Influence Maximization in Networks*

Daeyoung Jeong† Euncheol Shin‡

April 28, 2021

Abstract

We examine the optimal intervention strategy of an influence designer in the presence of a social learning process. Individuals form their opinions by taking a weighted average of their neighbors' opinions in a social network. Before learning begins, a designer with limited resources can intervene to change individuals' opinions. The designer tries to intervene in their opinions in order to lead agents with an initial opinion to have new opinions as close as possible to a target opinion, subject to budget constraints. We fully characterize the designer's optimal intervention in terms of *hub* and *authority* centrality of the influence matrix representing the underlying network structure (Kleinberg 1999). If the target opinion is embraced in terms of an individual's authority centrality, more opinion intervention should be injected for the individual. If the initial opinion is spread well in terms of an individual's hub centrality, a smaller amount of intervention should be injected for her. We also show that, when the designer has incomplete information, the optimal intervention in a large network is approximated and characterized by spectral clustering of the network. We present comparative static analyses in learning intensity and time.

JEL Classification: D83; D85.

Keywords: Davis-Kahan $\sin\Theta$ theorem; Singular value decomposition; Social learning; Social networks; Wedin $\sin\Theta$ theorem

1 Introduction

1.1 Overview

In many economic situations, individuals' choices are influenced by other individuals nearby such as friends, family, and neighbors. When one makes a purchasing decision on a product or a voting decision for a referendum, she may refer to the opinions of the people in her social network.[1] The effects of social

*We have benefited from conversations with Wonki Cho, Donggyu Kim, Sanghyun Kim, Semin Kim, Jong-Hee Hahn, KiEun Rhee, and Kirill Pogorelskiy.

†Division of Humanities and Social Sciences, Pohang University of Science and Technology; Email: daeyoung.jeong@gmail.com. Website: https://sites.google.com/site/daeyoungjeong/

‡KAIST College of Business, 85 Hoegiro, Dongdaemun-Gu, Seoul 02455, Republic of Korea. Email: eshin.econ@kaist.ac.kr. Website: https://sites.google.com/site/euncheolshin00/.

[1] People tend to conform to other people in a society. Psychologists explain this behavior with the term conformity (Cialdini and Goldstein 2004). This concept has been adopted and explored by various fields of research in social science, such as business marketing (e.g., Lascu and Zinkhan 1999) or political science (e.g., Sinclair 2012).

networks have been getting stronger with the development of the Internet and social media. Online shoppers are actively interacting through their social media pages as well as e-commerce websites, such as Amazon or eBay. Voters are eagerly sharing political opinions or related news articles on their own social media pages, such as Facebook or Twitter. These networks are redefining the market funnel and reshaping the political platform. Businesses are investing huge resources into the development of online marketing strategies,[2] and political parties are actively organizing online campaign strategies.[3]

In this study, we examine the optimal strategy of an influence designer in the presence of social learning in a network. We consider a social learning process in which individuals form their opinions by taking a weighted average of their neighbors' opinions in a social network (DeGroot 1974; Golub and Jackson 2010).[4] In our model, before learning begins, a designer with limited resources can intervene to change individuals' opinions. The cost of the intervention is increasing in the magnitude of the opinion intervention. The designer seeks to change the final outcome in favor of her objective: The designer injects the opinion intervention in order to lead agents with an initial opinion to have new opinions as close as possible to the target opinion, subject to budget constraints.

Businesses, as influence designers in an online marketplace, are actively engaging in "content marketing" and "influencer marketing." In the former, businesses directly post something on their own social media pages.[5] In the latter, they hire some outsiders, so-called *influencers*, who have power of influence in a certain online social network, such as Facebook or Instagram, and let them share some positive reviews of the firms' products.[6] In both forms of online marketing, or more specifically, of social media marketing, firms (influence designers) may control the content of the ads and/or the targeted individuals (potential customers) exposed to the ads.

[2]According to a report from eMarketer in 2020, advertisers in United States increased their spending on digital (online) advertisements spending on TV advertisements in 2016 by about 15% (eMarketer 2021).

[3]Since 2016, the new/social media executives have played important roles in the U.S. presidential campaigns. Stephen Bannon, the chief executive officer of Donald Trump's 2016 presidential campaign, was the executive chairman of Breitbart News, which is a conservative online news website founded in 2007. In 2020, the social media team of Joe Biden's 2020 presidential campaign recruited social media influencers to overcome a disadvantage in his online campaign against Donald Trump.

[4]This approach is called the DeGroot model of linear updating in the social learning literature. This approach is a simple heuristic learning rule in a social network, and it is widely used for tractability. Jackson (2010) and Bramoullé et al. (2016) excellently summarize the literature on learning models in networks. We refer to their books for other social learning models. We then discuss the DeGroot model of repeated linear updating. This theory employs a simple heuristic learning rule, delivering a fairly complete characterization of learning dynamics as a function of network structure.

[5]In the early stage of social media marketing, firms try to acquire more 'followers' for their social media pages. Then they ask how they can make the followers engage with their ads, by liking, sharing, commenting and clicking, and which content works best for that purpose. By analysing the advertising contents data from Facebook, Lee et al. 2018 examine the relationship between advertising contents in social media and followers' engagement activities.

[6]In 2015, Marriott, the international hotel chain, worked with five YouTube influencers, and released content videos to celebrate reaching one million check-ins on the Marriott mobile application.

In this context, consumers can usually be messengers of firms' ads by spreading the relevant posts. They can 'like' and/or 'share' the ads, intentionally or unintentionally, to their own 'followers.' Like the spread of disease in a pandemic, the spread of ads or information can be as severe and intense. So, we now have a neologism, 'infodemic.' Interestingly, unlike in a pandemic, in an infodemic, it is relatively easy to identify the intended party or designer of the information. Then, how does a designer optimally target interventions that reshape opinions in favor of his objective, and what are the consequences?

We fully characterize the designer's optimal intervention in terms of hub and authority centrality, or singular vectors, of the influence matrix representing the underlying network structure (Kleinberg 1999). We apply the singular value decomposition on the influence matrix. The right singular vectors are associated with the authority centrality, and the left singular vectors are associated with the hub centrality. In our framework, a good hub is an individual that spreads opinion to many good authorities; a good authority is an individual that embraces many opinions from good hubs.[7] Our main result Theorem 1 characterizes how these factors determine the optimal intervention.[8]

Note that, in some sense, the initial opinion of an individual is the obstacle for the designer to overcome, while the target opinion is the ultimate value for her to pursue. Under a particular network structure, if the target opinion is embraced in terms of an individual's authority centrality, more opinion intervention should be injected for the individual. If the initial opinion is spread well in terms of an individual's hub centrality, a smaller amount of intervention should be injected.

In reality, the information on an underling network structure may not be common knowledge or be completely known to the designer. Therefore, the analysis with Theorem 1, which requires complete information about the underlying network structure, may not be fully applicable for the real world situation. However, we have learned from the previous literature that the homophily of a social network is prevailing, and the few important "well-known" factors of a network can explain the link formation of the network. We show that, when the designer has incomplete information, the optimal intervention in a large network is approximated and characterized by spectral clustering of the network. We also present comparative static analyses in learning intensity and time.

[7] A good authority could be an "influencer" in the context of "influencer marketing."

[8] We decompose the designer's intervention into orthogonal singular vectors determined by the network and ordered according to their associated positive singular values. Since such singular value decomposition is always available, our analysis applies to asymmetric networks as well as symmetric networks.

1.2 Related Literature

The current paper is related to three strands of literature: social learning, intervention in networks, and clustering techniques in machine learning. First, there are two different approaches to social learning: (i) Bayesian updating models (Bala and Goyal 1998; Choi et al. 2005; Corazzini et al. 2012; Gale and Kariv 2003; Dasaratha and He 2020) and (ii) naive updating models (DeGroot 1974; Golub and Jackson 2010; Golub and Jackson 2012). This paper utilizes a naive updating model to focus on the influence designer's optimal intervention problem.

The intervention of an individual in networks can affect the neighbors of the individual through the network effect and/or the spillover effect as studied in the literature on network goods, where network externalities are generated in a social network (e.g., Rohlfs 1974; Fainmesser and Galeotti 2016; Galeotti et al. 2020; Radner et al. 2014; Shin 2017). In our framework, the network externalities are heterogeneous among the agents because they are heterogeneous in terms of their neighbors' connectivity as well as their connectivity.

The methodological contribution of this paper is to adopt the clustering techniques in the state-of-the-art machine learning literature to analyze the optimal intervention in networks. One closely related paper is Kleinberg (1999), which introduces the notion of hub and authorities in directed networks. These concepts, along with the singular vectors of the influence matrix, provides transparent intuition for our main results. Other related papers are the approximation techniques for stochastic block matrices. In particular, we use two theorems in the matrix perturbation theory (Stewart and Sun 1990), the Davis-Kahan $\sin\theta$ theorem and the Wedin $\sin\theta$ theorem, which are associated with convergence of singular vectors (Davis and Kahan 1970; Wedin 1972; Wedin 1983). Benefiting from these theorems, we prove that our approach can be extended to an incomplete information setting.

The rest of the paper is organized as follows. Section 2 builds the influence maximization problem. In Section 3, we explain how the singular value decomposition transforms the original influence maximization problem into a simple maximization problem. Then, in Section 4, we characterize the optimal intervention with economic intuition. In Section 5, we consider the situation in which the influence designer has incomplete information about the underlying network structure. Finally, in Section 6, we introduce particular forms of the optimal intervention as functions of model parameters of networks and agents' opinions. Section 7 concludes. All proofs are gathered in Appendix A.

2 Setup

2.1 Network and Spread of Information

Network. A *network* of n agents is represented by an $n \times n$ symmetric matrix $\mathbf{A}$ with each entry in $\{0, 1\}$.[9] $\mathbf{A}$ is called the *adjacency matrix*, and $\mathbf{A}_{ij} = 1$ represents that agent i and agent j are connected by a *link*. The *degree* of agent i is defined by $d_i(\mathbf{A}) = \mathbf{A}_i \mathbf{1}$, where $\mathbf{A}_i$ is the ith row of the adjacency matrix, and $\mathbf{1}$ is the (column) vector of ones.[10] Thus, the degree counts the number of agents sharing a link with agent i. The *degree matrix* $\mathbf{D}(\mathbf{A})$ is defined as $\mathbf{D}(\mathbf{A}) = diag(d_1(\mathbf{A}), \ldots, d_n(\mathbf{A}))$.[11] We assume that $d_i(\mathbf{A}) > 0$ for all i; that is, each agent is linked with at least one other agent.[12] The network is assumed to be *connected*; for any two agents $i, j \in N$, there is a sequence of neighbors who connect agent i to agent j.[13]

Influence in the network. We consider a social influence model in which agents form their opinions by taking a weighted average of their neighbors' opinions: agents respect opinion of their neighbors in the network and/or intend to conform with them (DeGroot 1974; Golub and Jackson 2010).[14] Let agents' initial private opinion vector be $\mathbf{b}^0 = (\mathbf{b}_1^0, \ldots, \mathbf{b}_n^0)^\mathsf{T} \in \mathbb{R}_+^n$, where $\mathbf{b}_i^0$ represents agent i's opinion, and $\mathbb{R}_+$ is the set of non-negative real numbers. We assume that agent i updates her opinion according to the following rule:

$$\mathbf{b}_i = \alpha \mathbf{b}_i^0 + (1 - \alpha) \sum_{j=1}^{n} \frac{\mathbf{A}_{ij}}{d_i} \mathbf{b}_j^0,$$

where $\alpha \in [0, 1]$ represents the relative importance of agent i's private opinion, and $\frac{\mathbf{A}_{ij}}{d_i}$ implies that she is uniformly influenced by her neighbors.[15]

We define the *influence matrix* $\mathbf{T}(\mathbf{A})$ by $\mathbf{T}(\mathbf{A}) = \alpha \mathbf{I} + (1 - \alpha) \mathbf{D}(\mathbf{A})^{-1} \mathbf{A}$, where $\mathbf{I}$ is the identity matrix of size n.[16] Thus, we write the opinion exchange system in the form of $\mathbf{b} = \mathbf{T}(\mathbf{A}) \mathbf{b}^0$. The

[9] We extend the model to asymmetric networks in Section 6.

[10] Throughout the paper, subscripts of a matrix represent a row vector of the matrix, and superscripts are for column vectors.

[11] That is, $\mathbf{D}(\mathbf{A})$ is a diagonal matrix in which its ith diagonal element is the degree of agent i.

[12] This assumption provides that the degree matrix $\mathbf{D}(\mathbf{A})$ is invertible.

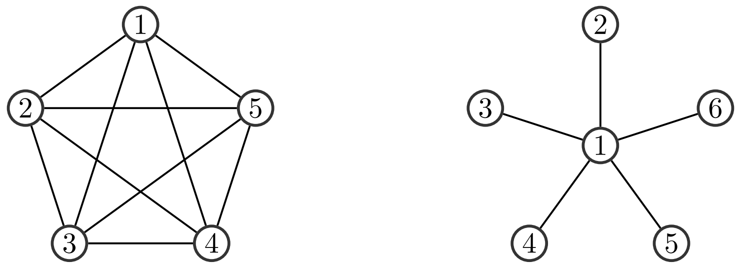
[13] Formally, a network is said to be connected if for any pair of agents (i, j), there is a sequence of agents say, $k_0 = i, k_1, \ldots, k_l = j$ such that $A_{k_s k_{s+1}} = 1$ for all $s = 0, \ldots, l - 1$. This assumption is to ensure that the inverse of the degree matrix is well-defined. The main theorem does not rely on this connectivity assumption, and we will precisely explain how this assumption plays a role in later results.

[14] We discuss this assumption further in Section 2.2.

[15] Instead, one may assume that α is heterogeneous among the agents, and our approach still incorporates this heterogeneity. Also, the uniform influence assumption of the neighbors is not crucial to derive our results.

[16] The influence matrix is well-defined because $\mathbf{D}(\mathbf{A})$ is invertiable.

following example illustrates the opinion exchange system in two benchmark networks.



(a) complete network (b) star network

Figure 1: Illustration of the two simple networks

Example 1 Consider the complete network in Figure 1-(a) consisting of five agents, where all agents are linked with each other. For $\alpha = \frac{1}{4}$, the corresponding adjacency, degree, and influence matrices are calculated as

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{D}(\mathbf{A}) = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}, \quad \mathbf{T}(\mathbf{A}) = \begin{pmatrix} \frac{1}{4} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & \frac{1}{4} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & \frac{3}{16} & \frac{1}{4} & \frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{4} & \frac{3}{16} \\ \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{4} \end{pmatrix}.$$

When the private opinion is given as $\mathbf{b}^0 = (1,1,1,0,0)^\mathsf{T}$, the updated opinion vector after the exchange of opinions is $\mathbf{b} = \mathbf{T}(\mathbf{A})\mathbf{b}^0 = \left(\frac{3}{8}, \frac{3}{8}, \frac{3}{8}, \frac{9}{16}, \frac{9}{16}\right)^\mathsf{T}$. As everyone exchanges opinions, the resulting opinions become similar to each other.

Now, consider the star network depicted in Figure 1-(b) consisting of five peripheral agents labeled by agents 2 to 6 and one central agent labeled as 1. For $\alpha = \frac{1}{4}$, the corresponding adjacency, degree, and influence matrices are

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{D}(\mathbf{A}) = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{T}(\mathbf{A}) = \begin{pmatrix} \frac{1}{4} & \frac{3}{20} & \frac{3}{20} & \frac{3}{20} & \frac{3}{20} & \frac{3}{20} \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 \\ \frac{3}{4} & 0 & 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}.$$

When the private opinion is $\mathbf{b}^0 = (0,1,0,0,0,0)^\mathsf{T}$, the resulting updated opinion is $\mathbf{b} = \mathbf{T}(\mathbf{A})\mathbf{b}^0 = \left(\frac{3}{16}, \frac{1}{4}, 0, 0, 0, 0\right)^\mathsf{T}$. Since agent 2 exchanges her opinion only with agent 1, the opinion of agents 3–6 remain the same as their initial private opinion.

Influence maximization. We consider a decision maker called the *influence designer*. Specifically, the influence designer interrupt agents' private opinions in order to move each agent's opinion closer to a target opinion $\mathbf{b}_i^* \in \mathbb{R}_+$. The designer's objective is to minimize the sum of the squared differences between the target opinion and the individual's opinion, $\sum_{i=1}^n (\mathbf{b}_i^* - \mathbf{b}_i)^2$, where $\mathbf{b}_i$ is agent i's opinion after an exchange of opinions with others in a given network.[17]

The designer injects a new set of private opinions $\mathbf{b}'$ into the network. This can be done, for example, by showing a personalized or targeted advertisement on a social media platform. Alternatively, a company may issue gift cards to consumers, and the amount in the gift card is heterogeneous among the consumers. However, we assume that the designer has limited power on the platform by assuming that he has no direct access to changing the structure of the network (i.e., $\mathbf{A}$) or the weight on private opinions (i.e., α). This assumption rules out possible scenarios in which is the influence designer can add or block certain opinion exchanges between social media users on a platform.[18]

This individual change of private opinion incurs a certain level of cost. For each agent i, the cost of influence is assumed to be $(\mathbf{b}_i' - \mathbf{b}_i^0)^2$, which is quadratic in the difference between the infused opinion $\mathbf{b}_i'$ and the initial opinion $\mathbf{b}_i^0$. As a result, the total sum of cost of influence is $\sum_{i=1}^n (\mathbf{b}_i' - \mathbf{b}_i^0)^2$. This cost can be interpreted as the cost of persuasion: The designer persuades an agent with an initial opinion $\mathbf{b}_i^0$ to believe $\mathbf{b}_i'$ by spending a certain amount of money $(\mathbf{b}_i' - \mathbf{b}_i^0)^2$.[19] The designer possesses a budget of $C > 0$, which the upper bound of the cost of influence. Consequently, the influence designer faces a budget constraint of $\sum_{i=1}^n (\mathbf{b}_i' - \mathbf{b}_i^0)^2 \leq C$.

The influence designer's problem is to solve the following optimization problem:

$$\begin{aligned} \min_{\mathbf{b}'} \quad & \sum_{i=1}^n (\mathbf{b}_i^* - \mathbf{b}_i)^2 && \text{(DP 1)} \\ \text{subject to} \quad & \mathbf{b} = \mathbf{T}(\mathbf{A})\mathbf{b}' \quad \text{exchange of opinions} \\ & \sum_{i=1}^n (\mathbf{b}_i' - \mathbf{b}_i^0)^2 \leq C \quad \text{budget constraint.} \end{aligned}$$

Note that the exchange of opinions in the network is based not on the initial opinion $\mathbf{b}^0$, but on the

[17] Our main results do depend on this quadratic objective function. In line with the decision maker's intervention on private opinions, one might interpret it as a reduced from expression of the persuation effect of personalized advertisements (e.g., Dixit and Norman 1978; Bagwell 2007).

[18] In reality, to a certain extent, the influence designer might reshape a network structure. For instance, Facebook regularly redesigns layouts of newsfeeds that the users see, and it affects users' communication structure on the platform (Source: https://techcrunch.com/2021/01/06/facebook-redesigns-pages-with-a-more-simplified-layout-and-no-like-button).

[19] Alternatively, one may interpret it as the cost of generating the misinformation or fake news: Here, $\mathbf{b}^0$ is the true state of nature and $\mathbf{b}'$ is the misinformation or fake news generated by the manipulator.

infused opinion $\mathbf{b}'$ chosen by the influence designer. The initial opinion $\mathbf{b}^0$ appears only in the budget constraint; the initial opinion works as the benchmark for calculating the adjustment cost of changing private opinion.

In order to sharpen predictions, we make two assumptions of the influence designer's problem. First, we assume that $\mathbf{b}_i^* > \mathbf{b}_i^0$ for all i, which means that every agent's private opinion is lower than the target opinion. Second, we assume that C is small. In other words, the designer's budget is too tight to infuse her ideal opinion $\mathbf{b}^*$ directly into the network:

$$\sum_{i=1}^{n}(\mathbf{b}_i^* - \mathbf{b}_i^0)^2 > C.$$

By these two assumptions, the influence designer's choice of $\mathbf{b}_i'$ is strictly higher than $\mathbf{b}_i^0$; otherwise, such choice will waste his budget.

2.2 Discussion of the Model

The designer's problem (DP 1) is similar to a network intervention problem in Galeotti et al. (2020). Instead of considering equilibrium behavior between agents in their model, agents interact with one another in the context of social learning in the current model. In particular, the agents update their opinions according to DeGroot's learning model (DeGroot 1974). One may require that the opinion of each individual is in $[0,1]$. In such case, it is necessary to assume that C is sufficiently small as in Galeotti et al. (2020); otherwise, there might exist an agent whose injected opinion chosen by the designer is strictly greater than 1, depending on the network structure. The assumption on the target opinion, $\mathbf{b}_i^* > \mathbf{b}_i^0$, implies that the designer has her own bias or directional motivation. However, this does not mean that she wants to unify the agents' opinions. Note that we still allow for fairly polarized opinions, as in $\mathbf{b}_i^* \approx 1$ and $\mathbf{b}_j^* \approx 0$ for some $i \neq j$.

We can think of many real-world applications of our theoretical approach. Consider a political party (or an organization) that wants to influence ballot casting decisions. Such a party often runs political campaigns to change voters' opinions as their decisions depend on their opinions on the issue at hand and their interests. Sometimes a party chooses to spread certain (dis)information. Obviously, the party's effort is costly, and there is a budget constraint determined by the campaign finances or the degree of misinformation. Voters are connected with one another and exchange their opinions. Each voter is less likely to vote for the party when her final opinion differs from her ideal opinion.

Obviously, each voter is equally important; the political party wants to minimize such possibility.[20]

In line with the word of mouth literature in management science, the current model also represents the mechanism of influencer marketing (e.g., Kanuri et al. 2018; Kempe et al. 2003; Lambrecht et al. 2018; Mallipeddi et al. 2021). In this context, influencers are internet celebrities who are actively engaged in social media platforms such as Facebook, Twitter, Instagram, Weibo, etc. They are often identified as users who have an enormous number of followers. By hiring (or paying off) some influencers, firms try to spread the word about their product in influencer reviews. This form of social media marketing is usually referred to as testimonial advertising.

3 Analysis

3.1 Singular Value Decomposition

Facts. We here gather mathematical facts related to the singular value decomposition, that plays the key role in our analysis.[21] For the influence matrix $\mathbf{T}(\mathbf{A}) = \alpha\mathbf{I} + (1-\alpha)\mathbf{D}(\mathbf{A})^{-1}\mathbf{A}$, the singular value decomposition provides that $\mathbf{T}(\mathbf{A}) = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, where:[22]

(i) $\mathbf{S} = diag(s_1, \ldots, s_n)$, where $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$ and $s_k \geq s_{k+1}$ for all k;[23]

(ii) $\mathbf{U}$ and $\mathbf{V}$ are normal matrices: $\mathbf{U}^\top\mathbf{U} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}$;

(iii) The kth column vector of $\mathbf{U}$, $\mathbf{u}^k$, is a right singular vector of $\mathbf{T}(\mathbf{A})$: $\mathbf{T}(\mathbf{A})\mathbf{v}^k = s_k\mathbf{u}^k$;

(iv) The kth column vector of $\mathbf{V}$, $\mathbf{v}^k$ is a left singular vector of $\mathbf{T}(\mathbf{A})$: $(\mathbf{u}^k)^\top\mathbf{T}(\mathbf{A}) = s_k(\mathbf{v}^k)^\top$.

Importantly, $\{\mathbf{u}^1, \ldots, \mathbf{u}^n\}$ forms an orthonormal basis of the column space of $\mathbf{T}(\mathbf{A})$, and $\{\mathbf{v}^1, \ldots, \mathbf{v}^n\}$ forms an orthonormal basis of the row space of $\mathbf{T}(\mathbf{A})$.

Note that the influence matrix $\mathbf{T}(\mathbf{A}) = \alpha\mathbf{I}+(1-\alpha)\mathbf{D}(\mathbf{A})^{-1}\mathbf{A}$ is not necessarily symmetric although the adjacency matrix $\mathbf{A}$ is assumed to be symmetric.[24] For this reason, the spectral decomposition technique used to study an optimal intervention in symmetric networks by Galeotti et al. (2020) is

[20]Fake news media can be another example. We have two different types of fake news media: One is a clickbait website that would like to maximize public attention by enticing agents to to click the link with misleading or sensationalized information. Another is politically biased media, which has a similar objective with a political party to directional motivation.

[21]We refer to Strang (2019) and Meyer (2010) for proofs and other applications of the singular value decomposition technique.

[22]$\mathbf{U}$ and $\mathbf{V}$ need not be unique for several reasons. First, singular values and their corresponding singular vectors can be permuted. Second, if $\mathbf{u}^k$ and $\mathbf{v}^k$ are singular vectors, then $-\mathbf{u}^k$ and $-\mathbf{v}^k$ also can be singular vectors for the same singular value; that is, each singular vector is unique only up to a sign. Third, repeated singular values may exist, and their singular vectors do not have to be unique.

[23]Since $d_i(\mathbf{A}) > 0$ for all i, at least one singular value is strictly greater than zero.

[24]More precisely, if $\mathbf{A}$ is symmetric, then $\mathbf{T}(\mathbf{A})$ is symmetric if and only if $d_i(\mathbf{A}) = d_j(\mathbf{A})$ for all i, j.

not directly applicable.[25]

Interpretation and examples. How do we interpret the $\mathbf{U}$, $\mathbf{S}$, and $\mathbf{V}$? We interpret these with a mutually reinforcing relationship discussed by Kleinberg (1999): an authority and a hub.[26] In our framework, a good hub is an agent that spreads opinion to many good authorities; a good authority is an agent that embraces opinions from many good hubs.

The singular value decomposition allows us to identify the hub and authority measure by breaking this circularity. Kleinberg (1999) calls the singular vectors $\mathbf{v}^1$ and $\mathbf{u}^1$ the hub and authority centrality measures, respectively. In our model, the hub and authority centrality measures are extended to n dimensions. That is, the left singular vector $\mathbf{v}^i$ represents how well the opinion spreads in the network, and its jth element, $\mathbf{v}^i_j$, represents the outward centrality of agent j in ith dimension. The right singular vector $\mathbf{u}^i$ represents how well the opinion is embraced in the network, and its kth element, $\mathbf{u}^i_k$, represents the inward centrality of agent k along dimension i. The diagonal matrix $\mathbf{S}$ captures the mutually reinforcing relationship between the hub and authority measures for different vectors. As the singular values are in descending order, the reinforcement effect of the singular vectors is decreasing in the index.

Example 2 Consider the complete network in Figure 1. We here focus on the first singular value and the corresponding singular vectors. Since all agents are symmetric in the complete network, the hub and authority measures are all equal, $\mathbf{U} = \mathbf{V}$ and $\mathbf{u}^1_i = \mathbf{u}^1_j = \frac{1}{\sqrt{5}}\mathbf{1}$ for all i and j. In the complete network, everyone's opinion spreads to the others in the same way, and everyone embraces each other's opinion in the same way.

For the star network, on the other hand, the two centrality measures are different. $\mathbf{v}^1 = (0.978, 0.104, 0.104, 0.104, 0.104)^\mathsf{T}$ is the hub centrality. The substantially high value of the first agent means that her opinion is dominantly spread to other agents. However, the authority centrality is measured by $\mathbf{u}^1 = (0.208, 0.489, 0.489, 0.489, 0.489)^\mathsf{T}$. This means that the central agent embraces relatively fewer opinions of other agents, and so her authority centrality is low.

[25] Since some papers on long-run behavior of the learning process focus on leading eigenvalues, orthogonality of the eigenvectors is not important in their analysis (e.g., Golub and Jackson 2010; Golub and Jackson 2012). In fact, the spectral decomposition provides an orthogonal decomposition of relevant interaction of the agents if the corresponding matrix is symmetric. However, as the influence matrix in our model is not necessarily symmetric, the singular value decomposition provides is required as it provides orthonormal decomposition of the matrix. For the relationships between the singular value decomposition and the spectral decomposition, see Strang (2019).

[26] In his paper, Kleinberg (1999) motivates this measure in the problem of searching on the world wide web and tries to measure webpages' centrality. As such, he also proposes an algorithm to calculate the hub and authority centrality measures effectively.

3.2 Transformation

We now transform the influence designer's problem. From here on, for a vector $\mathbf{b} \in \mathbb{R}^n_+$, we denote the projection of $\mathbf{b}$ onto the column space of $\mathbf{T}(\mathbf{A})$ by $\overline{\mathbf{b}} := \mathbf{U}^\mathsf{T}\mathbf{b}$, and the projection of $\mathbf{b}$ onto the row space by $\underline{\mathbf{b}} := \mathbf{V}^\mathsf{T}\mathbf{b}$. Using this notation system, when the designer chooses $\mathbf{b}'$, the resulting projected opinion of the agents after interactions can be written as $\overline{\mathbf{b}} = \mathbf{S}\underline{\mathbf{b}}'$, where $\overline{\mathbf{b}} := \mathbf{U}^\mathsf{T}\mathbf{b}$ and $\underline{\mathbf{b}}' = \mathbf{V}^\mathsf{T}\mathbf{b}'$.[27] Thus, $\underline{\mathbf{b}}'$ is the hub-centrality projected opinion of $\mathbf{b}'$. The $\overline{\mathbf{b}}$ is the resulting opinion projected in the space of authority-centrality.[28]

Along with the change of notation, the singular value decomposition also enables us to find simple alternative forms of mathematical expressions in the influence designer's problem. First, after plugging $\mathbf{b} = \mathbf{T}(\mathbf{A})\mathbf{b}'$ into the original objective function, we obtain

$$\sum_{i=1}^{n} (\mathbf{b}^*_i - \mathbf{T}(\mathbf{A})\mathbf{b}_i)^2 = (\overline{\mathbf{b}}^*)^\mathsf{T}\overline{\mathbf{b}}^* - 2(\overline{\mathbf{b}}^*)^\mathsf{T}\mathbf{S}\underline{\mathbf{b}}' + (\underline{\mathbf{b}}')^\mathsf{T}\mathbf{S}^2\underline{\mathbf{b}}'.$$

The budget constraint is

$$\sum_{i=1}^{n} \left(\mathbf{b}'_i - \mathbf{b}^0_i\right)^2 = \left(\underline{\mathbf{b}}' - \underline{\mathbf{b}}^0\right)^\mathsf{T} \left(\underline{\mathbf{b}}' - \underline{\mathbf{b}}^0\right) \leq C.$$

Therefore, the original problem (DP 1) now becomes

$$\begin{aligned} \min_{\underline{\mathbf{b}}'} \quad & (\overline{\mathbf{b}}^*)^\mathsf{T}\overline{\mathbf{b}}^* - 2(\overline{\mathbf{b}}^*)^\mathsf{T}\mathbf{S}\underline{\mathbf{b}}' + (\underline{\mathbf{b}}')^\mathsf{T}\mathbf{S}^2\underline{\mathbf{b}}' \\ \text{subject to} \quad & \left(\underline{\mathbf{b}}' - \underline{\mathbf{b}}^0\right)^\mathsf{T} \left(\underline{\mathbf{b}}' - \underline{\mathbf{b}}^0\right) \leq C. \end{aligned} \tag{DP 2}$$

In the original problem, the infused information of an agent i influences the opinions of the other agents, and vice versa. This interdependent influence is summarized by the influence matrix $\mathbf{T}(\mathbf{A})$. This inter-dependency of the infused information and the resulting opinions make the analysis and interpretation more complicated. Surprisingly, however, this dependency problem disappears when we transform the problem with the singular value decomposition. Agents' opinions are projected into the spaces having singular vectors as the base. Since singular vectors are orthogonal to each other, the change of one projected opinion does not affect any of other projected opinions. This convenience is

[27] To see why, note that $\mathbf{b} = \mathbf{T}(\mathbf{A})\mathbf{b}' = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{b}'$. Since $\mathbf{U}$ is normal, it follows that $\mathbf{U}^\mathsf{T}\mathbf{b} = \mathbf{T}(\mathbf{A})\mathbf{b}' = \mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{b}'$, which is equivalent to the expression in the main text.

[28] Geographically, the singular value decomposition of $\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ is a rotation $\mathbf{V}^\mathsf{T}$, followed by a scaling $\mathbf{S}$, followed by another rotation $\mathbf{U}$. So, $\underline{\mathbf{b}}'$ is the infused information rotated by $\mathbf{V}^\mathsf{T}$, and $\overline{\mathbf{b}}$ is the infused information rotated by $\mathbf{V}^\mathsf{T}$ and scaled by $\mathbf{S}$. Rotating $\overline{\mathbf{b}}$ by $\mathbf{U}$ gives the resulting opinion $\mathbf{b}$.

effectively presented by the following expression:

$$\overline{\mathbf{b}} = \mathbf{U}^\mathsf{T}\mathbf{b} = \begin{pmatrix} \overline{\mathbf{b}}_1 \\ \vdots \\ \overline{\mathbf{b}}_n \end{pmatrix} = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{pmatrix} \begin{pmatrix} \underline{\mathbf{b}}'_1 \\ \vdots \\ \underline{\mathbf{b}}'_n \end{pmatrix} = \mathbf{S}\underline{\mathbf{b}}' = \mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{b}'$$

A change in $\underline{\mathbf{b}}_i$ results in a change in $\overline{\mathbf{b}}_i$ only. As a result, the singular value decomposition enables us to treat $\underline{\mathbf{b}}_i$ and $\underline{\mathbf{b}}_j$ independently. Finally, if the optimal infused opinion is characterized by $\underline{\mathbf{b}}'$ in the projected space, then the inverse transformation $\mathbf{b}' = \mathbf{V}\underline{\mathbf{b}}'$ becomes the corresponding optimal solution that we are looking for.

4 Optimal Influence Design

In this section, we develop a characterization of the optimal solution to problem (DP 2) in terms of the hub and authority measures and examine their properties. Benefiting from the singular value decomposition, we treat $\underline{\mathbf{b}}_i$ and $\underline{\mathbf{b}}_j$ independently for all i and j. As the optimization problem is convex, the first-order condition fully characterizes the optimal solution. Specifically, the first-order equation for $\underline{\mathbf{b}}'_k$ is

$$(s_k\overline{\mathbf{b}}^*_k - s_k^2\underline{\mathbf{b}}'_k) = \mu(\underline{\mathbf{b}}'_k - \underline{\mathbf{b}}^0_k),$$

where $\mu > 0$ is the Lagrangian multiplier for the budget constraint. The left-hand side represents the marginal benefit of increasing $\underline{\mathbf{b}}'_k$, and the right-hand side is the marginal cost of it. By rearranging the equation, we find the optimal solution $\underline{\mathbf{b}}'_k$ as $\underline{\mathbf{b}}'_k = \frac{s_k}{s_k^2+\mu}\overline{\mathbf{b}}^*_k + \frac{\mu}{s_k^2+\mu}\underline{\mathbf{b}}^0_k \geq 0$. In a matrix form, we can find the closed form solution of $\underline{\mathbf{b}}'$ as $\underline{\mathbf{b}}' = \underline{\mathbf{b}}^0 + (\mathbf{S}^2 + \mu\mathbf{I})^{-1}\mathbf{S}(\overline{\mathbf{b}}^* - \mathbf{S}\underline{\mathbf{b}}^0)$.

We now find the solution of the original problem (DP 1) as

$$\begin{aligned} \mathbf{b}' &= \mathbf{b}^0 + \mathbf{V}(\mathbf{S} + \mu\mathbf{I})^{-1}\mathbf{S}\left(\mathbf{U}^\mathsf{T}\mathbf{b}^* - \mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{b}^0\right) \\ &= \mathbf{b}^0 + \sum_{k=1}^{n} \frac{s_k}{s_k^2 + \mu}\left(||\mathbf{b}^*||\cos(\mathbf{u}^k, \mathbf{b}^*) - s_k||\mathbf{b}^0||\cos(\mathbf{v}^k, \mathbf{b}^0)\right)\mathbf{v}^k, \end{aligned}$$

where $||\cdot||$ is the canonical Euclidean norm of vectors. Therefore, the degree of injected opinion vector, $\mathbf{b}' - \mathbf{b}^0$, is a weighted sum of hub centrality vectors (i.e., $\mathbf{v}^k$'s), where each weight is a function of a singular value (i.e., s_k), the *cosine similarity* of the target opinion (i.e., $\mathbf{b}^*$) to the authority centrality vectors, and the *cosine similarity* of the private opinion (i.e., $\mathbf{b}^0$) to the hub centrality vectors.

Furthermore, we can pin down the exact value of μ. As the budget constraint is binding at the optimal choice, μ solves the equation

$$\sum_{i=1}^{n}(\underline{\mathbf{b}}_i' - \underline{\mathbf{b}}_i^0)^2 = \sum_{k=1}^{n}\left(\frac{s_k}{s_k^2+\mu}\right)^2\left(\overline{\mathbf{b}}_k^* - s_k\underline{\mathbf{b}}_k^0\right)^2 = C.$$

There is a unique μ that satisfies the above equation as the summation is strictly decreasing in μ. This feature also implies that μ is decreasing in budget C. As such, μ is interpreted as the shadow price of the budget. The following theorem summarizes our analysis so far:

Theorem 1 *The solution of the influence maximization problem (DP 1) is*

$$\mathbf{b}' = \mathbf{b}^0 + \sum_{k=1}^{n}\frac{s_k}{s_k^2+\mu}\left(||\mathbf{b}^*||\cos(\mathbf{u}^k,\mathbf{b}^*) - s_k||\mathbf{b}^0||\cos(\mathbf{v}^k,\mathbf{b}^0)\right)\mathbf{v}^k,$$

where μ is a unique solution of the equation

$$\sum_{k=1}^{n}\left(\frac{s_k}{s_k^2+\mu}\right)^2\left(\overline{\mathbf{b}}_k^* - s_k\underline{\mathbf{b}}_k^0\right)^2 = C.$$

We provide the intuition behind Theorem 1. We can say that the influence designer injects the optimal opinion intervention $\mathbf{b}'$ in order to lead agents with the initial opinion $\mathbf{b}^0$ to have new opinions as close to the target opinion $\mathbf{b}^*$ as possible, subject to the budget constraint. In other words, the initial opinion is the obstacle for the designer to overcome, while the target opinion is the ultimate value for her to pursue.

The optimal degree of new opinion to agent i is

$$\mathbf{b}_i' - \mathbf{b}_i^0 = \sum_{k=1}^{n}\frac{s_k}{s_k^2+\mu}\left(||\mathbf{b}^*||\cos(\mathbf{u}^k,\mathbf{b}^*) - s_k||\mathbf{b}^0||\cos(\mathbf{v}^k,\mathbf{b}^0)\right)\mathbf{v}_i^k.$$

This is a weighted sum of the i'th entry $\mathbf{v}_i^k$ in the k'th left singular vector $\mathbf{v}^k$ (hub centrality in the k'th dimension). The entry $\mathbf{v}_i^k$ measures the importance of agent i in the left singular vector $\mathbf{v}^k$. If other things are equal, the designer's investment on agent i's opinion is proportional to the level of $\mathbf{v}_i^k$. It is weighted by $\frac{s_k}{s_k^2+\mu} > 0$, which captures the multiplier effect of the injected opinion. There are two other factors that matter beyond the hub centrality and the multiplier effect in k's dimension.

The factors include two similarities that depend on the network structure.[29] $\cos(\mathbf{u}^k,\mathbf{b}^*)$ measures the similarity of the k'th right singular vector, $\mathbf{u}^k$, and the target opinion, $\mathbf{b}^*$: As the target is similar to the k'th authority centrality, the weight increases. This similarity captures the benefit of the intervention in terms of the k'th authority centrality: If the targeted opinion could have been embraced well in terms of the k'th authority centrality, more opinion intervention should be injected. The other similarity, $\cos(\mathbf{v}^k,\mathbf{b}^0)$, is the similarity of k'th left singular vector, $\mathbf{v}^k$, and the initial opinion, $\mathbf{b}^0$: As

[29] Note that $||\mathbf{b}^*||$ and $||\mathbf{b}^0||$ are exogenously given and independent of the network structure.

the initial opinion is similar with the k'th hub centrality, the weight decreases. This similarity captures the cost of the intervention in terms of the k's hub centrality: If the initial opinion is spread well in terms of the k's hub centrality, a smaller amount of intervention should be injected.

We finally remark that for some singular vectors, its similarity to the target opinion or the initial opinion can be negative. Also, except the first left and right singular vectors, which are positive, all the other singular vectors contain at least one negative entry.[30] Hence, the amount of opinion intervention can be negative for some $\mathbf{v}^k$, but the final summation is positive. See Section 6 for examples.

5 Incomplete Information

Here we extend the previous singular value decomposition approach to the situation, where there is incomplete information about the underlying network structure. To solve the influence maximization problem under this situation, we introduce a technique from the matrix perturbation theory that enables us to approximate the optimal intervention for large networks.

5.1 Network Formation Model

Multi-type random network. We consider a simple network formation model that explains many observed network characteristics in reality.[31] Specifically, we consider the multi-type random networks proposed by Golub and Jackson (2012).[32] In the model, agents have types, which are the distinguishing features that affect their propensities to connect to each other. Examples of types are gender, age, race, education level, and so forth. We assume that there are m different types. Let $\mathbf{P}$ be a symmetric $m \times m$ matrix, whose entries in $(0, 1)$ describe the probabilities of links between types. The adjacency matrix is a realization of the random network in which entries $\mathbf{A}_{ij}$ with $i > j$ are independent Bernoulli random variables that take a value of 1 with probability $\mathbf{P}_{kl} \in (0, 1)$ when agent i is in group k and agent j is in group l. The other entries $\mathbf{A}_{ij}$ are automatically filled by $\mathbf{A}_{ij} = \mathbf{A}_{ji}$, and we let $\mathbf{A}_{ii} = 0$ for all i.[33]

[30] To see why, recall that the first left singular vector is positive, and other left singular vectors are orthogonal to the principal vector. Thus, there is at least one negative entry as singular vectors are not zero vectors. The same reasoning applies to the right singular vectors.

[31] See Jackson (2010) and Bramoullé et al. (2016) for empirical regularities of network characteristics and network formation models generating such properties. See Jackson and Rogers (2007), Shin (2021), and references therein for other network formation models explaining empirical regularities.

[32] In the statistics literature, this model is also called the *stochastic block model*. See Fan et al. (2020) and references therein for backgrounds and applications in the recent literature on big data analysis and machine learning.

[33] One may consider the case in which $\mathbf{A}_{ij} \neq \mathbf{A}_{ji}$. The approach in the current paper still applies with some variation in a few steps. See Section 6 for details.

For simplicity, we assume that (i) every group has the same number of agents, (ii) an agent only distinguishes between agents of her own type and agents of a different type, and (iii) every agent of the same type forms links with other agents in the same way. Assumptions (i) and (ii) hold for expositional simplicity, and the main results hold for a more general setting.[34] Assumption (iii) is the only crucial assumption. Golub and Jackson (2012) call this the *island model* with parameters (m, p_s, p_d) and it formalizes as the following:

(i) There are m islands of equal size $\frac{n}{m}$.

(ii) $\mathbf{P}_{kk} = p_s$ for all $k = 1, \ldots, m$.

(iii) $\mathbf{P}_{kl} = \mathbf{P}_{lk} = p_d$ for all $k \neq l$.

(iv) $p_s > p_d > 0$.

(v) $p_s, p_d \sim \mathcal{O}\left(\frac{\log n}{n}\right)$.

Condition (i) is for without loss of generality, and the main results of the current section do not rely on it. $p_s > p_d$ captures the idea of *homophily* that refers to the phenomenon that similar people tend to attach to each other more often than dissimilar ones (e.g., McPherson et al. 2001; Bramoullé et al. 2012).[35] Condition (v) is a necessary technical assumption required.[36] Figure 2 illustrates a possible realization from an island network of size 18 with parameters $(3, 1, \frac{1}{6})$. Each color represents a type of agent. All agents of the same type (color) are linked with one another as $p_s = 1$, but only a few of the agents in a group ($\frac{1}{3}$ of the agents of each type) are linked with a few of the agents in other groups.

5.2 Characterization of the Optimal Intervention

We consider the influence designer's problem when she does not know the exact structure of the network; instead, she knows that the probability of link formation between the agents follows the island model. Since we assume that the realized network is symmetric, the same optimal intervention strategy with the singular value decomposition is optimal if the designer knows the realized network.

A natural alternative candidate of the invention strategy is to utilize the information of the network generating process. Specifically, we let $\overline{\mathbf{A}} = \mathbb{E}[\mathbf{A}]$ be the expectation of the network, and let $\mathbf{A} = \overline{\mathbf{A}} + (\mathbf{A} - \overline{\mathbf{A}})$, where $(\mathbf{A} - \overline{\mathbf{A}})$ captures the noise (or deviation) from the expectation of the realized

[34]For instance, the main results hold for the case in which the number of agents with the same type is heterogeneous among the types. As long as the number of types is finite and independent of the network size, all the results in the paper hold.

[35]See Echenique and Fryer (2007) and references therein for homophily measurements.

[36]In other words, this assumption provides convergence of related eigenvalues and eigenvectors when the size of the network is sufficiently large. This assumption is frequently required for theoretical analysis for large networks and convergence results (e.g., Golub and Jackson 2012; Cho et al. 2017).

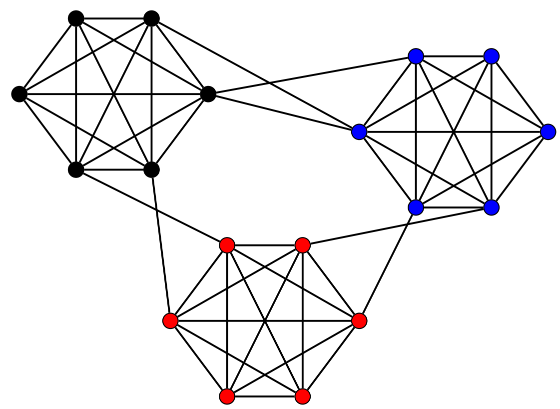
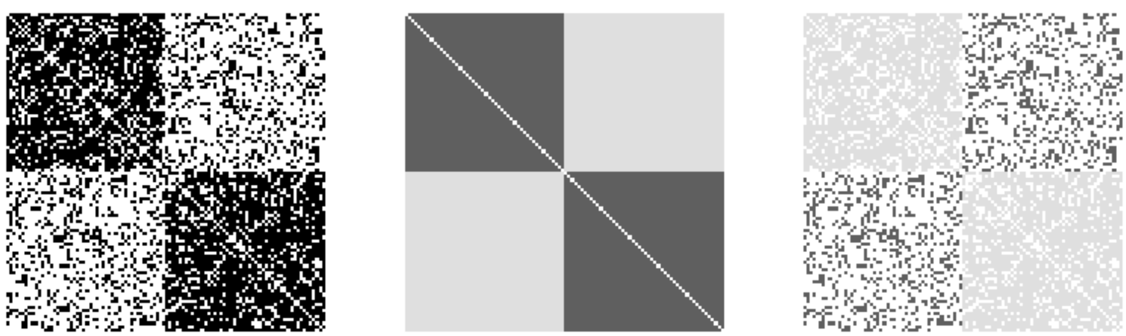


Figure 2: Illustration of the island network model

adjacency matrix. Figure 3-(a) is an illustration of the symmetric adjacency matrix of a realized network, which is generated by the island model of $(2, p_s, p_d)$ with $p_s = \frac{2}{3} > \frac{1}{3} = p_d$ when there are 100 agents. Black dot represents that there is a link between two agents.[37] This matrix can be decomposed as a sum of its expectation and the noise. Figure 3-(b) is the expectation of rank 2, and Figure 3-(c) is the noise.



(a) realization (b) expectation (c) noise

Figure 3: Illustration of matrix perturbation approach

One notable feature of this decomposition is that the expectation has rank 2 as

$$\overline{\mathbf{A}} = \begin{bmatrix} p_s \mathbf{1}_{n/2}\mathbf{1}_{n/2}^\intercal & p_d \mathbf{1}_{n/2}\mathbf{1}_{n/2}^\intercal \\ p_d \mathbf{1}_{n/2}\mathbf{1}_{n/2}^\intercal & p_s \mathbf{1}_{n/2}\mathbf{1}_{n/2}^\intercal \end{bmatrix} = \frac{p_s + p_d}{2}\mathbf{1}_{n/2}\mathbf{1}_{n/2}^\intercal + \frac{p_s - p_d}{2}\begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix}\begin{bmatrix} \mathbf{1}_{n/2}^\intercal & -\mathbf{1}_{n/2}^\intercal \end{bmatrix},$$

where $\mathbf{1}$ is the vector of ones with length $\frac{n}{2}$.[38] Independent of the number of types, the rank of the expectation matrix $\overline{\mathbf{A}}$ is 2 as the degree of freedom in the island model is 2.[39]

[37] Note that the diagonal entries are all white as no agent is linked to herself.

[38] $\frac{p_s+p_d}{2}$ is the first singular value, and $\frac{p_s-p_d}{2}$ is the second singular value. All the other singular values are zero.

[39] This observation implies that for the multi-type network having parameters up to m^2, and the rank of the expectation is at most m^2. Since our result only replies on the fixed rank size of the expectation matrix, our result holds for general multi-type networks as well as island model, which is a special case of the multi-type model. See Section 6 for more

We now explain a way to construct an approximated intervention $\overline{\mathbf{b}}(n)$ for the actual optimal intervention $\mathbf{b}'(n)$, as functions of network size n. We construct the expected influence matrix as $\mathbf{T}(\overline{\mathbf{A}}) = \alpha\mathbf{I} + (1-\alpha)\mathbf{D}(\overline{\mathbf{A}})^{-1}\,\overline{\mathbf{A}}$; then define the optimal strategy $\overline{\mathbf{b}}(n)$ by the singular value decomposition of $\mathbf{T}(\overline{\mathbf{A}})$ as in previous sections. Our goal is to show that as the network size n increases to infinity, $\mathbf{b}'(n)$ converges to $\overline{\mathbf{b}}(n)$ up to some normalization of population size under a proper convergence notion. In this regard, we define the following:

Definition 1 $\overline{\mathbf{b}}(n)$ is said to be *asymptotically optimal* if for any given $\varepsilon > 0$, there is N such that $n > N$ implies

$$\frac{1}{\sqrt{n}}||\mathbf{b}'(n) - \overline{\mathbf{b}}'(n)|| < \varepsilon$$

with probability at least $1-\varepsilon$.

We now explain why the approximated intervention $\overline{\mathbf{b}}(n)$ is asymptotically optimal for large networks. Let $\overline{s}_k$, $\overline{\mathbf{u}}^k$, and $\overline{\mathbf{v}}^k$ be the kth singular value and corresponding singular vectors of $\mathbf{T}(\overline{\mathbf{A}})$, respectively. Then, by definition, we have

$$\begin{aligned}
\overline{\mathbf{b}}'(n) &= \sum_{k=1}^{n} \frac{\overline{s}_k}{\overline{s}_k^2 + \overline{\mu}} \Big(||\mathbf{b}^*|| \cos(\overline{\mathbf{u}}^k, \mathbf{b}^*) - \overline{s}_k ||\mathbf{b}^0|| \cos(\overline{\mathbf{v}}^k, \mathbf{b}^0) \Big) \overline{\mathbf{v}}^k \\
&= \sum_{k=1}^{2} \frac{\overline{s}_k}{\overline{s}_k^2 + \overline{\mu}} \Big(||\mathbf{b}^*|| \cos(\overline{\mathbf{u}}^k, \mathbf{b}^*) - \overline{s}_k ||\mathbf{b}^0|| \cos(\overline{\mathbf{v}}^k, \mathbf{b}^0) \Big) \overline{\mathbf{v}}^k \\
&\quad + \sum_{l=3}^{n} \frac{\overline{s}_l}{\overline{s}_l^2 + \overline{\mu}} \Big(||\mathbf{b}^*|| \cos(\overline{\mathbf{u}}^l, \mathbf{b}^*) - \overline{s}_l ||\mathbf{b}^0|| \cos(\overline{\mathbf{v}}^l, \mathbf{b}^0) \Big) \overline{\mathbf{v}}^l,
\end{aligned}$$

where $\overline{\mu}$ is the Lagrangian multiplier of the maximization problem. Thus, we need to show that $\overline{s}_k$, $\overline{\mathbf{u}}^k$, and $\overline{\mathbf{v}}^k$ converge to the corresponding values and vectors of $\mathbf{T}(\mathbf{A})$ as n increases to infinity. To show these, we first show that $\mathbf{H} = \mathbf{T}(\mathbf{A}) - \mathbf{T}(\overline{\mathbf{A}})$ becomes arbitrarily small as the size of network increases to infinity. To this end, we find

$$||\mathbf{H}||_{\text{op}} = \Big|\Big| \left[\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})} - \frac{\overline{\mathbf{A}}_{ij}}{d_i(\overline{\mathbf{A}})} \right] \Big|\Big|_{\text{op}} \leq \Big|\Big| \left[\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})} \left(1 - \frac{d_i(\mathbf{A})}{d_i(\overline{\mathbf{A}})}\right) \right] \Big|\Big|_{\text{op}} + \Big|\Big| \left[\frac{(\mathbf{A}_{ij} - \overline{\mathbf{A}}_{ij})}{d_i(\overline{\mathbf{A}})} \right] \Big|\Big|_{\text{op}},$$

where $||\cdot||_{\text{op}}$ represents the operator norm of matrices. Then, by the standard techniques in the literature on social and economic networks (e.g., Golub and Jackson 2012), with a high probability, both terms on the right-hand side of the above inequality converge to zero, which relies on our assumption that $p_s, p_d \sim \mathcal{O}(\log n/n)$. Consequently, $||\mathbf{H}||_{\text{op}}$ converges to zero, and it implies that the $\overline{s}_k \to s_k$ as n increases to infinity for all k.

details.

Unfortunately, the convergence of $||\mathbf{H}||_{\text{op}}$ is not sufficient to ensure convergence of the singular vectors even if all singular values converge.[40] Thus, we need to separately show convergence of singular vectors, and this part relies on the Wedin $\sin\theta$ theorem.

We here briefly introduce the theorem. We note that the two leading singular values are strictly positive $\bar{s}_2 = \triangle > 0$. In addition, for any given $\varepsilon > 0$, $||\mathbf{H}||_{\text{op}} < \triangle\varepsilon$ with a high probability for large n. The Wedin $\sin\theta$ theorem states that with probability at least $1-\varepsilon$, the first two leading singular vectors of the expected influence matrix are close to the first two leading singular vectors of the realized influence matrix (Wedin 1972; Wedin 1983).[41] Specifically, let $\Theta(\mathbf{V},\widehat{\mathbf{V}})$ denote the $d \times d$ diagonal matrix whose jth diagonal entry is the jth singular angle, and let $\sin\Theta(\mathbf{V},\widehat{\mathbf{V}})$ be defined entry-wise. Then,

$$\max\{\sin\Theta(\mathbf{U}_0,\overline{\mathbf{U}}_0),\sin\Theta(\mathbf{V}_0,\overline{\mathbf{V}}_0)\} \leq \frac{||\mathbf{H}||_{\text{op}}}{\triangle} < \varepsilon,$$

where $\mathbf{U}_0$ and $\overline{\mathbf{U}}_0$ are the matrices having the two leading right singular vectors of $\mathbf{T}(\mathbf{A})$ and $\mathbf{T}(\overline{\mathbf{A}})$, respectively; and $\mathbf{V}_0$ and $\overline{\mathbf{V}}_0$ are the matrices having the two leading left singular vectors of $\mathbf{T}(\mathbf{A})$ and $\mathbf{T}(\overline{\mathbf{A}})$, respectively.

Therefore, by the Wedin $\sin\theta$ theorem, we have convergence of four the leading left and right singular vectors as well as their corresponding singular values of the expected influence matrix. Consequently, the approximated intervention is asymptotical optimal as the following theorem summarizes:

Theorem 2 *The approximated intervention* $\overline{\mathbf{b}}'(n)$ *is asymptotically optimal.*

6 Further Analysis

We now provide further analysis of the optimal intervention under different environments. In particular, we examine some situations under which the optimal intervention has a simple form. Then, we extend our analysis for directed networks and general multi-type networks.

6.1 Uniform and Clustered Interventions

We here consider the incomplete information setting and introduce special forms of the optimal intervention $\overline{\mathbf{b}}(n)$ and the shadow price $\overline{\mu}$. We also provide some comparative static results for them.

[40] A popular example of this problem is the following. Let $\varepsilon > 0$ and consider two matrices $\mathbf{I}$ and $\mathbf{M}$:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{M} = \mathbf{I} + \varepsilon\mathbf{1}\mathbf{1}^\intercal = \begin{pmatrix} 1+\varepsilon & \varepsilon \\ \varepsilon & 1+\varepsilon \end{pmatrix}.$$

$\mathbf{M}$ is a small perturbation of $\mathbf{I}$ as $||\mathbf{I}-\mathbf{M}||_{\text{op}} = 2\varepsilon$. They have similar eigenvalues as $\lambda(\mathbf{I}) = 1$ and $\lambda(\mathbf{M}) = (1+2\varepsilon), 1$. However, $\mathbf{M}$ has totally different eigenvectors of $(1,1)^\intercal$ and $(1,-1)^\intercal$.

[41] Yu et al. (2015) provide other variations of the theorem.

In order to obtain sharp results, we assume that there are two types among the agents. We also assume that $\alpha = 0$, and so $\mathbf{T}(\overline{\mathbf{A}}) = \mathbf{D}(\overline{\mathbf{A}})^{-1}\,\overline{\mathbf{A}}$; thus, every agent's opinion is solely determined as the average of their neighbors' opinions. Under this assumption, we investigate how the composition of optimal intervention $\overline{\mathbf{b}}(n)$ and the shadow price $\overline{\mu}$ are expressed as functions of two parameters determining network structure (p_s and p_d) and the target and initial opinions ($\mathbf{b}^*$ and $\mathbf{b}^0$) change.

Uniform intervention. We first consider the simplest example in which $\mathbf{b}^* = b^*\mathbf{1}$ and $\mathbf{b}^0 = b^0\mathbf{1}$ for some $0 \le b^* < b^0 \le 1$, where $\mathbf{1}$ is the vector of ones, and $\mathbf{0}$ is the vector of zeros of length n. Under this assumption, since influence matrix $\mathbf{T}(\overline{\mathbf{A}})$ is symmetric, $\cos(\overline{\mathbf{u}}^1, \mathbf{b}^*) = \cos(\overline{\mathbf{v}}^1, \mathbf{b}^0) = 1$ and $\cos(\overline{\mathbf{u}}^2, \mathbf{b}^*) = \cos(\overline{\mathbf{v}}^2, \mathbf{b}^0) = 0.$[42] In addition, for the island model, the largest singular value is $\overline{s}_1 = \frac{p_s+p_d}{2}$. Therefore, the optimal intervention is uniform as

$$\overline{\mathbf{b}}'(n) = \mathbf{b}^0 + \frac{\frac{p_s+p_d}{2}}{\frac{p_s+p_d}{2} + \overline{\mu}} \left(b^* - \frac{(p_s+p_d)b^0}{2} \right) \mathbf{1}.$$

The above uniform intervention arises from the facts that although there are two different types of agents, they are all symmetric in terms of (i) their hub and authority centralities, and (ii) they have the same cosine similarities to the target belief and the initial belief.

For the shadow price $\overline{\mu}$, we obtain

$$\left(\frac{\frac{p_s+p_d}{2}}{(\frac{p_s+p_d}{2})^2 + \overline{\mu}} \right)^2 \left(b^* - \frac{(p_s+p_d)b^0}{2} \right)^2 = C,$$

and it follows that $\overline{\mu}$ is increasing in p_s and p_d simultaneously because C is assumed to be small. That is, a uniform intervention becomes more valuable when agents become more tightly connected with one another independent of their types; consequently they behave as if they are all identical. The following proposition summarizes:

Proposition 1 $\mathbf{b}^* = b^*\mathbf{1}$ *and* $\mathbf{b}^0 = b^0\mathbf{1}$*. Then, the optimal intervention is uniform as*

$$\overline{\mathbf{b}}'(n) = \frac{\frac{p_s+p_d}{2}}{\frac{p_s+p_d}{2} + \overline{\mu}} \left(b^* - \frac{(p_s+p_d)b^0}{2} \right) \mathbf{1},$$

where $\overline{\mu}$ *is*

$$\left(\frac{\frac{p_s+p_d}{2}}{(\frac{p_s+p_d}{2})^2 + \overline{\mu}} \right)^2 \left(b^* - \frac{(p_s+p_d)b^0}{2} \right)^2 = C.$$

$\overline{\mu}$ *is increasing in* p_s *and* p_d*.*

[42] $\cos(\overline{\mathbf{u}}^1, \mathbf{b}^*) = \cos(\overline{\mathbf{v}}^1, \mathbf{b}^0) = 1$ follows from the fact that $\overline{\mathbf{u}}^1 = \overline{\mathbf{v}}^1 = \mathbf{1}$ by the Perron-Frobenuius theorem. Since all the other singular vectors are orthogonal to $\overline{\mathbf{u}}^1$ and $\overline{\mathbf{v}}^1$, we also have $\cos(\overline{\mathbf{u}}^k, \mathbf{b}^*) = \cos(\overline{\mathbf{v}}^k, \mathbf{b}^0) = 0$ for all $k \ge 2$.

Clustered intervention. We now consider the case in which agents are homogeneous in terms of their target belief, $\mathbf{b}^* = b^*\mathbf{1}$. However, agents of one group has a relatively higher initial opinion than the agents of the other type. Specifically, without loss of generality, we assume that

$$\mathbf{b}_i = \begin{cases} b_1^0 & \text{for } i \leq n/2 \\ b_2^0 & \text{for } i > n/2, \end{cases}$$

for some $0 \leq b_2^0 < b_1^0 < b^* \leq 1$. This assumption captures the idea that, for example, consumers of one group indexed by $i \leq n/2$ are loyal consumers, so that they have more favorable preference to the designer's target. In a political setting, it represents the situation under which one group has closer bliss points to the influence designer; but the others do not.

For this case, the optimal intervention is not uniform. To see why, we first find that the singular values and the singular vectors are the same as in the previous case for the uniform intervention. However, the cosine similarities are different as the cosine similarity of the initial opinion vector is not orthogonal to the second singular vectors. This feature requires us to calculate the second term of the expression of the optimal intervention.

Interestingly, the amount of optimal intervention is clustered as

$$\overline{\mathbf{b}}'(n) - \mathbf{b}^0 = (b_1\mathbf{1}_{n/2}, b_2\mathbf{1}_{n/2})^\top,$$

for some $b_1, b_2 > 0$ with $b_1 < b_2$, where $\mathbf{1}_{n/2}$ is the vector of one with length $n/2$. Thus, the influence designer put more efforts to the group of agents have lower initial opinion. To see why, note that

$$\overline{\mathbf{u}}^2 = \overline{\mathbf{v}}^2 = \frac{1}{\sqrt{n}}(\mathbf{1}_{n/2}, -\mathbf{1}_{n/2})^\top, \quad \overline{\mathbf{u}}^2 \cdot \mathbf{b}^* = 0, \quad \text{and} \quad \overline{\mathbf{v}}^2 \cdot \mathbf{b}^0 = \frac{n}{2}(b_1^0 - b_2^0) > 0.$$

The signs of $\overline{\mathbf{u}}^2$ indicates spectral cluster.[43] Since there are two groups in the network in terms of network formation, agents indexed $i \leq n/2$ have the same sign, and the other agents share the different sign. Since $\mathbf{b}^0$ is not proportional to $\mathbf{1}$, it has a positve similarity to the second singular vectors. Accordingly, the second term in the characterization of $\overline{\mathbf{b}}'(n)$ remains as non zero.

Since the group of agents in the first group have favorable initial opinion, for the designer, investment to those agents is less cost-effective than investment to the agents in the other group. In calculation of $\overline{\mathbf{b}}'(n)$, the second term is strictly positive for the second group, and it is negative to the first group. Since the first term is strictly positive to all the agents, this gap can be interpreted

[43] There is a literature on the topic how to group nodes in a network according to some criterion (e.g., Chung 1996; Liu and Stewart 2011; Luxburg 2007).

as adjustment between different group as function of their initial opinion. Of course, this adjustment does not exceed the different of the initial opinions, and the resulting opinion is still greater for the first group. The following proposition summarizes.

Proposition 2 $\mathbf{b}^* = b^*\mathbf{1}$ *and* $\mathbf{b}^0 = (b_1^0\mathbf{1}_{n/2}, b_2^0\mathbf{1}_{n/2})^\top$. *Then, the optimal intervention is clustered: there are* $b_1, b_2 \in (0,1)$ *such that*

$$\overline{\mathbf{b}}'(n) = \mathbf{b}^0 + (b_1\mathbf{1}_{n/2}, b_2\mathbf{1}_{n/2})^\top.$$

6.2 Extension to Directed Networks

We now consider directed networks under complete information: agent i takes into account agent j's opinion in her opinion formation, but agent j does not (i.e., $\mathbf{A}_{ij} \neq \mathbf{A}_{ji}$. For directed networks, we still require that $d_i(\mathbf{A}) = \sum_{j=1}^n \mathbf{A}_{ij} > 0$: the indegree of each agent is greater than one: for all $i \in \{1, \ldots, n\}$. Hence, every agent is influenced by at least one other agent. As for undirected networks, this assumption ensures that $\mathbf{T}(\mathbf{A})$ is well-defined as the inverse of $\mathbf{D}(\mathbf{A})$ exists.[44] Consequently, the singular value decomposition of $\mathbf{T}(\mathbf{A})$ exists, and the same characterization of the optimal policy emerges:

Proposition 3 *Let* $\mathbf{T}(\mathbf{A}) = \mathbf{USV}^\top$ *be a singular decomposition of* $\mathbf{T}(\mathbf{A})$. *Then, the optimal intervention is*

$$\mathbf{b}' = \mathbf{b}^0 + \sum_{k=1}^n \frac{s_k}{s_k^2 + \mu}\left(||\mathbf{b}^*|| \cos(\mathbf{u}^k, \mathbf{b}^*) - s_k||\mathbf{b}^0|| \cos(\mathbf{v}^k, \mathbf{b}^0)\right)\mathbf{v}^k,$$

where μ *is a unique solution of the equation*

$$\sum_{k=1}^n \left(\frac{s_k}{s_k^2 + \mu}\right)^2 \left(\overline{\mathbf{b}}_k^* - s_k\underline{\mathbf{b}}_k^0\right)^2 = C.$$

We now provide a simple example providing a clear interpretation of the optimal intervention.

Example 3 Consider the sandglass network in Figure 4 that consists of five agents. Agent 1's is the central hub agent in that he influence agents 2 and 4 who also influence other agents other than agent 1. On the contrary, agent 3 is not the authority because agents 3 and 5 who influence agent 1 are not hubs. Instead, agents 2 and 4 are the authorities as the agent 1 who influences the two agents is the

[44] This assumption is weaker than the connectivity assumption in an undirected network or the weak connectivity assumption in a directed network: An undirected network is connected if there is a path between a pair of vertices, and a directed network is weakly connected if replacing all of its edges to undirected ones makes it connected undirected network. Our focus is not the convergence of opinions, which requires the weakly connectivity assumption.

hub. We calculate the adjacency matrix, and the corresponding singular matrices as the following:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

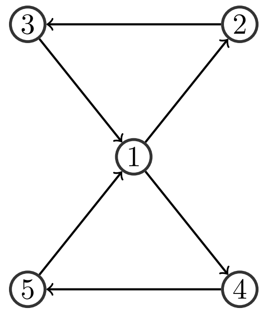


Figure 4: Illustration of the two simple networks

6.3 Extension to Multi-Type Network

We extend the characterization for the island network model to more general multi-type networks. There are two important deviations from the island model. First, we can consider unequal size of population. Second, we allows agents with different types to form links with other agents of different types with different probabilities. Importantly, in line with Section 6.2, the probability matrix $\mathbf{P}$ is not necessarily symmetric.

We assume that there are m types of the agents, and $(\tau_1, \dots, \tau_m)$ denotes types as a vector. We let $\mathbf{P}$ be an $m \times m$ matrix, whose entries in $(0, 1)$ describe the probabilities of links between types. The adjacency matrix is a realization of the random network in which entries $\mathbf{A}_{ij}$ with $i > j$ are independent Bernoulli random variables that take a value of 1 with probability $\mathbf{P}_{kl} \in (0, 1)$ when agent i is in group k and agent j is in group l. The entry represents that there is a link from agent j to agent i; that is, agent i is influenced by agent j. We still require that $\mathbf{A}_{ii} = 0$ for all i; however $\mathbf{A}$ is not necessarily symmetric because $\mathbf{P}$ is not symmetric. We let $|\tau_k|$ be the number of agents of type τ_k with $|\tau_1| + \cdots + |\tau_m| = n$. Therefore, a multi-type network of size n is denoted by a tuple, $(\mathbf{P}, (\tau_k)_{k=1}^m, (|\tau_k|)_{k=1}^m)$, which represents the link formation probability, types, and the size of types, respectively.

In order to have characterization of the optimal intervention, we assume the following:

Assumption 1 We assume the following for $(\mathbf{P}, (\tau_k)_{k=1}^m, (|\tau_k|)_{k=1}^m)$:

(i) $\mathbf{P}_{kl} \sim \mathcal{O}(\log n/n)$ for all $k, l \in \{1, \ldots, m\}$.

(ii) $|\tau_k| \sim \mathcal{O}(n)$.

The first assumption ensures that there are sufficiently many links between agents of different types when the network size n is large. The second assumption implies that the size of each type does not vanish as the network size n increases to infinity.

Since $\mathbf{P}$ is not symmetric, the influence matrix $\mathbf{T}(\overline{\mathbf{A}})$ with $\overline{\mathbf{A}} = \mathbb{E}[\mathbf{A}]$ is not symmetric. However, note that the singular decomposition $\mathbf{T}(\overline{\mathbf{A}})$ is not affected by this feature. In addition, since the Wedin $\sin\theta$ theorem does not reply on symmetric structure of the relevant matrices (Yu et al. 2015), it follows that the approximated intervention is asymptotically optimal. To be more precise, for large n, let $\overline{\mathbf{b}}(n)$ be the approximated intervention based on $\mathbf{T}(\overline{\mathbf{A}})$ as

$$\begin{aligned}\overline{\mathbf{b}}(n) = &\sum_{k=1}^{(m-1)^2} \frac{\overline{s}_k}{\overline{s}_k^2 + \overline{\mu}} \Big(||\mathbf{b}^*|| \cos(\overline{\mathbf{u}}^k, \mathbf{b}^*) - \overline{s}_k ||\mathbf{b}^0|| \cos(\overline{\mathbf{v}}^k, \mathbf{b}^0) \Big) \overline{\mathbf{v}}^k \\ &+ \sum_{l=(m-1)^2+1}^{n} \frac{\overline{s}_l}{\overline{s}_l^2 + \overline{\mu}} \Big(||\mathbf{b}^*|| \cos(\overline{\mathbf{u}}^l, \mathbf{b}^*) - \overline{s}_l ||\mathbf{b}^0|| \cos(\overline{\mathbf{v}}^l, \mathbf{b}^0) \Big) \overline{\mathbf{v}}^l.\end{aligned}$$

Note that the first summation includes $(m-1)^2$ terms because there are $(m-1)^2$ distinctive entries in $\mathbf{P}$. The second summation gathers reminder terms. As in a proof for Theorem 2, the singular vectors of $\mathbf{T}(\mathbf{A})$ converges to the singular vectors of $\mathbf{T}(\overline{\mathbf{A}})$ as long as the singular values are well-separated. Given this, the same convergence result emerges as follows.

Proposition 4 *The approximated intervention* $\overline{\mathbf{b}}'(n)$ *is asymptotically optimal.*

7 Conclusion

We examine the influence designer's intervention problem. The designer would like to lead agents with the initial opinion to have new opinions as close to the target opinion by injecting a new set of private opinion, subject to the budget constraint. We characterize the optimal intervention of the designer in terms of the hub and the authority centrality. We decompose the influence matrix into orthogonal singular vectors: The right singular vectors are associated with the authority centrality, and the left singular vectors are associated with the hub centrality. In Theorem 1, we characterize how these factors are considered in the optimal intervention: If the target opinion is embraced in terms of an individual's authority centrality, more opinion intervention should be injected for the individual. If the initial opinion is spread well in terms of an individual's hub centrality, a smaller amount of intervention should be injected for her.

We also examine the situation with the incomplete information about the underlying network structure. As for network formation process, we consider multi-type networks. We show that, when the influence designer has incomplete information, the optimal intervention in a large network is approximated and characterized by spectral clustering of the network (Theorem 2).

Two important factors should be taken into account in future research. The first factor is that firms dynamically intervene in consumers' opinions. For instance, at an early stage, the degree of intervention is severe (e.g., providing expensive gifts to specific consumers). The second factor is that firms may try to intervene in the network structure directly. For instance, a firm recommends a product reviewer's personal broadcasting channel (e.g., on YouTube or Instagram) to consumers, and once they subscribe to the broadcaster, they will be influenced by them from then on. Of course, a mix of these two factors is possible. Therefore, it is worth investigating these factors in a dynamic influence optimization model in future work.

A Proofs

Proof of Theorem 1

Proof. We first repeat the maximization problem as

$$\begin{aligned} \min_{\underline{\mathbf{b}}'} \quad & (\overline{\mathbf{b}}^*)^\mathsf{T}\overline{\mathbf{b}}^* - 2(\overline{\mathbf{b}}^*)^\mathsf{T}\mathbf{S}\underline{\mathbf{b}}' + (\underline{\mathbf{b}}')^\mathsf{T}\mathbf{S}^2\underline{\mathbf{b}}' \\ \text{subject to} \quad & \left(\underline{\mathbf{b}}' - \underline{\mathbf{b}}^0\right)^\mathsf{T}\left(\underline{\mathbf{b}}' - \underline{\mathbf{b}}^0\right) \le C. \end{aligned} \tag{DP 2}$$

As the above problem is convex, it suffices to solve the first order condition of the optimization: for all k, $(s_k\overline{\mathbf{b}}^*_k - s_k^2\underline{\mathbf{b}}'_k) = \mu(\underline{\mathbf{b}}'_k - \underline{\mathbf{b}}^0_k)$, where $\mu > 0$ is the Lagrangian multiplier. By rearrangements, we obtain $\underline{\mathbf{b}}'_k = \frac{s_k}{s_k^2+\mu}\overline{\mathbf{b}}^*_k + \frac{\mu}{s_k^2+\mu}\underline{\mathbf{b}}^0_k \ge 0$. Alternatively, in a matrix form, we obtain

$$\underline{\mathbf{b}}' = \underline{\mathbf{b}}^0 + (\mathbf{S}^2 + \mu\mathbf{I})^{-1}\mathbf{S}(\overline{\mathbf{b}}^* - \mathbf{S}\underline{\mathbf{b}}^0).$$

Now, we multiply $\mathbf{V}$ on both sides of the expression, and it follows that

$$\begin{aligned} \mathbf{b}' &= \mathbf{b}^0 + \mathbf{V}(\mathbf{S}^2 + \mu\mathbf{I})^{-1}\mathbf{S}\left(\mathbf{U}^\mathsf{T}\mathbf{b}^* - \mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{b}^0\right) \\ &= \mathbf{b}^0 + \left(\mathbf{v}^1|\cdots|\mathbf{v}^n\right)\begin{pmatrix} \frac{s_1}{s_1^2+\mu}\left((\mathbf{u}^1\cdot\mathbf{b}^*) - s_1(\mathbf{v}^1\cdot\mathbf{b}^0)\right) \\ \vdots \\ \frac{s_n}{s_n^2+\mu}\left((\mathbf{u}^n\cdot\mathbf{b}^*) - s_n(\mathbf{v}^n\cdot\mathbf{b}^0)\right) \end{pmatrix} \\ &= \mathbf{b}^0 + \sum_{k=1}^n \frac{s_k}{s_k^2+\mu}\left((\mathbf{u}^k\cdot\mathbf{b}^*) - s_k(\mathbf{v}^k\cdot\mathbf{b}^0)\right)\mathbf{v}^k \\ &= \mathbf{b}^0 + \sum_{k=1}^n \frac{s_k}{s_k^2+\mu}\left(||\mathbf{b}^*||\cos(\mathbf{u}^k,\mathbf{b}^*) - s_k||\mathbf{b}^0||\cos(\mathbf{v}^k,\mathbf{b}^0)\right)\mathbf{v}^k. \end{aligned}$$

Therefore, we obtain the following expression of $\mathbf{b}'$ in the theorem.

Note that the budget constraint is binding at $\mathbf{b}'$. Hence, μ solves the equation

$$\sum_{i=1}^n (\underline{\mathbf{b}}'_i - \underline{\mathbf{b}}^0_i)^2 = \sum_{k=1}^n \left(\frac{s_k}{s_k^2+\mu}\right)^2\left(\overline{\mathbf{b}}^*_k - s_k\underline{\mathbf{b}}^0_k\right)^2 = C.$$

There is a unique μ that satisfies the above equation because each term in the summation is strictly decreasing in μ. Therefore, the theorem is proven. ■

Proof of Theorem 2

Proof. To begin with, we first state the Wedin $\sin\theta$ theorem (Wedin 1972; Wedin 1983), which is an extension of the Davis-Kahan theorem for singular vectors (Davis and Kahan 1970).[45] Let $||\cdot||_\mathrm{F}$ be the

[45] See Yu et al. (2015) for general discussion of the theorems.

Frobenuous norm of matrices. For matrices $\mathbf{V}$ and $\widehat{\mathbf{V}}$, let $\Theta(\mathbf{V}, \widehat{\mathbf{V}})$ denote the $d \times d$ diagonal matrix whose jth diagonal entry is the jth singular angle, and let $\sin\Theta(\mathbf{V}, \widehat{\mathbf{V}})$ be defined entry-wise. Then, we have the following theorem:

Claim 1 (Wedin $\sin\theta$ theorem) *Let $\widehat{\mathbf{M}}$ be a perturbation of $\mathbf{M}$ as $\widehat{\mathbf{M}} = \mathbf{M} + \mathbf{H}$. Suppose that $s_r(\mathbf{M}) \geq a$ and $s_{r+1}(\widehat{\mathbf{M}}) \leq a - \triangle$ for some $\triangle > 0$. Then,*

$$\max\{\sin\Theta(\widehat{\mathbf{U}}_0, \mathbf{U}_0), \sin\Theta(\widehat{\mathbf{V}}_0, \mathbf{V}_0)\} \leq \frac{||\mathbf{H}||_{\mathrm{F}}}{\triangle},$$

where $\mathbf{U}_0$ and $\mathbf{V}_0$ represent the top-r singular subspaces of $\mathbf{M}$, and $\widehat{\mathbf{U}}_0$ and $\widehat{\mathbf{V}}_0$ represent the top-r singular subspaces of $\widehat{\mathbf{M}}$.

Note that the Frobenius norm in the theorem can be replaced by any other orthogonally invariant norm like the operator norm $||\cdot||_{\mathrm{op}}$. Also, the theorem provides that a similar inequality for $||\mathbf{v}_j - \widehat{\mathbf{v}}_j||$ holds, where $||\cdot||$ denotes the Euclidean norm (Yu et al. 2015). Hence, in the following proof, we apply the theorem for relevant norms in each step of the proof.

We prove the result for $\alpha = 0$ without loss of generality. Let $\mathbf{T}(\mathbf{A}) = \mathbf{D}(\mathbf{A})^{-1}\mathbf{A}$, $\mathbf{T}(\overline{\mathbf{A}}) = \mathbf{D}(\overline{\mathbf{A}})^{-1}\,\overline{\mathbf{A}}$, and $\mathbf{H} = \mathbf{T}(\mathbf{A}) - \mathbf{T}(\overline{\mathbf{A}})$. Note that $\mathbf{T}(\overline{\mathbf{A}})$ is row stochastic, and its rank is 2 because there are two model parameters, p_s and p_d. Note that $\mathbf{D}(\overline{\mathbf{A}}) = \frac{n(p_s+p_d)}{2}\mathbf{I}_n$, and so

$$\mathbf{T}(\overline{\mathbf{A}}) = \mathbf{D}(\overline{\mathbf{A}})^{-1}\,\overline{\mathbf{A}} = \frac{2}{n(p_s+p_d)}\begin{pmatrix} p_s\mathbf{E}_{n/2} & p_d\mathbf{E}_{n/2} \\ p_d\mathbf{E}_{n/2} & p_s\mathbf{E}_{n/2} \end{pmatrix}.$$

Since $\mathbf{H} = \mathbf{T}(\mathbf{A}) - \mathbf{T}(\overline{\mathbf{A}})$, we obtain

$$\mathbf{H}_{ij} = \frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})} - \frac{\overline{\mathbf{A}}_{ij}}{d_i(\overline{\mathbf{A}})},$$

where $d_i(\overline{\mathbf{A}}) = \frac{n(p_s+p_d)}{2}$. Thus, we have

$$||\mathbf{H}||_{\mathrm{op}} = \Big|\Big| \left[\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})} - \frac{\overline{\mathbf{A}}_{ij}}{d_i(\overline{\mathbf{A}})}\right] \Big|\Big|_{\mathrm{op}} \leq \Big|\Big| \left[\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})}\left(1 - \frac{d_i(\mathbf{A})}{d_i(\overline{\mathbf{A}})}\right)\right] \Big|\Big|_{\mathrm{op}} + \Big|\Big| \left[\frac{(\mathbf{A}_{ij} - \overline{\mathbf{A}}_{ij})}{d_i(\overline{\mathbf{A}})}\right] \Big|\Big|_{\mathrm{op}}. \tag{A.1}$$

We now bound each term on the right-hand side of the above inequality (A.1). First, we have

$$\Big|\Big| \left[\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})}\left(1 - \frac{d_i(\mathbf{A})}{d_i(\overline{\mathbf{A}})}\right)\right] \Big|\Big|_{\mathrm{op}} \leq \Big|\Big|\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})}\Big|\Big|_{\mathrm{op}} \times \Big|\Big| \left[\Big|\left(1 - \frac{d_i(\mathbf{A})}{d_i(\overline{\mathbf{A}})}\right)\Big|\right] \Big|\Big|_{\mathrm{op}}.$$

Note that because $p_s \geq p_d \geq K\frac{\log n}{n}$ for some K. By Theorem 3.6 in Chung et al. (2004), for any given $\delta > 0$, it follows that $|d_i(\mathbf{A}) - d_i(\overline{\mathbf{A}})| \leq \delta d_i(\overline{\mathbf{A}})$ for all i with high probability. Thus, we obtain

$$\Big|\Big| \left[\Big|\left(1 - \frac{d_i(\mathbf{A})}{d_i(\overline{\mathbf{A}})}\right)\Big|\right] \Big|\Big|_{\mathrm{op}} \leq \frac{\delta}{3},$$

which results in

$$\left|\left| \left[\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})}\left(1-\frac{d_i(\mathbf{A})}{d_i(\overline{\mathbf{A}})}\right)\right] \right|\right|_{\text{op}} \le \left|\left|\frac{\mathbf{A}_{ij}}{d_i(\mathbf{A})}\right|\right|_{\text{op}} \times \frac{\delta}{3} \le \frac{\delta}{3},$$

where the last inequality follows from the fact that $\mathbf{T}(\mathbf{A})$ is row-stochastic.

We now bound the second term on the right-hand side of (A.1) as

$$\left|\left| \left[\frac{(\mathbf{A}_{ij}-\overline{\mathbf{A}}_{ij})}{d_i(\overline{\mathbf{A}})}\right] \right|\right|_{\text{op}} = \frac{1}{d_i(\overline{\mathbf{A}})}\left|\left|\mathbf{A}_{ij}-\overline{\mathbf{A}}_{ij}\right|\right|_{\text{op}} = \frac{1}{\frac{n(p_s+p_d)}{2}}\left|\left|\mathbf{A}_{ij}-\mathbb{E}[\mathbf{A}_{ij}]\right|\right|_{\text{op}} \le \frac{1}{\frac{n(p_s+p_d)}{2}}\sqrt{np\log n},$$

which becomes less than or equal to $\frac{\delta}{3}$ for sufficiently large n.

We finally apply the Wedin $\sin\theta$ theorem. First, let $\triangle = s_2(\mathbf{T}(\overline{\mathbf{A}})) > 0$. Second, the singular values of $\mathbf{T}(\mathbf{A})$ converge to the singular values of $\mathbf{T}(\overline{\mathbf{A}})$ as shown in Golub and Jackson (2012). Third, for any given $\varepsilon > 0$, $||\mathbf{H}||_{\text{op}} < \triangle\varepsilon$ for sufficiently large n. In sum, the following holds with at least probability $1-\varepsilon$:

$$\max\{\sin\Theta(\widehat{\mathbf{U}}_0,\mathbf{U}_0),\sin\Theta(\widehat{\mathbf{V}}_0,\mathbf{V}_0)\} \le \frac{||\mathbf{H}||_{\text{op}}}{\triangle} < \varepsilon.$$

We finally prove that $||\mathbf{b}'(n)-\overline{\mathbf{b}}'(n)|| \to 0$ as $n\to\infty$ with high probability.

$$\begin{aligned}
&\frac{1}{\sqrt{n}}||\mathbf{b}'(n)-\overline{\mathbf{b}}'(n)|| \\
&\le \frac{1}{\sqrt{n}}\Big|\Big|\sum_{k=1}^{2}\frac{s_k}{s_k^2+\mu}\Big(||\mathbf{b}^*||\cos(\mathbf{u}^k,\mathbf{b}^*)-s_k||\mathbf{b}^0||\cos(\mathbf{v}^k,\mathbf{b}^0)\Big)\mathbf{v}^k \\
&\quad -\sum_{k=1}^{2}\frac{\overline{s}_k}{\overline{s}_k^2+\overline{\mu}}\Big(||\mathbf{b}^*||\cos(\overline{\mathbf{u}}^k,\mathbf{b}^*)-\overline{s}_k||\mathbf{b}^0||\cos(\overline{\mathbf{v}}^k,\mathbf{b}^0)\Big)\overline{\mathbf{v}}^k\Big|\Big| \\
&\quad +\frac{1}{\sqrt{n}}\underbrace{\Big|\Big|\sum_{k=3}^{n}\frac{s_k}{s_k^2+\mu}\Big(||\mathbf{b}^*||\cos(\mathbf{u}^k,\mathbf{b}^*)-s_k||\mathbf{b}^0||\cos(\mathbf{v}^k,\mathbf{b}^0)\Big)\mathbf{v}^k\Big|\Big|}_{\text{remainder}}
\end{aligned}$$

By the previous arguments with the Wedin $\sin\theta$ theorem, for a given $\varepsilon > 0$, it follows that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\Big|\Big|\sum_{k=1}^{2}\frac{s_k}{s_k^2+\mu}\Big(||\mathbf{b}^*||\cos(\mathbf{u}^k,\mathbf{b}^*)-s_k||\mathbf{b}^0||\cos(\mathbf{v}^k,\mathbf{b}^0)\Big)\mathbf{v}^k \\
&\quad -\sum_{k=1}^{2}\frac{\overline{s}_k}{\overline{s}_k^2+\overline{\mu}}\Big(||\mathbf{b}^*||\cos(\overline{\mathbf{u}}^k,\mathbf{b}^*)-\overline{s}_k||\mathbf{b}^0||\cos(\overline{\mathbf{v}}^k,\mathbf{b}^0)\Big)\overline{\mathbf{v}}^k\Big|\Big| \\
&< \frac{\varepsilon}{2}.
\end{aligned}$$

For the reminder, for sufficiently large n,

$$\frac{1}{\sqrt{n}}\Big|\Big|\sum_{k=3}^{n}\frac{s_k}{s_k^2+\mu}\Big(||\mathbf{b}^*||\cos(\mathbf{u}^k,\mathbf{b}^*)-s_k||\mathbf{b}^0||\cos(\mathbf{v}^k,\mathbf{b}^0)\Big)\mathbf{v}^k\Big|\Big|$$

$$
\begin{aligned}
&\leq \sum_{k=3}^{n} \underbrace{\frac{1}{s_k^2+\mu}}_{\geq \mu} s_k \underbrace{\left|\left|\left(\frac{||\mathbf{b}^*||}{\sqrt{n}} + \frac{||\mathbf{b}^0||}{\sqrt{n}}\right)\right|\right|}_{\leq 1} \times \underbrace{\left|\left|\mathbf{v}^k\right|\right|}_{\leq 1} \\
&\leq \frac{1}{\mu}\sum_{k=3}^{n} s_k \\
&< \frac{1}{\mu}\text{Trace}(\mathcal{L}(\mathbf{W}\mathbf{W}^\intercal)) \\
&= \frac{1}{\mu} n \frac{p^2+q^2}{n^2(p+q)^2} \\
&< \frac{\varepsilon}{2}.
\end{aligned}
$$

Therefore, the statement is proven. ■

Proof of Proposition 1

Proof. By the discussion in the main text, a proof of the proposition is straightforward. ■

Proof of Proposition 2

Proof. By the discussion in the main text, a proof of the proposition is straightforward. ■

Proof of Proposition 3

Proof. By the discussion in the main text, a proof of the proposition is straightforward. ■

Proof of Proposition 4

Proof. To show the result for general multi-type networks, there are two particular challenges. Since $\mathbf{T}(\mathbf{A})$ and $\mathbf{T}(\overline{\mathbf{A}})$ are not necessarily symmetric, we need to consider singular vectors, not eigenvectors. This challenge is solved by applying the Wedin $\sin\theta$ theorem instead of the Davis-Kahan $\sin\theta$ theorem. Second, there are $(m-1)^2$ parameters. Since the Wedin $\sin\theta$ theorem does not rely on the number of singular values, the same result applies. ■

References

Bagwell, K. (2007). “Chapter 28 The Economic Analysis of Advertising”. Ed. by M. Armstrong and R. Porter. Vol. 3. Handbook of Industrial Organization. Elsevier, pp. 1701–1844.

Bala, V. and S. Goyal (1998). “Learning from Neighbours”. *The Review of Economic Studies* 65.3, pp. 595–621.

Bramoullé, Y. et al. (2012). “Homophily and Long-Run Integration in Social Networks”. *Journal of Economic Theory* 5.147, pp. 1754–1786.

Bramoullé, Y., A. Galeotti, and B. Rogers, eds. (2016). *The Oxford Handbook of the Economics of Networks*. 1st Edition. Oxford University Press.

Cho, J., D. Kim, and K. Rohe (2017). “Asymptotic Theory for Estimating the Singular Vectors and Values of a Partially Observed Low Rank Matrix with Noise”. *Statistica Sinica* 27, pp. 1921–1948.

Choi, S., D. Gale, and S. Kariv (2005). “Behavioral aspects of learning in social networks: an experimental study”. *Experimental and Behavioral Economics*. Ed. by J. Morgan. Emerald Group Publishing. Chap. 2, pp. 25–61.

Chung, F., L. Lu, and V. Vu (2004). “The Spectra of Random Graphs with Given Expected Degrees”. *Internet Mathematics* 1.3, pp. 257–275.

Chung, F. R. K. (1996). *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society.

Cialdini, R. B. and N. J. Goldstein (2004). “Social influence: Compliance and conformity”. *Annu. Rev. Psychol.* 55, pp. 591–621.

Corazzini, L. et al. (2012). “Influential listeners: An experiment on persuasion bias in social networks.” *European Economic Review* 6, p. 1276.

Dasaratha, K. and K. He (2020). “Network structure and naive sequential learning”. *Theoretical Economics* 15.2, pp. 415–444.

Davis, C. and W. M. Kahan (1970). “The Rotation of Eigenvectors by a Perturbation. III”. *SIAM Journal on Numerical Analysis* 7.1, pp. 1–46.

DeGroot, M. H. (1974). “Reaching a Consensus”. *Journal of the American Statistical Association* 69.345, pp. 118–121. ISSN: 01621459.

Dixit, A. and V. Norman (1978). “Advertising and Welfare”. *Bell Journal of Economics* 9.1, pp. 1–17.

Echenique, F. and R. Fryer (2007). "A Measure of Segregation Based on Social Interactions". *Quarterly Journal of Economics* 122.2, pp. 441–485.

eMarketer (Apr. 2021). *US Digital Ad Spending 2021: Investments in Video and Performance-Oriented Ads Drive Pandemic Gains*. Ed. by eMarketer.

Fainmesser, I. P. and A. Galeotti (2016). "Pricing network effects". *Review of Economic Studies* 83, pp. 165–198.

Fan, J. et al. (2020). *Statistical Foundations of Data Science*. 1st Edition. Chapman & Hall/CRC Data Science Series. Chapman and Hall/CRC.

Gale, D. and S. Kariv (2003). "Bayesian learning in social networks". *Games and Economic Behavior* 45.2. Special Issue in Honor of Robert W. Rosenthal, pp. 329–346.

Galeotti, A., B. Golub, and S. Goyal (2020). "Targeting Interventions in Networks". *Econometrica*.

Golub, B. and M. O. Jackson (2010). "Naïve Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics* 1, p. 112.

Golub, B. and M. O. Jackson (2012). "How Homophily Affects the Speed of Learning and Best-Response Dynamics". *Quarterly Journal of Economics* 127.3, pp. 1287–1338.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. and B. W. Rogers (2007). "Meeting strangers and friends of friends: How random are social networks?" *American Economic Review* 97.3, pp. 890–915.

Kanuri, V. K., Y. Chen, and S. Sridhar (2018). "Scheduling Content on Social Media: Theory, Evidence, and Application". *Journal of Marketing* 82.6, pp. 89–108.

Kempe, D., J. Kleinberg, and É. Tardos (2003). "Maximizing the Spread of Influence through a Social Network". *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD'03. Association for Computing Machinery, pp. 137–146.

Kleinberg, J. (1999). "Authoritative Sources in a Hyperlinked Environment". *Journal of the ACM* 46.5, pp. 604–632.

Lambrecht, A., C. E. Tucker, and C. Wiertz (2018). "Advertising to early trend propagators: Evidence from Twitter". *Marketing Science* 37.2, pp. 177–199.

Lascu, D.-N. and G. Zinkhan (1999). "Consumer Conformity: Review and Applications for Marketing Theory and Practice". *Journal of Marketing Theory and Practice* 7.3, pp. 1–12.

Lee, D., K. Hosanagar, and H. S. Nair (2018). "Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook." *Management Science* 64.11, p. 5105.

Liu, N. and W. J. Stewart (2011). "Markov Chains and Spectral Clustering". Ed. by K. A. Hummel, H. Hlavacs, and W. Gansterer. Springer, pp. 87–98.

Luxburg, U. von (2007). "A tutorial on spectral clustering". *Statistics and Computing* 17, pp. 395–416.

Mallipeddi, R. R. et al. (2021). "A Framework for Analyzing Influencer Marketing in SocialNetworks: Selection and Scheduling of Influencers". *Management Science*.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). "Birds of a Feather: Homophily in Social Networks". *Annual Review of Sociology* 27, pp. 415–444.

Meyer, C. D. (2010). *Matrix analysis and applied linear algebra*. SIAM: Society for Industrial and Applied Mathematics.

Radner, R., A. Radunskaya, and A. Sundararajan (2014). "Dynamic pricing of network goods with boundedly rational consumers". *Proceedings of the National Academy of Sciences of the United States of America* 111.1, pp. 99–104.

Rohlfs, J. (1974). "A Theory of Interdependent Demand for a Telecommunications Service". *The Bell Journal of Economics and Management Science* 5.1, pp. 16–37.

Shin, E. (2017). "Monopoly pricing and diffusion of social network goods". *Games and Economic Behavior* 102, pp. 162–178.

Shin, E. (2021). "Social Network Formation and Strategic Interaction in Large Networks". *Mathematical Social Sciences*.

Sinclair, B. (2012). *The social citizen: Peer networks and political behavior*. University of Chicago Press.

Stewart, G. W. and J.-g. Sun (1990). *Matrix Perturbation Theory*. Academic Press.

Strang, G. (2019). *Linear Algebra and Learning from Data*. 1st Edition. Wellesley-Cambridge Press.

Wedin, P.-Å. (1972). "Perturbation bounds in connection with singular value decomposition". *BIT* 12, pp. 99–111.

Wedin, P.-Å. (1983). "Matrix Pencils". Ed. by B. Kågstrõm and A. Ruhe. Springer-Verlag. Chap. On angles between subspaces, pp. 263–285.

Yu, Y., T. Wang, and R. J. Samworth (2015). "A useful variant of the Davis–Kahan theorem for statisticians". *Biometrika* 102.2, pp. 315–323.