^1 + \dots + r^{t-1} \leq r$  and by the induction hypothesis we know that  $f^t(r^0, r^1, \dots, r^{t-1}) = r^{t-1}$ . We therefore have:

$$\begin{aligned} x_{ij}(t, G) &= \sum_{(r^0, r^1, \dots, r^{t-1})} p(r^0, r^1, \dots, r^{t-1}) \alpha f^t(r^0, r^1, \dots, r^{t-1}) \\ &= \sum_{(r^0, r^1, \dots, r^{t-1}) \neq (R^0, R^1, \dots, R^{t-1})} p(r^0, r^1, \dots, r^{t-1}) \alpha r^{t-1} + p(R^0, R^1, \dots, R^{t-1}) \alpha f^t(R^0, R^1, \dots, R^{t-1}) \end{aligned}$$

and since  $p(R^0, R^1, \dots, R^{t-1}) > 0$ ,

$$f(R^0, R^1, \dots, R^{t-1}) = R^{t-1}$$

and hence the claim is true for  $H[r + 1]$ .

This proves both inductions. ■

Theorem 1 shows that  $dc_{ij}(t, G)$  corresponds to the expected transmission of messages from  $i$  to  $j$  in an anonymous network diffusion process if, and only if, individuals at period  $t$  transmit as many messages as they received at period  $t - 1$  to each of their neighbors. This means, in particular, that messages received before the previous period do not matter - as if agents had one-period recall - and that agents who receive several messages also transmit several messages. Under these conditions, diffusion centrality,



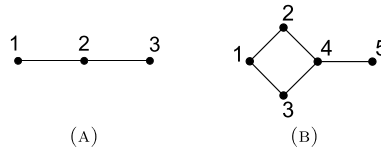


Fig. 1. Examples of networks.

which sums  $dc_{ij}(t, G)$  over all recipients, can indeed be interpreted as the expected number of times anyone receives a message from the sender.

The following examples illustrate this concept by considering alternative diffusion processes and explaining why they cannot serve as the foundation for diffusion centrality.

**Example 1:** Consider the following diffusion process:

$$f^{t, CC}(\mathcal{H}_k^{t-1}) := \begin{cases} 1, & \text{if } \sum_{s=1}^{t-1} R_k^s > 0; \\ 0, & \text{otherwise.} \end{cases}$$

This diffusion process corresponds to the communication mechanism underlying the communication centrality described in the first part of Banerjee et al. (2013). Within this process, agents have perfect recall, and any agent who has been informed at any point transmits the information during every period. Furthermore, there is a singular retransmission of multiple signals: an agent who has received the information multiple times retransmits it only once to each of his neighbors in each period.

Consider the line in Fig. 1A with  $T = 2$ . It is easy to see that the expected number of messages received by an agent is different under this diffusion process than under the diffusion process in (7). Under (7), we have  $n_{11}(2, G) = \alpha^2$ ,  $n_{12}(2, G) = \alpha$  and  $n_{13}(2, G) = \alpha^2$ , leading to a diffusion centrality of  $n_1(2, G) = \alpha + 2\alpha^2$  for agent 1. By contrast, under  $f^{t, CC}$ , agent 1 may retransmit to agent 2 at period 2. This now yields  $n_{12}(2, G) = 2\alpha$  and  $n_1(2, G) = 2\alpha + 2\alpha^2$ .

**Example 2:** Take yet another diffusion processes

$$f^{t, UM}(\mathcal{H}_k^{t-1}) := \begin{cases} 1, & \text{if } R_k^{t-1} > 0; \\ 0, & \text{otherwise.} \end{cases}$$

This diffusion process is similar to the diffusion process  $f^{t, DC}$  in (7) except that it assumes unique retransmission of multiple signals: an agent who received the information multiple times in the previous period transmits it only once to each of her neighbors in each period. In contrast, under (7) agents transmit as many messages as they receive in the previous period.

Take the network depicted in Fig. 1B and  $T = 3$ . When agent 1 initially sends the information, at time 2 agent 4 may receive messages from both agent 2 and agent 3. Under (7) we get  $n_{15}(3, G) = \alpha^3[G^3]_{15} = 2\alpha^3$ , since agent 4 retransmits both messages in period 3. By contrast,  $n_{15}(3, G)$  would have been  $\alpha^3$  under  $f^{3, UM}$ .

Theorem 1 characterizes the unique anonymous network diffusion process that underlies the concept of diffusion centrality. This theorem allows us to correct an imprecision in the descriptions of the model in Banerjee et al. (2013, 2019) (see Appendix A). Another important underlying assumption of diffusion centrality is the assumption that all agents play a similar role in retransmission. This is the assumption we will now relax.

## 4. Targeting centrality

### 4.1. Motivation

The premise of anonymous information transmission, previously mentioned, seems entirely plausible in the context of factual “external” news, such as announcing “a new microcredit organization in town.” However, this presumption becomes questionable in situations where transmitted information explicitly references the sender’s or recipient’s identities, or both.

Consider the example of cyberbullying.<sup>7</sup> Anonymity would require the victim to transmit derogatory messages about themselves. This assumption defies logic, and it would be more reasonable to assume that victims would not retransmit the gossip that concerns them.

Another scenario involves *favor requests* of the form: “Agent  $i$  requires assistance.” In this context, it becomes important to question why anyone would feel motivated to relay this information back to agent  $i$ . A more realistic assumption might be that this information does not undergo retransmission to, and consequently by, the sender  $i$ .

Finally, consider *targeted requests*, exemplified by the statement: “Agent  $i$  requires a favor from agent  $j$ .” A real-world illustration can be found in the realm of political intermediation as described in Cruz et al. (2017) wherein agent  $j$  represents a politician who accedes to requests upon reception (for a comprehensive discussion, see Appendix A). In the framework of diffusion centrality,

<sup>7</sup> A Pew Research Center survey finds that 59% of U.S. teens have personally experienced abusive online behavior, with a third (32%) of teens saying that someone has spread false rumors about them on the internet (Vogels 2022).



the inherent information diffusion process assumes that the request is propagated both to the original sender,  $i$ , and by the politician,  $j$ , during the information's dissemination. However, a more logical assumption suggests that the request would not undergo retransmission by either its target or its sender throughout the diffusion process.

This paper introduces a class of centrality measures that we call *targeting centralities* in which a sender  $i$ 's request is not retransmitted by either target  $j$ , sender  $i$ , or both during information diffusion.<sup>8</sup> The other assumptions underlying diffusion centrality remain unchanged:  $i$ 's request for  $j \neq i$  diffuses in the network under  $f^{t,DC}$ , and the focus remains on the expected frequency with which the target receives the sender's request.

When we sum this metric over all potential senders for a given target, we obtain the agent's *targeting centrality*. This metric gauges an agent's susceptibility to being targeted. Conversely, if we sum over all conceivable targets for a given sender, we derive the sender's *centralities to target*, their power to target others within the network.<sup>9</sup>

#### 4.2. Definitions and notation

We denote by  $\tilde{x}_{ij}(s, G)$  the *expected number of times agent  $j$  receives the message originally sent by  $i$  at exactly time  $s$*  when target  $j$  does not retransmit the request. Similarly,  $\tilde{y}_{ij}(s, G)$  denotes the *expected number of times agent  $j$  receives  $i$ 's message at exactly time  $s$  when sender  $i$  does not retransmit the request*. Finally,  $\tilde{z}_{ij}(s, G)$  is the same object when neither sender  $i$  nor target  $j$  retransmits the request. (Appendix B provides the LaTeX code to create this notation.)

For each of the no-retransmission assumptions, the corresponding *expected number of times agent  $j$  hears  $i$ 's request between period 1 and  $t \geq 1$*  is then given by:

$$\tilde{n}_{ij}(t, G) = \sum_{s=1}^t \tilde{x}_{ij}(s, G), \quad \tilde{y}_{ij}(t, G) = \sum_{s=1}^t \tilde{y}_{ij}(s, G) \quad \& \quad \tilde{z}_{ij}(t, G) = \sum_{s=1}^t \tilde{z}_{ij}(s, G). \tag{8}$$

We define the measures of *targeting centrality of  $j$*  for a finite time  $t$  corresponding to the different no-retransmission assumptions as

$$\tilde{n}_{\cdot j}(t, G) = \sum_{i \neq j} \tilde{n}_{ij}(t, G), \quad \tilde{y}_{\cdot j}(t, G) = \sum_{i \neq j} \tilde{y}_{ij}(t, G) \quad \& \quad \tilde{z}_{\cdot j}(t, G) = \sum_{i \neq j} \tilde{z}_{ij}(t, G). \tag{9}$$

The targeting centrality of  $j$  corresponds to the expected number of times  $j$  hears about any other individual's request by time  $t$  under each no-retransmission assumption.<sup>10</sup>

Alternatively, the centrality of  $i$  to target other agents is obtained by summing over possible recipients for a given sender:

$$\tilde{n}_{i \cdot}(t, G) = \sum_{j \neq i} \tilde{n}_{ij}(t, G), \quad \tilde{y}_{i \cdot}(t, G) = \sum_{j \neq i} \tilde{y}_{ij}(t, G) \quad \& \quad \tilde{z}_{i \cdot}(t, G) = \sum_{j \neq i} \tilde{z}_{ij}(t, G). \tag{10}$$

Letting the number of periods tend to infinity,  $t \rightarrow \infty$ , the *expected number of times agent  $j$  hears  $i$ 's request* is given by:

$$\tilde{n}_{ij}(G) = \sum_{s=1}^{\infty} \tilde{x}_{ij}(s, G), \quad \tilde{y}_{ij}(G) = \sum_{s=1}^{\infty} \tilde{y}_{ij}(s, G) \quad \& \quad \tilde{z}_{ij}(G) = \sum_{s=1}^{\infty} \tilde{z}_{ij}(s, G). \tag{11}$$

The corresponding targeting centralities are as follows:

$$\tilde{n}_{\cdot j}(G) = \sum_{i \neq j} \tilde{n}_{ij}(G), \quad \tilde{y}_{\cdot j}(G) = \sum_{i \neq j} \tilde{y}_{ij}(G) \quad \& \quad \tilde{z}_{\cdot j}(G) = \sum_{i \neq j} \tilde{z}_{ij}(G). \tag{12}$$

The targeting centrality of  $j$  corresponds to the expected number of times  $j$  receives the information originating from any other individual under each no-retransmission assumption. It captures their centrality as a target. The centrality of  $i$  to target others is

$$\tilde{n}_{i \cdot}(G) = \sum_{j \neq i} \tilde{n}_{ij}(G), \quad \tilde{y}_{i \cdot}(G) = \sum_{j \neq i} \tilde{y}_{ij}(G) \quad \& \quad \tilde{z}_{i \cdot}(G) = \sum_{j \neq i} \tilde{z}_{ij}(G). \tag{13}$$

Note that targeting centrality under no retransmission by the sender and the target, like diffusion centrality, is symmetric when  $G$  is undirected:  $\tilde{n}_{\cdot k}(t, G) = \tilde{n}_{k \cdot}(t, G)$  and  $\tilde{n}_{\cdot k}(G) = \tilde{n}_{k \cdot}(G)$ .

In contrast, equivalent measures of centrality under the assumption that only the target does not retransmit, or only the original sender does not retransmit, are generally not symmetric ( $\tilde{y}_{\cdot k}(t, G) \neq \tilde{y}_{k \cdot}(t, G)$  and  $\tilde{y}_{\cdot k}(G) \neq \tilde{y}_{k \cdot}(G)$  and  $\tilde{z}_{\cdot k}(t, G) \neq \tilde{z}_{k \cdot}(t, G)$  and  $\tilde{z}_{\cdot k}(G) \neq \tilde{z}_{k \cdot}(G)$ )

<sup>8</sup> The assumptions of no-retransmission by the sender or to the sender are almost equivalent. In either case, diffusion of information from  $i$  to  $j \neq i$  occurs along the walks connecting  $i$  and  $j$  and where  $i$  only appears at the beginning. Both assumptions yield the same values of targeting centrality and centrality to target others. The only difference is whether the sender keeps hearing about the information during the process.

<sup>9</sup> We do not need to think of all targeted messages being initiated at the same time for these centralities to be relevant. Indeed, normalizing centralities by dividing them by  $n-1$  yields averages that may be more intuitive if each targeted request or gossip has the same probability of being initiated.

<sup>10</sup> We sum over all senders that are not  $j$ ,  $i \neq j$ . Note that in the case of diffusion and Katz-Bonacich centrality summing is usually over all individuals including  $j$  themselves. Doing so would simply add  $\tilde{n}_{jj}(t, G) = \tilde{y}_{jj}(t, G) = \tilde{z}_{jj}(t, G) = 1$  to the measure.

$\check{n}_{\cdot k}(G) \neq \check{n}_{k \cdot}(G)$ ). However, when  $G$  is undirected, a symmetry does exist across the sender-only and target-only measures:  $\check{n}_{\cdot k}(t, G) = \check{n}_{k \cdot}(t, G)$ ,  $\check{n}_{\cdot k}(G) = \check{n}_{k \cdot}(G)$ ,  $\check{n}_{\cdot k}(t, G) = \check{n}_{k \cdot}(t, G)$  and  $\check{n}_{\cdot k}(G) = \check{n}_{k \cdot}(G)$ .

We will use a similar notation for the sets of walks. Recall that  $W_{ij}(l, G)$  denotes the set of walks of length  $l$  connecting  $i$  to  $j$ . We now define  $\check{W}_{ij}(l, G)$  as the set of walks of length  $l$  connecting  $i$  to  $j$  and such that  $j$  only appears at the end of the walk. Likewise,  $\bar{W}_{ij}(l, G)$  denotes the set of walks of length  $l$  connecting  $i$  to  $j$  and such that  $i$  only appears at the beginning of the walk. Finally,  $\check{\bar{W}}_{ij}(l, G)$  denotes the set of walks of length  $l$  connecting  $i$  to  $j$  and such that  $i$  only appears at the beginning of the walk and  $j$  only appears at the end of the walk.  $W_{ij}(G)$ ,  $\bar{W}_{ij}(G)$ ,  $\check{W}_{ij}(G)$  and  $\check{\bar{W}}_{ij}(G)$  refer to the set of walks of all lengths.

Finally, we denote by  $G[-i]$  the network over  $n - 1$  nodes obtained by removing  $i$  and her links.

### 4.3. Infinite time

This section shows that when the number of periods is infinite, targeting centralities can be obtained by applying a simple correction to diffusion centrality.

**Proposition 1.** *Suppose that  $i$ 's request for a favor from  $j \neq i$  diffuses in the network under  $f^{t,DC}$ . The expected number of times  $i$ 's request reaches  $j$  is equal to*

$$\check{n}_{ij}(G) = \frac{n_{ij}(G)}{1 + n_{jj}(G)}, \quad \check{n}_{ij}(G) = \frac{n_{ij}(G)}{1 + n_{ii}(G)} \quad \& \quad \check{\check{n}}_{ij}(G) = \frac{\check{n}_{ij}(G)}{1 + n_{jj}(G[-i])} = \frac{\check{\check{n}}_{ij}(G)}{1 + n_{ii}(G[-j])}$$

if  $i$ 's request to  $j$  is not retransmitted by target  $j$  or by sender  $i$  or by both during information diffusion.

**Proof.** Recall from (6) and Theorem 1 that under diffusion process  $\{f^{t,DC}\}_t$

$$n_{ij}(G) = \sum_{w \in W_{ij}(G)} \alpha^{l(w)}.$$

Let  $C_i(G)$  denote the union of the set of all cycles originating at  $i$  in network  $G$  (a cycle originating at  $i$  is a walk from  $i$  to  $i$ ) and of the empty cycle, and similarly for  $C_j(G)$ .

We can uniquely decompose each walk from  $i$  to  $j$  into: a cycle from  $i$  to  $i$  (possibly empty) and a walk from  $i$  to  $j$  where  $i$  only appears at the beginning. This decomposition implies that:

$$n_{ij}(G) = \sum_{w \in W_{ij}(G)} \alpha^{l(w)} = \sum_{c \in C_i(G), w' \in \check{W}_{ij}(G)} \alpha^{l(c)+l(w')} = \left( \sum_{c \in C_i(G)} \alpha^{l(c)} \right) \left( \sum_{w' \in \check{W}_{ij}(G)} \alpha^{l(w')} \right).$$

Given the intimate connection between walks and information diffusion, it should be clear that if the sender does not retransmit messages:

$$\check{n}_{ij}(G) = \sum_{w \in \check{W}_{ij}(G)} \alpha^{l(w)}.$$

Moreover, since  $C_i(G) = W_{ii}(G) \cup \{\emptyset\}$ ,  $\sum_{c \in C_i(G)} \alpha^{l(c)} = 1 + \sum_{w \in W_{ii}(G)} \alpha^{l(w)} = 1 + n_{ii}$ . It follows that

$$n_{ij}(G) = (1 + n_{ii}(G))\check{n}_{ij}(G) \Rightarrow \check{n}_{ij}(G) = \frac{n_{ij}(G)}{1 + n_{ii}(G)}.$$

A symmetric argument can be made to show that if only the target does not retransmit:

$$\check{\check{n}}_{ij}(G) = \frac{n_{ij}(G)}{1 + n_{jj}(G)}.$$

Finally, consider the case where both sender and target do not retransmit. Each walk from  $i$  to  $j$  can be uniquely decomposed into: a cycle from  $i$  to  $i$  (possibly empty), a walk where  $i$  only appears at the beginning and  $j$  only appears at the end, and a cycle from  $j$  to  $j$  which does not go through  $i$  (possibly empty). This decomposition implies that:

$$\begin{aligned} n_{ij}(G) &= \sum_{w \in W_{ij}(G)} \alpha^{l(w)} = \sum_{c \in C_i(G), w' \in \check{\bar{W}}_{ij}(G), c' \in C_j(G[-i])} \alpha^{l(c)+l(w')+l(c')} \\ &= \left( \sum_{c \in C_i(G)} \alpha^{l(c)} \right) \left( \sum_{w' \in \check{\bar{W}}_{ij}(G)} \alpha^{l(w')} \right) \left( \sum_{c' \in C_j(G[-i])} \alpha^{l(c')} \right). \end{aligned}$$

Since  $\check{n}_{ij}(G) = \sum_{w \in \check{W}_{ij}(G)} \alpha^{l(w)}$  and  $\sum_{c' \in C_j(G[-i])} \alpha^{l(c')} = 1 + n_{jj}(G[-i])$ , we get that

$$\check{n}_{ij}(G) = \frac{n_{ij}(G)}{(1 + n_{ii}(G))(1 + n_{jj}(G[-i]))} = \frac{\check{n}_{ij}(G)}{1 + n_{jj}(G[-i])}.$$

Similarly, each walk from  $i$  to  $j$  in  $G$  can be uniquely decomposed into: a cycle from  $i$  to  $i$  which does not go through  $j$  (possibly empty), a walk where  $i$  only appears at the beginning and  $j$  only appears at the end, and a cycle from  $j$  to  $j$  (possibly empty), leading to the last equality. ■

Proposition 1 shows that the expected number of times  $i$ 's request reaches the target  $j$  when the request is not retransmitted by  $j$  or to  $i$  is simply related to this number in the absence of constraints on retransmission. The discounted number of walks connecting  $i$  to  $j$  when  $i$  appears only at the beginning and  $j$  appears only at the end is equal to the discounted number of unconstrained walks connecting  $i$  to  $j$  divided by the discounted number of cycles starting at  $i$  and by the discounted number of cycles starting at  $j$  in  $G[-i]$ . Thus, if  $j \neq i$  and  $\alpha < 1/\lambda_1(G)$ ,<sup>11</sup>

$$\check{n}_{ij}(G) = \frac{[\alpha G(I - \alpha G)^{-1}]_{ij}}{[I - \alpha G]_{ii}^{-1}[I - \alpha G[-i]]_{jj}^{-1}}. \tag{14}$$

Following (12), the targeting centrality of  $j$  is then given by  $\check{n}_{\cdot j}(G) = \sum_{i \neq j} \check{n}_{ij}(G)$ . We see clearly that targeting centrality differs conceptually from diffusion centrality  $n_{\cdot j}(G) = \sum_j n_{ij}(G)$ . In Section 5.3, we will show that targeting centrality and diffusion centrality also differ in practice.

A natural question is whether targeting centrality is computationally more demanding than diffusion centrality. The computation of targeting centrality relies on elements of inverse matrices of the kind  $(I - \alpha G)^{-1}$ . Under no retransmission by the sender only or by the target only, the computation involves the inverse of one matrix of size  $n$  by  $n$ , as with diffusion centrality. Under no retransmission by both sender and target, it involves the inverses of one matrix of size  $n$  by  $n$  and  $n$  matrices of size  $n - 1$  by  $n - 1$ . Thus, targeting centrality has a similar computational complexity to diffusion centrality.

#### 4.4. Finite time

It is also possible to derive exact formulas for targeting centrality in finite time. This section does just that, after introducing a few more pieces of notation.

**Notation.** For convenience of notation, we set  $x_{ij}(0, G) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$ .<sup>12</sup>

Now, for each pair of distinct sender  $i$  and recipient  $j$ , let

$$X_{ij}(t, G) = \begin{bmatrix} x_{ij}(1, G) \\ \dots \\ x_{ij}(t, G) \end{bmatrix}, \check{X}_{ij}(t, G) = \begin{bmatrix} \check{x}_{ij}(1, G) \\ \dots \\ \check{x}_{ij}(t, G) \end{bmatrix}, \check{X}_{ij}(t, G) = \begin{bmatrix} \check{x}_{ij}(1, G) \\ \dots \\ \check{x}_{ij}(t, G) \end{bmatrix}, \check{X}_{ij}(t, G) = \begin{bmatrix} \check{x}_{ij}(1, G) \\ \dots \\ \check{x}_{ij}(t, G) \end{bmatrix}.$$

Moreover, for each agent  $j$ , we denote by  $M_j(t, G)$  the following lower triangular matrix:

$$M_j(t, G) = \begin{bmatrix} x_{jj}(0, G) & 0 & 0 & \dots & 0 \\ x_{jj}(1, G) & x_{jj}(0, G) & 0 & \dots & 0 \\ x_{jj}(2, G) & x_{jj}(1, G) & x_{jj}(0, G) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ x_{jj}(t, G) & x_{jj}(t-1, G) & x_{jj}(t-2, G) & \dots & x_{jj}(0, G) \end{bmatrix} \\ = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \alpha^2 G_{jj}^2 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \alpha^t G_{jj}^t & \alpha^{t-1} G_{jj}^{t-1} & \alpha^{t-2} G_{jj}^{t-2} & \dots & 1 \end{bmatrix}.$$

Note that this triangular matrix is invertible, as it has a diagonal of ones.

<sup>11</sup> A direct implication of Proposition 1 is that for any  $i \neq j$ ,

$$[I - \alpha G]_{ii}^{-1}[I - \alpha G[-i]]_{jj}^{-1} = [I - \alpha G]_{jj}^{-1}[I - \alpha G[-j]]_{ii}^{-1}.$$

To our knowledge, this provides a novel result in matrix analysis.

<sup>12</sup> We can make this assumption because diffusion centrality and targeting centrality will only take into account messages received at time 1 or higher.

**Target does not resend.**

We start with the case where the target does not resend the message. To derive formulas in finite time, we can use decomposition arguments similar to those used for infinite time in Section 4.3. However, we must now keep track of the exact length of the walks.

Each walk  $w$  from  $i$  to  $j$  of length  $t$  ( $w \in \bar{W}_{ij}(t)$ ) can be uniquely decomposed into: a walk of length  $l$  in  $\bar{W}_{ij}(l)$  and a cycle of length  $t - l$  from  $j$  to  $j$  (or the empty cycle). Let  $C_j(l', G)$  denote the set of cycles of length  $l'$  originating at  $j$  in network  $G$  if  $l' \geq 1$  and the empty cycle if  $l' = 0$ .

This decomposition implies that:

$$\begin{aligned} x_{ij}(t, G) &= \sum_{w \in \bar{W}_{ij}(t, G)} \alpha^t = \sum_{l=1}^t \sum_{w' \in \bar{W}_{ij}(l, G), c \in C_j(t-l, G)} \alpha^t, \\ &= \sum_{l=1}^t \left( \sum_{w' \in \bar{W}_{ij}(l, G)} \alpha^l \right) \left( \sum_{c \in C_j(t-l, G)} \alpha^{t-l} \right), \end{aligned}$$

so that

$$x_{ij}(t, G) = \sum_{l=0}^t \check{x}_{ij}(l, G) x_{jj}(t-l, G) \quad \forall i \neq j. \tag{15}$$

It follows that

$$X_{ij}(t, G) = M_j(t-1, G) \check{X}_{ij}(t, G) \Rightarrow \check{X}_{ij}(t, G) = M_j(t-1)^{-1} X_{ij}(t, G).$$

This tells us that a simple correction can be applied to  $X_{ij}(t, G)$  to obtain  $\check{X}_{ij}(t, G)$ .

**Sender does not resend.**

It is straightforward to check that if the sender does not resend the message, then:

$$\check{X}_{ij}(t, G) = M_i(t-1, G)^{-1} X_{ij}(t, G) \text{ for } i \neq j.$$

The details are available in Appendix C.

**Both target and sender do not resend.**

Now assume that both sender  $i$  and target  $j \neq i$  do not retransmit the request. To calculate  $\check{\check{x}}(t, G)$ , notice that we can decompose each walk from  $i$  to a (distinct)  $j$  of length  $t$  where  $j$  does not retransmit ( $w \in \bar{\bar{W}}_{ij}(t, G)$ ) uniquely into: a cycle of size  $l$  from  $i$  to  $i$  in the network where we exclude the target  $G[-j]$  and a walk of length  $t - l$  in  $\bar{W}_{ij}(t-l, G)$ . Similarly, walks in  $\bar{\bar{W}}_{ij}(t, G)$  can be decomposed into walks in  $\bar{W}_{ij, G}$  and cycles in  $C_j(G[-i])$ . This means that for  $i \neq j$ :

$$\check{\check{x}}_{ij}(t, G) = \sum_{l=0}^{t-1} x_{ii}(l, G[-j]) \check{x}_{ij}(t-l, G); \tag{16}$$

$$\check{\check{x}}_{ij}(t, G) = \sum_{l=1}^t \check{x}_{ij}(l, G) x_{jj}(t-l, G[-i]). \tag{17}$$

Equations (16) and (17) then imply that, for  $i \neq j$

$$\check{\check{X}}_{ij}(t, G) = M_i(t-1, G[-j]) \check{X}_{ij}(t, G)$$

$$\check{\check{X}}_{ij}(t, G) = M_j(t-1, G[-i]) \check{X}_{ij}(t, G),$$

so that

$$\begin{aligned} \check{\check{X}}_{ij}(t, G) &= M_i(t-1, G[-j])^{-1} \check{\check{X}}_{ij}(t, G) = M_i(t-1, G[-j])^{-1} M_j(t-1)^{-1} X_{ij}(t, G) \\ &= M_j(t-1, G[-i])^{-1} \check{\check{X}}_{ij}(t, G) = M_j(t-1, G[-i])^{-1} M_i(t, G)^{-1} X_{ij}(t, G). \end{aligned}$$

The targeting centrality of individual  $j$  from (9) is then given by

$$\check{\check{n}}_{\cdot j}(t, G) = \sum_{i \neq j} \check{\check{n}}_{ij}(t, G) = \sum_{i \neq j} \sum_{s=1}^t \check{\check{x}}_{ij}(s, G)$$

**From finite to infinite time.**

The previous section presents exact formulas for targeting centralities in finite time. These formulas exclusively rely on matrix inversions, like diffusion centrality, and are easy to program. Yet, their computational intensity surpasses that of the uncomplicated adjustment for infinite time delineated in Proposition 1. In cases where either the sender or the target does not retransmit, these

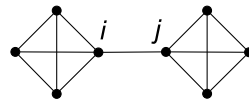


Fig. 2. Targeting vs diffusion centrality of bridges.

formulas hinge on the computation of the inverse of a matrix sized  $t + 1$  by  $t + 1$ , a matrix constructed from successive powers of  $G$  up to  $G^t$ . In scenarios where both the sender and the target do not retransmit, the formulas necessitate the inversion of two matrices, both of dimensions  $t + 1$  by  $t + 1$ . Consequently, the computational burden of calculating targeting centrality within a finite span increases in  $t$ .

Fortunately, the expressions for infinite time offer inherent approximations. Take, for instance, the situation where the target refrains from retransmission. A natural question is how well simply dividing  $n_{ij}(t, G)$  by the number of cycles  $jj$  of length less than or equal to  $t$  (including the empty cycle),  $1 + n_{jj}(t, G)$  would approximate  $\check{n}_{ij}(t, G)$ . Detailed in Appendix E, Proposition 4 shows that, with increasing  $t$ , the two expressions promptly converge.

**Further extensions.**

In this section, we proposed to modify the benchmark diffusion process underlying diffusion centrality by assuming no retransmission of the sender and/or the target. Other variations of this benchmark process may be relevant in some contexts. Consider, for instance, targeted requests for favors appearing in political intermediation, and suppose that the request is not retransmitted to the sender and by the target. Our maintained assumption, so far, is that the person requesting a favor from the politician sends the information only once, initially. Alternatively, the sender could try and reinforce their demand by repeatedly sending it. If the sender resends their request for favors in every period, the expected number of times target  $j$  hears about it is now equal to  $\check{n}_{ij}(t, G) + \check{n}_{ij}(t - 1, G) + \dots + \check{n}_{ij}(2, G) + \check{n}_{ij}(1, G)$ . Proposition 1 can then be used to compute this number and corresponding extensions of targeting centrality.

**5. Comparison between targeting and diffusion centrality**

By this juncture, the reader may wonder whether it makes a difference whether we measure diffusion centrality or targeting centrality. While there will be clear differences in levels – with the number of expected messages being higher under diffusion centrality – a key question is whether the choice between these centralities leads to varying results when comparing nodes. To elucidate this, this section assumes an infinite time frame and starts with a practical example, followed by formal results.

**5.1. An example**

We start with an example showcasing the contrast between targeting and diffusion centralities. Imagine a scenario in which the target does not retransmit messages, such as in the case of bullying. Intuitively, whether the target transmits or not should be particularly important if the target is a bridging agent linking separate groups of individuals. Consider the graph in Fig. 2. Link  $ij$  is a bridge and the source of asymmetry between these two agents and the others.

Recall that  $\alpha$  represents the probability of information transmission along a link. When  $\alpha$  is small, information flow primarily relies on direct links, and as a result, the distinctions between targeting centrality and diffusion centrality are minimal. In the graph depicted in Fig. 2 for instance, the correlation and rank correlation between diffusion and targeting centrality across the nodes of the graph in Fig. 2 is 1 for small values of  $\alpha$ , with  $i$  and  $j$  being more central than the others.

In contrast, let's consider high values of  $\alpha$ . Since diffusion centrality is only well-defined for  $\alpha < \alpha_{\max} = 1/\lambda_1(G)$ , we choose  $\alpha = 0.999 \alpha_{\max} \approx 0.302$ .

As expected, the diffusion centrality of nodes  $i$  and  $j$  remains the highest. However, for high values of  $\alpha$ , their targeting centrality becomes the lowest among the nodes in the network. Consequently, the correlation and rank correlation between targeting and diffusion centrality both become  $-1$ . This indicates that individuals  $i$  and  $j$  possess a greater ability to control the spread of information within the network, effectively limiting the flow of gossip. Thanks to their bridging position,  $i$  and  $j$  would be relatively powerful bullies while being less vulnerable to being bullied themselves.

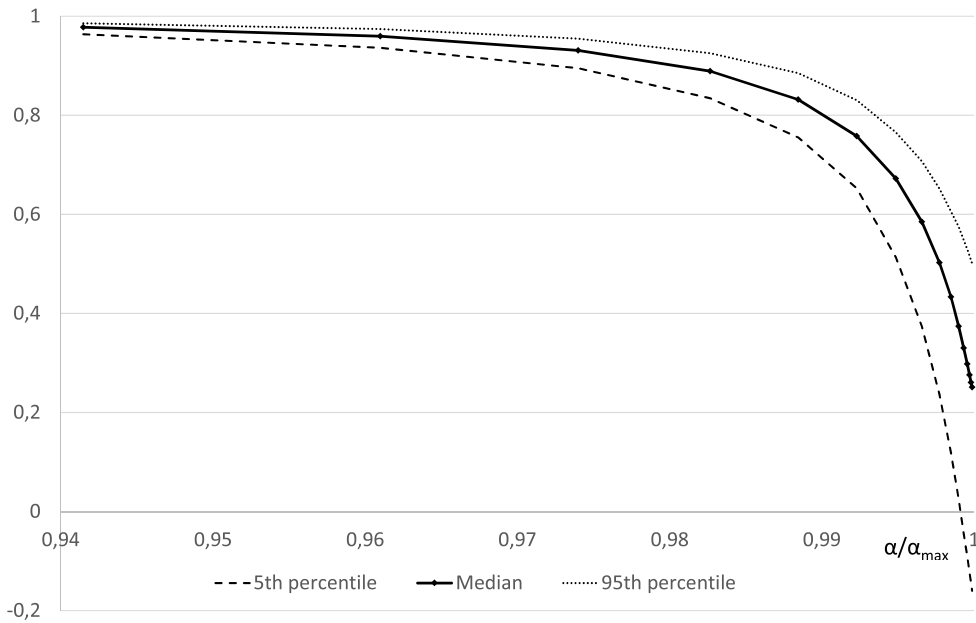
**5.2. Formal comparison**

We will now demonstrate that this is far from an isolated example. In fact, our analysis reveals that rankings of nodes according to their targeting and diffusion centrality tend to diverge for high values of  $\alpha$  when the target does not resend, and to converge when the sender does not retransmit.

**Target does not transmit.**

We start with scenarios in which the target does not transmit. We rely on the notion of primitive matrices, familiar from the Perron-Frobenius Theorem. A non-negative matrix  $G$  is primitive if there exists  $s > 0$  such that  $\forall i, j, [G^s]_{ij} > 0$ . The adjacency matrix of the network is primitive only when the network is strongly connected, i.e., if any two nodes are connected by a walk.<sup>13</sup>

<sup>13</sup> Theorems 2 and B.1 in Banerjee et al. (2019) also require the matrices to be primitive.



**Fig. 3.** CORRELATION BETWEEN TARGETING AND DIFFUSION CENTRALITY. We take 1000 Erdős-Renyi random graphs with  $n = 50$  agents and probability of link formation  $p = 0.2$ . The figure plots the 5<sup>th</sup> percentile, median, and 95<sup>th</sup> percentile of the correlation coefficients between targeting and diffusion centrality across agents within the graphs for the value of  $\alpha/\alpha_{\max}$  on the horizontal axis.

**Proposition 2.** Suppose that  $G$  is a primitive matrix. As  $\alpha$  converges to  $1/\lambda_1(G)$ :

- (1) The rank correlation between diffusion and targeting centrality converges to  $-1$ ;
- (2) The rank correlation between diffusion and the sender’s centrality to target others converges to  $1$ .

Proposition 2 reveals a strong asymmetry between targeting centrality and centrality to target others, when the target only does not retransmit. This asymmetry stands in stark contrast with diffusion centrality, which remains symmetric if  $G$  is undirected:  $n_{\cdot k} = n_{k \cdot}$ . The agent with the highest centrality in terms of hearing the information and is also the most central in terms of information diffusion (Banerjee et al. (2013)).

When targets do not propagate information concerning them, the symmetry disappears. The centrality of agents as targets departs from their centrality at targeting others. In a bullying context, for instance, it means that an individual ability to bully others may significantly differ from their vulnerability to being bullied themselves.

**Sender does not retransmit.**

When  $G$  is undirected, the correspondence across the sender-only and target-only measures,  $\vec{n}_{\cdot k} = \vec{n}_{k \cdot}$  and  $\vec{n}_{k \cdot} = \vec{n}_{\cdot k}$ , allows us to rely on the result from the previous section.

This implies that, under no retransmission by the sender alone, the rank correlation between diffusion and targeting centrality converges to  $1$  as  $\alpha$  converges to  $1/\lambda_1$ , while the rank correlation between diffusion centrality and the centrality to target others converge to  $-1$ . This property holds for any primitive  $G$ ; the proof for directed networks is similar to the proof of Proposition 2 and omitted for clarity.

**Both sender and target do not retransmit.**

What if both senders and recipients refrain from retransmitting information? This scenario presents a hybrid of the preceding two sections, where conflicting dynamics come into play. The direction of the rank correlation between diffusion and targeting centrality for elevated  $\alpha$  values hinges on the network’s structure. Yet, as demonstrated through simulations in the forthcoming section, it becomes evident that diffusion and targeting centralities possess the capability to yield markedly distinct rankings.

**5.3. Random graphs under both sender and target no-transmission**

Consider Erdős-Renyi random graphs with  $n = 50$  agents and a link formation probability of  $p = 0.2$ . We randomly select 1,000 graphs and, for each graph, calculate the correlation between the two measures across nodes for various values of  $\alpha$ . We know that for small values of  $\alpha$  comparing nodes according to their targeting centrality or their diffusion centrality yields equivalent results. The question is: what happens as  $\alpha$  increases? The findings are depicted in Fig. 3 that represents the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles of the distribution of correlations as the value of  $\alpha/\alpha_{\max}$  approaches 1.

The close relation between  $n_j$  and  $\check{n}_j$  breaks down when  $\alpha$  approaches its maximal value. Correlation is generally decreasing and concave in  $\alpha$ .<sup>14</sup> For values close to  $1/\lambda_1(G)$ , the correlation between the two measures tends to be quite low, about 0.23 for the median network, and displays significant dispersion, from  $-0.12$  at the 5<sup>th</sup> percentile to 0.50 at the 95<sup>th</sup> percentile. Fig. 4 in the Appendix shows that the rank correlation between the two measures exhibits very similar patterns: an initially strong rank correlation between diffusion and targeting centrality plunges and comes close to 0 as  $\alpha/\alpha_{\max}$  approaches 1.

This example demonstrates that the individuals with the highest values on targeting centrality are not necessarily the same as the individuals with the highest values on diffusion centrality, for values of  $\alpha$  close to  $\alpha_{\max}$ .<sup>15</sup>

### 6. Reachability

Both diffusion centrality (Section 3) and targeting centrality (Section 4) count the expected total number of messages from a sender to a recipient. For instance, Cruz et al. (2017) assume that  $j$  grants a favor to  $i$  each time he receives  $i$ 's request.

However, in some contexts it may not be the number of times someone hears the information that matters, but whether the information reaches the target. Say that  $j$  will satisfy  $i$ 's request if he hears about it, and that this is a one-time favor. The relevant question is then whether  $i$ 's request reaches  $j$ . In this case, we are interested in the *reachability* of  $j$ . Denote by  $p_{ij}(t, G) \in [0, 1]$  the probability of  $i$ 's request for a favor successfully reaching  $j$  within the first  $t$  periods and  $p_{ij}(G) = \lim_{t \rightarrow \infty} p_{ij}(t, G)$ .

We define the *reachability* of  $j$  as  $\sum_i p_{ij}(G)$ , the overall expected number of favors provided by  $j$  when every agent sends a request. The computation of these probabilities depends on details of the underlying process of request transmission. We consider the unique anonymous diffusion process compatible with diffusion centrality,  $f^{t,DC}$ , in what follows; our result below can be extended to alternative assumptions.

Not surprisingly, reachability does not display additive properties. For instance, suppose that  $i$  has a direct connection with  $j$  and an indirect connection through a common friend. In this case,  $p_{ij} = \alpha + (1 - \alpha)\alpha^2 = \alpha + \alpha^2 - \alpha^3$ . We next derive a general formula based on the inclusion-exclusion principle, which could be used for algorithmic implementation. Let us introduce some notions and notations. Denote by  $w_1 \cap_b w_2 \cap_b w_3 \dots \cap_b w_k$  the intersection of the beginning of the  $k$  walks  $w_1, w_2, \dots, w_k$ . Define  $L(w_1, \dots, w_k)$  as follows:

$$L(w_1, \dots, w_k) = \sum_{s_1=1}^k l(w_{s_1}) - \sum_{s_1 < s_2} l(w_{s_1} \cap_b w_{s_2}) + \dots + (-1)^{k+1} l(w_1 \cap_b \dots \cap_b w_k).$$

The general idea here is to count common beginnings only once. As soon as two walks separate, however, we add the lengths of the remaining segments. For instance, with  $k = 2$ ,  $L(w_1, w_2) = l(w_1) + l(w_2) - l(w_1 \cap_b w_2)$ . Finally, write  $w_1 \neq \dots \neq w_k$  to denote that the  $k$  walks  $w_1, \dots, w_k$  are distinct.

**Proposition 3.** Consider a model of political intermediation under  $f^{t,DC}$  and the assumption that  $i$ 's request for a favor from  $j$  is not retransmitted to  $i$  or by  $j$ . The probability of  $i$ 's request successfully reaching  $j$  within the first  $t$  periods is equal to

$$p_{ij}(t, G) = \sum_{k=1}^{|\cup_{s \leq t} \check{W}_{ij}(s, G)|} (-1)^{k+1} \sum_{w_1 \neq \dots \neq w_k \in \cup_{s \leq t} \check{W}_{ij}(s, G)} \alpha^{L(w_1, \dots, w_k)}.$$

Note that the first term in the formula is equal to  $\sum_{w \in \cup_{s \leq t} \check{W}_{ij}(s, G)} \alpha^{l(w)}$  which is equal to  $\check{n}_{ij}(t, G)$ . Proposition 3 then clarifies the difference between the number of times  $j$  is expected to hear about  $i$ 's request and the probability that  $j$  will hear about it. This formula is combinatorially complex, which confirms that computing these probabilities in practice is computationally hard. Applied researchers therefore have two options. One, straightforward but computationally intensive, is to rely on numerical simulations to obtain approximate values of these probabilities, as follows. Simply generate  $N$  realizations of information diffusion at random and count the number of times  $K$  that  $i$ 's request for a favor reaches  $j$ . Then,  $K/N$  converges to  $p_{ij}$  as  $N$  tends to infinity. A similar numerical procedure underlies the structural estimations in Banerjee et al. (2013).

Alternatively, and as proposed by Banerjee et al. (2013), researchers can rely on simpler proxies which are easier to compute and likely to be highly correlated with these probabilities. Proxies vary in their usefulness, however, and the literature still lacks formal results on why and when we should expect diffusion centrality to perform well empirically. In a context of targeted requests, such as political intermediation, we conjecture that targeting centrality may provide a significantly better proxy than diffusion centrality.

### 7. Conclusion

This paper contributes in three significant ways.

<sup>14</sup> This result does not rely on the use of Erdős-Renyi random graphs. Similar patterns can be shown within graphs possessing small-world characteristics—a small average path length and substantial clustering, as discussed in Watts (2004), see Fig. 5 in the Appendix.

<sup>15</sup> Banerjee et al. (2013, 2019) assume that  $\alpha = 1/\lambda_1(G)$  in their empirical implementation of diffusion centrality with  $t < \infty$ . Cruz et al. (2017) also assume that  $\alpha = 1/\lambda_1(G)$ .



First, we have characterized the information diffusion process that underlies diffusion centrality. Diffusion centrality hinges on tallying the volume of messages dispatched from a particular sender to a given recipient when the initial message gets transmitted through the network. We call a diffusion process *anonymous* if each agent treats each of their neighbors the same way and conditions their transmission of information only on the history of messages received, not their origin. We show that the unique anonymous diffusion process compatible with diffusion centrality is the following: in each period, each agent transmits exactly the amount of messages they received in the previous period to each of their neighbors.

Crucially, this process operates under the assumption that both the initial sender and the recipient subsequently retransmit the message. Within the domain of targeted communications, messages containing the identity of the sender and/or target, we question this assumption that the sender or the target or both will retransmit the information.

Our second contribution is thus to introduce the concept of *targeting centrality*. This concept relies on measures of the expected number of messages from a sender to a given recipient when the sender and/or the target do not retransmit the request. We provide explicit formulas for these measures across both finite and infinite time horizons. Through formal results and simulations, we show that targeting centrality and diffusion centrality can be quite different when there is a high probability of transmission.

Lastly, we define an agent's *reachability* as the probability of their receipt of information disseminated by fellow agents. We provide an explicit formula to compute this measure.

Which measure provides a better proxy of the measure of influence that researchers are interested in depends on the specific context studied and is therefore an empirical question. We conjecture that targeting centrality and reachability will perform better than diffusion centrality in contexts where the transmitted information explicitly mentions the identities of the sender and/or the target.

In particular, we believe that bullying and cyberbullying would be promising areas to apply these concepts. Cook et al. (2010) reviews individual and environmental characteristics predicting bullying behavior. Future research could study how the distribution of these characteristics interacts with one's position in the network to predict vulnerability to bullying and the potential for a bully to cause harm.

Our work opens the door to other research questions. Targeting centrality and reachability still share some limitations with diffusion centrality, and more work can be done in the future to deepen our understanding of diffusion in networks. It would notably be interesting to introduce concepts of trust in our framework. This could be done, for instance, by relaxing the assumption that the probability of receiving a message is the same across all pairs of senders and recipients. Instead, these probabilities could differ across pairs, with  $\alpha_{ij}$  representing the level of trust between the two agents.

Relaxing the assumption of multiple retransmissions of multiple messages in the network is another natural, and challenging, direction of research. Introducing strategic behavior in information transmission would be even more challenging and rewarding. For instance, with targeted requests, agents could aim at transmitting information along the shortest paths toward the target.

## CRedit authorship contribution statement

**Yann Bramoullé:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Garance Genicot:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Declaration of competing interest

Yann Bramoullé acknowledges financial support from the European Research Council under consolidator grant 616442, the French National Research Agency under grant ANR-17-EURE-0020, and the Excellence Initiative of Aix-Marseille University, A\*MIDEX and declares that he (and his family) has no relevant or material financial interests that relate to the research described in this paper.

Garance Genicot acknowledges financial support from the National Science Foundation under grant SES-22442288 and declares that she (and her family) has no relevant or material financial interests that relate to the research described in this paper.

## Data availability

No data was used for the research described in the article.

## Appendix A. Discussion around two papers

### A.1. Banerjee et al. (2013)

Theorem 1 shows that the neat interpretation of diffusion centrality only holds under the specific anonymous diffusion process  $f^{t,DC}$  specified in (7). This highlights a difference between the models of information transmission underlying the two centrality notions proposed by Banerjee et al. (2013).

The paper starts by clearly stating that the process of information diffusion that underlies communication centrality and their structural estimations is based on an assumption of perfect recall: “In each subsequent period, households that have been informed

in previous periods pass information to each of their neighbors, independently, with probability  $q_P$  if they are participants and with probability  $q_N$  if they are not.”, (Banerjee et al. 2013, 1236498-2).

Then, on diffusion centrality, Banerjee et al. (2013) p.1236498-6 state regarding their model that “at each iteration every informed node tells each neighbor with probability  $q$ ”. It could easily be inferred that “informed node” still refers to ever informed node. However, as we clarify in Section 3.3, the assumption underlying that concept of diffusion centrality is that only informed nodes receiving the information at  $t - 1$  can transmit it at  $t$ .

The model description of Banerjee et al. (2019), p.2472 suffers from similar imprecision: “In each period, with probability  $w_{ij} \in (0, 1]$ , independently across pairs of neighbors and history, each informed node  $i$  informs each of its neighbors  $j$  of the piece of information and the identity of its original source.”

### A.2. Cruz et al. (2017)

Cruz et al. (2017) build a model of political intermediation, where a citizen’s request for a favor is transmitted to the elected politician through the social network. Two candidates  $A$  and  $B$  compete for votes. Voter  $i$  derives utility  $U_i^j$  from the clientelistic goods and services received from candidate  $j$  if elected. Requests for clientelistic goods (or services) are assumed to travel through the social network to the politicians, and voters are assumed to get utility  $b$  each time they obtain a good.

In their model, the expected benefit that voter  $i$  gets from politician  $j$  is equal to  $bn_{ij}$  where  $n_{ij}$  is the number of time the politicians hears the request (1). Implicitly, this means that politicians deliver a good each time they hear the request. It then follows in their model that  $j$ ’s vote share is an affine function of  $n_{.j} = \sum_i n_{ij}$  under diffusion process  $f^{t,DC}$ . The vote share simply depends on the candidates’ Katz-Bonacich centrality. This is the central theoretical prediction that the authors bring to data.

Note in passing that there are a couple of slight inconsistencies in this description. The implicit assumption that every successful request translates into a new favor does not seem consistent with the assumption of voter  $i$  asking for a specific favor. In addition, Cruz et al. (2017) further assume that  $\alpha = 1/\lambda_1(G)$  where  $\lambda_1(G)$  is  $G$ ’s largest eigenvalue. In that case, the number of successful requests diverges to infinity and this model predicts that politicians will provide an infinite number of favors.<sup>16</sup>

More to the point of this paper: in a context in which politician  $j$  grants a favor to  $i$  each time  $j$  receives the information, it is strange to assume that  $j$  is retransmitting the request. In fact, when a request specifically mentions the identity of the recipient and the target, it seems more sensible to assume that the request would not be retransmitted by the target nor by the person initiating the request during diffusion. This is what targeting centrality and reachability do.

### Appendix B. Latex notation

Add this code to the preamble of your LaTeX document to create symbols like  $\overleftarrow{x}$  or  $\overleftarrow{n}$ :

```
\documentclass{article}
\usepackage{tikz, mathtools}
\usetikzlibrary{arrows.meta}
\tikzset{tail/.tip={Straight Barb[reversed, length=1.5pt]}}

\newcommand{\limstyle}[1]{\mathclap{\scriptsize{#1}}}
\newcommand{\overrighttail}[2][\tikz[baseline,anchor=base]{\node[inner sep=0pt,#1](a){$#2$};
\draw[tail-]([shift={(1pt,2pt)}]a.north west)--([shift={(-1pt,2pt)}]a.north east);}]
\newcommand{\overlefttail}[2][\tikz[baseline,anchor=base]{\node[inner sep=0pt](a){$#2$};
\draw[-tail]([shift={(1pt,2pt)}]a.north west)--([shift={(-1pt,2pt)}]a.north east);}]
\newcommand{\overlefttrightrighttail}[2][\tikz[baseline,anchor=base]{\node[inner sep=0pt](a){$#2$};
\draw[tail-tail]([shift={(1pt,2pt)}]a.north west)--([shift={(-1pt,2pt)}]a.north east);}]
```

### Appendix C. Targeting centrality in finite time

#### The sender does not retransmit.

Alternatively, each walk  $w$  from  $i$  to  $j$  of length  $t$  ( $w \in W_{ij}(t)$ ) can be uniquely decomposed into: a cycle of length  $l$  from  $i$  to  $i$  (possibly  $l = 0$ ) and a walk of length  $t - l$  in  $\overleftarrow{W}_{ij}(t - l)$ . This decomposition implies that:

$$x_{ij}(t, G) = \sum_{w \in W_{ij}(t, G)} \alpha^t = \sum_{l=0}^t \sum_{c \in C_l(t, G), w' \in \overleftarrow{W}_{ij}(t-l, G)} \alpha^t$$

<sup>16</sup> Cruz et al. (2017) incorrectly claim on p. 3011: “For this particular value of  $\alpha$ , Katz centrality is equal to eigenvector centrality”. As  $\alpha \rightarrow 1/\lambda_1(G)$ , Katz-Bonacich centrality  $n_A$  diverges to infinity. It is the ratio  $n_A/n_B$  which converges to the ratio of the eigenvector centralities of the two candidates.

$$= \sum_{l=1}^t \left( \sum_{c \in C_l(l,G)} \alpha^l \right) \left( \sum_{w' \in \check{W}_{ij}(t-l,G)} \alpha^{t-l} \right).$$

Hence,

$$x_{ij}(t, G) = \sum_{l=0}^t x_{ii}(l) \check{x}_{ij}(t-l). \tag{18}$$

**Both sender and target do not retransmit.**

If both sender  $i$  and target  $j \neq i$  do not retransmit, then we can decompose each walk from  $i$  to  $j$  of length  $t$  uniquely into: a cycle of length  $0 \leq l_1 < t$  from  $i$  to  $i$  (the empty set corresponding to  $l_1 = 0$ ), a walk in  $\check{W}_{ij}(l_2)$  of length  $1 \leq l_2 \leq t - l_1$  and a cycle of length  $t - l_1 - l_2$  from  $j$  to  $j$  in  $G[-i]$  (the empty set corresponding to a cycle of length 0). This implies that:

$$\begin{aligned} x_{ij}(t, G) &= \sum_{w \in W_{ij}(t,G)} \alpha^t \\ &= \sum_{l_1=0}^{t-1} \sum_{c \in C_{l_1}(l_1,G), w' \in \check{W}_{ij}(l_2,G), c' \in C_{j(t-l_1-l_2,G[-i])}} \alpha^t \\ &= \sum_{l_1=0}^{t-1} \sum_{l_2=1}^{t-l_1} \left( \sum_{c \in C_{l_1}(l_1,G)} \alpha^{l_1} \right) \left( \sum_{w' \in \check{W}_{ij}(l_2,G)} \alpha^{l_2} \right) \left( \sum_{c' \in C_{j(t-l_1-l_2,G[-i])}} \alpha^{t-l_1-l_2} \right) \\ &= \sum_{l_1=0}^{t-1} \sum_{l_2=1}^{t-l_1} x_{ii}(l_1, G) \check{x}_{ij}(l_2) x_{jj}(t-l_1-l_2, G[-i]). \end{aligned}$$

When both sender and target are the same, excluding retransmission of messages by sender and recipients is the same, so we let  $\check{x}_{ii} = \check{x}_{ii}$ .

**Appendix D. Proofs**

**Proof of Proposition 2.** Our proof makes use of arguments appearing in the proof of Theorem B.1 in the Online Appendix of Banerjee et al. (2019). A main difference is that we consider limits as  $\alpha$  tends to  $1/\lambda_1$  for  $t = \infty$  rather than limits as  $t$  tends to  $\infty$  for a fixed  $\alpha$ .

Suppose first that matrix  $G$  is positive and diagonalizable. Let  $\lambda_1$  be the spectral radius of  $G$  and  $\lambda_2, \dots, \lambda_n$  be the other eigenvalues. By Perron-Frobenius Theorem,  $\lambda_1 > |\lambda_k|$  for  $k \neq 1$ . Then,

$$G = V \Lambda V^{-1}$$

with

$$V = (v_1^R, \dots, v_n^R) \\ V^{-1} = \begin{pmatrix} v_1^L \\ \dots \\ v_n^L \end{pmatrix}$$

and  $v_k^R$  is a column  $n$  by  $1$  vector, the right eigenvector associated with  $\lambda_k$ , while  $v_k^L$  is a row  $1$  by  $n$  vector, the left eigenvector associated with  $\lambda_k$ . Indeed,  $GV = V\Lambda$  and  $V^{-1}G = \Lambda V^{-1}$ , and hence  $Gv_k^R = \lambda_k v_k^R$  and  $v_k^L G = \lambda_k v_k^L$ . Moreover,  $\forall i, v_1^R(i) > 0$  and  $v_1^L(i) > 0$ .

We have:

$$G^t = V \Lambda^t V^{-1}$$

$$G^t = \lambda_1^t v_1^R v_1^L + \dots + \lambda_n^t v_n^R v_n^L$$

and hence

$$[\alpha G + \dots + \alpha^t G^t]_{ij} = (\alpha \lambda_1 + \dots + \alpha^t \lambda_1^t) v_1^R(i) v_1^L(j) + \dots + (\alpha \lambda_n + \dots + \alpha^t \lambda_n^t) v_n^R(i) v_n^L(j)$$

$$[\alpha G + \dots + \alpha^t G^t]_{ij} = \sum_{k=1}^n \left( \sum_{s=1}^t \alpha^s \lambda_k^s \right) v_k^R(i) v_k^L(j).$$

First, let us consider diffusion centrality with infinite time,  $n_i(G) = \sum_j n_{ij}(G)$ , which converges when  $\alpha \lambda_1 < 1$ :

$$n_{ij}(G) = \sum_{k=1}^n \left( \sum_{s=1}^{\infty} \alpha^s \lambda_k^s \right) v_k^R(i) v_k^L(j),$$

$$n_{ij}(G) = \sum_{k=1}^n \frac{\alpha \lambda_k}{1 - \alpha \lambda_k} v_k^R(i) v_k^L(j).$$

This yields

$$n_{i.}(G) = \sum_j n_{ij}(G) = \sum_{k=1}^n \frac{\alpha \lambda_k}{1 - \alpha \lambda_k} v_k^R(i) \left( \sum_j v_k^L(j) \right)$$

As  $\alpha$  converges to  $1/\lambda_1$ , the first term diverges to infinity while all other terms are finite since  $|\lambda_k| < \lambda_1$ . This means that

$$n_{ij}(G) \sim \frac{\alpha \lambda_1}{1 - \alpha \lambda_1} v_1^R(i) v_1^L(j),$$

$$n_{i.}(G) \sim \frac{\alpha \lambda_1}{1 - \alpha \lambda_1} v_1^R(i) \left( \sum_j v_1^L(j) \right),$$

and hence

$$\frac{n_{i.}(G)}{n_{i'.}(G)} \rightarrow \frac{v_1^R(i)}{v_1^R(i')}.$$

Next, consider targeting centrality. Recall, if the target does not retransmit,  $\check{n}_{.j} = \sum_{i \neq j} \check{n}_{ij}$  denotes the targeting centrality of  $j$ . From Proposition 1,

$$\check{n}_{.j}(G) = \frac{\sum_{i \neq j} n_{ij}(G)}{1 + n_{jj}(G)}.$$

Using the formulas above,

$$\sum_{i \neq j} n_{ij}(G) \sim \frac{\alpha \lambda_1}{1 - \alpha \lambda_1} \left( \sum_{i \neq j} v_1^R(i) \right) v_1^L(j),$$

$$1 + n_{jj}(G) \sim \frac{\alpha \lambda_1}{1 - \alpha \lambda_1} v_1^R(j) v_1^L(j),$$

and hence

$$\check{n}_{.j}(G) \rightarrow \frac{\sum_{i \neq j} v_1^R(i)}{v_1^R(j)}.$$

If we adopt the normalization  $\sum_i v_1^R(i) = 1$ , which we can do, this yields

$$\check{n}_{.j}(G) \rightarrow \frac{1}{v_1^R(j)} - 1.$$

This implies that the rank correlation between diffusion and targeting centrality converges to  $-1$  as  $\alpha$  converges to  $1/\lambda_1$ .

Consider instead the centrality of  $i$  to target others  $\check{n}_{i.}(G) = \sum_{j \neq i} \check{n}_{ij}(G)$  under the assumption that the target does not retransmit.

We get

$$\check{n}_{i.}(G) = \sum_{j \neq i} \frac{n_{ij}(G)}{1 + n_{jj}(G)},$$

$$\frac{n_{ij}(G)}{1 + n_{jj}(G)} \sim \frac{v_1^R(i) v_1^L(j)}{v_1^R(j) v_1^L(j)} = \frac{v_1^R(i)}{v_1^R(j)},$$

$$\check{n}_{i.}(G) \rightarrow v_1^R(i) \left[ \sum_{j \neq i} \frac{1}{v_1^R(j)} \right].$$

Suppose we adopt here the normalization  $\sum_i \frac{1}{v_1^R(i)} = 1$ . Then,  $\check{n}_{i.}(G) \rightarrow v_1^R(i) \left( 1 - \frac{1}{v_1^R(i)} \right) = v_1^R(i) - 1$  and the rank correlation converges to 1.

Finally, suppose that  $G$ , the adjacency matrix of the network, is primitive. Then,  $\lambda_1$  the spectral radius of  $G$  is a simple eigenvalue of  $G$  and for any other eigenvalue  $\lambda_k$ ,  $\lambda_1 > |\lambda_k|$ . Moreover,  $\forall i, v_1^R(i) > 0$  and  $v_1^L(i) > 0$ . We know that we can find matrices  $H$  that

are positive, diagonalizable with the same spectral radius, and arbitrarily close to  $G$ . For these matrices  $H$ , we have  $\frac{n_i(G)}{n_i'(G)} \rightarrow \frac{v_1^R(i)}{v_1^R(i')}$ ,  $\tilde{n}_{\cdot j}(G) \rightarrow \frac{1}{v_1^R(j)} - 1$  and  $\tilde{n}_{i \cdot}(G) \rightarrow v_1^R(i) - 1$  as  $\alpha$  tends to  $1/\lambda_1$ . By continuity, all these statements also hold for  $G$ . ■

**Proof of Proposition 3.** We consider requests eventually reaching  $j$  from a walk in  $\cup_{s \leq t} \tilde{W}_{ij}(s, G)$  as events. We know that the request reaches its target if and only if it reaches it through such a walk. We can then apply the principle of inclusion-exclusion to these events. This implies that

$$p_{ij}(t) = \sum_{k=1}^{|\cup_{s \leq t} \tilde{W}_{ij}(s, G)|} (-1)^{k+1} \sum_{w_1 \neq \dots \neq w_k \in \cup_{s \leq t} \tilde{W}_{ij}(s, G)} p(w_1 \wedge \dots \wedge w_k)$$

where  $p(w_1 \wedge \dots \wedge w_k)$  is the probability of the info reaching  $j$  through all walks  $w_1, w_2, \dots, w_k$ . Next, let us show recursively that  $p(w_1 \wedge \dots \wedge w_k) = \alpha^{L(w_1, \dots, w_k)}$ .

For any walk  $w$ , the probability that a request will travel all the way through  $w$  is  $p(w) = \alpha^{l(w)}$ . Next, take two walks  $w \neq w'$  originating in  $i$ . The chain rules tell us that  $p(w \wedge w') = p(w) p(w'|w)$  where  $p(w'|w)$  is the probability that  $i$ 's request will go through  $w'$  conditional on having gone through  $w$ .

$$p(w'|w) = p(w' \setminus (w \cap_b w') | w) = \alpha^{l(w') - J(w, w')}$$

with  $J(w, w')$  denoting the number of links initially common to  $w$  and  $w'$ :  $J(w, w') = l(w \cap_b w')$ . Hence,

$$p(w \wedge w') = \alpha^{l(w) + l(w') - l(w \cap_b w')} = \alpha^{L(w, w')}.$$

Assume now that we have proven that  $p(w'_1 \wedge \dots \wedge w'_{k-1}) = \alpha^{L(w'_1, \dots, w'_{k-1})}$  for any set of  $k-1$  walks originating in  $i$  ( $k \geq 2$ ) and that we have constructed  $J((w'_1 \wedge \dots \wedge w'_{k-2}), w)$  for any  $\{w_1, \dots, w_{k-2}\}$  and  $w$  that originates in  $i$ .

Take a set of  $k$  walks originating in  $i$ :  $\{w_1, \dots, w_k\}$ . Clearly,

$$p(w_1 \wedge \dots \wedge w_k) = p(w_1 \wedge \dots \wedge w_{k-1}) p(w_k | w_1 \wedge \dots \wedge w_{k-1}). \tag{19}$$

The probability that a request will go through  $w_k$  conditional on having gone through all walks in  $\{w_1, \dots, w_{k-1}\}$  is the probability that it will go through the remainder of  $w_k$  once we remove the initial links that may have been accounted for:

$$p(w_k | w_1 \wedge \dots \wedge w_{k-1}) = \alpha^{l(w_k) - J((w_1 \wedge \dots \wedge w_{k-1}), w_k)} \tag{20}$$

where  $J((w_1 \wedge \dots \wedge w_{k-1}), w_k)$  is the number of links initially common to  $(w_1 \wedge \dots \wedge w_{k-1})$  and  $w_k$ . Note that  $J((w_1 \wedge \dots \wedge w_{k-1}), w_k)$  equals the number of links initially common to both  $w_k$  and  $(w_1 \wedge \dots \wedge w_{k-2})$  plus the number of links initially common to both  $w_k$  and  $w_{k-1}$  minus the double counting<sup>17</sup>:

$$J((w_1 \wedge \dots \wedge w_{k-1}), w_k) = J((w_1 \wedge \dots \wedge w_{k-2}), w_k) + J(w_{k-1}, w_k) - J((w_1 \wedge \dots \wedge w_{k-2}), w_{k-1} \cap_b w_k). \tag{21}$$

Using (20) in (19) along with our induction hypothesis tells us that

$$p(w_1 \wedge \dots \wedge w_k) = \alpha^{L(w_1, \dots, w_{k-1})} \alpha^{J(w_k) - J((w_1 \wedge \dots \wedge w_{k-1}), w_k)}.$$

Expanding the terms in (21), we get

$$\begin{aligned} l(w_k) - J((w_1 \wedge \dots \wedge w_{k-1}), w_k) &= l(w_k) - \sum_{s \in \{1, \dots, k-1\}} l(w_s \cap_b w_k) + \dots + (-1)^{k-1} l(w_1 \cap_b \dots \cap_b w_k) \\ &= L(w_1, \dots, w_k) - L(w_1, \dots, w_{k-1}). \end{aligned}$$

It follows that  $p(w_1 \wedge \dots \wedge w_k) = \alpha^{L(w_1, \dots, w_k)}$ . ■

### Appendix E. From finite to infinite time

As discussed within the main text, the exact formulas for targeting centrality in finite time, while straightforwardly programmable, exhibit greater computational demands compared to the infinite time formula, particularly for larger values of  $t$ . A crucial question becomes whether we can employ the formula for targeting centrality in infinite time to effectively approximate the formula for targeting centrality in finite time, especially for large values of  $t$ .

Let's consider the scenario where the target refrains from retransmission. To what extent does the ratio between  $n_{ij}(t, G)$  and the count of cycles  $jj$  having lengths less than or equal to  $t$  (including the null cycle),

<sup>17</sup> The number of links common to  $(w_1 \wedge \dots \wedge w_{k-2})$  and the initial intersection between  $w_k$  and  $w_{k-1}$ .

$$\frac{n_{ij}(t, G)}{1 + n_{jj}(t, G)},$$

serve as a suitable approximation for  $\tilde{n}_{ij}(t, G)$  as  $t$  becomes large? Proposition 4 below confirms the affirmative stance, demonstrating that  $\Delta_{ij}(t) := \tilde{n}_{ij}(t)(1 + n_{jj}(t, G)) - n_{ij}(t, G)$  rapidly converges to 0 as  $t$  grows.

**Definition.** We say that sequence  $\Delta_{ij}(t, G) \geq 0$  converges geometrically towards 0 as  $t \rightarrow \infty$  if there exists  $L > 0$  and  $\varepsilon < 1$  such that  $\Delta_{ij}(t, G) \leq L\varepsilon^t$ .

**Proposition 4.**  $\Delta_{ij}(t, G) := \tilde{n}_{ij}(t, G)(1 + n_{jj}(t, G)) - n_{ij}(t, G) \geq 0$  converges geometrically towards 0 as  $t \rightarrow \infty$ .

**Proof.** By convention, set  $n_{jj}(0) = 0$ . It can be checked that

$$n_{ij}(t, G) = \sum_{s=0}^t \tilde{x}_{ij}(s) n_{jj}(t-s, G). \tag{22}$$

Indeed,

$$\begin{aligned} n_{ij}(t, G) &= \sum_{l=0}^t \sum_{s=0}^l \tilde{x}_{ij}(s, G) x_{jj}(l-s) = \sum_{s=0}^t \sum_{l=s}^t \tilde{x}_{ij}(s, G) x_{jj}(l-s, G) \\ &= \sum_{s=0}^t \sum_{s'=0}^{t-s} \tilde{x}_{ij}(s, G) x_{jj}(s', G) = \sum_{s=0}^t \tilde{x}_{ij}(s, G) n_{jj}(t-s, G). \\ \Delta_{ij}(t, G) &= \tilde{n}_{ij}(t, G)(1 + n_{jj}(t, G)) - n_{ij}(t, G) = \sum_{s=0}^t \tilde{x}_{ij}(s, G)(1 + n_{jj}(t, G)) \\ &\quad - \sum_{s=0}^t \tilde{x}_{ij}(s, G) n_{jj}(t-s, G) = \sum_{s=0}^t \tilde{x}_{ij}(s, G) (1 + n_{jj}(t, G) - n_{jj}(t-s, G)) \end{aligned}$$

are all positive elements.

Recall that  $1 + n_{jj}(t, G) = \sum_{s=0}^t \alpha^s [G^s]_{jj}$ , such that  $n_{jj}(t, G) - n_{jj}(t-s, G) = \sum_{l=t-s}^t \alpha^l [G^l]_{jj}$ .

Moreover  $\tilde{x}_{ij}(s, G) \leq x_{ij}(s, G) = \alpha^s [G^s]_{ij}$ .

Using these two inequalities in  $\Delta_{ij}(t)$  gives us

$$\Delta_{ij}(t, G) \leq \sum_{s=0}^t \alpha^s [G^s]_{ij} \left( \sum_{l=t-s}^t \alpha^l [G^l]_{jj} \right).$$

Let  $\rho(G)$  be the spectral norm of matrix  $G$ , i.e., its largest eigenvalue. Using the facts that  $G_{ij} \leq \rho(G)$  and  $\rho(G^l) = \rho(G)^l$ , we have:

$$\Delta_{ij}(t) \leq \sum_{s=0}^t \alpha^s \rho(G^s) \left( \sum_{l=t-s}^t \alpha^l \rho(G^l) \right) \leq \sum_{s=0}^t \alpha^s \rho(G)^s \frac{\alpha^{t-s} \rho(G)^{t-s}}{1 - \alpha \rho(G)} \leq \alpha^t \rho(G)^t \frac{t+1}{1 - \alpha \rho(G)}$$

Now, for any  $\varepsilon$  such that  $\alpha \rho(G) < \varepsilon < 1$ , we can check that  $\frac{(t+1)\alpha^t \rho(G)^t}{\varepsilon^t}$  converges towards 0. This shows that there exists  $L > 0$  such that  $\Delta_{ij}(t) \leq L\varepsilon^t$ . ■

Appendix F. Additional figures

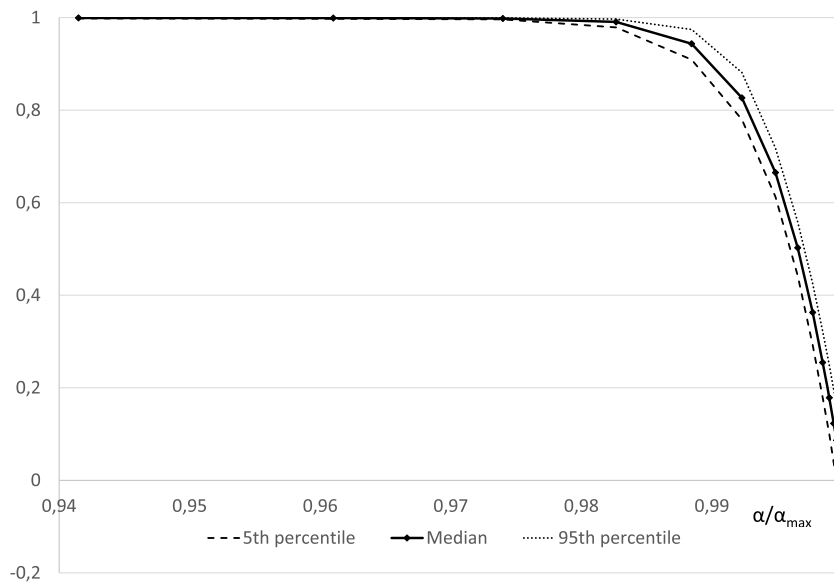


Fig. 4. RANK CORRELATION BETWEEN TARGETING AND DIFFUSION CENTRALITY. We take 1000 Erdős-Renyi random graphs with  $n = 50$  agents and probability of link formation  $p = 0.2$ . The figure plots the 5<sup>th</sup> percentile, median, and 95<sup>th</sup> percentile of the rank correlation between targeting and diffusion centrality across agents within graphs for the value of  $\alpha/\alpha_{max}$  on the horizontal axis.

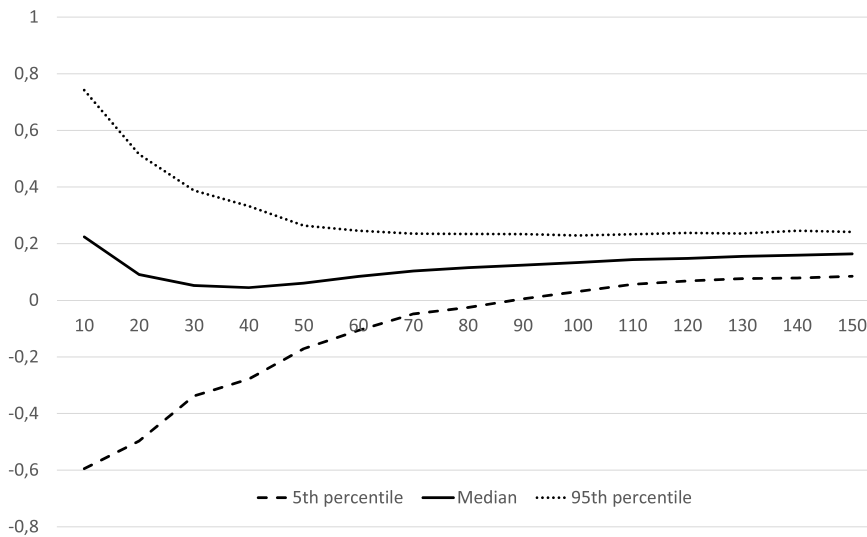


Fig. 5. RANK CORRELATION BETWEEN TARGETING AND DIFFUSION CENTRALITY. We take 1000 small world graphs where  $l$  links within 5 fully connected groups of 10 agents are randomly rewired across groups. The figure plots the 5<sup>th</sup> percentile, median, and 95<sup>th</sup> percentile of the rank correlation between targeting and diffusion centrality across agents within the graphs as  $l$  increases.

References

Atay, Ata, Mauleon, Ana, Schopohl, Simon, Vannetelbosch, Vincent, 2022. Key players in bullying networks. UB Economics–Working Papers E22/422. Universitat de Barcelona.  
 Ballester, Coralio, Calvó-Armengol, Antoni, Zenou, Yves, 2006. Who’s who in networks. Wanted: the key player. *Econometrica* 74 (5), 1403–1417.  
 Ballester, Coralio, Zenou, Yves, 2012. Key player policies when contextual effects matter. *J. Math. Sociol.* 38.  
 Banerjee, Abhijit, Chandrasekhar, Arun G., Duflo, Esther, Jackson, Matthew O., 2013. The diffusion of microfinance. *Science* 341.



- Banerjee, Abhijit, Chandrasekhar, Arun G., Duflo, Esther, Jackson, Matthew O., 2019. Using gossips to spread information: theory and evidence from two randomized controlled trials. *Rev. Econ. Stud.* 86 (6), 2453–2490.
- Beaman, Lori, BenYishay, Ariel, Magruder, Jeremy, Mobarak, Ahmed Mushfiq, 2021. Can network theory-based targeting increase technology adoption? *Am. Econ. Rev.* 111 (6), 1918–1943.
- Bloch, Francis, Jackson, Matthew O., Tebaldi, Pietro, 2023. Centrality measures in networks. *Soc. Choice Welf.*, 1–41.
- Bonacich, Phillip, 1987. Power and centrality: a family of measures. *Am. J. Sociol.* 92 (5), 1170–1182.
- Cook, Clayton R., William, Kirk R., Guerra, Nancy G., Kim, Tia E., Sadek, Shelly, 2010. Predictors of bullying and victimization in childhood and adolescence: a meta-analytic investigation. *School Psychol. Q.* 25 (2), 65–83.
- Cruz, Cesi, Labonne, Julien, Querubin, Pablo, 2017. Politician family networks and electoral outcomes: evidence from the Philippines. *Am. Econ. Rev.* 107 (10), 3006–3037.
- Dequiedt, Vianney, Zenou, Yves, 2017. Local and consistent centrality measures in parameterized networks. *Math. Soc. Sci.* 88, 28–36.
- Drago, Francesco, Mengel, Friederike, Traxler, Christian, 2020. Compliance behavior in networks: evidence from a field experiment. *Am. Econ. J. Appl. Econ.* 12 (2), 96–133.
- Duarte, Raúl, Finan, Frederico, Larreguy, Horacio, Schechter, Laura, 2019. Brokering Votes with Information Spread via Social Networks. NBER Working Papers 26241. National Bureau of Economic Research, Inc.
- Duncan, A., 2004. Powers of the adjacency matrix and the walk matrix. *Collection*, 4–11.
- Guilbeault, Douglas, Centola, Damon, 2021. Topological measures for identifying and predicting the spread of complex contagions. *Nat. Commun.* 12 (1), 1–9.
- Jackson, Matthew O., Rogers, Brian W., 2007. Relating network structure to diffusion properties through stochastic dominance. *B. E. J. Theor. Econ.* 7 (1), 1–13.
- Jackson, Matthew O., Rogers, Brian W., Zenou, Yves, 2017. The economic consequences of social-network structure. *J. Econ. Lit.* 55 (1), 49–95.
- Katz, Leo, 1953. A new status index derived from sociometric analysis. *Psychometrika* 18 (1), 39–43.
- Kaufman, Tessa M.L., Huitsing, Gijs, Bloemberg, Rick, Veenstra, René, 2021. The systematic application of network diagnostics to monitor and tackle bullying and victimization in schools. *Int. J. Bullying Prev.* 3 (1), 75–87.
- King, Maia, 2024. Walk-independence probabilities and WIP centrality: a new heuristic for diffusion probabilities in networks. *Soc. Netw.* 78, 173–183.
- Lamberson, P.J., 2016. Diffusion in networks. In: Bramoullé, Y., Rogers, B., Galeotti, A. (Eds.), *Oxford Handbook of the Economics of Networks*. Oxford University Press.
- López-Pintado, Dunia, 2008. Diffusion in complex social networks. *Games Econ. Behav.* 62 (2), 573–590.
- Rose, Christiern, 2019. Optimal injection points for information diffusion. *Econ. Lett.* 175, 67–70.
- Sadler, Evan, 2022. Ordinal centrality. *J. Polit. Econ.* 130 (4), 926–955.
- Vogels, Emily A., 2022. Teens and Cyberbullying 2022. Report. Pew Research Center.
- Watts, D.J., 2004. *Six Degrees: The Science of a Connected Age*. W. W. Norton.