Games on Endogenous Networks*

Benjamin Golub† and Evan Sadler‡

May 17, 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 first condition concerns whether links create positive or negative payoff spillovers. Our second condition concerns whether actions and links are strategic complements or substitutes. Together, these conditions allow for a taxonomy of how network structure depends on economic primitives. 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

Social connections influence our behavior, but our behavior in turn affects the connections we form. For instance, a good study partner might lead a student to exert more effort in school, but studying harder also creates incentives for the partner to collaborate. Understanding both directions of influence is crucial for policy design. Indeed, as Carrell et al. [2013] show, if one neglects that the relevant networks are endogenous, then well-intended interventions can backfire.

In their study, the researchers designed peer groups in a military academy, aiming to improve the academic performance of low-skilled freshmen through peer effects. They first estimated these effects using data from random assignment to peer groups before the intervention and found that additional high-skilled peers did in fact improve the grades of freshmen with less preparation. Extrapolating from these estimates, the authors subsequently

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

†Northwestern University, benjamin.golub@northwestern.edu

‡Columbia University, es3668@columbia.edu

designed peer groups, placing low-skilled freshmen with a higher proportion of high-skilled freshmen. While the initial estimates suggested this would have a positive effect on the low-skilled students' performance, the intervention ultimately produced a comparably sized *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 occur 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. We adapt the definitions of equilibrium (for actions) and pairwise stability (for network formation) to propose two nested solution concepts. These mirror existing concepts in the network formation literature and extend them to our setting with action choices. Intuitively, in a solution to a network game with network formation, players should have an incentive to change neither their actions nor their links. More precisely, our notion of a stable outcome requires that no player benefits from changing her action,[1] holding the graph fixed, nor from 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 depends only on the identities and actions of the two players involved, 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 degree[2] in the network. Actions and links are substitutes if these returns decrease with one's degree. 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.

Three additional contributions emerge as we connect the predictions in the table to common social and economic situations. First, we shed light on the counter-intuitive empirical result of Carrell et al. [2013]. In an academic setting, students who study together create benefits for their peers, but more time studying makes link formation and maintenance more costly. Consequently, we study an example with strategic complements and link-action substitutes. In our example, students have one of three innate ability levels—low, medium, or high. Although replacing just one medium-ability student with a high-ability students bene-

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

[2]Number of neighbors.

		Interaction between neighbors' action incentives	
		Positive Spillovers	**Negative Spillovers**
Interaction between links and actions	**Complements**	Nested split graph, higher degree implies higher action	Ordered overlapping cliques, neighbors take similar actions
	Substitutes	Ordered overlapping cliques, neighbors take similar actions	Nested split graph, higher degree implies lower action

Table 1: Summary of main result.

fits the low-ability students in the group, additional replacements can reverse this effect. The reason is that small changes in group composition do not affect the set of stable graphs—in our example, the complete graph is uniquely stable, so all students benefit from each others' efforts—but larger changes cause the group to fragment. When the group divides into two cliques, one with high-ability students and one with low-ability students, low-ability students no longer experience peer effects from high-ability students, so the benefits disappear.

Second, we study a model of "status games" based on Immorlica et al. [2017]. Intuitively, competitions for status combine action-link complements with negative spillovers. For example, imagine people competing for social status through conspicuous consumption while simultaneously forming relationships. Those with more friends have a greater incentive to flaunt their wealth (action-link complements). On the other hand, those who do so are less attractive friends, since linking with them creates negative comparisons (negative spillovers). In this setting, our model predicts that individuals will sort into cliques with members that invest similar amounts in status signaling—a finding consistent with stylized facts from sociological studies. Subsequent analysis highlights that those in larger groups—popular individuals—engage in more conspicuous consumption due to heightened competition, and an increase in status concerns causes the social graph to fragment into smaller cliques.

This point underscores a third 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 share information bilaterally in addition to their existing team relationships), there are natural conditions under which cliques are still the predicted outcome.

Our analysis generates new insights on strategic network formation while simultaneously unifying and organizing existing knowledge. In our applications, we emphasize cells in our table that do not appear in earlier work—to the best of our knowledge, ordered cliques are a novel prediction in the literature on strategic network formation. Though nested split graphs arise in existing models, we identify the essential payoff properties that produce these structures. In both cases, the predicted structures are clearly much more rigid than what we observe in real networks. Following our applications, we discuss two ways to accommodate more complex networks before concluding with brief remarks.

1.1 Related Work

Our analysis sits at the intersection of two strands of work in network theory: network games and strategic network formation. Within the network games literature, 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. Sadler [2020] explores the robustness of equilibrium characterizations based on centrality measures. Our analysis is closest in spirit to this 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].[3] 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. This contrasts with 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. Motivated by observed patterns

[3] Hellmann [2013] shows that pairwise stable graphs exist if payoffs are convex in own links, and others' links are complements to own links. These conditions are jointly equivalent to our definition of quasi-convexity in links (see Definition 4).

in interbank lending and trade networks, the authors seek to explain the prevalence of hierarchical, nested structures. The underlying incentives satisfy positive spillovers and link-action complements, and accordingly the stochastically stable outcomes are nested split graphs.

Turning to our applications and illustrations, we note several other connections. Most obviously, we highlight how our results can explain counter-intuitive findings from studies on peer effects [Carrell et al., 2013] and derive new insights on the effects of status competitions Immorlica et al. [2017]. For two of the cells in Table 1, our results state that stable structures consist of ordered cliques, and the members of a clique share similar attributes. In some cases, these cliques are disjoint. One can view this result as providing a microfoundation for group matching models. In these models, players choose what group to join rather than what links to form, so it is assumed ex ante that the graph is a collection of disjoint cliques. For instance, Baccara and Yariv [2013] study a model in which players choose to join in groups before investing in public projects, finding that stable groups exhibit homophily.[4] Our analysis extends this finding, and one can use our results to find conditions under which the group matching assumption is without loss of generality.

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.[5]
- 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

[4] In other related work, Bandyopadhyay and Cabrales [2020] study pricing for group membership in a similar setting, and Chade and Eeckhout [2018] study the allocation of experts to disjoint teams.

[5] We 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 $\{i, j\}$.

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.

- The action profile $\mathbf{s}$ 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 $u_j(G + ij, \mathbf{s}) \geq u_j(G, \mathbf{s})$ with at least one strict inequality.

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

- The action profile $\mathbf{s}$ 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 $u_i\left(G - \{ik : k \in H\}, \mathbf{s}\right) > u_i(G, \mathbf{s})$.
- There is no link $ij \notin G$ such that both $u_i(G + ij, \mathbf{s}) \geq u_i(G, \mathbf{s})$ and $u_j(G + ij, \mathbf{s}) \geq u_j(G, \mathbf{s})$ with at least one strict inequality.

Both of these solution concepts reflect that link formation requires mutual consent. An outcome is pairwise stable if $\mathbf{s}$ 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. Whenever a player considers a link deviation, she takes the action profile $\mathbf{s}$ 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.[6]

Note that standard models of network games and strategic network formation are nested in our framework. To represent a network game on a fixed graph G, one can take the utility function from the network game and add terms so it is strictly optimal for all players to include exactly the links in G. Pairwise stable outcomes in the corresponding network game with network formation correspond to Nash equilibria in the original network game. To represent a model of network formation, simply impose that each S_i is a singleton.

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.

[6]Of course, it may also create issues for existence.

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

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

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 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

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

with b_i and c strictly positive.[7] 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.[8]

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

[7] This 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.

[8] One 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.

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 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 a definition.

Definition 2. Fixing a strategy profile $\mathbf{s}$ and two networks G and G' that differ in exactly one edge, with $G' = G + ij$. We say G' **dominates** G given $\mathbf{s}$ if both $u_i(G', \mathbf{s}) \geq u_i(G, \mathbf{s})$ and $u_j(G', \mathbf{s}) \geq u_j(G, \mathbf{s})$, with at least one strict inequality. Conversely, we say G dominates G' given $\mathbf{s}$ if either $u_i(G, \mathbf{s}) > u_i(G', \mathbf{s})$ or $u_j(G, \mathbf{s}) > u_j(G', \mathbf{s})$.

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' dominates G given $\mathbf{s}$.

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 $\mathbf{s}$—if a link is removed, the potential increases as long as at least one player is made strictly better off, and if a link is added, the potential increases as long as both players are weakly better off and one is strictly better off. 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 link $ij \in G$ such that one of i or j strictly benefits from its removal, and there is no link $ij \notin G$ such that i and j both benefit from its addition, with one strictly benefiting. We conclude that $(G, \mathbf{s})$ is pairwise stable. □

To illustrate the application of this result, 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, Proposition 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 complete 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 t* if, whenever $x' \geq x$ and $t' \geq t$, it satisfies

$$f(x', t) \geq (>)\, f(x, t) \quad \Longrightarrow \quad f(x', t') \geq (>)\, f(x, t').$$

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

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

Definition 3. 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 3 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 definitions introduce monotonicity conditions on the marginal benefit from adding a link.

Definition 4. A network game with network formation has **positive spillovers** if for each edge $ij \notin G$, any $s'_j \geq s_j$, and any s_{-j}, we have

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

with strict inequality if $s'_j > s_j$. The game has **weakly positive spillovers** if

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

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

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

A network game with network formation has **(weakly) negative spillovers** if the game with all u_i replaced by $-u_i$ has (resp., weakly) positive spillovers.

A network game with network formation is **convex** in links if for each edge $ij \notin G' \supset G$ and any $\mathbf{s}$, we have

$$u_i(G + ij, \mathbf{s}) - u_i(G, \mathbf{s}) \leq u_i(G' + ij, \mathbf{s}) - u_i(G', \mathbf{s}).$$

The game is **quasi-convex** in links if

$$u_i(G + ij, \mathbf{s}) \geq u_i(G, \mathbf{s}) \quad \Longrightarrow u_i(G' + ij, \mathbf{s}) - u_i(G', \mathbf{s}).$$

A network game with network formation is **(quasi-)concave** in links if the game with all u_i replaced by $-u_i$ is (resp., quasi-)convex in links.

Positive spillovers means that, all else equal, if player j takes a higher action, she becomes a more attractive neighbor. Convexity in links means that adding other links to the graph can only increase the benefit from forming a link with j. Negative spillovers reverses the first statement—higher actions make someone a *less* attractive neighbor—and concavity in links reverses the second—additional links in the graph weakly *decrease* the benefit of linking. Analogous to the distinction between increasing differences and the single crossing property, weak positive/negative spillovers, and quasi-convexity/concavity in links, are corresponding ordinal conditions.

Proposition 2. If a network game with network formation exhibits strategic complements, has weakly positive spillovers, and is quasi-convex in links, then there exist minimal and maximal pairwise stable outcomes. Moreover, the minimal pairwise stable outcome is pairwise Nash stable.[9]

[9] If in the definition of strategic complements one replaces single crossing in s_i and G with single crossing in s_i and $-G$, so best responses *decrease* as links are added, then one can obtain an analogous existence result assuming weakly negative spillovers and quasi-convexity in links.

Proof. We carry out the existence 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 $k \geq 1$, we take $ij \in G^{(k)}$ if and only if we have both $u_i(G^{(k-1)} + ij, \mathbf{s}^{(k-1)}) \geq u_i(G^{(k-1)} - ij, \mathbf{s}^{(k-1)})$ and $u_j(G^{(k-1)} + ij, \mathbf{s}^{(k-1)}) \geq u_j(G^{(k-1)} - ij, \mathbf{s}^{(k-1)})$, 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. To see that the minimal graph is in fact pairwise Nash stable, note that for any player i, if we remove any subset of links $S \subseteq G_i$ one by one, reversing order in which they were added above, quasi-convexity in links implies that each removal decreases i's utility, so removing the links in S decreases i's utility.

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.

3.3 Selection via improvement paths

The notion of an improvement path can help us select among different pairwise stable outcomes.

Definition 5. 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,

$$u_i(G^{(k+1)}, \mathbf{s}^{(k)}) \geq u_i(G^{(k)}, \mathbf{s}^{(k)}) \quad \text{and} \quad u_j(G^{(k+1)}, \mathbf{s}^{(k)}) \geq u_j(G^{(k)}, \mathbf{s}^{(k)}),$$

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, a network improvement path is *maximal* (i.e., it cannot be further extended) if and only if its last two entries constitute a pairwise stable outcome.

This definition suggests a selection criterion when 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.[10] 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 Propositions 1 and 2,

[10]Given 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.

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

4.1 Separability

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 statistics of the two players on either end of the link. Throughout this section, we assume that each action set S_i is partially ordered with order $\geq$.

Definition 7. Let $\mathcal{G}_{-i}$ denote the set of all graphs with vertex set $N \setminus \{i\}$, and given a graph G, write G_{-i} for the subgraph of G with vertex i removed. A **statistic** for player i is a function $h_i : \mathcal{G}_{-i} \times S_i \to \mathbb{R}$ that is strictly increasing in s_i. A network game with network formation is **separable** if there exist statistics $\{h_i\}_{i \in N}$ for each player and a function $g : \mathbb{R}^2 \to \mathbb{R}$ such that

$$u_i(G + ij, \mathbf{s}) - u_i(G, \mathbf{s}) = g\left(h_i(s_i, G_{-i}), h_j(s_j, G_{-j})\right)$$

for all players i and j and all graphs G with $ij \notin G$. The game is **strongly separable** if the corresponding statistics do not depend on the graph G.

We say that **actions and links are complements** if, whenever $h_i' > h_i$, we have

$$g(h_i, h_j) \geq 0 \implies g(h_i', h_j) > 0.$$

We say that **actions and links are substitutes** if, whenever $h_i' > h_i$, we have

$$g(h_i, h_j) \leq 0 \implies g(h_i', h_j) < 0.$$

We note additionally that, applying Definition 4, a separable game has positive (negative) spillovers if g is increasing (decreasing) in its second argument.

Since the statistics $\{h_i\}_{i \in N}$ are player specific, Definition 7 allows linking incentives to depend on essentially any idiosyncratic attributes of the players as long as we can reduce them to a one-dimensional summary statistic. Moreover, separability is compatible with any underlying network game holding the graph fixed—the game can exhibit strategic complements or substitutes (or neither), and players can have arbitrary private incentives to take high or low actions. The main substantive restrictions are that i) the incentive for i and j to form a link cannot depend directly on a third player k's action or attributes,[11] and ii) factors that influence linking incentives can be reduced to a one-dimensional scale.

[11] To the extent that player k's action or attributes affect player i's action in equilibrium, it can have an indirect effect on i's incentives to link to j.

A pair of examples helps illustrate the kinds of incentives that separable games can capture. In the simplest case, the incentive to form a link depends only on the two players' actions. Suppose payoffs take the form

$$u_i(G, \mathbf{s}) = v_i(s_i) + \sum_{j \in G_i} w(s_i, s_j) - cd_i,$$

in which $d_i = |G_i|$ is player i's degree. Here, the function w describes a link benefit that depends on the two players' actions, and linking incurs some fixed cost c. This game is separable with index $h_i(G_{-i}, s_i) = s_i$ and $g(h_i, h_j) = w(h_i, h_j) - c$. The public goods game from Section 2.2 takes this form, and the complementary-effort game is similarly separable if the cost function $c(d_i, s_i)$ in equation (1) is linear in d_i, holding s_i fixed.

As a second example, suppose actions are real-valued, each player i has a type $t_i \in [0, 1]$, and payoffs take the form

$$u_i(G, \mathbf{s}) = v_i(s_i) + \sum_{j \in G_i} w(s_i + t_i, s_j + t_i)$$

Here, the value of a link between i and j depends on a combination of the two players' actions and idiosyncratic types, and the game is separable with index $h_i(G_{-i}, s_i) = s_i + t_i$ and $g(h_i, h_j) = w(h_i, h_j)$. Using this functional form, we can capture situations in which higher types are more attractive partners—take w increasing in the second argument—situations in which higher types seek out more partners—take w increasing in the first argument—or even situations in which players seek out partners with similar types—take w decreasing in the distance between its arguments.

Both of these examples are in fact strongly separable as player statistics do not depend on the graph. While statistics in general can vary with the network, note that the dependence of player i's statistic on G is limited to the subgraph G_{-i} induced by i's removal. In assessing the value of a link ij, the value of player i's statistic cannot depend on the presence or absence of link ij. Nevertheless, a link between i and k can still indirectly affect the value of link ij through its effect on i's equilibrium action.

In a strongly separable game, we can strengthen our two existence results because the value of a link depends only on players' actions, not on other links present in the graph: under the conditions of Propositions 1 or 2, a pairwise Nash stable outcome exists.

Proposition 3. If a network game with network formation is strongly 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 strong separability because the marginal gain from deleting a link does not depend on the presence or absence of other links. □

4.2 Main result

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 with statistics $\{h_i\}_{i \in N}$, 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 $h_i(G_{-i}, s_i) > h_j(G_{-j}, s_j)$ implies $G_j \subseteq G_i \cup \{i\}$.

(b) If the game has weakly positive spillovers, and links and actions are substitutes, then if $h_i(G_{-i}, s_i) > h_j(G_{-j}, s_j) > h_k(G_{-k}, s_k)$ 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 $h_i(G_{-i}, s_i) > h_j(G_{-j}, s_j) > h_k(G_{-k}, s_k)$ and $ik \in G$, then $ij, jk \in G$.

(d) If the game has weakly negative spillovers and links and actions are substitutes, then $h_i(G_{-i}, s_i) > h_j(G_{-j}, s_j)$ implies $G_i \subseteq G_j \cup \{j\}$.

Proof. To simplify notation, throughout this proof we suppress the dependence of the statistic h_i on G_{-i} and s_i. We begin with part (a). Suppose $jk \in G$ and $h_i > h_j$. Since $g(h_k, h_j) \geq 0$, weakly positive spillovers imply that $g(h_k, h_i) > 0$, and since $g(h_j, h_k) \geq 0$, action and link complements imply $g(h_i, h_k) > 0$. 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 $g(h_i, h_k) \geq 0$. We make two observations from this. First, since $h_i > h_j$, action-link substitutes imply $g(h_j, h_k) > 0$. Since $h_i > h_k$, weakly positive spillovers now imply that $g(h_j, h_i) > 0$. Second, since $h_j > h_k$, weakly positive spillovers imply $g(h_i, h_j) > 0$. Since $h_i > h_k$, action-link substitutes now imply that $g(h_k, h_j) > 0$. 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 $g(h_k, h_i) \geq 0$. Since $h_j > h_k$, action-link complements imply $g(h_j, h_i) > 0$, and since $h_i > h_k$, we can now infer $g(h_j, h_k) > 0$ from weakly negative spillovers. Moreover, since $h_i > h_j$ and $g(h_k, h_i) \geq 0$, weakly negative spillovers imply that $g(h_k, h_j) > 0$. Action-link complements now imply $g(h_i, h_j) > 0$. 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 $h_i > h_j$, then since $g(h_k, h_i) \geq 0$, weakly negative spillovers imply $g(h_k, h_j) > 0$, and since $g(h_i, h_k) \geq 0$, action-link substitutes imply $g(h_j, h_k) > 0$. We conclude that $jk \in G$. $\square$

The characterization in Theorem 1 is stark. In cases (a) and (d), if one player has a higher statistic than another, then the two neighborhoods are ordered by set inclusion. In cases (b) and (c), a link between two players, one with a higher statistic than the other, implies that the set of players with intermediate statistics 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 we assume that $(G, \mathbf{s})$ is strictly pairwise stable in the hypothesis of Theorem 1, then all statements remain true if we replace strict inequalities with weak inequalities (i.e., assume throughout only that $h_i \geq h_j \geq h_k$).[12] This in turn implies that stable graphs have particular neighborhood structure. To state our result, we first define two classes of graphs.

Definition 9. Given a graph G, let $\mathcal{D} = (D_0, D_1, ..., D_k)$ denote its degree partition—players are grouped according to their degrees, and those in the (possibly empty) element D_0 have degree 0. The graph G with degree partition $\mathcal{D}$ is a **nested split graph** if for each $\ell = 1, 2, ..., k$ and each $i \in D_\ell$ we have

$$G_i = \begin{cases} \bigcup_{j=1}^{\ell} D_{k+1-j} & \text{if } \ell \leq \lfloor \frac{k}{2} \rfloor \\ \bigcup_{j=1}^{\ell} D_{k+1-j} \setminus \{i\} & \text{if } \ell > \lfloor \frac{k}{2} \rfloor. \end{cases}$$

In particular, if $d_i \leq d_j$, then $G_i \subset G_j \cup \{j\}$.

A graph G consists of **ordered overlapping cliques** if we can order the players $\{1, 2, ..., n\}$ such that $G_i \cup \{i\}$ is an interval for each i, and the endpoints of this interval are weakly increasing in i.

In a nested split graph, players' neighborhoods are totally ordered through set inclusion, resulting in a strong hierarchical structure. In a graph with ordered overlapping cliques, the order on the players induces an order on the set of maximal cliques. Each maximal clique consists of an interval of players, and both end points of these cliques are strictly increasing. Any graph in which every component is a clique is special case of this structure. As a corollary of Theorem 1, we note that stable graphs have one of these two structures.

Corollary 1. Suppose a network game with network formation is separable with statistics $\{h_i\}_{i \in N}$, and $(G, \mathbf{s})$ is a strictly pairwise stable outcome. Then:

(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 $h_i(G_{-i}, s_i) > h_j(G_{-j}, s_j)$.

(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 $h_i(G_{-i}, s_i) > h_j(G_{-j}, s_j)$.

[12] Analogous 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.

(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 G consists of ordered overlapping cliques with players ordered according to their statistics.

Proof. Consider claim (a) first. Part (a) of Theorem 1 immediately implies that if $|G_i| \geq |G_j|$, it must be that $h_i \geq h_j$, and $G_j \subset G_i \cup \{i\}$. Hence, players with the same degree share the same neighbors, and for players with different degrees, neighborhoods are ordered by set inclusion. Part (a) of the corollary follows, and claim (b) holds by a similar argument.

Moving to claim (c), parts (b) and (c) of Theorem 1 imply that if we order players according to their statistics, neighborhoods are intervals—if i is linked to any j with $h_j < h_i$, then i is linked to all k with $h_j \leq h_k \leq h_i$, and similarly if i is linked to j with $h_j > h_i$, then i is linked to all k with $h_j \geq h_k \geq h_i$. The endpoints are necessarily increasing in the order because the Theorem also implies that player k is linked to player j in each case. The result follows. □

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 organized into overlapping cliques of players. In many natural examples, the statistics are simply players' actions, and in this case players in a clique necessarily take similar actions.

Implicit in this result is a novel characterization of structures that arise in more standard network formation games. If action sets are singletons, then the assumptions of positive or negative spillovers, and action-link complements or substitutes, translate into assumptions about idiosyncratic attributes of the players. Imagine that each player has a one-dimensional type. Positive spillovers means that, all else equal, higher types are more attractive neighbors, while negative spillovers means that higher types are less attractive neighbors. Similarly, link-action complements means that higher types desire more links, while link-action substitutes means that higher types desire fewer links. Work on strategic network formation has thus far struggled to obtain general results on the structure of pairwise stable graphs, and Corollary 1 highlights non-trivial conditions that yield sharp predictions.

5 Discussion

We now discuss several applications of our results.

5.1 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 inform the assignment of new cadets to squadrons. Based on the first wave of randomly assigned squadrons, the authors concluded that being in a squadron with higher performing peers[13] led to better academic performance among less prepared cadets. In the second wave, incoming cadets with less preparation were systematically placed in squadrons with more high ability peers. While 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 cadets 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. Player i's net benefit from adding a link to player j is

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

which satisfies the single crossing property in Definition 7. As s_i increases, this eventually turns negative—those who invest a lot of effort find linking too costly. 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

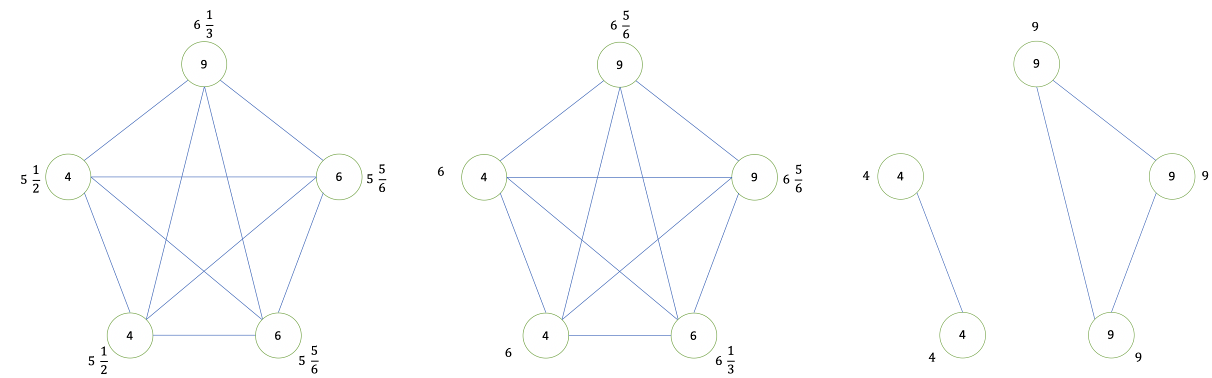
$$s_i = \frac{1}{d_i + 1} \left(b_i + \alpha \sum_{j \in G_i} s_j \right)$$

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

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

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

[13] Specifically, those entering with relatively high SAT verbal scores.



(a) Squadron 1 (b) Squadron 2 (c) Squadron 3

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.

Now assume that $\alpha = 1$, and that 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[14] 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:

- 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—low ability cadets benefit from this small change in group composition.

What happens when we add another high ability cadet? In squadron 3, the unique coordination-free outcome (one reachable by adding links from an empty graph) 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

[14] We use "ability" as a shorthand for aptitude and preparation.

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 collaboration may restore the beneficial relationship we saw in the shift from squadron 1 to squadron 2.

5.2 Status games and ordered cliques

Competitions for status feature action-link complements and negative spillovers. For instance, conspicuous consumption among friends tends to increase one's own conspicuous consumption in response, but those who flaunt expensive possessions are less attractive as friends. Jackson [2019] argues that many social behaviors (e.g., binge drinking) have the same properties: those with more friends find these behaviors more rewarding, but they exert negative health externalities across neighbors. More generally, this pattern applies to any domain in which friends' achievement drives one to excel, but there is disutility from negative comparisons among friends. Our theory suggests that these interactions drive the formation of social cliques ordered according to their relative status.

This prediction agrees with anthropological and sociological studies documenting the pervasiveness of ranked cliques. For instance, 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.[15] Adler and Adler [1995] conduct an ethnographic study of older elementary-school children that highlights the prevalence of cliques. The authors argue that status differentiation is clear across cliques, and there are unambiguous orderings, 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 negative externalities arising from dominance contests within cliques, consistent with our negative spillovers assumption. Building on this ethnographic work, Gest et al. [2007] carry out a detailed quantitative examination of the social structures in a middle school, with a particular focus on gender differences. The authors' summary confirms the ethnographic narrative: "girls and boys were similar in their tendency to form same-sex peer groups that were distinct, tightly knit, and characterized by status hierarchies."

Within the economics literature, Immorlica et al. [2017] introduce a framework in which players exert inefficient effort in a status-seeking activity and earn disutility from network neighbors who exert higher effort—we can view this as a model of conspicuous consumption with upward-looking comparisons. The authors assume an exogenous network and explore how the network structure influences individual behavior. Formally, the authors take $S_i = \mathbb{R}_+$ for each player i, and payoffs are

$$u_i(\mathbf{s}) = b_i s_i - \frac{s_i^2}{2} - \sum_{j \in G_i} g_{ij} \max\{s_j - s_i, 0\},$$

[15] Davis 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.

in which $g_{ij} \geq 0$ for each $ij \in G$. The paper shows that an equilibrium partitions the players into classes making the same level of effort, and the highest class consists of the subset of players that maximizes a measure of group cohesion. Our framework makes it possible to endogenize the network in this model—under a natural extension of the payoff function, the classes that emerge in equilibrium form distinct cliques in the social graph.

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

$$u_i(G, \mathbf{s}) = bs_i - \frac{s_i^2}{2} + \sum_{j \in G_i} \left(1 - \delta \max\{s_j - s_i, 0\}\right).$$

In this game, player i earns a unit of utility for each neighbor, but suffers a loss $s_j - s_i$ if neighbor j invests more effort. To highlight the role of network formation, rather than individual incentives, we also specialize the model so that all players have the same private benefit b for effort. The game is clearly separable with negative spillovers, and links and actions are (weak) complements. Hence, stable outcomes consist of ordered overlapping cliques, and we can only have $ij \in G$ if $|s_i - s_j| \leq \frac{1}{\delta}$. For the purposes of this example, we restrict attention to outcomes in which the cliques partition the players. Moreover, following Immorlica et al. [2017], we focus on maximal equilibria of the status game, with players taking the highest actions they can sustain given the graph. Since all players have the same private benefit b, all players in a clique play the same action, and the maximal action in a clique of size k is $b + (k-1)\delta$.

Two features of stable outcomes stand out. First, those in large groups take higher actions—popular individuals invest more in status signaling. Second, as status concerns increase, the graph can fragment. Let c^* denote the smallest integer such that $c^*\delta \geq \frac{1}{\delta}$—this is the unique integer satisfying $\delta \in [1/\sqrt{c^*}, 1/\sqrt{c^*-1})$. If i and j are in different cliques, we must have $|s_i - s_j| \geq \frac{1}{\delta}$, which implies the cliques differ in size by at least c^*. The larger c^* is, the more cohesive stable networks are. If there are n players in total, and $\delta < \frac{1}{\sqrt{n-2}}$, then the complete graph is the only stable outcome. As δ increases, meaning there are greater status concerns, then stable outcomes can involve more fragmented graphs. If $\delta \geq 1$, then separate cliques need only differ in size by one player, and the maximal number of cliques is the largest integer k such that $\frac{k(k+1)}{2} \leq n$. This simple example highlights how our framework enables meaningful study of comparative statics for stable networks.

5.3 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—crucially, payoffs are invariant to permutations of agents within groups. In essence, these models constrain the network that can form, assuming disjoint cliques. For example, Baccara and Yariv [2013] study a setting in which individuals join groups (e.g., social clubs) and then choose how much to contribute to an activity within the group. These contributions affect the payoffs of other group members symmetrically.

Similarly, Chade and Eeckhout [2018] model the allocation of experts to teams. These experts share information within their teams, benefiting all team members equally, but not across teams.

The interactions motivating these models are not so constrained in reality—there is no reason why pairs cannot meet outside the groups, and in many cases a person could choose to join multiple groups. However, assuming that interactions happen in groups allows simplifications that are essential to the tractability of these models. To what extent are these restrictions without loss of generality? Our results allow us to provide simple sufficient conditions.

For this section, we assume all players have the same action set S, which is a closed interval in $\mathbb{R}$, and each player has one of finitely many types—write $t_i \in T$ for player i's type. The types capture all heterogeneity in payoffs across players, and utility is continuous in the action profile. Moreover, we assume that payoffs are strongly separable and exhibit a weak preference for conformity—this means that optimal actions result in player statistics that lie between some type-specific benchmark and neighbors' statistics.

More formally, a graph $G = (N, E)$ is *type-isomorphic* to another graph $G' = (N, E')$ if there exists a bijection $\pi : N \to N$ such that $t_i = t_{\pi(i)}$ and $ij \in E$ if and only if $\pi(i)\pi(j) \in E'$. We have

$$u_i\left(G, \pi(\mathbf{s})\right) = u_{\pi(i)}\left(G', \mathbf{s}\right)$$

for all such bijections π, all such graphs G and G', and all players i. We further suppose that players have unique best responses, holding the graph and other players' actions fixed, so there exists a unique action s_t^* that a type t player would take if isolated with no neighbors. This is the *privately optimal action.* Since the game is strongly separable we can write the statistic for each player i as $h(s_i, t_i)$ for some function $h : S \times T \to \mathbb{R}$. We write $h_t^* = h(s_t^*, t)$ for the statistic corresponding to a type t player's privately optimal action. Payoffs exhibit a *weak preference for conformity* if player i's statistic when playing a best response always lies somewhere in between her privately optimal benchmark and the statistics of her neighbors. That is, for $\hat{s} = \arg\max_{s_i \in S} u_i(G, s_i, s_{-i})$, we have

$$h(\hat{s}, t_i) \in \left[\min\{h_{t_i}^*, \min_{j \in G_i}\{h(s_j, t_j)\}\}, \max\{h_{t_i}^*, \max_{j \in G_i}\{h(s_j, t_j)\}\}\right]$$

for all i and G.

We say that types form *natural cliques* if there exists a partition $\{T_1, T_2, ..., T_K\}$ of T such that

- $g\left(h_t^*, h_{t'}^*\right) \geq 0$ for any $t, t' \in T_k$ and any k.
- $g\left(h_t^*, h_{t'}^*\right) \leq 0$ for any $t \in T_k$ and $t' \in T_\ell$ with $k \neq \ell$.

In words, this means that if all players were to choose their privately optimal actions, and form the network taking those actions as given, then disjoint cliques based on the partition of types would be pairwise stable. If payoffs exhibit a weak preference for conformity, these same cliques remain pairwise stable when players can change their actions.

Proposition 4. Under the above assumptions, suppose the game exhibits either positive spillovers and link-action substitutes or negative spillovers and link-action complements. If types form natural cliques, then there exists a pairwise stable outcome in which the players are partitioned into disjoint cliques.

Proof. We carry out the proof assuming positive spillovers and link-action substitutes—the other case is analogous. Since types form natural cliques, there is a partition $\{T_1, T_2, ..., T_K\}$ of types such that, when playing the privately optimal actions, players have an incentive to link if and only if their types are in the same element of the partition. Suppose this graph forms. We show it is part of a pairwise stable outcome.

For each T_k let $\underline{h}_k$ and $\overline{h}_k$ denote the lowest and highest values respectively of h_t^* for some type $t \in T_k$. Weak preference for conformity implies that there exists an equilibrium in actions in which $h_i \in [\underline{h}_k, \overline{h}_k]$ for every player i with type $t_i \in T_k$. Given two such players i and j, we have

$$g(h_i, h_j) \geq g(h_i, \underline{h}_k) \geq g(\overline{h}_k, \underline{h}_k) \geq 0,$$

in which the first inequality follows from positive spillovers, and the second follows from link-action substitutes. Hence, these two players have an incentive to link.

For two partition elements T_k and T_ℓ, with $k \neq \ell$, assume without loss of generality that $\underline{h}_\ell \geq \overline{h}_k$. For player i with type $t_i \in T_k$ and j with type $t_j \in T_\ell$ we have

$$g(h_i, h_j) \leq g(\underline{h}_k, h_j) \leq g(\underline{h}_k, \overline{h}_\ell) \leq 0, \text{ and}$$

$$g(h_j, h_i) \leq g(\underline{h}_\ell, h_i) \leq g(\underline{h}_\ell, \overline{h}_k) \leq 0,$$

so the players have no incentive to link. □

Under natural assumptions, stable networks preserve natural cleavages between identifiable types of individuals, and players endogenously organize themselves into disjoint cliques as assumed in group matching models. Even if the natural cleavages are not so stark, our results show that much of the simplifying structure remains: individuals can be part of multiple groups, but each group is a clique, and there is a clear ordering among the cliques. Imposing this slightly weaker assumption in models of group matching may allow for richer analysis while preserving the tractability in existing models.

5.4 Complex Network Structure

Our predictions about the structure of stable networks are stark. Real networks are typically not organized precisely into ordered cliques, nor are neighborhoods perfectly ordered via set inclusion. Nevertheless, our results provide a starting point to better understand how incentives affect the complex structures we observe in real networks. There are at least two natural directions to extend our analysis. One is to layer different relationships on top of one another in a "multiplex" network—stark patterns across different layers can combine to form more realistic arrangements. A second is to introduce noise.

Consider a simple example with two 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. Both entail positive spillovers, but Work exhibits action-link substitutes—forming friendships takes time that could be devoted to production—while church exhibits action-link complements—attendance makes it easier to form ties. Assuming suitable heterogeneity in ability or preferences, a non-trivial network will form through each activity. In the work network, we get ordered cliques. In the church network, we get 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.

König et al. [2014] demonstrate the second approach, describing a dynamic process in which agents either add or delete one link at a time, and the underlying incentives exhibit positive spillovers and link-action complements. If agents always make the myopically optimal link change, the graph is a nested split graph at every step of the process. However, if agents sometimes make sub-optimal changes, then all graphs appear with positive probability, but the distribution is still heavily skewed towards those with a nested structure. This allows the authors to fit the model to real-world data. Based on our analysis, one could adapt this model to study peer effects or status games, obtaining a noisy version of our ordered cliques prediction.

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 side 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 that emerge—an explanation of cliquish behavior, a reason why we should expect certain efforts at group design to fail. A promising direction for future work is to examine how combining different types of relationships (multiplexing), or allowing small amounts of noise in linking decisions, can give rise to more complex and realistic network structures.

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.

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.