DISCORD AND HARMONY IN NETWORKS

ANDREA GALEOTTI, BENJAMIN GOLUB, SANJEEV GOYAL, AND RITHVIK RAO

Abstract. Consider a coordination game played on a network, where agents prefer taking actions closer to those of their neighbors and to their own ideal points in action space. We explore how the welfare outcomes of a coordination game depend on network structure and the distribution of ideal points throughout the network. To this end, we imagine a benevolent or adversarial planner who intervenes, at a cost, to change ideal points in order to maximize or minimize utilitarian welfare subject to a constraint. A complete characterization of optimal interventions is obtained by decomposing interventions into principal components of the network's adjacency matrix. Welfare is most sensitive to interventions proportional to the last principal component, which focus on *local* disagreement. A welfare-maximizing planner optimally works to reduce local disagreement, bringing the ideal points of neighbors closer together, whereas a malevolent adversary optimally drives neighbors' ideal points apart to decrease welfare. Such welfare-maximizing/minimizing interventions are very different from ones that would be done to change some traditional measures of discord, such as the cross-sectional variation of equilibrium actions. In fact, an adversary sowing disagreement to *maximize* her impact on welfare will *minimize* her impact on global variation in equilibrium actions, underscoring a tension between improving welfare and increasing global cohesion of equilibrium behavior.

Date Printed. February 26, 2021.
Joey Feffer and Zoë Hitzig provided exceptional research assistance. Andrea Galeotti gratefully acknowledges financial support from the European Research Council through the ERC-consolidator grant (award no. 724356) and the European University Institute through the Internal Research Grant. Benjamin Golub gratefully acknowledges financial support from The Pershing Square Fund for Research on the Foundations of Human Behavior and the National Science Foundation (SES-1658940, SES-1629446). Galeotti: Department of Economics, London Business School, agaleotti@london.edu. Golub: Departments of Economics and Computer Science, Northwestern University, benjamin.golub@northwestern.edu. Goyal: Faculty of Economics and Christ's College, University of Cambridge, sg472@cam.ac.uk. Rao: School of Engineering and Applied Sciences, Harvard University, rithvikrao@college.harvard.edu.

1. Introduction

Consider a simple coordination game played on a network. Each player takes an action, having an incentive to bring this action closer to both a personal *ideal point* and to the actions of neighbors. In the absence of any coordination concerns, each player would set their actions equal to their ideal points; we thus also call an ideal point a *favorite action*. Coordination concerns typically change this, pulling an agent's choices in equilibrium toward the ideal points of network neighbors, as well as of those farther away with whom the agent interacts only indirectly. A number of examples motivate our setup. The action may be declaring political opinions or values in a setting where it is costly to disagree with friends, but also costly to distort one's true position from the ideal point of sincere opinion. Alternatively, an action might be a choice in a technological space. For instance, in a software company, designer preferences inform tradeoffs between usability and power in the tools they use, but all are better off when their tools are more compatible with those of their colleagues.[1] In this example, the network is determined by collaboration relationships, i.e. which designers work together.[2]

The broad question we are concerned with is how the favorite actions and the network jointly determine welfare. Given a network, how do changes in agents' ideal points affect the efficiency of equilibrium outcomes? When can relatively small changes in these ideal points have large welfare impacts? We operationalize this question by imagining a planner who can, at a cost, change favorite actions. Supposing an adversary can undertake costly influence activities and change people's views, how would she do so if her goal was to increase miscoordination? Turning to the organization example, if managers can exert influence, provide encouragement, and offer incentives to change agents' inclinations, what changes would a benevolent manager undertake to maximize welfare? By understanding what such planners would do, we can understand how the relationship between favorite points and the network determines welfare. Such insights will also be relevant for problems concerning the composition of a team; rather than directly manipulating a particular person's incentives, a planner may instead choose *whom* to put in a certain organizational role or position. Such interventions require careful analysis of the welfare implications of the joint arrangement of ideal points and network links. Our results shed light on these issues.

[1]This interpretation of actions as choices in a technological space aligns with standard models in the literature on organizations—see, e.g., Calvó-Armengol, De Martí and Prat (2015).

[2]In these examples, and throughout, we take the network to be exogenous to the decisions in question, which is often realistic in the short run. Endogenous network formation is, as always, an important concern.

To analyze this intervention problem, we take a spectral approach. That is, we write the relevant optimization problems in terms of functions of eigenvalues and eigenvectors of the network, which are important invariants often used to capture various aspects of network structure. Working in a "principal component" basis permits legible characterizations of equilibrium outcomes and optimal interventions. (In contrast, in a natural basis the solutions to our optimization problems would be unwieldy and would not shed much light on the relationship between structural features of the network and the optimal intervention.) Our main findings are characterizations of the optimal intervention using certain eigenvectors and substantive implications for what such a planner focuses on.

Our main result, Theorem 1, is that the most welfare-consequential changes in favorite actions focus primarily (in a sense we make precise) on the *last* eigenvector of the network: the one associated with its lowest (typically most negative) eigenvalue. Beyond this, there is a monotonicity to the structure of interventions: principal components with lower eigenvalues receive less focus in optimal interventions. In special cases that we describe, the focus on the lowest principal component can be exclusive: at the optimal intervention, all disagreement in favorite actions is loaded onto this one principal component.[3] Our results also imply that explicit functions of certain eigenvalues can summarize the sensitivity of equilibrium welfare to optimal perturbations of ideal points. This gives an answer to the question posed at the beginning about how sensitive welfare is to the configuration of ideal points.

Going beyond a characterization in terms of a canonical graph statistic, we interpret the implications in terms of more intuitive aspects of graph structure. A key distinction we emphasize is between *local discord*—creating disagreement[4] at the "street level," between neighbors—and *global discord*—which creates disagreement between separate regions. Our result implies that optimal—i.e., welfare-maximizing or minimizing—interventions have a very local focus in a precise sense. An adversary seeks to amplify disagreement between neighbors, pushing neighbors' favorite points apart.

Notably, the interventions that best achieve this are quite distinct from those that best create global discord in the network. Indeed, creating global discord is in tension with reducing welfare. When an adversary optimally sows discord in ideal points to reduce welfare, this leads to a low level of variation across the population in equilibrium behavior, in a sense we make precise. Relatedly, if there is a certain amount of diversity (cross-sectional variation) in ideal points, it turns out that agents are best off when they are arranged so that

[3]We will use *principal components* and *eigenvectors* interchangeably.
[4]Throughout, we use "disagreement" to refer to differences in actions across the network.

they agree with their immediate neighbors and disagree with those distant from them in the network. This naturally leads to societies sustaining more diversity in equilibrium behavior and appearing more divided in a global sense. To summarize, our main results deliver stark predictions about which aspects of the configuration of ideal points matter for welfare, and these are quite different from what we might expect from standard intuitions about discord (as we elaborate on in our discussion of related literature below).

Finally, a conceptual point in our analysis is that intervention problems can be useful metaphors for understanding what structural features matter for welfare in a game. In some cases, a planner may indeed be intervening quite explicitly. For instance, an adversary may be seeking to use social media to sow division in opinions and cause costly tensions between neighbors.[5] But in many other problems, an analyst may simply be interested in understanding which shifts in exogenous primitives most affect welfare; hypothetical intervention problems shed light on this even when an intervention is not literally being designed.

1.1. **Related work.** Broadly, we are situated in the economics literature on network games, surveyed, for example by Jackson and Zenou (2014); see also the bibliography of Bramoullé, Kranton and D'Amours (2014). This literature, in terms of techniques and many of the measures that are relevant, is also related to the literature on opinion updating and social learning in networks, going back to DeGroot (1974) and surveyed by Acemoglu and Ozdaglar (2011) and Golub and Sadler (2016).

Within this broad literature, our project is distinguished by two aspects of our substantive focus. First, we are interested in a welfare objective. While most works in the economics literature on network games of course touch on efficiency and welfare considerations, the main outcome of interest is often an overall level of activity or knowledge—as, for instance, in Ballester, Calvó-Armengol and Zenou (2006) and Kempe, Kleinberg and Tardos (2015).[6] There are fewer that are focused on social welfare. An early contribution, with a price of anarchy approach, is Bindel, Kleinberg and Oren (2011), who give bounds on the welfare difference between equilibrium and a social optimum under the Friedkin and Johnsen (1999) social learning model. Another closely related contribution is due to Angeletos and Pavan

[5]See U.S. House of Representatives Permanent Select Committee on Intelligence (2018) for a report on such activities.

[6]Spectral methods play a significant role in the study of global influence, which is closely connected to the Perron vector (*eigenvector centrality*), as in Ballester, Calvó-Armengol and Zenou (2006); Acemoglu, Carvalho, Ozdaglar and Tahbaz-Salehi (2012). Different eigenvectors matter in our analysis because we are not concerned with first moments of behavior but rather variation and disagreement across agents.

(2007), who study fundamental structural properties of equilibrium welfare in beauty contests among other classes of games. In macroeconomics, the welfare implications of shocks are studied by Baqaee and Farhi (2019) and Baqaee and Farhi (2020). Galeotti, Golub and Goyal (2020) and King and Allouch (2019) are perhaps the closest in that they consider welfare-optimal interventions.[7] However, the class of games considered is very different: investment or public goods games. These involve quite different externalities from the ones that are relevant for coordination games and discord, which is what we focus on (as discussed by Angeletos and Pavan (2007)).[8]

Issues of miscoordination and discord are touched on in another thread of literature. This work analyzes how the configuration of agents' attributes (initial opinions, ideal points, etc.) affects the dynamics and ultimate outcomes of processes in social networks. The connection of these outcomes to spectral aspects of the network was noted by DeMarzo, Vayanos and Zwiebel (2003), and further developed by Golub and Jackson (2012), which highlighted the relation to spectral clustering.[9] An important recent contribution on discord is Gaitonde, Kleinberg and Tardos (2020), which studies maximizing and minimizing particular measures of discord in Friedkin and Johnsen (1999) updating processes (which, mathematically, are closely related to our games). Crucially, in all these projects, the notion of discord that is of interest is a particular, exogenously given measure, rather than welfare in the game. Criteria of interest include the duration of disagreement in an updating process, average disagreement across individuals in the network, etc. In our work, we get the objective from the preferences of the players themselves, maximizing utilitarian welfare. Thus, while the principal component approach overlaps methodologically with many of these studies, the welfare-oriented questions we ask lead to insights quite different from those in the prior literature. Indeed, a theme in the prior literature is that *global* discord between loosely connected regions is most important in slowing down agreement (DeMarzo, Vayanos and Zwiebel, 2003; Golub and Jackson, 2012). The component of disagreement that most strongly remains after a long period of updating opinions is proportional to the second eigenvector. As we will show, our results deliver a starkly different message. The classical spectral cut component—the second eigenvector that partitions the network into pieces that are relatively loosely connected to each other—is the

[7]Targeting of interventions more broadly is studied, e.g., in Albert, Jeong and Barabási (2000); Valente (2012); Kempe, Kleinberg and Tardos (2015).

[8]Bramoullé, Kranton and D'Amours (2014) focus on stability of equilibrium rather than targeting, but find that eigenvectors related to the ones we study matter in public goods games.

[9]For more on various segregation measures that come up in various related contexts, see Morris (2000); DeMarzo, Vayanos and Zwiebel (2003); Currarini, Jackson and Pin (2009); Golub and Jackson (2012); Spielman and Teng (2007).

least consequential for welfare in our setting. Gaitonde, Kleinberg and Tardos (2020) has more subtle results showing that there is no clear ordering of how an adversary focuses effort on various spectral components of disagreement. This is natural in view of their wider class of objectives. We show that for standard welfare-oriented objectives in coordination games, there is a clear ordering, with the last eigenvector being of primary importance. Finally, our Theorem 1 imposes less structure on the class of possible interventions than, e.g., Golub and Jackson (2012) or Gaitonde, Kleinberg and Tardos (2020); we allow perturbations around an arbitrary status quo and, for small interventions, can deal with a large class of intervention cost functions.

2. Model, Basic Facts, and Definitions

In this section, we state the model and definitions we need. We also mention some standard results on the structure of equilibrium that serve as a foundation for our subsequent results.

2.1. **Coordination game.** We consider a one-shot game played between individuals $\mathcal{N} = \{1, \ldots, n\}$, with a typical individual denoted i. Each individual takes an *action*[10] $a_i \in \mathbb{R}$. We are given a *favorite action* $f_i \in \mathbb{R}$ for each agent and a network with a weighted adjacency matrix $\boldsymbol{G} \in \mathbb{R}^{n \times n}$. An agent's payoff is determined by her favorite action and the actions of her neighbors in $\boldsymbol{G}$. We write the vector of actions as $\boldsymbol{a} \in \mathbb{R}^n$, and the vector of favorite actions as $\boldsymbol{f} \in \mathbb{R}^n$. Individual i chooses a_i, while $\boldsymbol{f}$ and $\boldsymbol{G}$ are exogenous.

We will assume that $\boldsymbol{G}$ is row-stochastic and symmetric, and that each i meets and interacts with j with probability g_{ij}. The payoff to an agent i of interacting with agent j is given by:

$$v_i(a_i, a_j) = -\underbrace{\beta(a_i - a_j)^2}_{\text{miscoordination}} - \underbrace{(1-\beta)(a_i - f_i)^2}_{\text{distance from favorite action}}, \tag{1}$$

where $\beta \in [0, 1)$ determines the relative payoff weight of miscoordination with other individuals and distance from an individual's favorite action. The expected payoff of individual i given action profile $\boldsymbol{a}$ is

$$V_i(\boldsymbol{a}) = \sum_j g_{ij} v_i(a_i, a_j).$$

Utilitarian welfare is defined by

$$V(\boldsymbol{a}) = \sum_i V_i(\boldsymbol{a}).$$

[10]The one-dimensional space is for simplicity: our analysis extends without much change to actions in an arbitrary Euclidean space.

2.2. **Nash equilibrium: A formula and a few basic properties.** Here we review a few standard facts about the Nash equilibrium.

Fixing $\boldsymbol{f}$ and $\boldsymbol{G}$, the first-order condition characterizing the Nash equilibrium action profile is given by

$$a_i^* = \beta \sum_j g_{ij} a_j^* + (1-\beta) f_i,$$

and this can be rewritten in vector notation to show that any Nash equilibrium action profile $\boldsymbol{a}^*$ must satisfy

$$\boldsymbol{a^*} = (1-\beta)(\boldsymbol{I} - \beta \boldsymbol{G})^{-1} \boldsymbol{f}. \tag{2}$$

We make the following two assumptions, the first of which has already been mentioned above.

Assumption 1. The adjacency matrix $\boldsymbol{G}$ is row-stochastic and symmetric.

Assumption 1 is implied by the description of $\boldsymbol{G}$ as meeting probabilities. It implies that the largest eigenvalue of $\boldsymbol{G}$ is 1 and ensures that (2) characterizes a unique, stable Nash equilibrium (Ballester, Calvó-Armengol and Zenou, 2006; Bramoullé, Kranton and D'Amours, 2014). Indeed, we have the following fact:

Fact 1. The game has a unique Nash equilibrium, which is in pure strategies and given by (2). In this equilibrium, each a_i is a (possibly different) weighted average of the f_j.

Proof. It is straightforward to check that the second-order conditions for optimization hold, so the first-order condition is necessary and sufficient. Assumption 1 ensures $\beta \boldsymbol{G}$ has spectral radius less than 1 and so we may rewrite (2) by the Neumann series as

$$\boldsymbol{a}^* = \underbrace{\left(\sum_{t=0}^{\infty} (1-\beta) \beta^t \boldsymbol{G}^t \right)}_{\boldsymbol{W}} \boldsymbol{f}. \tag{3}$$

Letting $\boldsymbol{W}$ be the matrix in parentheses, we see that it is a weighted average (with weights $(1-\beta)\beta^t$) of stochastic matrices $\boldsymbol{G}^t$, so $\boldsymbol{W}$ is itself stochastic. Thus, $a_i = \boldsymbol{W}_{i\bullet} \boldsymbol{f}$, where $\boldsymbol{W}_{i\bullet}$ is row i of $\boldsymbol{W}$. □

To illustrate the implications of Fact 1, consider Figure 1. There, we take a particular vector $\boldsymbol{f}$ where half of the agents (those in the bottom left) have a favorite action of $+1$, while those in the top right have a favorite action of -1. We calculate equilibrium using (2) for a particular value of β. We can then see the structure of equilibrium asserted in Fact 1:

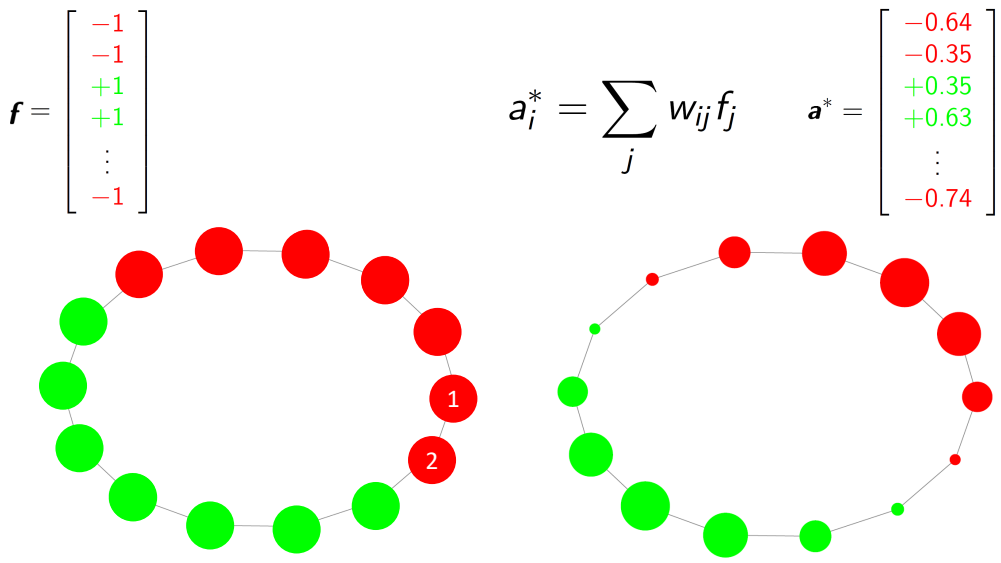

FIGURE 1. An illustration of equilibrium for a given network (the circle, where $g_{ij} = 0.5$ whenever i and j are adjacent). The node labels are as shown, continuing clockwise. On the left we depict a particular vector $\boldsymbol{f}$. When we depict a node-indexed vector (such as $\boldsymbol{f}$ or $\boldsymbol{a}$) visually, our convention is that entries with positive sign are indicated by making the node green, while entries with negative sign are indicated by making the node red. The size of each node corresponds to the magnitude of its entry. On the left side we write and illustrate $\boldsymbol{f}$, while on the right side we calculate $\boldsymbol{a}^*$ using Fact 1 and illustrate it in the same type of diagram.

everyone's action is a weighted average of $+1$'s and -1's, with closer agents weighted more and farther agents weighted less.[11]

While the favorite actions exhibit a very stark difference between groups, the equilibrium actions *attenuate* the diversity of favorite actions. "Boundary" agents average together roughly as many $+1$'s as -1's, and end up with equilibrium actions close to 0. They are quite far from their favorite actions, though they are fairly closely coordinated with their neighbors. Agents deep in the bottom left or top right end up with actions that are much more extreme, and therefore closer to their ideal points. Even they, however, end up with actions less extreme than the extremes of $\boldsymbol{f}$, illustrating the attenuation property of best responses.[12]

[11]Note also from the form of (3) that the Nash equilibrium can be seen as the average of $\boldsymbol{G}^t\boldsymbol{f}$ for $t \in \{0, 1, 2, \ldots\}$, which are the outcomes of DeGroot (1974) or Friedkin and Johnsen (1999) learning or myopic updating at various times t; see Golub and Jackson (2012). This explains the close connection between properties of equilibria in network games and the dynamics of certain updating/learning processes in networks; see also Gaitonde, Kleinberg and Tardos (2020).

[12]This can be seen by noting from (3) that in a connected graph, each agent puts positive weight on all others, and thus even the most extreme agents become less extreme. The higher β is, the stronger the attenuation.

We will make a final, technical, assumption to simplify the statement of some results. This holds generically (over the choice of weights in the symmetric matrix).

Assumption 2. All eigenvalues of $\boldsymbol{G}$ are distinct.

2.3. **Planner interventions and objective.** Our main interest is in understanding how, in examples such as the one just discussed, welfare is affected by changes in the favorite actions of various players. We investigate this question by considering a planner who can modify the vector of favorite actions: the favorite actions $\hat{\boldsymbol{f}}$ are modified by some perturbation vector $\boldsymbol{\delta} \in \mathbb{R}^n$. Formally, the planner's problem is given by

$$\begin{aligned} \max_{\boldsymbol{\delta}} \quad & \gamma V(\boldsymbol{a}^*) \\ \text{s.t.} \quad & \boldsymbol{f} = \hat{\boldsymbol{f}} + \boldsymbol{\delta} \\ & \boldsymbol{a^*} = (1-\beta)(\boldsymbol{I} - \beta \boldsymbol{G})^{-1}\boldsymbol{f}, \\ & c(\boldsymbol{\delta}) \leq C. \end{aligned} \tag{4}$$

The parameter γ scaling the objective is $+1$ or -1, corresponding to the planner being benevolent or malevolent, respectively. The constraint $c(\boldsymbol{\delta}) \leq C$ limits the feasible interventions. The cost function $c(\cdot)$ is for now taken to be arbitrary. For various results, we will give specific cost functions: for example, constraining interventions to a ball of fixed size around the status quo. In our most general results in Section 4.2, we study classes of cost functions satisfying certain assumptions, such as that interventions have (at least locally) convex costs. The number $C \geq 0$ is called the *budget*.

2.4. **Principal components: Definitions and notation.** We introduce notation for the key objects that play a role in our approach: the principal components of the network $\boldsymbol{G}$. We write the spectral decomposition of $\boldsymbol{G}$ as follows:

$$\boldsymbol{G} = \underbrace{\begin{bmatrix} | & & | \\ \boldsymbol{u}^1 & \dots & \boldsymbol{u}^n \\ | & & | \end{bmatrix}}_{\boldsymbol{U}\text{: eigenvectors}} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{\boldsymbol{\Lambda}\text{: eigenvalues}} \underbrace{\begin{bmatrix} — & (\boldsymbol{u}^1)^\intercal & — \\ & \vdots & \\ — & (\boldsymbol{u}^n)^\intercal & — \end{bmatrix}}_{\boldsymbol{U}^\intercal\text{: eigenvectors}}. \tag{5}$$

Here, $\boldsymbol{U}$ gives an orthonormal basis of eigenvectors. We adopt the convention that the eigenvectors and eigenvalues are arranged so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We will refer to the eigenvector corresponding to λ_ℓ as $\boldsymbol{u}^\ell$. For any vector $\boldsymbol{z} \in \mathbb{R}^n$, let $\underline{\boldsymbol{z}} = \boldsymbol{U}^\intercal \boldsymbol{z}$. We will refer to

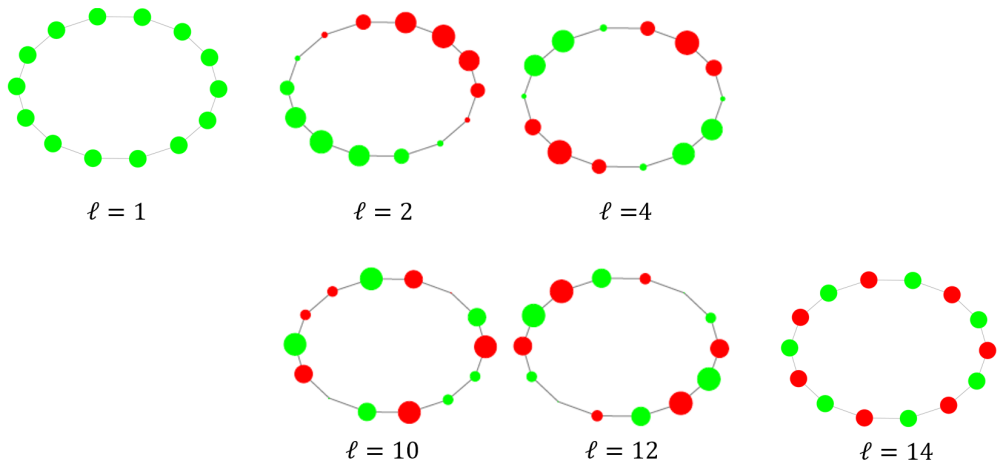

FIGURE 2. Six eigenvectors of a circle network. The eigenvector $\boldsymbol{u}^\ell$ corresponding to the ℓ^{th}-highest eigenvalue λ_ℓ, is depicted using the same visual convention we introduced in Figure 1. Note that the eigenvectors higher eigenvalues (higher ℓ) vary "more slowly" over the circle than those with lower eigenvalues (higher ℓ).

$\underline{z}_\ell$ as the projection of $\boldsymbol{z}$ onto the ℓ^{th} principal component, or the magnitude of $\boldsymbol{z}$ in that component.

Figure 2 illustrates some principal components of an example network.

Throughout, we use $\langle \boldsymbol{y}, \boldsymbol{z} \rangle = \sum_{i \in \mathcal{N}} y_i z_i$ to denote the Euclidean dot product, and we let $\|\boldsymbol{z}\| = \langle \boldsymbol{z}, \boldsymbol{z} \rangle^{1/2}$ denote the Euclidean norm. Since the eigenvectors are normalized, they satisfy $\|\boldsymbol{z}\| = 1$.

3. Two Simple Planner Problems and Two Distinguished Principal Components

Certain principal components will play an important role in our analysis. For instance, the *last* principal component, the eigenvector $\boldsymbol{u}^n$ corresponding to the lowest eigenvalue λ_n will correspond to the direction in which interventions are most consequential for welfare. It will also be helpful to contrast it with another eigenvector, $\boldsymbol{u}^2$, the one that corresponds to the second-highest eigenvalue λ_2. This eigenvector, which has been important in prior studies of segregation and homophily (DeMarzo, Vayanos and Zwiebel, 2003; Golub and Jackson, 2012), turns out to describe *least* welfare-consequential interventions, and so it will serve as an important foil or contrast for some of our results.

To show the role these eigenvectors play in optimization problems, we define a special case of the planner's problem, in which the planner chooses any $\boldsymbol{f}$ on a sphere of radius 1

to maximize or minimize welfare. This corresponds to holding the cross-sectional variation of favorite actions fixed, and distributing a "fixed" amount of disagreement to achieve the objective. In this section, we dispense with $\boldsymbol{\delta}$ and work with choosing the vector $\boldsymbol{f}$ directly, since the simplicity of the problem makes this change straightforward. Thus, we can simply consider how the planner decides to allocate disagreement in her choice of $\boldsymbol{f}$, subject to a norm constraint.

The optimization problem of interest is defined by

$$\begin{aligned} \max_{\boldsymbol{f}} \quad & \gamma V(\boldsymbol{a}^*) \\ \text{s.t.} \quad & \boldsymbol{a^*} = (1-\beta)(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f} \\ & \|\boldsymbol{f}\| = 1. \end{aligned} \tag{6}$$

Proposition 1. Fixing β, there is an increasing function $\zeta : \mathbb{R} \to \mathbb{R}$ such that:

(1) The optimum of (6) for the malevolent planner ($\gamma = -1$) is achieved by $\boldsymbol{f}^* = \boldsymbol{u}^n$ and is equal to $\zeta(\lambda_n)$.
(2) The optimum of (6) for the benevolent planner ($\gamma = 1$) is achieved by $\boldsymbol{f}^* = \boldsymbol{u}^2$ and is equal to $\zeta(\lambda_2)$.

Proof. We begin by writing the formula for equilibrium welfare in terms of an inner product expression depending on $\boldsymbol{f}$ and $\boldsymbol{G}$.

$$\begin{aligned} V^* &= -\sum_i \left((1-\beta)(a_i^* - f_i)^2 + \sum_j g_{ij}\beta(a_i^* - a_j^*)^2 \right) \\ &= -\langle \boldsymbol{a}^*, ((1+\beta)\boldsymbol{I} - 2\beta\boldsymbol{G})\boldsymbol{a}^* \rangle + (1-\beta)\langle \boldsymbol{f} - 2\boldsymbol{a}^*, \boldsymbol{f} \rangle \\ &= -(1-\beta)\left[\langle \boldsymbol{f}, \boldsymbol{f} \rangle + (1-\beta)\langle (\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}, ((1+\beta)\boldsymbol{I} - 2\beta\boldsymbol{G})(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f} - 2\boldsymbol{f} \rangle \right] \end{aligned}$$

We now switch into the basis of principal components. Recall $\underline{\boldsymbol{z}} = \boldsymbol{U}^\intercal \boldsymbol{z}$. Then

$$\boldsymbol{a} = (1-\beta)(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}$$

if and only if

$$\underline{\boldsymbol{a}} = (1-\beta)(\boldsymbol{I}-\beta\boldsymbol{\Lambda})^{-1}\underline{\boldsymbol{f}}.$$

Moreover, we may replace all vectors and matrices in the above expression for $-W^*$ by their versions in the new basis. All matrices involved are diagonal, so this greatly simplifies the expression; indeed, as shown in Lemma A.1 in the appendix, this yields the following

expression

$$V^* = \sum_{\ell=1}^{n} \zeta(\lambda_\ell) \underline{f}_\ell^2,$$

for some increasing, nonnegative function $\zeta(\lambda)$, with $\zeta(1) = 0$ (so that the λ_1 term drops out, since $\lambda_1 = 1$). Note also that because the change of basis is orthonormal, the constraint set for $\boldsymbol{f}$ does not change.

Because ζ is increasing in λ, the optimum for $\gamma = -1$ is achieved by $\boldsymbol{f}^* = \boldsymbol{u}^n$ and is equal to $\zeta(\lambda_n)$. The optimum for $\gamma = 1$ is achieved by $\boldsymbol{f}^* = \boldsymbol{u}^2$ and is equal to $\zeta(\lambda_2)$. □

Proposition 1 shows that when $\boldsymbol{f}$ is constrained to a sphere, extremal welfare in the minimization problem depends on $\boldsymbol{G}$ only through an extreme eigenvalue, λ_n or λ_2. Indeed, it remains true if we replace the constraint by $\|\boldsymbol{f}\| \leq C$, for $C > 0$, as long as we make the adjustment that $\zeta(\cdot)$ is replaced by $C\zeta(\cdot)$. Thus, $\zeta(\lambda_n)$ captures the sensitivity of welfare to the size of the invervention when the intervention is chosen optimally.[13]

In terms of the form of intervention, loading all the diversity in favorite actions onto the last principal component is the most effective way of reducing welfare subject to an upper bound on the norm of the favorite actions. This is the first manifestation of the idea that the last principal component is the one to which welfare is most sensitive.

In contrast, the second part of the result highlights that welfare is, in a sense, *least* sensitive to disagreement along the second principal component. For fixed norm of $\boldsymbol{f}$, if we load all disagreement onto $\boldsymbol{u}^2$, welfare turns out to be the *least* negative—least changed from a baseline of 0 when there is no disagreement.

Finally, it is worth remarking on the fact that $\boldsymbol{u}^1$ plays no role in the characterization. Note that in this problem $\boldsymbol{u}^1$ is a constant vector, because $\boldsymbol{G}$ is row-stochastic.[14] Thus, changes in $\underline{f}_1$ correspond to constant shifts in favorite actions, which, by Fact 1 translate into the same constant shifts in equilibrium actions. These shifts do not affect welfare, and so are never used by the planner.

[13]We reproduce the function ζ here for convenience, from Lemma A.1:

$$\zeta(\lambda) = -\beta(1-\beta)\frac{(1-\lambda)[2-\beta(1+\lambda)]}{(1-\beta\lambda)^2}.$$

[14]In general, this vector gives the agents' *eigenvector centralities* in the network, which measure the global influence of each agent. Because of the symmetry of interactions and the fact that each agent has the same endowment of total interaction probability, there is no heterogeneity in this, but our analysis can be extended to settings where there is heterogeneity in interaction quantity.

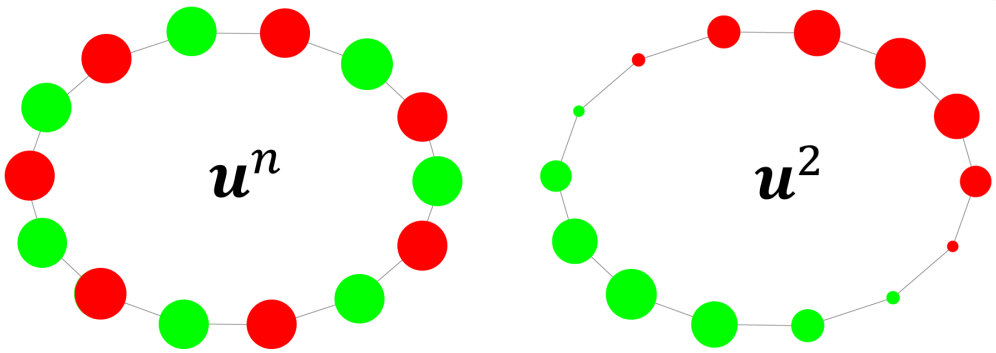

FIGURE 3. The eigenvectors of a circle network corresponding to the n^{th} largest and 2^{nd} largest eigenvalues, respectively. The former maximizes local heterogeneity and separates neighbors, while the latter finds a global cut.

3.1. **Local and global disagreement at the optima.** Next, we are interested in describing how the eigenvectors identified in Proposition 1 relate to the network, and what qualitative comparisons we can make between equilibrium behavior at the two configurations analyzed. We will show that, in a suitable sense, the last eigenvector $\boldsymbol{u}^n$ is the one that maximizes local disagreement, while the second eigenvector $\boldsymbol{u}^2$ maximizes global disagreement, subject to a constraint on norm.

We make a few definitions. Let $\mathcal{D}_{\text{R}}$ be the uniform distribution on the set $\{(i,j) \in \mathcal{N} \times \mathcal{N} \text{ s.t. } i \neq j\}$. This corresponds to drawing a random pair. Let $\mathcal{D}_{\boldsymbol{G}}$ be the distribution on the same set obtained by drawing the pair (i,j) with probability g_{ij}/n.

Definition 1. Fix a vector $\boldsymbol{z} \in \mathbb{R}^n$ such that $\sum_{i \in \mathcal{N}} z_i = 0$.

(1) The *covariance of a random pair* for $\boldsymbol{z}$ is defined to be $\mathbb{E}_{(i,j)\sim\mathcal{D}_{\text{R}}}[z_i z_j]$
(2) The *covariance of neighbors* for $\boldsymbol{z}$ is defined to be $\mathbb{E}_{(i,j)\sim\mathcal{D}_{\boldsymbol{G}}}[z_i z_j]$.

Now we use the covariance of the actions of a pair of neighbors selected at random[15] as a measure of local disagreement, and the covariance of the actions of a random pair of agents as a measure of global disagreement. In each case, the more negative the number, the more disagreement there is of the relevant kind.

Proposition 2. Let $\mathcal{F}$ be the set of vectors $\boldsymbol{f}$ satisfying $\sum_{i \in \mathcal{N}} f_i = 0$ and $\|\boldsymbol{f}\| = 1$. The values of $\boldsymbol{f}$ in this set that maximize and minimize each quantity below are given by the following table:

[15]According to the same distribution that selects partners to play the bilateral game in our model.

Statistic for eq'm actions $\boldsymbol{a}^*(\boldsymbol{f})$	maximizer	minimizer
covariance of neighbors	$\boldsymbol{u}^2$	$\boldsymbol{u}^n$
covariance of random pair	$\boldsymbol{u}^n$	$\boldsymbol{u}^2$

Proof. We show each covariance result separately.

The covariance of neighbors for equilibrium actions $\boldsymbol{a}^*(\boldsymbol{f})$ is given by

$$\frac{1}{n}\left(\sum_{i,j\in\mathcal{N}} g_{ij}a_i^*a_j^*\right) = \frac{1}{n}\langle \boldsymbol{a}^*, \boldsymbol{G}\boldsymbol{a}^*\rangle,$$

because we sample an agent i uniformly at random from $\mathcal{N}$, and a second agent j incident to i (the probability of an agent k being sampled is g_{ik}). As with Proposition 1, we can rewrite this expression in the principal component basis as

$$\sum_{\ell=1}^{n} \eta(\lambda_\ell)\underline{f}_\ell^2,$$

for an increasing function $\eta(\lambda)$. (This is the content of Lemma A.1 in the appendix.) Because each summand is increasing in λ_ℓ, this expression achieves its minimum at $\boldsymbol{f}^* = \boldsymbol{u}^n$ and its maximum at $\boldsymbol{f}^* = \boldsymbol{u}^2$.

The covariance of a random pair for equilibrium actions $\boldsymbol{a}^*(\boldsymbol{f})$ is given by

$$\frac{1}{n^2}\left(\sum_{i,j\in\mathcal{N}} a_i^*a_j^* - \sum_{i\in\mathcal{N}}(a_i^*)^2\right),$$

because we sample an agent i uniformly at random from $\mathcal{N}$, and we sample a second agent uniformly at random from $\mathcal{N}\setminus\{i\}$. Because $\boldsymbol{G}$ is row-stochastic, its Perron vector is the all-ones vector, with eigenvalue 1. Thus the projection operator onto the eigenspace associated with eigenvalue $\lambda_1 = 1$ is $\boldsymbol{P}_{(1)} = \mathbf{1}\mathbf{1}^\intercal$. We can then rewrite the above expression as

$$\langle \boldsymbol{a}^*, \boldsymbol{P}_{(1)}\boldsymbol{a}^*\rangle - \langle \boldsymbol{a}^*, \boldsymbol{a}^*\rangle = \langle \boldsymbol{a}^*, (\boldsymbol{P}_{(1)} - \boldsymbol{I})\boldsymbol{a}^*\rangle.$$

The average equilibrium action is a constant times the average of $\boldsymbol{f}$. Thus, $\boldsymbol{P}_{(1)}\boldsymbol{a}^* = \mathbf{0}$. It follows that the covariance-minimizing $\boldsymbol{a}$ maximizes $\langle \boldsymbol{a}, \boldsymbol{a}\rangle$. This expression can be written in the principal component basis as

$$\sum_{\ell=1}^{n} \nu(\lambda_\ell)\underline{f}_\ell^2,$$

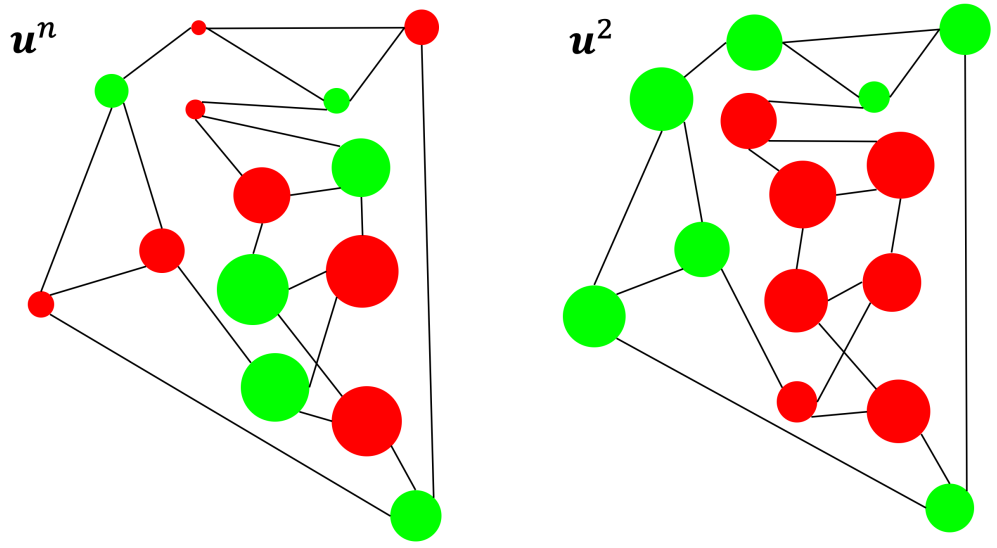

FIGURE 4. The 2nd and n^{th} eigenvectors of a more complex network. In this setting, $\boldsymbol{u}^n$ and $\boldsymbol{u}^2$ still respectively recover local and global structure. In particular, $\boldsymbol{u}^n$ separates neighbors by creating heterogeneity between actions (both in sign and in magnitude), while $\boldsymbol{u}^2$ globally creates two nearly-homogenous groups.

for a decreasing function $\nu(\lambda)$. (See Lemma A.1 for the explicit function.) Because each summand is decreasing in λ_ℓ, this expression achieves its minimum by $\boldsymbol{f}^* = \boldsymbol{u}^n$ and its maximum by $\boldsymbol{f}^* = \boldsymbol{u}^2$.

Note that $\boldsymbol{u}^1$ does not optimize either function for the same reason as mentioned previously: $\boldsymbol{u}^1$ is a constant vector for a row-stochastic $\boldsymbol{G}$, and the attenuation process keeps it constant, so any such intervention has no effect on welfare. □

Proposition 2 can be seen as a *local-global disagreement tradeoff*: the $\boldsymbol{f}$ that maximizes local disagreement in equilibrium also minimizes global disagreement.[16]

3.2. **Interpretation and discussion.** Imagine that an adversary manipulates individual ideal points in a community to reduce its members' welfare. Naively, one might expect that the consequences of this adversary's activity would be to cause global discord: to make it likely that two randomly chosen individuals would disagree strongly. Our results show that, in fact, for a given amount of cross-sectional variation in favorite points, the adversary in a sense accomplishes the opposite. We now explore this somewhat counterintuitive phenomenon.

[16]If β is not too small, we can obtain equivalent results defining disagreement as the expectation of $(a_i - a_j)^2$ under the appropriate distribution of i, j (random pair or neighbor).

First, let us formalize what we said in the previous paragraph. Proposition 1 says that the malevolent planner chooses $\boldsymbol{f} = \boldsymbol{u}^n$. Proposition 2 says that this choice causes the *least* global disagreement: it creates the highest possible covariance of equilibrium actions between random pairs of individuals.

To understand the forces behind these results, we again consider the example network in Figure 3, showing $\boldsymbol{u}^2$ and $\boldsymbol{u}^n$ for a circle network. As a warm-up exercise, let us discuss a limit case. Suppose that β is positive but quite small, so that, by (3) in Fact 1, $\boldsymbol{a}^* \approx \boldsymbol{f}$. Then studying statistics of the favorite points $\boldsymbol{f}$ is the same as studying the corresponding statistics of $\boldsymbol{a}^*$. Let $\mathcal{F}$ be the set of vectors $\boldsymbol{f}$ satisfying $\sum_{i \in \mathcal{N}} = 0$ and $\|\boldsymbol{f}\| = 1$. We will now note that the extreme eigenvector $\boldsymbol{u}^n$ achieves extreme levels of both covariance between neighbors and disagreement disutility. For intuition, consider Figure 3. It is clear that under $\boldsymbol{f} = \boldsymbol{u}^n$, each agent has a favorite point that is the opposite of those of its neighbors. It is then intuitive that neighbor covariance is as negative as possible: each person disagrees with a random neighbor for sure. Because the costs of disagreement are convex, it is also intuitive that this configuration creates maximum disutility from miscoordination (relative to one where neighbors were closer to each other, as in $\boldsymbol{u}^2$). Indeed, by making $\boldsymbol{f}$ "vary gradually" (changing as little as possible between connected nodes), as in $\boldsymbol{u}^2$, we achieve the opposite effect and minimize both disutility and covariance.

These effects are intuitive. However, they do not exhaust the story: to understand how much disutility players experience, we must understand their actions in equilibrium. And, as we have already remarked in presenting Fact 1 and in Figure 1, for β not too close to 0, these involve substantial attenuation relative to favorite actions. We now turn to explaining this aspect of the result.

Using (2) and rewriting the condition in the principal component basis, we have

$$\underline{a}^*_\ell = \frac{1-\beta}{1-\beta\lambda_\ell} f_\ell.$$

When only one principal component is represented in the favorite actions, as when $\boldsymbol{f} = \boldsymbol{u}^n$ or $\boldsymbol{f} = \boldsymbol{u}^2$, the same is true for equilibrium actions. In other words, in these cases $\boldsymbol{a}^*$ is a *scaling* of $\boldsymbol{f}$. But the scaling is nontrivial: in best-responding to each other, the disagreement in favorite points is attenuated to a smaller disagreement in equilibrium actions. Indeed, because players best-respond to their neighbors, under $\boldsymbol{f} = \boldsymbol{u}^n$ they have a strong reason to bring actions closer to zero, in order to coordinate with neighbors.

Thus, the result involves both forces described above having to do with the structure of $\boldsymbol{f}$ alone (which are present even in the $\beta \approx 0$ case) as well as the equilibrium attenuation of $\boldsymbol{a}$ (which has a substantial effect only when β is far from zero); these forces may pull in opposite directions.

Our result shows that attenuation is *not* enough to overcome the harm done by the strong local disagreement induced by $\boldsymbol{u}^n$. One reason for this is that even when players benefit from attenuation by miscoordinating less with neighbors, under $\boldsymbol{f} = \boldsymbol{u}^n$ they also suffer by being farther from ideal points. It turns out that the planner maximizes their pain by making near neighbors disagree strongly. This pattern, presented in an exteremely simple way for the circle, generalizes to more complex networks as shown in Figure 4.

On the other hand, a planner who is concerned with creating global disagreement (i.e., minimizing the covariance of a random pair for equilibrium actions) is not at all concerned with making neighbors disagree. For this planner, minimizing attenuation turns out to be the dominant consideration: the planner wants to make sure that as much of the "size" of initial disagreement remains in the final equilibrium actions. It is intuitive that this is accomplished by making neighbors *agree* as often as possible. Then strategic forces will not lead them to moderate their behavior by much relative to $\boldsymbol{f}$. Of course, the requirement (imposed by definition of $\mathcal{F}$) that $\boldsymbol{f}$ have a positive norm, along with the normalization that the average of $\boldsymbol{f}$ is equal to zero, requires heterogeneity across society in favorite points. The best way for a planner to place this heterogeneity is to put the polarization along a "cut" such as that depicted in the vectors $\boldsymbol{u}^2$ of Figure 3. Here disagreement is designed to be as small as possible across most links, and at the optimum, $\boldsymbol{f}$ (and, consequently $\boldsymbol{a}^*$) will be quite similar for most nodes at short distances. As we have already noted, the configuration $\boldsymbol{u}^2$ finds cohesive areas in the network and keeps their $\boldsymbol{f}$ similar, while making relatively "faraway" regions disagree with each other. Especially in networks that have good cuts, with large groups that interact fairly little, this is natural: if the global disagreement in $\boldsymbol{f}$ is experienced across few links, then it makes little difference to welfare. The vector $\boldsymbol{u}^2$ can be seen in a network more interesting than the circle in Figure 4.

We have spoken informally of $\boldsymbol{u}^n$ tending to make neighbors take opposite signs, whereas $\boldsymbol{u}^2$ divides the network into cohesive regions. These notions have been extensively formalized in the graph theory literature: see Desai and Rao (1994), Alon and Kahale (1997), and Urschel (2018) for some examples.

4. Generalizations: General Initial Conditions and Cost Functions

We return to the general case of the planner's problem stated in (4):

$$\begin{aligned} \max_{\boldsymbol{\delta}} \quad & \gamma V(\boldsymbol{a}^*) \\ \text{s.t.} \quad & \boldsymbol{f} = \hat{\boldsymbol{f}} + \boldsymbol{\delta} \\ & \boldsymbol{a^*} = (1-\beta)(\boldsymbol{I} - \beta \boldsymbol{G})^{-1}\boldsymbol{f}, \\ & c(\boldsymbol{\delta}) \leq C. \end{aligned}$$

The previous section showed that for very simple planner's constraints, there is a simple description of the most welfare-consequential interventions. However, we worked under many simplifying assumptions: $\hat{\boldsymbol{f}}$ was taken to be $\mathbf{0}$, and the constraint on interventions was to choose one in a ball or on a sphere.

It is worthwhile to relax both restrictions: we want to consider a status quo that is more flexible. We want to understand to what extent the intuitions extend to more general cost functions. In this section, we address these issues.

To state results, we need to make a definition measuring the similarity of various vectors to principal components of the underlying network. For this, we use the notion of *cosine similarity*.

Definition 2 (Cosine Similarity)**.** The *cosine similarity* of two nonzero vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ is

$$\rho(\boldsymbol{y}, \boldsymbol{z}) = \frac{\boldsymbol{y} \cdot \boldsymbol{z}}{\|\boldsymbol{y}\| \, \|\boldsymbol{z}\|}.$$

A canonical interpretation of cosine similarity is that it gives the cosine of the angle between the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ in the plane determined by $\boldsymbol{y}$ and $\boldsymbol{z}$. When $\rho(\boldsymbol{y}, \boldsymbol{z}) = 1$ (resp., -1), the vector $\boldsymbol{z}$ is a positive (resp., negative) rescaling of $\boldsymbol{y}$. A cosine similarity of 0 implies that $\boldsymbol{y}$ is orthogonal to $\boldsymbol{z}$.

4.1. A monotonicity result.

We are now ready to characterize optimal interventions for a quadratic planner's adjustment cost and arbitrary status quo vector.

Recall the earlier finding that in the simple planner's problem with $\gamma = -1$ (malevolent planner) and a constraint of the form $\|\boldsymbol{f}\| \leq 1$, the planner focused *only* on the lowest principal component. The substance of the next result is that in a suitable sense, this finding generalizes: the planner intervenes more on the principal components with lower eigenvalues.

Theorem 1 (Characterization of Optimal Interventions)**.** Suppose[17] $c(\boldsymbol{\delta}) = \|\boldsymbol{\delta}\|^2$. Also suppose that either $\gamma = -1$ or C is small enough that $W(\boldsymbol{a}^*) = \mathbf{0}$ is not feasible for the planner. For generic $\hat{\boldsymbol{f}}$, the similarity between $\boldsymbol{\delta}^*$ and principal component $\boldsymbol{u}^\ell(\boldsymbol{G})$ satisfies, for $\ell \geq 2$,

$$\rho(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell) = \rho(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell) \cdot m(\lambda_\ell),$$

where the **multiplier function** m is such that $|m(\lambda)|$ is decreasing in λ.

Proof. Let $\boldsymbol{f}^*$ give the optimal choice of $\boldsymbol{f}$, so that $\boldsymbol{\delta}^* = \boldsymbol{f}^* - \hat{\boldsymbol{f}}$. Define

$$x_\ell = \frac{\underline{f}_\ell - \underline{\hat{f}}_\ell}{\underline{\hat{f}}_\ell}.$$

Then we can rewrite the optimization problem in the principal component basis as follows, for an increasing, negative function $\zeta(\lambda)$:

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \gamma \sum_\ell \zeta(\lambda_\ell)(1 + x_\ell)^2 \underline{\hat{f}}_\ell^2 \\ \text{s.t.} \quad & \sum_\ell \underline{\hat{f}}_\ell^2 x_\ell^2 \leq C. \end{aligned} \tag{7}$$

By our assumption that either $\gamma = -1$ or achieving no miscoordination is infeasible, the budget constraint binds. Thus, letting μ be the Lagrange multiplier on the budget constraint, the Karush-Kuhn-Tucker necessary condition for optimization is

$$2\gamma \underline{\hat{f}}_\ell^2 \cdot \zeta(\lambda_\ell)(1 + x_\ell^*) = 2\underline{\hat{f}}_\ell^2 \cdot \mu x_\ell^*.$$

Solving for x_ℓ^*, we get $\gamma\zeta(\lambda_\ell) = x_\ell^*(\mu + \gamma\zeta(\lambda_\ell))$, and since the left-hand side is clearly nonzero whenever $\lambda_\ell \neq 1$, it follows that the right-hand side is nonzero too, and we may write

$$\frac{\gamma\zeta(\lambda_\ell)}{\mu + \gamma\zeta(\lambda_\ell)} = x_\ell^*. \tag{8}$$

We note a few facts about the solution. From (7) it follows that the x_ℓ are all positive at an optimum if $\gamma = -1$ and all negative at an optimum if $\gamma = 1$ (by the same argument as in the proof of Theorem 1 of Galeotti, Golub and Goyal (2020)).[18] Lemma A.1 gives us that ζ is a negative, increasing function of its argument. Thus, the denominator $\mu + \gamma\zeta(\lambda_\ell)$ in the solution for x_ℓ^* is always positive, and $|x_\ell^*|$ is decreasing in λ_ℓ.

[17]Note that we can accommodate any scaling of such a function by suitably adjusting C.

[18]The intuition is that at $x_\ell = 0$, the marginal returns of increasing any x_ℓ are nonzero, while the marginal costs are arbitrarily low.

Note that

$$x_\ell^* = \frac{\|\boldsymbol{\delta}^*\| \, \rho\left(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell(\boldsymbol{G})\right)}{\left\|\hat{\boldsymbol{f}}\right\| \rho\left(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell(\boldsymbol{G})\right)}$$

by definition of cosine similarity, so the previous display (8) becomes

$$\frac{\gamma\zeta(\lambda_\ell)}{\mu + \gamma\zeta(\lambda_\ell)} = \frac{\|\boldsymbol{\delta}^*\| \, \rho\left(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell(\boldsymbol{G})\right)}{\left\|\hat{\boldsymbol{f}}\right\| \rho\left(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell(\boldsymbol{G})\right)}.$$

Rearranging the previous expression gives

$$\rho\left(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell(\boldsymbol{G})\right) = \rho\left(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell(\boldsymbol{G})\right) \cdot \frac{\gamma\zeta(\lambda_\ell)}{\mu + \gamma\zeta(\lambda_\ell)} \frac{\left\|\hat{\boldsymbol{f}}\right\|}{\|\boldsymbol{\delta}^*\|}.$$

By our earlier remark about the monotonicity of x_ℓ^* the claim of the proposition follows. □

It is worth remarking on a few features of the key expression

$$\rho(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell) = \rho(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell) \cdot m(\lambda_\ell).$$

First, the "status quo term" $\rho(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell)$ reflects that the nature of interventions depends on the status quo. For example, if the planner is benevolent and $\hat{\boldsymbol{f}}_\ell$ is zero or nearly zero, then there is very little disagreement in that principal component and thus very little to remove; therefore, the planner will not devote a lot of resources to reducing disagreement in that component. The multiplier term captures that components with lower eigenvalues have a bigger welfare impact, and so a planner will care more about adjusting them.

Crucially, under the assumptions of the theorem, this is true whether the planner is malevolent or benevolent. In the malevolent case, the intuition is exactly the same as that of Proposition 1: intensifying disagreement in that component has the greatest impact on the disutility of miscoordination, and the planner will want to take advantage of that to increase this disutility. But, under our assumptions that a benevolent planner cannot reach her bliss point of no misscoordination, the intuition applies in the other direction, too: reducing disagreement in the lowest-eigenvalue component is the most effective use of resources to *reduce* disutility.[19]

4.2. **General cost functions and small budgets.** A quadratic adjustment cost is a restrictive assumption. Here we show that we can relax this assumption and obtain a version

[19]Note that the result in Proposition 1(2) was about a constraint with a *fixed* amount of disagreement, and thus there is no conflict between that result and this intuition.

of our result for small budgets C, with a simpler characterization of the multiplier function m.

We first make a few assumptions on the structure of the cost function $c(\cdot)$.

Assumption 3 (Properties of the Cost Function)**.** The cost function $c(\cdot)$ satisfies the following assumptions: it is twice differentiable; invariant to permutations of the entries of its argument $\boldsymbol{\delta}$; nonnegative on its domain; has the value $c(\mathbf{0}) = \mathbf{0}$; and has nonsingular Hessian at $\boldsymbol{\delta} = \mathbf{0}$.

Making these assumptions implies by standard arguments the approximation

$$c(\boldsymbol{\delta}) = k \left\|\boldsymbol{\delta}\right\|^2 + o(\|\boldsymbol{\delta}\|^2).$$

Proposition 3 (Characterization of Small Interventions)**.** Suppose Assumption 3 holds. Then for generic $\hat{\boldsymbol{f}}$, the similarity between $\boldsymbol{\delta}^*$ and principal component $\boldsymbol{u}^\ell(\boldsymbol{G})$ satisfies, for $\ell \geq 2$,

$$\rho(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell) = \rho(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell) \cdot m(\lambda_\ell)$$

where

$$\lim_{C \to 0} \frac{m(\lambda_\ell)}{m(\lambda_{\ell'})} = \frac{\zeta(\lambda_\ell)}{\zeta(\lambda_{\ell'})}.$$

The result follows immediately from Theorem 1 by the same argument as in Galeotti, Golub and Goyal (2020, OA3.3).

Because we have an explicit form for ζ in Lemma A.1, this result gives a complete description of the optimal intervention. All the cosine similiarities for an orthonormal basis fully pin down the direction of the intervention, and its magnitude is found by exhausting the budget.

4.3. **An implication for networks with homophily.** We emphasized in Section 3.2 that interventions for global discord are extremely different in their form from those for welfare reasons. We can now sketch an application of this to assess whether an intervention is in fact optimal in a practical setting. Our point will be that the characterization permits some simple insights, building on what is known about the spectral structure of real social networks.

Suppose a planner faces a network such as the one shown in Figure 5, with a certain value of λ_2, say $\lambda_2 \geq 0.9$ in a homophilous network.[20] Because $\zeta(\lambda_2)$ is small for large λ_2, the proposition immediately implies a bound on the cosine similarity $\rho(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell)$: if $m(\lambda_\ell)$ is small, then $\rho(\boldsymbol{\delta}^*, \boldsymbol{u}^\ell)$ is small irrespective of the value of $\hat{\boldsymbol{f}}$, since the $\rho(\hat{\boldsymbol{f}}, \boldsymbol{u}^\ell)$ factor in Proposition 3 is bounded by 1.

[20]See Golub, Jackson et al. (2012) for more details.

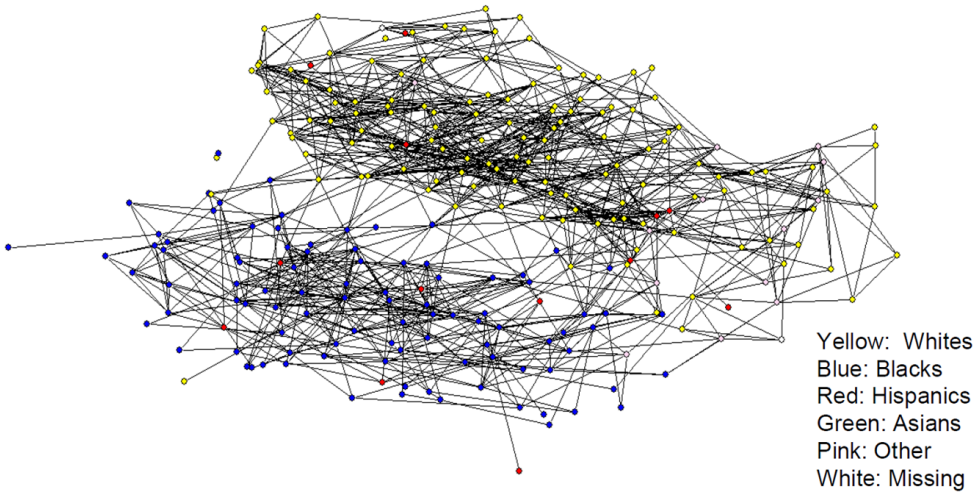

FIGURE 5. A school network from Currarini, Jackson and Pin (2009).

It follows that if a purportedly optimal intervention has a substantial correlation with $\boldsymbol{u}^2$, it is not in fact optimal.[21] In other words, welfare-optimal interventions cannot have significant correlation with the main spectral cut of a homophilous network ($\boldsymbol{u}^2$).

5. Conclusion

There is a useful duality between the theory of network games and the study of network structure. A familiar pattern goes as follows. We fix a game—e.g., a canonical coordination game—and ask a natural economic question about it, such as what perturbations of agents' ideal points result in large welfare changes. Sometimes, a particular family of network statistics (in this case, the lowest eigenvalue and its associated eigenvector) emerges as an important part of a characterization. Then we have learned both an answer to our economic question and a new interpretation of certain statistics—as well as a new reason to be attentive to the statistics in some situations.

In this paper, the statistics that emerge from this procedure are λ_n and $\boldsymbol{u}^n$, as well as other low eigenvalues and eigenvectors. The eigenvalue λ_2 and the eigenvector $\boldsymbol{u}^2$ have been made famous in both applied mathematics and economics by studies of spectral clustering, homophily, and opinion polarization (Spielman and Teng, 2007; DeMarzo, Vayanos and Zwiebel, 2003). But we have spent less time with λ_n, $\boldsymbol{u}^n$, and their friends at the low

[21]In practice, $\boldsymbol{u}^2$ is highly correlated with demographic covariates (in this example, race), as discussed in Golub, Jackson et al. (2012). So a substantial correlation with race would imply a substantial correlation with $\boldsymbol{u}^2$. Thus, one can refute that an intervention is optimal even without detailed network data, as long as we know that racial homophily is strong.

end of the spectrum. Our analysis here has emphasized their importance for coordination, complementing the findings of some recent studies such as Bramoullé, Kranton and D'Amours (2014); King and Allouch (2019) and Galeotti, Golub and Goyal (2020). More generally, the spectral method for analyzing welfare functionals should be useful for enriching our understanding of the interplay between economic interactions and the networks in which they are embedded.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016, https://doi.org/10.3982/ECTA9623.

Acemoglu, D., and A. Ozdaglar (2011): "Opinion dynamics and learning in social networks," *Dynamic Games and Applications*, 1, 3–49.

Albert, R., H. Jeong, and A.-L. Barabási (2000): "Error and attack tolerance of complex networks," *Nature*, 406, 378–382.

Alon, N., and N. Kahale (1997): "A Spectral Technique for Coloring Random 3-Colorable Graphs," *SIAM Journal on Computing*, 26, 1733–1748.

Angeletos, G.-M., and A. Pavan (2007): "Efficient Use of Information and Social Value of Information," *Econometrica*, 75, 1103–1142, https://doi.org/10.1111/j.1468-0262.2007.00783.x.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's Who in Networks. Wanted: The Key Player," *Econometrica*, 74, 1403–1417, https://doi.org/10.1111/j.1468-0262.2006.00709.x.

Baqaee, D. R., and E. Farhi (2019): "The macroeconomic impact of microeconomic shocks: beyond Hulten's Theorem," *Econometrica*, 87, 1155–1203.

——— (2020): "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 135, 105–163.

Bindel, D., J. Kleinberg, and S. Oren (2011): "How Bad is Forming Your Own Opinion?" in *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, 57–66, 10.1109/FOCS.2011.43.

Bramoullé, Y., R. Kranton, and M. D'Amours (2014): "Strategic Interaction and Networks," *American Economic Review*, 104, 898–930, 10.1257/aer.104.3.898.

Calvó-Armengol, A., J. De Martí, and A. Prat (2015): "Communication and influence," *Theoretical Economics*, 10, 649–690.

Currarini, S., M. O. Jackson, and P. Pin (2009): "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 77, 1003–1045, https://doi.org/10.3982/ECTA7528.

DeGroot, M. H. (1974): "Reaching a Consensus," *Journal of the American Statistical Association*, 69, 118–121.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions," *The Quarterly Journal of Economics*, 118, 909–968, 10.1162/00335530360698469.

Desai, M., and V. Rao (1994): "A Characterization of the Smallest Eigenvalue of a Graph," *Journal of Graph Theory*, 18, 181–194.

Friedkin, N. E., and E. C. Johnsen (1999): "Social Influence Networks and Opinion Change," *Advances in Group Processes*, 16, 1–29.

Gaitonde, J., J. Kleinberg, and E. Tardos (2020): "Adversarial Perturbations of Opinion Dynamics in Networks," in *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, 471–472, New York, NY, USA: Association for Computing Machinery, 10.1145/3391403.3399490.

Galeotti, A., B. Golub, and S. Goyal (2020): "Targeting Interventions in Networks," *Econometrica*, 88, 2445–2471, https://doi.org/10.3982/ECTA16173.

Golub, B., M. O. Jackson et al. (2012): "Does homophily predict consensus times? Testing a model of network structure via a dynamic process," *Review of Network Economics*, 11, 1–31.

Golub, B., and M. O. Jackson (2012): "How Homophily Affects the Speed of Learning and Best-Response Dynamics," *The Quarterly Journal of Economics*, 127, 1287–1338, 10.1093/qje/qjs021.

Golub, B., and E. Sadler (2016): "Learning in social networks," in *The Oxford Handbook of the Economics of Networks* ed. by Bramoullé, Y., Galeotti, A., Rogers, B., and Rogers, B.: Oxford University Press, Chap. 19, 504–542.

Jackson, M. O., and Y. Zenou (2014): "Games on Networks," in *Handbook of Game Theory* ed. by Young, P., and Zamir, S.: Elsevier Science, Chap. 3, 95–163.

Kempe, D., J. Kleinberg, and E. Tardos (2015): "Maximizing the Spread of Influence through a Social Network," *Theory of Computing*, 11, 105–147, 10.4086/toc.2015.v011a004.

King, M., and N. Allouch (2019): "A network approach to welfare," *BSG Working Paper Series*, https://www.bsg.ox.ac.uk/sites/default/files/2019-02/

BSG-WP-2019-027.pdf.

Morris, S. (2000): "Contagion," *The Review of Economic Studies*, 67, 57–78, 10.1111/1467-937X.00121.

Spielman, D. A., and S.-H. Teng (2007): "Spectral partitioning works: Planar graphs and finite element meshes," *Linear Algebra and its Applications*, 421, 284–305, https://doi.org/10.1016/j.laa.2006.07.020, Special issue in honor of Miroslav Fiedler.

Urschel, J. C. (2018): "Nodal decompositions of graphs," *Linear Algebra and its Applications*, 539, 60–71, https://doi.org/10.1016/j.laa.2017.11.003.

U.S. House of Representatives Permanent Select Committee on Intelligence (2018): "Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements," https://intelligence.house.gov/social-media-content/.

Valente, T. W. (2012): "Network Interventions," *Science*, 337, 49–53, http://www.jstor.org/stable/41585201.

Appendix A. Functions Used in Spectral Forms of Objectives

Lemma A.1. The following functions give the welfare, covariance of neighbors, and covariance of a random pair of agents in the principal component basis.

(1) Welfare is given by

$$\sum_{\ell=1}^{n} \zeta(\lambda_\ell) \underline{f}_\ell^2,$$

where

$$\zeta(\lambda) = -\beta(1-\beta)\frac{(1-\lambda)[2-\beta(1+\lambda)]}{(1-\beta\lambda)^2}.$$

(2) Covariance of neighbors is given by

$$\sum_{\ell=1}^{n} \eta(\lambda_\ell) \underline{f}_\ell^2,$$

where

$$\eta(\lambda) = \frac{(1-\beta)^2\lambda}{(1-\beta\lambda)^2 n}.$$

(3) Covariance of a random pair is given by

$$\sum_{\ell=1}^{n} \nu(\lambda_\ell) \underline{f}_\ell^2,$$

where

$$\nu(\lambda) = \frac{-(1-\beta)^2}{(1-\beta\lambda)^2 n^2}.$$

Proof. The welfare function is given by

$$V^* = -(1-\beta)\left[\langle \boldsymbol{f}, \boldsymbol{f}\rangle + (1-\beta)\langle (\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}, ((1+\beta)\boldsymbol{I}-2\beta\boldsymbol{G})(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}-2\boldsymbol{f}\rangle\right].$$

The covariance of neighbors is given by

$$\frac{1}{n}\langle \boldsymbol{a}^*, \boldsymbol{G}\boldsymbol{a}^*\rangle = \frac{1}{n}\langle (1-\beta)(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}, (1-\beta)\boldsymbol{G}(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}\rangle.$$

The covariance of a random pair is given by

$$-\frac{1}{n^2}\langle \boldsymbol{a}^*, \boldsymbol{a}^*\rangle = -\frac{1}{n^2}\langle (1-\beta)(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}, (1-\beta)(\boldsymbol{I}-\beta\boldsymbol{G})^{-1}\boldsymbol{f}\rangle.$$

The ζ, η, and ν functions are then immediate by calculation. □