Games on Endogenous Networks

Benjamin Golub and Evan Sadler* February 3, 2021

Abstract

We study network games in which players both create spillovers for one another and choose with whom to associate. The endogenous outcomes include both the strategic actions (e.g., effort levels) and the network in which spillovers occur. We introduce a framework and two solution concepts that extend standard approaches—Nash equilibrium in actions and pairwise (Nash) stability in links. Our main results show that under suitable monotonicity assumptions on incentives, stable networks take simple forms. Our central conditions concern whether actions and links are strategic complements or substitutes, as well as whether links create positive or negative payoff spillovers. We apply our model to understand the consequences of competition for status, to microfound matching models that assume clique formation, and to interpret empirical findings that highlight unintended consequences of group design.

1 Introduction

Consider a situation in which agents' effort has spillovers on peers. For example, if students in an academic program study together, then one student's studying effort changes the incentives of other students. The spillover might exhibit strategic complements, so that more effort results in greater incentives for others to study, or it might exhibit strategic substitutes, so that students have an incentive to free-ride on the efforts of others. How does the economic environment (e.g., individuals' private incentives and the nature of spillovers) shape outcomes? The analysis of such problems is important for designing policies, such as deciding which students to place in a peer group together. If effort decisions are strategic complements, a planner may want to incentivize some group members to work harder, inducing positive spillovers on others. Such a planner might also introduce a new member to a low effort group in order to increase its productivity. Indeed, policy often affects both the incentives of individual members and the composition of peer groups that engage in production, education, or other group activities. These kinds of network spillovers are of

^{*}Golub gratefully acknowledges support from the National Science Foundation under grant SES-1629446. Rithvik Rao provided excellent research assistance.

first-order importance in a broad array of settings including schools, workplaces, military cohorts, and other teams. A literature on *network games* studies these strategic interactions in a given network to understand how different networks and exogenous conditions (e.g., incentive schemes) translate into outcomes.

An obvious but often neglected factor is that the relevant network is endogenous. In addition to choosing actions, such as a productive effort, individuals also engage in strategic network formation, selecting those with whom they partner. These endogenous relationships affect whose effort within a group is complementary—who can collaborate or who can influence each other—and the joint presence of network formation and effort choice is critical. While network games and strategic network formation are mature and largely distinct areas of study, there are clear reasons to combine the two.

Indeed, endogenous linking decisions can overturn predictions based on an exogenous network. To illustrate this, consider an important example from Carrell et al. [2013]. In this study, researchers designed peer groups in a military academy, aiming to improve the academic performance of low-skilled freshmen through network spillovers. They first estimated peer effects using data from random assignment to peer groups before the intervention. This analysis showed that low-skilled students received positive spillovers from high-skilled freshmen. Extrapolating from these estimates, the authors subsequently designed peer groups, placing low-skilled freshmen in groups with a higher proportion of high-skilled freshmen. While the initial estimates suggested this would help low-skilled students perform better, the intervention ultimately produced a negative effect on its intended beneficiaries. The authors interpret this as a consequence of endogenous friendship and collaboration networks within administratively assigned groups: the interactions generating positive spillovers in the random groups did not form in the designed groups. This example underscores the need for a simultaneous analysis of network formation and peer effects.

We introduce a general framework to study games with network spillovers together with strategic link formation. Our theoretical contribution is twofold. First, our framework nests standard models of each type of interaction on its own, and we adapt the definitions of equilibrium and pairwise stability to obtain an appropriate solution concept. Our notion of a stable outcome captures the intuition that no player benefits from changing her action¹ or unilaterally removing links, and no pair of players can jointly benefit from creating a link between them. We subsequently establish existence results through standard methods.

Second, we identify key payoff properties ensuring that stable networks have simple structures. When a game is separable, meaning that the value of a link is a function of players' equilibrium actions, we obtain sharp characterizations that depend on two kinds of strategic monotonicity. The first concerns the nature of spillovers. We say a game has positive spillovers if players taking higher actions are more attractive neighbors; correspondingly, a game has negative spillovers if players taking higher actions are less attractive neighbors. The second kind of monotonicity concerns the relationship between actions and links. Actions and links are complements if the returns from taking higher actions increase with one's number of links. Actions and links are substitutes if these returns decrease with one's num-

¹We use this term from now on to mean the strategic action *other* than the link choice.

ber of links. Given a pair of monotonicity assumptions, our main result characterizes the corresponding structure of both actions and links in equilibrium. Table 1 summarizes our findings. The result demonstrates that these two natural properties—the nature of spillovers and whether actions and links are complements or substitutes—provide a useful way to organize our understanding of games on endogenous networks.

Interaction between neighbors' action incentives

		Positive Spillovers	Negative Spillovers
Interaction between links and actions	Complements	Nested split graph, higher degree implies higher action	1 / 0
	Substitutes	Cliques, neighbors take similar actions	Nested split graph, higher degree implies lower action

Table 1: Summary of main result.

Three additional contributions emerge from applications of our results. First, we connect the predictions of the table to common social and economic situations. Some of the cells in our table concern combinations not previously studied in the networks formation literature, and we argue that our results can explain intuitive social phenomena. As one example, we analyze "status games," which we define as the combination of action-link complements and negative spillovers. For example, imagine people competing for social status by investing in wealth signaling through conspicuous consumption while simultaneously forming relationships. Those with more friends have a greater incentive to consume conspicuously; on the other hand, those with higher actions are less attractive friends, since adding them creates negative comparisons. In this setting, our model predicts that individuals will sort into cliques with members that invest similar amounts in signaling—a finding consistent with stylized facts from sociological studies.

A similar situation arises with positive spillovers and link-action substitutes—in an academic setting, students who study together create benefits for their peers, but more time studying makes link formation and maintenance more costly. This point underscores a second application of our results, providing a microfoundation (under suitable circumstances) for "club" or "group matching" models. Theories of endogenous matching for public goods or team production often assume a clique structure in the incidence of spillovers, which is critical for tractability. We show that even when agents can arrange their interactions into more complex structures if they wish (e.g., a pair could deviate to collaborate in addition to their existing team relationships), cliques are still the predicted outcome in the relevant cells of the above table.

Our third application uses the model to shed light on the counterintuitive empirical result of Carrell et al. [2013]. We work with a setting with strategic complements and link-action

substitutes—which, as noted above, fits a setting of educational effort where time is scarce. We find that replacing some medium-ability students with high-ability students can make the low-ability students worse off because the cliques that form exclude the low-ability students. The reduction of a general network formation model to a much simpler one of matching into cliques drastically simplifies the analysis.

1.1 Related Work

Our analysis sits at the intersection of two strands of work in network theory: network games and strategic network formation. Some of the most widely-used and tractable models feature real-valued actions and best replies that are linear in opponents' strategies [Ballester et al., 2006, Bramoullé and Kranton, 2007, Bramoullé et al., 2014]; many of our examples are based on these models. Chen et al. [2018] extends this work, studying a model in which players take two distinct actions that may be either complements or substitutes for one another, and Sadler [2020] explores the robustness of equilibrium characterizations based on centrality measures. Our analysis is closest in spirit to this last paper as our findings derive from order properties of the payoff functions and do not rely on particular functional forms.

Within the network formation literature, Jackson and Watts [2001], Hellmann [2013], and Hellmann [2020] each provide antecedents to a corresponding result in our paper. Jackson and Watts [2001] present an existence result for pairwise stable networks based on a potential function—our first existence result extends this to a setting in which players take strategic actions in addition to forming links. Likewise, our second existence result extends the main finding in Hellmann [2013]. Hellmann [2020] studies a network formation game in which all players are ex-ante identical and uses order properties of the payoff functions to characterize the architecture of stable networks. A key result shows that if more central players are more attractive linking partners, then stable networks are nested split graphs. By specifying an appropriate network game, one can view this finding as a special case of the link-action complements and positive spillovers cell in our table.

We are aware of two papers that also study strategic interactions together with endogenous network formation—in both cases, the decision to form a link is made unilaterally (in contrast to our model, in which stability is based on mutual consent). Galeotti and Goyal [2009] study a game in which players invest in information gathering and simultaneously choose links to form. Linked players share the information that they gather. Though link formation is unilateral, and the proposer of a link incurs the cost, information flows in both directions. Equilibrium networks involve a core-periphery structure. In Herskovic and Ramos [2020], agents receive exogenous signals and form links to observe others' signals, and they subsequently play a beauty contest game. In this game, a player whose signal is observed by many others exerts greater influence on the average action, which in turn makes this signal more valuable to observe. The equilibrium networks have a hierarchical structure closely related to nested split graphs. In a related but distinct effort, König et al. [2014] study a dynamic network formation model in which agents myopically add and delete links. The underlying incentives satisfy positive spillovers and link-action complements, and accordingly the stochastically stable outcomes are nested split graphs.

Turning to applications and illustrations, there are several other connections to note. First, recall that under action-link substitutes and positive spillovers, our results state that the stable structure are cliques, with all members of a clique exerting symmetric externalities and strategic spillovers on one another. This result can be seen as providing a microfoundation for assumptions that take such clique-matching as an assumed feature of the technology. Examples include Baccara and Yariv [2013] on the formation of social clubs, Chade and Eeckhout [2018] on information provision in teams, and Bandyopadhyay and Cabrales [2020] on matching for public goods provision. We flesh out these connections in Section 5.1.

Another important connection is to theoretical and empirical work on competition for status. Immorlica et al. [2017] present a theoretical model that features negative spillovers and action-link complements in an exogenous network. They use cohesion properties of this fixed network to characterize the status hierarchy. In contrast, we study an endogenous network and find that different status levels endogenously organize themselves into disjoint cliques. On the empirical side, an extensive literature in social science documents cliquish patterns of interaction and their relation to competition for status; we survey this in 5.3.

2 Framework

A network game with network formation is a tuple $\langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$ consisting of the following data:

- There is a finite set N of players; we write \mathcal{G} for the set of all simple, undirected graphs on N.²
- For each player $i \in N$, there is a set S_i of actions; we write $S = \prod_{i \in N} S_i$ for the set of all action profiles.
- For each player $i \in N$, there is a payoff function $u_i : \mathcal{G} \times S \to \mathbb{R}$. This gives player i's payoff as a function of a graph $G \in \mathcal{G}$ and a profile of players' actions.

A pair $(G, \mathbf{s}) \in \mathcal{G} \times S$ is an **outcome** of the game. Given a graph G, we write G_i for the neighbors of player i, we write G + E for the graph G with the links E added, and we write G - E for the graph G with the links E removed.

2.1 Solution concepts

Intuitively, in a solution to a network game with network formation, players should have an incentive to change neither their actions nor their links. We propose two nested solution concepts. These mirror, and extend to our setting with action choices, existing concepts in the network formation literature.

Definition 1. An outcome (G, \mathbf{s}) is **pairwise stable** if the following conditions hold.

²We identify a graph with its set E of edges or links—an edge is an unordered pair of players. We write ij for the edge $\{i, j\}$.

- The action profile **s** is a Nash equilibrium of the game $\langle N, (S_i)_{i \in N}, (u_i(G, \cdot))_{i \in N} \rangle$ in which G is fixed and players only choose actions s_i .
- There is no link $ij \in G$ such that $u_i(G ij, \mathbf{s}) > u_i(G, \mathbf{s})$.
- There is no link $ij \notin G$ such that both $u_i(G+ij, \mathbf{s}) \geq u_i(G, \mathbf{s})$ and $u_j(G+ij, \mathbf{s}) \geq u_i(G, \mathbf{s})$ with at least one strict inequality.

An outcome (G, \mathbf{s}) is **pairwise Nash stable** if the following conditions hold.

- The action profile **s** is a Nash equilibrium of the game $\langle N, (S_i)_{i \in N}, (u_i(G, \cdot))_{iN} \rangle$ in which G is fixed and players only choose actions s_i .
- There is no subset $H \subseteq G_i$ of i's neighbors such that $u_i(G \{ik : k \in H\}, \mathbf{s}) > u_i(G, \mathbf{s})$.
- There is no link $ij \notin G$ such that both $u_i(G+ij, \mathbf{s}) \geq u_i(G, \mathbf{s})$ and $u_j(G+ij, \mathbf{s}) \geq u_i(G, \mathbf{s})$ with at least one strict inequality.

Both of these solution concepts reflect that link formation requires mutual consent. An outcome is pairwise stable if **s** is a Nash equilibrium given the graph, no player wants to unilaterally delete a link, and no pair of players jointly wish to form a link. Pairwise Nash stability adds the stronger requirement that no player benefits from unilaterally deleting some subset of her links.

Note that whenever a player considers a link deviation, she takes the action profile **s** as given; implicit in our definition is that players assess links and actions separately. Allowing the possibility of joint deviations, in which a player or pair of players simultaneously adjust links and actions, would further refine these solution concepts.³

2.2 Two examples

Inspired by common applications of network models, we highlight two examples that we use throughout the paper to illustrate key findings.

A complementary-effort game: The first is a game of strategic complements that generalizes standard peer effects models. In this example, we take the action space $S_i = \mathbb{R}_+$ to be the set of nonnegative real numbers for each $i \in N$. We refer to s_i as an effort level. Letting $d_i = |G_i|$ denote the degree of player i, the payoffs are

$$u_i(G, \mathbf{s}) = b_i s_i + \alpha s_i \sum_{j \in G_i} s_j - c(d_i, s_i).$$
(1)

We interpret the action s_i as a level of effort in some activity—studying, crime, attendance at religious services—and links represent relationships that convey spillovers. Each player

³Of course, it may also create issues for existence.

derives some standalone benefit b_i per unit of effort, and assuming $\alpha > 0$, each player also benefits from the efforts of her neighbors.

The last term captures the costs of effort and link maintenance; we assume c is increasing in both arguments. Further assumptions on the cost function should reflect the setting we have in mind. A key dimension is whether d_i and s_i are strategic complements or substitutes. Suppose we are studying peer effects in education, and s_i describes time spent studying. If maintaining a collaborative friendship requires some activities that are distinct from studying, then it would be natural that the marginal cost of a link c(d+1,s) - c(d,s) is increasing in s—the more time a student spends studying, the less time is left over to sustain friendships. Alternatively, if the effort s_i describes direct social investment, such as regular attendance at events, then this effort would facilitate new relationships, and we should expect the marginal cost of a link to decrease with s.

A public goods game: Our second example is a game of strategic substitutes. Players invest in creating local public goods—for instance, gathering information or learning through experience how to use a new technology—and can form links to enjoy the benefit of others' efforts. Again assume $S_i = \mathbb{R}_+$ for each $i \in N$, and the payoffs are

$$u_i(G, \mathbf{s}) = b_i s_i + \sum_{j \in G_i} f(s_i, s_j) - \frac{1}{2} s_i^2 - c d_i$$
 (2)

with b_i and c strictly positive.⁴ The parameter b_i captures the private benefit from investing for player i, while the constant c captures the cost of linking. The function $f(s_i, s_j)$ describes the benefit from linking to j. For this example, we assume $f(s_i, s_j) = f(s_j, s_i) > 0$ is symmetric, increasing and concave in each argument, and the cross partial $\frac{\partial^2 f}{\partial s_i \partial s_j} < 0$ is negative—this implies that neighbors' actions are strategic substitutes.⁵

The form of (2) implies that the net gains from forming a link are shared equally between the two players involved. We note, however, that since monotonic transformations of u_i do not affect best responses, this is less restrictive than it seems: differences in relative benefits can be reflected in the ratio of f to b_i , which can vary across the players due to differences in b_i . The substantive assumption here is that any link either benefits both players or harms both players. In the context of a game in which players gather information and share it with their neighbors, we might imagine that a player who has invested much more effort than a neighbor still benefits from explaining her findings.

3 Existence and a selection criterion

As in the literature on strategic network formation, pairwise stable or pairwise Nash stable outcomes need not exist. We provide conditions, covering some important applications, under

⁴This example is very similar to the model of Galeotti and Goyal [2010]. The main differences are i) the impact of neighbors' actions on a player's payoff is additively separable, and ii) link formation requires mutual consent.

⁵One example of such a function is $f(s_i, s_j) = g(s_i + s_j)$, in which g is an increasing and concave function.

which existence of equilibria is guaranteed. Each builds on canonical techniques for showing existence. The first derives from potential games [Monderer and Shapley, 1996], while the second is based on monotonicity and strategic complements.

3.1 Potentials

We begin with two definitions.

Definition 2. Given a network game with network formation, a **network improvement path** is an alternating sequence of graphs and action profiles $(G^{(0)}, \mathbf{s}^{(0)}, G^{(1)}, \mathbf{s}^{(1)}, ...)$ such that the following conditions hold

- (i) For every k, the action profile $\mathbf{s}^{(k)}$ is a Nash equilibrium holding $G^{(k)}$ fixed.
- (ii) The graphs $G^{(k)}$ and $G^{(k+1)}$ differ in exactly one edge.
- (iii) If $ij \in G^{(k+1)}$ but $ij \notin G^{(k)}$, then both players weakly benefit from the link, $u_i(G^{(k+1)}, \mathbf{s}^{(k)}) \ge u_i(G^{(k)}, \mathbf{s}^{(k)})$ and $u_j(G^{(k+1)}, \mathbf{s}^{(k)}) \ge u_j(G^{(k)}, \mathbf{s}^{(k)})$,

with at least one strict inequality.

(iv) If $ij \notin G^{(k+1)}$ but $ij \in G^{(k)}$, then at least one player strictly prefers to delete the link:

$$u_i(G^{(k+1)}, \mathbf{s}^{(k)}) > u_i(G^{(k)}, \mathbf{s}^{(k)}) \quad \text{or} \quad u_j(G^{(k+1)}, \mathbf{s}^{(k)}) > u_j(G^{(k)}, \mathbf{s}^{(k)}).$$

Note some simple implications of this definition: First, an outcome (G, \mathbf{s}) is pairwise stable if and only if there is no network improvement path of the form (G, \mathbf{s}, G') . Second, If we take a maximal network improvement path (i.e., one that cannot be extended at the end), its last two entries constitute a pairwise stable outcome. Hence, if all network improvement paths are finite, then a pairwise stable outcome exists. A necessary condition for finiteness of all network improvement paths is that no improvement path contains a cycle, which is true if the game has an (ordinal) potential.

Definition 3. A function $\phi: \mathcal{G} \times S \to \mathbb{R}$ is a potential for $\langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$ if it satisfies the following conditions for all values of the arguments.

- The inequality $\phi(G, s_i', s_{-i}) > \phi(G, s_i, s_{-i})$ holds if and only if $u_i(G, s_i', s_{-i}) > u_i(G, s_i, s_{-i})$.
- Whenever G and G' differ in exactly one edge, the inequality $\phi(G', \mathbf{s}) > \phi(G, \mathbf{s})$ holds if and only if (G, \mathbf{s}, G') is a network improvement path.

The first condition says that for any fixed G, the game $\langle N, (S_i)_{i \in N}, (u_i(G, \cdot))_{i \in N} \rangle$ is an ordinal potential game with potential $\phi(G, \cdot)$. The second condition says that $\phi(\cdot, \mathbf{s})$ ordinally represents the benefit of any single link change to the players involved, holding fixed the action profile s—the potential increases as long as at least one player is made strictly better off. Our first result shows that if a network game with network formation has a potential, and that potential attains a maximum, then a pairwise stable outcome must exist.

Proposition 1. If a network game with network formation $\langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$ has a potential ϕ , and ϕ attains a maximum at (G, \mathbf{s}) , then (G, \mathbf{s}) is a pairwise stable outcome—in particular, a pairwise stable outcome exists.

Proof. If ϕ attains a maximum at (G, \mathbf{s}) , then \mathbf{s} is clearly a Nash equilibrium holding G fixed as $\phi(G, \cdot)$ is an ordinal potential for the corresponding game. Moreover, since ϕ attains a maximum at (G, \mathbf{s}) , there is no G' differing from G in exactly one edge such that $\phi(G', \mathbf{s}) > \phi(G, \mathbf{s})$, which (by definition of the potential) means there is no improvement path of the form (G, \mathbf{s}, G') . We conclude that (G, \mathbf{s}) is pairwise stable.

An immediate corollary of Proposition 1 is that a potential implies the existence of a pairwise stable outcome as long as there are finitely many Nash equilibria in each game with the network held fixed.

Corollary 1. Suppose that a network game with network formation has a potential ϕ , and for each $G \in \mathcal{G}$, the game $(N, \{S_i\}_{i \in N}, \{u_i(G, \cdot)\}_{i \in N})$ has finitely many Nash equilibria. A pairwise stable outcome exists.

Proof. The potential is strictly increasing along any improvement path. Since there are finitely many action profiles that can appear in an improvement path—recall each of them must be a Nash equilibrium holding the graph fixed—and finitely many graphs, the potential must attain a maximum.

To illustrate the application of these results, consider the second example from Section 2.2, with $s_i \in \mathbb{R}_+$ and payoffs given by (2). We can write a potential for this game as

$$\phi(G, \mathbf{s}) = \sum_{i \in N} \left(b_i s_i - \frac{1}{2} s_i^2 \right) + \sum_{ij \in G} \left(f(s_i, s_j) - c \right).$$

Concavity of f implies there is a bound on each player's best response correspondence, so we can treat the action sets as compact, and ϕ must attain a maximum. Hence, Theorem 1 tells us that a pairwise stable outcome exists.

3.2 Strategic complements

Our second existence result requires that the game exhibits strategic complements. Here we assume that each action set S_i is a lattice with order \geq . Recall that a function $f: X \times T \to \mathbb{R}$, in which X is a lattice and T partially ordered, has the *single crossing property* in X and X if, whenever $X' \geq X$ and $X' \geq X$, it satisfies

$$f(x',t) \ge (>) f(x,t) \implies f(x',t') \ge (>) f(x,t').$$

The function is quasisupermodular in x if for any $x, x' \in X$,

$$f(x,t) - f(x \wedge x',t) \ge (>) \, 0 \quad \implies \quad f(x \vee x',) - f(x',t) \ge (>) \, 0.$$

Definition 4. A network game with network formation exhibits **strategic complements** if:

- Each action set S_i is a complete lattice.
- For each $i \in N$ and $G \in \mathcal{G}$, the payoff $u_i(G, \mathbf{s})$ is quasisupermodular in s_i and has the single crossing property in s_i and s_{-i} .
- For each $i \in N$ and $s_{-i} \in S_{-i}$, the payoff $u_i(G, \mathbf{s})$ has the single crossing property in s_i . and G

The first two parts of Definition 4 ensure that $\langle N, (S_i)_{i \in N}, (u_i(G, \cdot))_{i \in N} \rangle$ is a game of strategic complements for each graph G. Holding G fixed, each player's best response correspondence is increasing (in the strong set order) in opponents' actions, and standard results imply that minimal and maximal Nash equilibria exist. The single crossing property in s_i and G implies that, holding opponent actions fixed, player i wants to choose a higher action if we add links to the graph—adding a link intensifies complementarities.

Our next definition introduces monotonicity conditions on the marginal benefit from adding a link.

Definition 5. A network game with network formation has **positive spillovers** if for each edge ij, any $G \subseteq G'$ with $ij \notin G'$, any $s'_i \ge s_j$, and any s_{-j} , we have

$$u_i(G'+ij,s'_j,s_{-j})-u_i(G',s'_j,s_{-j}) \ge u_i(G+ij,s_j,s_{-j})-u_i(G,s_j,s_{-j}),$$

with strict inequality if $s'_j > s_j$. The game has **weakly positive spillovers** if for each edge ij, any $G \subseteq G'$ with $ij \notin G'$, any $s'_j \ge s_j$, and any s_{-j} , we have

$$u_i(G+ij, s_j, s_{-j}) \ge u_i(G, s_j, s_{-j}) \implies u_i(G'+ij, s'_j, s_{-j}) \ge u_i(G', s'_j, s_{-j}),$$

with strict inequality if $s'_j > s_j$, and

$$u_i(G+ij, s_j, s_{-j}) > u_i(G, s_j, s_{-j}) \implies u_i(G'+ij, s'_j, s_{-j}) > u_i(G', s'_j, s_{-j}).$$

in which the first implication yields strict inequality if $s'_j > s_j$. A game has (weakly) negative spillovers if the game with all u_i replaced by $-u_i$ has (resp., weakly) positive spillovers.

Positive spillovers encompasses two distinct properties. First, all else equal, if player j takes a higher action, she becomes a more attractive neighbor. Second, adding other links to the graph can only increase the benefit from forming a link with j. Negative spillovers reverses both statements: higher actions make someone a less attractive neighbor, and additional links weakly decrease the benefit of linking. Analogous to the distinction between increasing differences and the single crossing property, weak positive spillovers requires only that if i wishes to add link ij to G when j's action is s_j , then i still wishes to form this link if the graph becomes more connected and j takes a higher action. This property, together with strategic complements, ensures that higher actions and more links are mutually reinforcing, and this ensures the existence of pairwise stable outcomes.

Proposition 2. If a network game with network formation exhibits strategic complements and has weakly positive spillovers, then there exist minimal and maximal pairwise stable outcomes.

Proof. We carry out the argument for the minimal outcome; the argument for the maximal outcome is analogous. We can find the minimal pairwise stable outcome via the following algorithm:

- (a) Let $G^{(0)}$ be the empty graph, and let $\mathbf{s}^{(0)}$ be the minimal Nash equilibrium of the game holding the graph $G^{(0)}$ fixed.
- (b) For each $k \ge 1$, we take $ij \in G^{(k)}$ if and only if we have both $u_i(G^{(k-1)} + ij, \mathbf{s}^{(k-1)}) \ge u_i(G^{(k-1)} ij, \mathbf{s}^{(k-1)})$ and $u_j(G^{(k-1)} + ij, \mathbf{s}^{(k-1)}) \ge u_j(G^{(k-1)} ij, \mathbf{s}^{(k-1)})$, with at least one strict inequality.
- (c) For each $k \geq 1$, the action profile $\mathbf{s}^{(k)}$ is the minimal Nash equilibrium of the game holding the graph $G^{(k)}$ fixed.

The action profile $\mathbf{s}^{(k)}$ is always well-defined because the game on a fixed graph is one of strategic complements, which implies it has a minimal and maximal equilibrium in pure strategies. Additionally, since u_i has the single crossing property in s_i and G for each $i \in N$, we know that if \mathbf{s} and \mathbf{s}' are the minimal Nash equilibria associated with G and G' respectively, then $\mathbf{s} \leq \mathbf{s}'$ whenever $G \subseteq G'$. Since $G^{(0)} \subseteq G^{(1)}$, we know that $\mathbf{s}^{(0)} \leq \mathbf{s}^{(1)}$, and weakly positive spillovers now imply that $G^{(1)} \subseteq G^{(2)}$ —by induction, we conclude that $G^{(k-1)} \subseteq G^{(k)}$ for each $k \geq 1$. Since there are finitely many possible graphs, the algorithm must terminate, and the resulting graph is pairwise stable.

For the maximal pairwise stable outcome, start from the complete graph instead of the empty graph, choose the maximal equilibrium instead of the minimal one, and in steb (b) of the algorithm, always include link ij if it weakly benefits both players.

Our first example from Section 2.2 is a case to which Theorem 2 applies. Looking at the payoff function (1), since α is positive, the game exhibits strategic complements. The benefit to i of adding link ij is

$$\alpha s_i s_j - c(d_i + 1, s_i) + c(d_i, s_i).$$

This tells us that if c(d+1,s) - c(d,s) is non-increasing in d for each s, then the game has positive spillovers. As long as action sets are effectively bounded—either due to constraints on effort or because the cost c(d,s) is sufficiently convex in s—Theorem 2 ensures that minimal and maximal pairwise stable outcomes exist.

The notion of a network improvement path suggests a selection criterion in the event that there are multiple pairwise stable outcomes. Suppose there exists a network improvement path starting from an empty graph and ending at a pairwise stable outcome. This means that a group can arrive at this outcome by changing a single link at a time, requiring no coordination beyond bilateral agreements to form a new connection.

Definition 6. A pairwise stable outcome (G, \mathbf{s}) of a network game with network formation is **coordination-free** if there exists an improvement path $(G^{(0)}, \mathbf{s}^{(0)}, ... G^{(k)}, \mathbf{s}^{(k)})$ such that $G^{(0)}$ is the empty graph, and $(G^{(k)}, \mathbf{s}^{(k)}) = (G, \mathbf{s})$.

If the setting we have in mind involves a group of people with no prior relationships forming a new network, then choosing a coordination-free outcome seems like a reasonable selection criterion.⁶ For instance, in the study by Carrell et al. [2013], the vast majority cadets do not know one another when they are first assigned to squadrons—we make use of the coordination-free selection in a later example based on this study.

Implicit in Definition 6 is the idea that links are harder to change than actions; the definition captures this by effectively positing that players can only adjust one link at a time between spells of adjusting action. After each link change, we should imagine players finding a new equilibrium of the underlying game, holding the graph fixed, and only then reconsidering their links again. Note that under the conditions of Theorems 1 and 2, pairwise stable outcomes that are coordination-free must exist. If there is a potential, then we can always reach some local maximum starting from the empty graph. If the game exhibits strategic complements and positive spillovers, then the minimal pairwise stable outcome is the unique coordination-free pairwise stable outcome. Moving past it to a higher outcome requires more than the deviations that some players would want to make bilaterally.

4 The structure of stable graphs

How do properties of the payoff functions $(u_i)_{i\in N}$ affect stable network structures? Our analysis here focuses on separable games, meaning that the payoff to a player from forming a link depends only on her own action and that of the other player. Throughout this section, we assume that all players share the same action set S, and that S is partially ordered, with order \geq .

Definition 7. A network game with network formation is **separable** if there is a function $f: S \times S \to \mathbb{R}$ such that

$$u_i(G+ij,\mathbf{s}) - u_i(G,\mathbf{s}) = f(s_i,s_j)$$

for all players i and j and all graphs G with $ij \notin G$. We say that **actions and links are complements** if, whenever $s'_i > s_i$, we have

$$f(s_i, s_j) \ge 0 \implies f(s_i', s_j) > 0.$$

We say that actions and links are substitutes if, whenever $s'_i > s_i$, we have

$$f(s_i, s_j) \le 0 \implies f(s_i', s_j) < 0$$

⁶Given information on prior relationships, we can still provide a selection based on the same principle, choosing a pairwise stable outcome that can be reached via a network improvement path starting from the initial graph we observe.

We note additionally that, applying Definition 5, a separable game has positive (negative) spillovers if f is increasing (decreasing) in s_j .

Although separability places substantive restrictions on incentives to form links, it nevertheless permits essentially any type of underlying network game. The game can exhibit strategic complements or substitutes (or neither), and players can have arbitrary idiosyncratic incentives to take high or low actions. Because the value of a link depends only on players' actions, not their identities and not on other links present in the graph, in separable games we can strengthen our two existence results: under the conditions of Propositions 1 or 2, a pairwise Nash stable outcome exists.

Proposition 3. If a network game with network formation is separable, then any pairwise stable outcome is pairwise Nash stable.

Proof. We need only check that if there are no profitable single link deletions, then there is no way to profit from multiple link deletions. This is immediate from separability because the marginal gain from deleting a link does not depend on the presence or absence of other links. \Box

Turning now to our main question, we obtain a sharp characterization of the network structures that can arise as stable outcomes in separable games.

Theorem 1. Suppose a network game with network formation is separable, and (G, \mathbf{s}) is a pairwise stable outcome. Then:

- (a) If the game has weakly positive spillovers, and links and actions are complements, then $s_i > s_j$ implies $G_j \subseteq G_i$.
- (b) If the game has weakly positive spillovers, and links and actions are substitutes, then if $s_i > s_j > s_k$ and $ik \in G$, then $ij, jk \in G$.
- (c) If the game has weakly negative spillovers, and links and actions are complements, then if $s_i > s_j > s_k$ and $ik \in G$, then $ij, jk \in G$.
- (d) If the game has weakly negative spillovers and links and actions are substitutes, then $s_i > s_j$ implies $G_i \subseteq G_j$.

Proof. We begin with part (a). Suppose $jk \in G$ and $s_i > s_j$. Since $f(s_k, s_j) \ge 0$, weakly positive spillovers imply that $f(s_k, s_i) > 0$, and since $f(s_j, s_k) \ge 0$, action and link complements imply $f(s_i, s_k) > 0$. Players i and k strictly benefit from linking, so a link must exist.

For part (b), if $ik \in G$, then stability of G implies $f(s_i, s_k) \ge 0$. We make two observations from this. First, since $s_i > s_j$, action-link substitutes imply $f(s_j, s_k) > 0$. Since $s_i > s_k$, weakly positive spillovers now imply that $f(s_j, s_i) > 0$. Second, since $s_j > s_k$, weakly positive spillovers imply $f(s_i, s_j) > 0$. Since $s_i > s_k$, action-link substitutes now imply that $f(s_k, s_j) > 0$. We conclude that ij and jk are both in G.

For part (c), suppose actions and links are complements, but the game has weakly negative spillovers. If $ik \in G$, then $f(s_k, s_i) \geq 0$. Since $s_j > s_k$, action-link complements imply $f(s_j, s_i) > 0$, and since $s_i > s_k$, we can now infer $f(s_j, s_k) > 0$ from weakly negative spillovers. Moreover, since $s_i > s_j$ and $f(s_k, s_i) \geq 0$, weakly negative spillovers imply that $f(s_k, s_j) > 0$. Action-link complements now imply $f(s_i, s_j) > 0$. We conclude that ij and jk are both in G.

Finally, for part (d), suppose actions and links are substitutes, and the game has weakly negative spillovers. If $ik \in G$ and $s_i > s_j$, then since $f(s_k, s_i) \ge 0$, weakly negative spillovers imply $f(s_k, s_j) > 0$, and since $f(s_i, s_k) \ge 0$, action-link substitutes imply $f(s_j, s_k) > 0$. We conclude that $jk \in G$.

The characterization in Theorem 1 is stark. In cases (a) and (d), if one player takes a higher action than another, then the two neighborhoods are ordered by set inclusion. In cases (b) and (c), a link between two players, one taking a higher action than the other, implies that the set of players taking intermediate actions forms a clique. Strict comparisons play an important role as any link ij need not be in G if both i and j are indifferent about adding it.

By strengthening the stability concept slightly, we can obtain a stronger characterization.

Definition 8. The outcome (G, \mathbf{s}) is *strictly pairwise stable* if no player is indifferent about keeping any link in G, and no two players are both indifferent about adding a link between them.

If the hypothesis of Theorem 1 is strengthened to say that (G, \mathbf{s}) is a strictly pairwise stable outcome, then all statements in the theorem remain true if we replace all strict inequalities among actions with weak inequalities (i.e., assume throughout only that $s_i \geq s_j \geq s_k$).⁷

If the action set S is totally ordered, we then obtain the following corollary, which relates neighborhood structures to action levels.

Corollary 2. Suppose the action set S is totally ordered, and (G, \mathbf{s}) is a strictly pairwise stable outcome.

- (a) If the game has weakly positive spillovers and actions and links are complements, then G is a nested split graph, and $|G_i| > |G_j|$ implies $s_i > s_j$.
- (b) If the game has weakly negative spillovers and actions and links are substitutes, then G is a nested split graph, and $|G_i| < |G_j|$ implies $s_i > s_j$.
- (c) If the game has weakly positive spillovers and actions and links are substitutes, or if the game has weakly negative spillovers and actions and links are complements, then every component of G is a clique. Moreover, these cliques are ordered such that every player in a higher clique takes a higher action than every player in a lower clique.

⁷Analogous to the distinction between single crossing differences and strict single crossing differences, in this case we can also weaken the definitions of positive/negative spillovers and action and link complements/substitutes.

Barring indifference, there are essentially two network structures that can arise in stable outcomes. Either neighborhoods are nested, with the order depending on whether we are in case (a) or (b), or the network is segregated into cliques of players taking similar actions.

5 Discussion

5.1 Foundations for group-matching models

Models of endogenous matching that go beyond pairwise interactions often posit that individuals belong to a group of others. Externalities and strategic interactions then occur within or across groups, but crucially, all assumptions about payoffs are invariant to permutations of agents within groups. In our model, this corresponds to interaction happening in cliques by assumption. For example Baccara and Yariv [2013] study a setting where individuals join groups (e.g., social clubs) and then choose an activity to contribute to within the group. Agents' payoffs and incentives are affected by the choices in their group, symmetrically. In Chade and Eeckhout [2018], agents who have information are allocated to teams; again, information sharing occurs within teams, benefitting all members equally, but not across teams. In Bandyopadhyay and Cabrales [2020], agents signal a privately known type, then match into groups in which they provide public goods.

In all these cases, the real interactions motivating the models are not constrained to be in groups by technological facts: bilateral interactions outside the groups are possible. For example, in the example of social clubs, pairs could meet outside groups to share in an activity. In information sharing, some pairs could exchange some, but not all, information with those not in their groups. However, modeling interactions at the group level facilitates simplifications that are essential to the tractability of these models.

Our results give conditions on which, even in an environment that allows much richer interactions (e.g., arbitrarily organized bilateral partnerships), equilibrium interactions would nevertheless endogenously organize themselves into cliques satisfying the symmetry assumptions we have mentioned. Thus, our results can be seen as a microfoundation for, or robustness check on, group-matching assumptions.

5.2 Perverse consequences of group design: An example based on Carrell et al. [2013]

Carrell et al. [2013] estimated academic peer effects among first year cadets at the US Air Force Academy and then used these estimates to design the assignment of new cadets to squadrons. Based on the first wave of measurements using random assignment, the authors concluded that being in a squadron with higher performing peers⁸ led to better academic performance among less prepared cadets. In the second wave, incoming less prepared cadets were placed in squadrons with more high ability peers than under random assignment. While

⁸Specifically, those entering with relatively high SAT verbal scores.

the researchers' goal was to improve the performance of the less prepared cadets, the intervention ultimately backfired: these students performed significantly worse. In this section, we present a simple example showing that our theory can simultaneously explain two peculiar features of the Air Force study:

- (a) When peer group composition changes slightly, low ability are better off when they have more high ability peers, and
- (b) Larger changes in peer group composition eliminate or even reverse this effect.

Consider a network game with network formation in which $S_i = \mathbb{R}_+$ for each player i, and payoffs take the form

$$u_i(G, \mathbf{s}) = b_i s_i + \alpha s_i \sum_{j \in G_i} s_j - \frac{1}{2} (1 + d_i) s_i^2,$$

in which $d_i = |G_i|$ is player *i*'s degree and $\alpha > 0$. Holding the graph fixed, this is a standard linear-quadratic network game of strategic complements. There are positive spillovers, as an increase in s_j makes a link to player *j* more valuable. Moreover, links and actions are substitutes for high enough actions: as player *i*'s net benefit from adding a link to player *j* is

$$\alpha s_i s_j - \frac{1}{2} s_i^2 = s_i \left(\alpha s_j - \frac{1}{2} s_i \right).$$

As s_i increases, this eventually turns negative. From this expression, one can check that in a pairwise stable outcome, players i and j are neighbors only if $\frac{s_j}{2\alpha} \leq s_i \leq 2\alpha s_j$.

A pairwise stable outcome satisfies the first order condition

$$s_i = \frac{1}{d_i + 1} \left(b_i + \alpha \sum_{j \in G_i} s_j \right)$$

for each $i \in N$. Writing \tilde{G} for a matrix with entries $\tilde{g}_{ij} = \frac{1}{d_i+1}$ if $ij \in G$ and 0 otherwise, and $\tilde{\mathbf{b}}$ for a column vector with entries $\frac{b_i}{d_i+1}$, we can write this in matrix notation as

$$\mathbf{s} = \tilde{\mathbf{b}} + \alpha \tilde{G} \mathbf{s} \implies \mathbf{s} = (I - \alpha \tilde{G})^{-1} \tilde{\mathbf{b}}.$$

As long as α is small enough (e.g., $\alpha \leq 1$), the solution for **s** is well-defined for any graph G, and it is an equilibrium of the game holding G fixed.

Now assume $\alpha = 1$, and the game consists of five players, with b_i taking the values 4, 6, or 9. We understand players with $b_i = 4$ as cadets with low ability and those with $b_i = 9$ as cadets with high ability. Given an outcome (G, \mathbf{s}) , we interpret the action s_i as the academic performance of cadet i, and we interpret links as friendships through which peer effects can operate.

We now assess stable outcomes for three different squadron compositions:

⁹We use "ability" as a shorthand for aptitude and preparation.

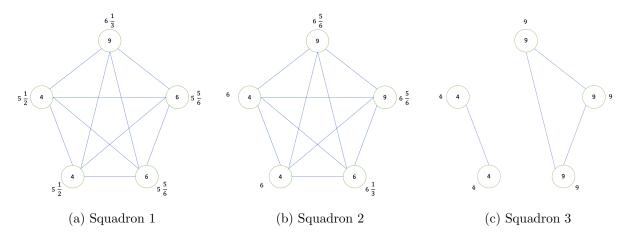


Figure 1: An illustration of the stable outcomes for the three squadrons. Ability levels b_i appear inside each node, while equilibrium actions s_i are next to the node.

• Squadron 1: $\mathbf{b} = (4, 4, 6, 6, 9)$

• Squadron 2: $\mathbf{b} = (4, 4, 6, 9, 9)$

• Squadron 3: $\mathbf{b} = (4, 4, 9, 9, 9)$

In each successive model, we replace a cadet of intermediate ability with a high ability cadet, and we are interested in how the actions of the low ability cadets change.

In squadron 1, the unique pairwise stable outcome is a complete graph with action vector $\mathbf{s} = \left(5\frac{1}{2}, 5\frac{1}{2}, 5\frac{5}{6}, 5\frac{5}{6}, 6\frac{1}{3}\right)$. In squadron 2, the unique pairwise stable outcome again involves a complete graph, and the action vector is $\mathbf{s} = \left(6, 6, 6\frac{1}{3}, 6\frac{5}{6}, 6\frac{5}{6}\right)$. From this we see that adding a second high ability cadet to the squadron increases the performance of low ability cadets from $5\frac{1}{2}$ to 6—there is a benefit to the low ability cadets from this small change in group composition.

What happens when we add another high ability cadet? In squadron 3, the unique coordination-free outcome involves two separate cliques—the two low ability cadets form one clique, the three high ability cadets form the other, and the action vector is $\mathbf{s} = (4, 4, 9, 9, 9)$. A larger change in the group composition results in a marked decline in performance for the low ability cadets.

Although the coordination-free refinement is particularly compelling in this setting—cadets generally do not know one another beforehand—note that the segregated outcome is not uniquely pairwise stable for squadron 3. In fact, the complete graph is still part of a pairwise stable outcome—the corresponding action vector is $\mathbf{s} = \left(6\frac{1}{2}, 6\frac{1}{2}, 7\frac{1}{3}, 7\frac{1}{3}, 7\frac{1}{3}\right)$. This suggests that a more coordinated effort to facilitate friendships may restore the beneficial relationship we saw in the shift from squadron 1 to squadron 2.

5.3 Status games and ordered cliques

Recall that status games involve the combination of action-link complements and negative spillovers—we have already mentioned the example of conspicuous consumption, which is complementary to social activity for the person doing it but makes someone a less attractive friend. The same description applies to many other behaviors such as dominance displays, in which individuals exert a costly effort in explicit competitions to secure status relative to others. Jackson [2019] argues that social behaviors such as binge drinking have the same properties: they are more rewarding for the more socially connected, but they create negative externalities across neighbors because they induce an inefficient amount of unhealthy behavior through social complementarities.

These examples dovetail with anthropological and sociological studies on the pervasiveness of cliques. For example, Davis and Leinhardt [1967] formalize the theory of Homans [1950] asserting that small or medium-sized groups (e.g., departments in workplaces, grades in a school) are often organized into cliques with a clear ranking among them. Adler and Adler [1995] is an ethnographic study of older elementary-school children focused on the clique phenomenon. It argues that status differentiation is clear across cliques and that there are unambiguous orderings between cliques, with one clique occupying the "upper status rung of a grade," and "identified by members and nonmembers alike as the 'popular clique." This study also emphasizes the negative externalities occurring through dominance contests within cliques, consistent with our modeling of negative spillovers. Building on this ethnographic work, Gest et al. [2007] is a detailed quantitative examination of the social structures of students in a middle school (with a particular focus on gender differences). They summarize their findings as confirming the ethnographic narratives: "girls and boys were similar in their tendency to form same-sex peer groups that were distinct, tightly knit, and characterized by status hierarchies."

5.4 Multiplex Networks

Our predictions of stable networks structure are stark. Real networks are typically not organized into entirely disjoint cliques, nor do they have the stark nesting of neighborhoods exhibited by nested split graphs. Nevertheless, by layering the different relationships on top of each other in a "multiplex" network, these stark patterns can combine into something closer to the network patterns that are observed.

Consider a simple example with two types of activities: work on the weekdays—in which the activity is production—and religious services on the weekends—in which the activity is attendance and engagement. Work is characterized by action-link substitutes (forming friendships takes time that could be devoted to production) and positive spillovers. Church is characterized by action-link complements (since attendance makes forming more ties easier) and positive spillovers. Assuming suitable heterogeneity in ability at the work and religious activity, in each activity individually there will be a nontrivial network. In the work network,

¹⁰Davis and Leinhardt [1967] discuss purely graph-theoretic principles that guarantee some features of a ranked-cliques graph, but do not have a model of choices.

this takes the form of disjoint cliques. In the church network, this takes the form of a nested split graph, with the more committed members more broadly connected. Layering these networks on top of each other can produce a complex network with aspects of both "centralization," mediated by the religious ties, and homophily, driven by the work ties. This description ties into Simmel's account, subsequently developed by many scholars, of cross-cutting cleavages.

6 Final Remarks

From academic peer effects to social status to trading networks, the connections people and firms choose to form affect the strategic actions they take and vice versa. As we have seen through examples, restricting attention to one half of the story can lead to misguided predictions and counter-productive interventions. We offer a formal framework that unites two previously distinct areas of study, and we identify simple conditions that allow a sharp characterization of the ensuing network structures. Several widely studied applications fit neatly within this framework, and our discussion highlights new insights—an explanation of cliquish behavior, a reason why we should expect certain efforts at group design to fail—that emerge. Additional applications to better understand how combining different types of relationships can give rise to more complex network structures appear to be a promising direction for future work.

References

- Patricia A Adler and Peter Adler. Dynamics of inclusion and exclusion in preadolescent cliques. Social psychology quarterly, pages 145–162, 1995.
- Mariagiovanna Baccara and Leeat Yariv. Homophily in peer groups. *American Economic Journal: Microeconomics*, 5(3):69–96, 2013.
- C. Ballester, A. Calvó-Armengol, and Y. Zenou. Who's who in Networks. Wanted: the Key Player. *Econometrica*, 74(5):1403–1417, 2006.
- Siddhartha Bandyopadhyay and Antonio Cabrales. Pricing group membership. 2020.
- Yann Bramoullé and Rachel Kranton. Public Goods in Networks. *Journal of Economic Theory*, 135:478–494, 2007.
- Yann Bramoullé, Rachel Kranton, and Martin D'Amours. Strategic Interaction and Networks. *American Economic Review*, 104(3):898–930, 2014.
- Scott Carrell, Bruce Sacerdote, and James West. From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica*, 81(3):855–882, 2013.

- Hector Chade and Jan Eeckhout. Matching information. *Theoretical Economics*, 13(1): 377–414, 2018.
- Ying-Ju Chen, Junjie Zhou, and Yves Zenou. Multiple Activities in Networks. *American Economic Journal: Microeconomics*, 10(3):34–85, 2018.
- James A Davis and Samuel Leinhardt. The structure of positive interpersonal relations in small groups. 1967.
- Andrea Galeotti and Sanjeev Goyal. A Theory of Strategic Diffusion. Rand Journal of Economics, 40:509–532, 2009.
- Andrea Galeotti and Sanjeev Goyal. The Law of the Few. American Economic Review, 100 (4):1468–1492, 2010.
- Scott D Gest, Alice J Davidson, Kelly L Rulison, James Moody, and Janet A Welsh. Features of groups and status hierarchies in girls' and boys' early adolescent peer networks. *New Directions for Child and Adolescent Development*, 2007(118):43–60, 2007.
- Tim Hellmann. On the Existence and Uniqueness of Pairwise Stable Networks. *International Journal of Game Theory*, 42(1):211–237, 2013.
- Tim Hellmann. Pairwise Stable Networks in Homogeneous Societies with Weak Link Externalities. Forthcoming in European Journal of Operational Research, 2020.
- Bernard Herskovic and João Ramos. Acquiring Information Through Peers. American Economic Review, 110(7):2128–2152, 2020.
- George C Homans. The Human Group. Harcourt, New York, 1950.
- Nicole Immorlica, Rachel Kranton, Mihai Manea, and Greg Stoddard. Social Status in Networks. American Economic Journal: Microeconomics, 9(1):1–30, 2017.
- Matthew Jackson and Alison Watts. The Existence of Pairwise Stable Networks. Seoul Journal of Economics, 14(3):299–321, 2001.
- Matthew O Jackson. The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, 127(2):777–818, 2019.
- Michael König, Claudio Tessone, and Yves Zenou. Nestedness in Networks: A Theoretical Model and Some Applications. *Theoretical Economics*, 9(3):695–752, 2014.
- Dov Monderer and Lloyd Shapley. Potential Games. Games and Economic Behavior, 14: 124–143, 1996.
- Evan Sadler. Ordinal Centrality. Working Paper, 2020.