Games on Endogenous Networks*

Evan Sadler[†] Benjamin Golub[‡]

September 12, 2022

Abstract

We study network games in which players choose both the partners with whom they associate and an action level (e.g., effort) that creates spillovers for those partners. We introduce a framework and two solution concepts, extending standard approaches for analyzing each choice in isolation: Nash equilibrium in actions and pairwise stability in links. Our main results show that, under suitable order conditions on incentives, stable networks take simple forms. The first condition concerns whether links create positive or negative payoff spillovers. The second concerns whether actions are strategic complements to links, or strategic substitutes. Together, these conditions yield a taxonomy of the relationship between network structure and economic primitives organized around two network architectures: *ordered overlapping cliques* and *nested split graphs*. 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 contacts influence people's behavior, and that behavior in turn affects the connections they form. For instance, a good study partner might lead a student to exert more effort in school, and that student's increased effort may incentivize the partner to maintain the collaboration. Understanding how networks and actions mutually influence one another is crucial for policy design. Nevertheless, a general theoretical framework for such situations is missing.

While many studies analyze peer effects assuming an exogenously fixed social network, an important paper by Carrell et al. (2013) shows this can lead to mistaken predictions and

*We are grateful to Krishna Dasaratha for a very helpful discussion. We are also grateful to Drew Fudenberg, Kevin He, Navin Kartik, Mihai Manea, Eran Shmaya, and seminar participants for useful comments. Golub gratefully acknowledges support from the National Science Foundation under grant SES-1629446. Zoë Hitzig and Rithvik Rao provided excellent research assistance.

†Columbia University, es3668@columbia.edu

‡Northwestern University, bgolub@northwestern.edu

interventions that backfire. The authors first estimated academic peer effects among cadets at the U.S. Air Force Academy and subsequently designed peer groups to improve the performance of academically weaker cadets. Extrapolating from peer effects estimates based on randomly assigned first-year peer groups, the authors expected that low-skilled freshmen would benefit from being placed in groups with higher proportions of high-skilled freshmen. Instead, the intervention produced a comparably sized *negative* effect. The authors interpret this as a consequence of endogenous friendship and collaboration networks *within* administratively assigned groups. Friendships between low- and high-skilled freshmen were much less likely to form in the designed groups than in the randomly assigned ones. Therefore, the positive spillovers that the design sought to maximize failed to happen. To account for such effects, researchers need models that permit a simultaneous analysis of network formation and peer effects.

We introduce a framework to study games with network spillovers together with strategic link formation. Our theoretical contribution is twofold. First, we propose a model that nests standard analyses of each type of interaction on its own. Players choose actions (e.g., effort levels) from an ordered set, and they also have preferences over links given actions. We adapt the definitions of Nash equilibrium (for actions) and pairwise stability (for network formation) to define our solution concepts. Intuitively, in a solution to a network game with network formation, players should have an incentive to change neither their actions[1] nor their links. More precisely, our notion of a stable outcome requires that no player benefits from changing her action, holding the network fixed, nor from unilaterally removing links, and no pair of players can jointly benefit from creating a link between them.[2]

Second, we identify payoff properties under which stable networks have simple structures. Here, we first focus on a large class of separable network games that nests essentially all prior models of consensual network formation and action choice. We obtain sharp characterizations of outcomes using two kinds of order conditions. 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.[3] The second type of condition concerns the relationship between action incentives and links. The game exhibits *action–link complements* if the returns from taking higher actions increase with one's degree[4] in the network. The game exhibits *action–link substitutes* if these returns decrease with one's degree. Our main result characterizes the structure of both actions and links for any combination of order conditions (one of each type). Table 1 summarizes our findings. The result demonstrates that these conditions provide a useful way to organize our understanding of games on endogenous networks.

Figure 1 illustrates the two types of graphs our model predicts. Nested split graphs are

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

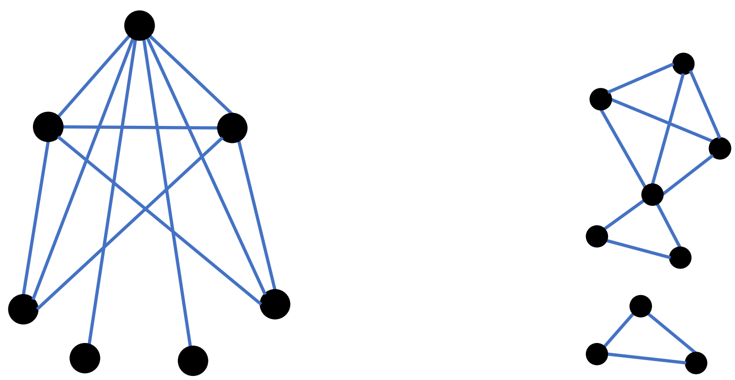
[2]Mirroring standard definitions in network formation, we consider both pairwise stability, in which players may drop only one link at a time, and pairwise Nash stability, in which players may drop many links simultaneously.

[3]Note these spillover properties are distinct from strategic complements or substitutes *in actions*. Our properties concern levels of utility rather than how others' actions affect incentives to take higher actions.

[4]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 results; each cell indicates which network and action configurations are stable under the corresponding pair of assumptions on payoffs.



(a) Nested split graph

(b) Overlapping cliques

Figure 1: Examples of the two types of networks that are characterized by our main result.

strongly hierarchical networks: nodes are partitioned into classes according to their degree, and a higher class's neighborhood is a strict superset of any lower class's neighborhood. Ordered overlapping cliques means that we can order nodes such that every neighborhood is an interval, and the endpoints of a node's neighborhood are increasing in the node's own position in the order—imagine a ranking of nodes, and each is connected to all others' that are close enough in rank.[5] In each case, equilibrium action levels are ordered in a way that corresponds to the network structure. Going beyond the separable games on which we focus, we also identify ordinal properties of linking incentives—*consistency* and *alignment*—that yield this dichotomy in a much larger class of games.

We subsequently specialize our framework to interpret the counterintuitive findings of

[5]For recent econometric work concerning the estimation of related models, see Chandrasekhar and Jackson (2021).

Carrell et al. (2013). We represent their environment as one of positive spillovers combined with action–link substitutes—we assume that students who study together create benefits for their peers, but more time studying makes link formation and maintenance more costly. We identify conditions under which a complete graph is part of a stable outcome and further conditions under which it is uniquely so. The overarching message of these characterizations is that the complete graph becomes harder to sustain as types get more spread out—types in the middle are necessary to induce types at the extremes to link with one another. We then look at a simple example in which students have one of three exogenous ability levels—low, medium, or high.[6] Although replacing just one medium-ability student with a high-ability student benefits the low-ability students in the group, additional replacements—representing the larger changes induced by the deliberate design of Carrell et al. (2013)—can reverse this effect. The reason is that small changes in group composition do not affect the set of stable networks—in our example, the complete network is uniquely stable at first, so all students benefit from each other's 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 of group design disappear. We discuss how the mechanism in our model closely tracks the interpretation that Carrell et al. (2013) give for their results.

To further demonstrate the usefulness of our framework, we briefly study two additional applications. First, we introduce a model of "status games" based on Immorlica et al. (2017). Competitions for status entail a combination of action–link complements and negative spillovers. Imagine, for example, people who compete for social status through investments in conspicuous consumption and also choose their network connections. 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. Moreover, those in larger friend 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.

Our final application provides a microfoundation for "club" or "group matching" models. Theories of endogenous matching for public goods or team production often *assume* that spillovers occur in disjoint cliques, which is critical for tractability. The question is then: which cliques form? We show that even if agents can arrange their interactions into more complex structures if they wish, there are natural conditions under which cliques are still the predicted outcome.

Following our applications, we address several important questions that are tangential to our main analysis. First, we report conditions under which pairwise stable outcomes are guaranteed to exist—while existence is assured in all of our examples, it is not immediate in general because the presence of a link is a discrete outcome, and our solution concept

[6] Interpreted as a mix of aptitude and preparation before their further investment in skills at the academy.

considers joint deviations. Second, as our predicted structures are more rigid that what we observe in real networks, we discuss two ways to accommodate more complex structures. Finally, we provide a microfoundation for our solution concepts based on dynamic adjustment with forward-looking players.

Our analysis generates new insights on strategic network formation while simultaneously unifying and organizing existing work. In our applications, we emphasize predictions not familiar from existing equilibrium models of joint link formation and action choice—in particular, ordered overlapping cliques. Identifying this type of outcome as a robust prediction under certain qualitative assumptions on the strategic interaction constitutes one of our main contributions. Nested split graphs, on the other hand, have received considerable attention in earlier equilibrium models of network formation under specific payoff functions, and are thought to capture key features of inter-bank lending networks and other trading networks (König et al., 2014). By identifying ordinal payoff properties that produce these structures, we help clarify and generalize the conditions under which nested split graphs are likely to emerge, unifying several earlier results.

1.1 Related Work

Our analysis sits at the intersection of two strands of work in network theory: games on fixed networks 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 (2021) explores the robustness of equilibrium characterizations based on centrality measures when payoffs are more general. In the same spirit, our analysis derives predictions from order properties of the payoff functions (jointly with the action space) rather than particular functional forms.

The literature thus far on strategic interactions in endogenous networks is small. Jackson and Watts (2002) observed that combining network formation and action choice could produce different predictions from those arising from either process separately.[7] They demonstrated this in a two-action coordination game with a simple linear linking cost, focusing on stochastically stable play. In many subsequent papers, the decision to form a link is made unilaterally. 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. Similarly, the model of Baetz (2015) entails unilateral link formation together with a linear-quadratic game of strategic complements, leading to strongly hierarchical network structures. One key difference, however, is that decreasing marginal returns to linking cause those at the top of the hierarchy to refrain from linking with one another. In Herskovic and

[7] Similar concerns motivate Cabrales et al. (2011), who study network formation with a single global search investment. We focus on settings in which linking efforts are more precisely directed.

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 again have a hierarchical structure closely related to nested split graphs. Unilateral link formation in these papers contrasts with our model, in which stability is based on mutual consent.

Joshi and Mahmud (2016), Hiller (2017), and Badev (2021) are closer to our approach. Badev (2021) studies a binary action coordination game with endogenous link formation, proposing a solution concept that interpolates between pairwise stability and pairwise Nash stability. This is a parametric model for estimation and simulation procedures in a high-dimensional environment; the goal is to study empirical counterfactuals rather than deriving theoretical results. Hiller (2017) studies a game in which each player chooses a real-valued effort level and simultaneously proposes a set of links—a link forms if and only if both players propose it. The author then refines the set of Nash equilibria by ruling out pairwise deviations in which two players create a link between them and simultaneously adjust their actions. In the underlying game, players have symmetric payoff functions that exhibit strategic complements and positive spillovers. The setting therefore falls into the first cell in our table, and even though the solution concept differs slightly from ours, the resulting outcomes are indeed nested split graphs.[8] Joshi and Mahmud (2016) take a slightly different approach, modeling link proposals followed by action choices in a canonical linear-quadratic game. Their analysis includes both local and global interaction terms, but it still produces nested split graphs in the relevant cells.[9] Relative to this work, we significantly relax parametric and symmetry assumptions on players' payoffs, highlighting more fundamental properties that lead to this structure.

In a related but distinct effort, König et al. (2014) study a dynamic network formation model in which agents choose strategic actions and myopically add and delete links. Motivated by observed patterns in interbank lending and trade networks, the authors seek to explain the prevalence of hierarchical, nested structures. The underlying incentives satisfy positive spillovers and action–link complements, and accordingly the stochastically stable outcomes are nested split graphs.

Several other literatures connect to our applications. Most obviously, we highlight how our results can explain counterintuitive 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

[8] Within the pure network formation literature, without action choice, 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 action–link complements and positive spillovers cell in our table.

[9] Bolletta (2021) offers a more recent example, studying a myopic-adjustment network formation dynamic under a particular linear peer effects specification.

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 clubs (i.e., cliques) before investing in club goods, finding that stable clubs exhibit homophily.[10] 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.[11]
- For each player $i \in N$, there is a set S_i of actions; we write $\mathcal{S} = \prod_{i\in N} S_i$ for the set of all action profiles.
- For each player $i \in N$, there is a payoff function $u_i : \mathcal{G} \times \mathcal{S} \to \mathbb{R}$. This gives player i's payoff as a function of a graph $G \in \mathcal{G}$ and a profile of players' actions.

A pair $(G, \mathbf{s}) \in \mathcal{G} \times \mathcal{S}$ 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 $G - E$ 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 reflecting this idea. These parallel existing concepts in the network formation literature and extend them to our setting with action choices.[12]

Definition 1. An outcome $(G, \mathbf{s})$ is **strictly pairwise stable**[13] if the following conditions hold.

[10]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.

[11]The set N is fixed, so 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 $\{i, j\}$.

[12]See Bloch and Jackson (2006) and Jackson (2008, Chapter 6).

[13]We say the outcome is **pairwise stable** if in the second bullet we replace the weak inequality with a strict inequality (there can exist a link ij such that i is indifferent about deleting it), and in the third bullet we require one of the two inequalities to be strict (two players may both be indifferent about adding the missing link between them).

- 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 $u_i(G - ij, \mathbf{s}) \geq u_i(G, \mathbf{s})$.
- There is no link $ij \notin G$ such that both $u_i(G + ij, \mathbf{s}) \geq u_i(G, \mathbf{s})$ and $u_j(G + ij, \mathbf{s}) \geq u_j(G, \mathbf{s})$.

The outcome $(G, \mathbf{s})$ is **strictly pairwise Nash stable** if additionally there is no pair (s_i', H_i), consisting of an action $s_i' \in S_i$ and a subset of i's neighbors, $H_i \subseteq G_i$, such that $u_i\left(G - \{ik : k \in H_i\}, (s_i', s_{-i})\right) > u_i(G, \mathbf{s})$.

Note that our primary definitions require strict preferences over links. This is technically convenient because it facilitates the cleanest characterization of stable networks, but one could obtain weaker results using analogous solution concepts that permit indifference.

More substantively, both of these solution concepts reflect that link formation requires mutual consent. An outcome is strictly 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 simultaneously changing her action and deleting some subset of her links. Whenever two players consider adding a link between them, they take the action profile $\mathbf{s}$ as given. Section 7.2 discusses this assumption and provides foundations for it.

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 functions 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 make each S_i a singleton (and let the payoffs correspond to those in the network formation model).

2.2 Separable network games

We now introduce some structure on payoffs; the environment we study nests canonical models of network games and permits a simple statement of sufficient conditions for our characterization of network structures.

Suppose all players have the same action set $S_i = S \subseteq \mathbb{R}$, and payoffs take the form

$$u_i(G, \mathbf{s}) = v_i(\mathbf{s}) + \sum_{j \in G_i} g(s_i, s_j) \tag{1}$$

for each player i. Here, the function $v_i : \mathbb{R}^n \to \mathbb{R}$ is idiosyncratic to player i and captures strategic incentives that are independent of the network. These payoffs embed two substantive assumptions about the incremental value of any link to its participants, namely that this value: (i) is invariant to permutations of players' labels (linking incentives are *anonymous*); and (ii) depends only on the actions of the two players involved (linking incentives are *separable*). These properties imply that the direct effects of links on agents' payoffs are determined bilaterally, which is a canonical feature of network games.[14] For formalizations of conditions (i) and (ii) and a statement that these imply the payoff form (1), see Appendix A.1.

These qualitative assumptions nest all models of which we are aware that study consensual link formation and action choice together. For instance, this nests the popular linear-quadratic form introduced by Ballester et al. (2006), as well as elaborations that include linking costs, such as Joshi and Mahmud (2016). Note that g can readily incorporate a cost of linking that depends arbitrarily on player i's own action s_i. Nevertheless, the payoff form (1) is considerably more restrictive than what our results require, and thus the next subsection introduces weaker, purely ordinal assumptions.

Within the class of separable network games, we can state simple order conditions that determine the structure of stable networks, and these cover all of our applications. In a network game with network formation, we can view higher actions and additional links as two different activities in which a player can invest. Given payoffs of the form (1), we say the game exhibits **action-link complements** if g satisfies a single-crossing condition in its first argument:

$$g(s,r) > (\geq)\ 0 \quad\implies\quad g(s',r) > (\geq)\ 0$$

whenever $s' > s$. Actions and links are complements if taking higher actions makes forming additional links more attractive. The game exhibits **action-link substitutes** if the above implication holds whenever $s' < s$.[15] Actions and links are substitutes if taking higher actions makes forming additional links less attractive.

Analogously, we say the game exhibits **positive spillovers** if g satisfies a single-crossing condition in its second argument:

$$g(s,r) > (\geq)\ 0 \quad\implies\quad g(s,r') > (\geq)\ 0$$

whenever $r' > r$, and the game exhibits **negative spillovers** if the above implication holds whenever $r' < r$.[16] If the game exhibits positive spillovers, players who take higher actions, all else equal, are more attractive neighbors. This assumption naturally captures situations in which the action s_i represents effort that benefits neighbors (e.g., studying or gathering information). If the game exhibits negative spillovers, then players who take higher actions are less attractive neighbors.

Remark 1. We emphasize that whether a network game with network formation exhibits action-link complements/substitutes, or positive/negative spillovers, says nothing about

[14] This refers only to the part of the payoff that affects linking incentives, since $v_i(\mathbf{s})$ is completely arbitrary. Importantly, there may be *indirect* effects of other players' actions on a player's payoffs, but these are mediated through the action choices of neighbors.

[15] This definition readily extends to arbitrary network games with network formation—a game exhibits link-action complements if $u_i(G+ij,\mathbf{s})-u_i(G-ij,\mathbf{s}) \geq (>)\ 0$ implies $u_i(G+ij,s_i',s_{-i})-u_i(G-ij,s_i',s_{-i}) \geq (>)\ 0)$ whenever $s_i' > s_i$ and link-action substitutes if the implication holds for $s_i' < s_i$.

[16] Again, these definitions naturally extend to arbitrary games—a game exhibits positive spillovers if $u_i(G+ij,\mathbf{s}) - u_i(G-ij,\mathbf{s}) \geq (>)\ 0$ implies $u_i(G+ij,s_j',s_{-j}) - u_i(G-ij,s_j',s_{-j}) \geq (>)\ 0$ whenever $s_j' > s_j$ and negative spillovers if the implication holds for $s_j' < s_j$.

whether there are strategic complements or substitutes in *actions*: the actions s_i and s_j of two different players i and j could be (strategic) complements, substitutes, or neither. This should be clear as the term $v_i(\mathbf{s})$ in (1), which is independent of the graph, can depend on the entire profile of actions $\mathbf{s}$ in a completely arbitrary way. Moreover, there is no complementarity or substitutability assumed within the function g. Action-link complements/substitutes tells us how taking higher actions affects a given player i's *incentive to form links*. Similarly, positive/negative spillovers tells us how other players' actions affect player i's linking incentives.

2.3 Ordinal assumptions

The single-crossing properties introduced in the previous subsection yield orderings on players capturing their preferences for linking and their desirability as partners. These orderings are ultimately the key inputs allowing us to characterize stable outcomes. Separable payoffs are helpful for interpretation, but they are not necessary to obtain these orderings. This subsection introduces weaker, more primitive order conditions that distill precisely what is needed for our results to hold. Taking one single-crossing condition from each category (action–link complements or substitutes, positive or negative spillovers) implies the ordinal assumptions sufficient for our main characterizations of stable outcomes.

In the following definitions, we write

$$\Delta_{ij} u_i(G, \mathbf{s}) = u_i(G + ij, \mathbf{s}) - u_i(G - ij, \mathbf{s})$$

for the marginal value of link ij to player i at outcome $(G, \mathbf{s})$, and we write

$$S_i^+(G, \mathbf{s}) = \{j \in N \,:\, \Delta_{ij} u_j(G, \mathbf{s}) \geq 0\}$$

for the set of players with a weak incentive to link with i.

Definition 2. Given a network game with network formation, linking incentives are **consistent** if there is no outcome $(G, \mathbf{s})$, and a corresponding collection of players i, j, k, ℓ, such that

$$\Delta_{ik} u_i(G, \mathbf{s}) \geq 0, \quad \Delta_{i\ell} u_i(G, \mathbf{s}) < 0, \quad \text{and}$$
$$\Delta_{jk} u_j(G, \mathbf{s}) < 0, \quad \Delta_{j\ell} u_j(G, \mathbf{s}) \geq 0.$$

In words, linking incentives are consistent if, given any outcome, the players agree on who is a more desirable linking partner. There are no players i, j, k, ℓ such that i wishes to link with k but not ℓ, and j wishes to link with ℓ but not k. Note this does not preclude heterogeneity in preferences over links—far from it—but if some player i wishes to link with k but not ℓ, then any player who wishes to link with ℓ necessarily also wishes to link with k.

Definition 3. Linking incentives are **aligned** if for any outcome $(G, \mathbf{s})$, any three players i, j, k such that

$$S_i^+(G, \mathbf{s}) \subseteq S_j^+(G, \mathbf{s}) \subseteq S_k^+(G, \mathbf{s}),$$

and any fourth player ℓ, we have

$$\Delta_{i\ell} u_i(G, \mathbf{s}) > (\geq)\ 0 \quad \text{and} \quad \Delta_{k\ell} u_k(G, \mathbf{s}) > (\geq)\ 0 \quad \implies \quad \Delta_{j\ell} u_j(G, \mathbf{s}) > (\geq)\ 0, \quad \text{and}$$

$$\Delta_{i\ell} u_i(G, \mathbf{s}) < 0, \quad \text{and} \quad \Delta_{k\ell} u_k(G, \mathbf{s}) < 0 \quad \implies \quad \Delta_{j\ell} u_j(G, \mathbf{s}) < 0.$$

Alignment relates a player's desirability as a linking partner to her own linking incentives. Suppose k is a more desirable neighbor than j, and j is a more desirable neighbor than i, so that j's desirability is between that of i and k. Alignment says that then j's desire for links is also between that of i and k. That is, if both k and i wish to link with ℓ, then j also wishes to link with ℓ, and if both k and i do not wish to link with ℓ, then j also does not wish to link with ℓ.

Together, the two properties of consistency and alignment allow us to measure players' desirability as partners, and their inclination to form links, on the same one-dimensional scale.

Lemma 1. Suppose a network game with network formation has consistent and aligned linking incentives. At any outcome $(G, \mathbf{s})$, there exist weak orders $\succeq_{\text{in}}$ and $\succeq_{\text{out}}$ on the players such that

$$\Delta_{ij} u_i(G, \mathbf{s}) > (\geq)\ 0 \implies \Delta_{ik} u_i(G, \mathbf{s}) > (\geq)\ 0$$

whenever $k \succeq_{\text{in}} j$, and

$$\Delta_{ij} u_j(G, \mathbf{s}) > (\geq)\ 0 \implies \Delta_{ik} u_k(G, \mathbf{s}) > (\geq)\ 0$$

whenever $k \succeq_{\text{out}} j$. That is, if $i \succeq_{\text{in}} j$, then every player who wants to link with j wants to link with i as well, and if $i \succeq_{\text{out}} j$, then i wants to link with every player with whom j wants to link. Moreover, the two orders are either identical or directly opposed.

Proof. See Appendix. □

Consistency and alignment may seem like they introduce a lot of structure, but they are implied by the natural single-crossing conditions introduced in the last subsection. If a game with payoffs of the form (1) exhibits action-link complements or substitutes, and positive or negative spillovers, then linking incentives are necessarily consistent and aligned. Moreover, the orders $\succeq_{\text{in}}$ and $\succeq_{\text{out}}$, expressing players' desirability as neighbors and desire for neighbors, follow the order of the action set S. This result does not require $S \subseteq \mathbb{R}$—the same conclusion holds for any linearly ordered action set.

Lemma 2. Suppose a network game with network formation has payoffs of the form (1). If the game exhibits action-link complements or substitutes, and positive or negative spillovers, then linking incentives are consistent and aligned at any outcome. Moreover:

(a) If there are action-link complements and positive spillovers, then $i \succeq_{\text{in}} j$ and $i \succeq_{\text{out}} j$ in any outcome with $s_i \geq s_j$.

(b) If there are action-link complements and negative spillovers, then $i \preceq_{\text{in}} j$ and $i \succeq_{\text{out}} j$ in any outcome with $s_i \geq s_j$.

(c) If there are action-link substitutes and positive spillovers, then $i \succeq_{\text{in}} j$ and $i \preceq_{\text{out}} j$ in any outcome with $s_i \geq s_j$.

(d) If there are action-link substitutes and negative spillovers, then $i \preceq_{\text{in}} j$ and $i \preceq_{\text{out}} j$ in any outcome with $s_i \geq s_j$.

Proof. By definition, if the game exhibits action-link complements and positive spillovers, then any player i desires more links when s_i is higher and is a more attractive neighbor when s_i is higher. We immediately see from (1) that $i \succeq_{\text{in}} j$ and $i \succeq_{\text{out}} j$ whenever $s_i \geq s_j$, and the result follows. The other three cases are analogous. □

3 The structure of stable graphs

How do properties of the payoff functions $(u_i)_{i \in N}$ affect stable network structures? This section derives our main results on the taxonomy of stable graphs. To state the results, we first define two classes of graphs—recall the illustration in Figure 1.

Definition 4. A graph G is a **nested split graph** if $d_j \geq d_i$ implies that $G_i \subseteq G_j \cup \{j\}$.

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

In a nested split graph, neighborhoods are ordered through set inclusion, resulting in a strong hierarchical structure.[17] In a graph with ordered overlapping cliques, the player order induces an order on the set of maximal cliques. Each maximal clique consists of an interval of players, and both endpoints of these cliques are strictly increasing. Any graph in which every component is a clique is special case of this structure.

[17] See König et al. (2014) and Belhaj et al. (2016) for more on the structure and other economic properties of these networks. A common alternative characterization is the following. Given a graph G, let $\mathcal{D} = (D_0, D_1, \ldots, D_k)$ 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 and only if for each ℓ and each $i \in D_\ell$, we have

$$G_i = \left[\bigcup_{j=1}^{\ell} D_{k+1-j} \right] \cap N_{-i},$$

in which N_{-i} denotes players other than i—taking the intersection with this set is necessary because for $\ell > k/2$, the union includes i's own partition element.

3.1 General characterization

Our first theorem shows that, if linking incentives are consistent and aligned, then stable graphs have one of the two structures we have defined.

Theorem 1. Suppose a network game with network formation has consistent and aligned linking incentives, and $(G, \mathbf{s})$ is a strictly pairwise stable outcome. Then:

(a) If the orders $\succeq_{\text{in}}$ and $\succeq_{\text{out}}$ are identical, the graph G is a nested split graph in which players higher in the two orders have higher degrees.

(b) If the orders $\succeq_{\text{in}}$ and $\succeq_{\text{out}}$ are opposed, the graph G consists of ordered overlapping cliques with respect to either order.

Proof. We begin with part (a). Fixing the outcome $(G, \mathbf{s})$, suppose $j \succeq_{\text{in}} i$ and $j \succeq_{\text{out}} i$. This means that every k that wants to link with i also wants to link with j, and j wants to link with every k with whom i wants to link. Since $(G, \mathbf{s})$ is strictly pairwise stable, there can be no indifference about links, so any $k \neq j$ that is a neighbor of i must be a neighbor of j.

For part (b), we show that if $i \succeq_{\text{in}} j \succeq_{\text{in}} k$ and $ik \in G$, then also $ij \in G$ and $jk \in G$—note this implies $i \preceq_{\text{out}} j \preceq_{\text{out}} k$. Since $ik \in G$, we know i wants to link with k, and since $j \succeq_{\text{in}} k$, this means i wants to link with j. Since i wants to link with j, and $k \succeq_{\text{out}} i$, we know k wants to link with j. Similarly, since i wants to link with k and $j \succeq_{\text{out}} i$, we know j wants to link with k. Since $i \succeq_{\text{in}} k$, this implies j wants to link with i. Since $(G, \mathbf{s})$ is strictly pairwise stable, there can be no indifference about links, so we conclude that $ij \in G$ and $jk \in G$. □

The characterization in Theorem 1 is stark. There are essentially two network structures that can arise in stable outcomes: either neighborhoods are nested, or the network is organized into overlapping cliques of players. In case (a), if one player is ranked higher than another in the two orders, then the two neighborhoods are ordered by set inclusion. In case (b), a link between two players implies that the set of players ranked in between the two 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.

Remark 2. Implicit in this result is a novel characterization of structures that arise in pure network formation games—the definitions of consistent and aligned link preferences do not change if action sets are singletons. In this case, it is as if each player has a one-dimensional type, and linking incentives depend on these types—higher types are more attractive neighbors, and a higher ranked player either always has a stronger incentive to form links or always has a weaker incentive to form links. Work on strategic network formation has thus far faced challenges in obtaining general results on the structure of pairwise stable graphs, and Theorem 1 highlights non-trivial conditions that yield sharp predictions.

3.2 Separable network games

Recall the separable games of Section 2.2. Lemma 2 showed that single-crossing conditions in separable games imply consistent and aligned linking incentives. Combining this with Theorem 1, we obtain the following characterization of stable network structures in separable games.

Corollary 1. Suppose a network game with network formation has payoff functions of the form (1). If $(G, \mathbf{s})$ is a strictly pairwise stable outcome, then:

(a) If the game exhibits action-link complements and positive spillovers, then G is a nested split graph in which players with higher degrees take higher actions.

(b) If the game exhibits action-link substitutes and negative spillovers, then G is a nested split graph in which players with higher degrees take lower actions.

(c) If the game exhibits action-link complements and negative spillovers, or action-link substitutes and positive spillovers, then G consists of ordered overlapping cliques with respect to the order of players' actions.

Proof. The result is immediate from Theorem 1 and Lemma 2. □

Remark 3. While payoffs of the form (1) encompass all of our applications in the next section, we want to highlight that Theorem 1 applies much more broadly. There are at least two natural extensions of this class of games for which an analogous result is immediate. First, one could replace the function $g(s_i, s_j)$ in the sum with a term of the form $g(s_i, s_j) - c_i(s_i)$—having added an idiosyncratic cost of linking, the conclusions of Corollary 1 continue to hold for $g(s_i, s_j)$ increasing/decreasing in each argument. Second, one could replace $g(s_i, s_j)$ with a term of the form $g\left(h_i(s_i), h_j(s_j)\right)$, in which $\{h_i\}_{i\in N}$ are arbitrary idiosyncratic functions of the players' actions. The orders $\succeq_{\text{in}}$ and $\succeq_{\text{out}}$ would then follow the order of the function values $\{h_i\}_{i\in N}$ rather than the order of players' actions.

4 Perverse consequences of group design

We now turn to our first application, using the tractability our characterizations afford to incorporate a network formation analysis into the peer effects model of Carrell et al. (2013). Carrell et al. (2013) estimated academic peer effects among first-year cadets at the US Air Force Academy (using a standard model that took networks as exogenous) and then used these estimates to inform the assignment of new cadets to squadrons—administratively designed peer groups of about 30 cadets. Based on a first cohort of randomly assigned squadrons, the authors concluded that being in a squadron with higher-performing peers[18] led to better academic performance among less prepared cadets. In the treatment group of a later cohort, incoming cadets with less preparation were systematically placed in squadrons

[18]Specifically, those entering with relatively high scores on the verbal section of the SAT exam.

with larger numbers of high-ability peers. While the researchers' goal was to improve the performance of the less prepared cadets,[19] the intervention ultimately backfired: these students performed significantly worse. In this section, we present a model showing that our theory can simultaneously explain two distinctive 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.

Broadly, our results show that stable graphs become more fragmented if private incentives or abilities are more heterogeneous. Thus, placing cadets of high and low abilities together, without cadets of middle ability to bridge the gap, can result in isolated cliques that eliminate the desired spillovers.

Methodologically, this application illustrates that our main theorems can apply to a natural specification of link formation and action choice suited to a practical setting. Our characterization implies an overlapping cliques structure that makes it tractable to analyze these outcomes and their welfare implications both numerically and analytically.

4.1 A model

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 $\alpha \in [0, 1]$. Holding the graph fixed, this is a standard linear-quadratic network game of strategic complements. Taking $v_i(\mathbf{s}) = b_i s_i - \frac{1}{2} s_i^2$ and $g(s_i, s_j) = \alpha s_i s_j - \frac{1}{2} s_i^2$, we see it also falls into the class (1). There are positive spillovers, as an increase in s_j makes a link to player j more valuable. Moreover, links and actions are substitutes as $g(s_i, s_j)$ satisfies the requisite single crossing property—as s_i increases, the benefit to i of linking to j decreases and eventually turns negative, implying those who invest a lot of effort find linking too costly.[20,21] One can readily check that in a pairwise stable outcome, players i and j are neighbors only if $\frac{s_j}{2\alpha} \leq s_i \leq 2\alpha s_j$.

[19] The precise objective they were maximizing was the performance of the bottom third of cadets.

[20] A natural interpretation is that studying and socializing each take time away from the other activity. While studying together can also strengthen social ties, the substantive assumption here is that a marginal hour studying together is less conducive to friendship formation than that same hour spent together on a leisure activity.

[21] One might alternatively use a payoff function with a hard resource constraint split between studying and socializing—our structural results would still apply—but we believe a flexible allocation is more realistic. Cadets spend time on other activities, such as sleep and solitary leisure, that can also be reallocated.

A pairwise stable outcome satisfies the first-order condition

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

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 express this in matrix notation as

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

For $\alpha \in [0,1]$, the solution for $\mathbf{s}$ is unique and well-defined in any graph G, and it is an equilibrium of the game holding G fixed. Notice that when i plays her best response s_i, her payoff in the graph G is

$$u_i = \frac{1}{2}(1+d_i)s_i^2.$$

Hence, if $(G, \mathbf{s})$ and $(G', \mathbf{s}')$ are two pairwise stable outcomes, player i is better off under $(G', \mathbf{s}')$ if and only if $(1+d_i')(s_i')^2 > (1+d_i)s_i^2$.

4.2 The structure of stable outcomes

How do private incentives b_i, and the strength of spillovers α, affect the set of stable outcomes? As a general rule, stable graphs become more fragmented if spillovers α are small, if private incentives b_i are more spaced out, and if the population size n is large. As a first step towards our results for this model, a lemma characterizes equilibrium actions for players connected in a clique.

Lemma 3. Suppose $\alpha \leq 1$, and the players in a set C form a clique —meaning i and j are linked for every $i, j \in C$—with no other links. Writing $\overline{b}_C = \frac{1}{|C|}\sum_{j \in C} b_j$ for the average private incentive, each player $i \in C$ has a unique equilibrium action

$$s_i = \frac{1}{\alpha + |C|}\left(b_i + \frac{\alpha|C|\overline{b}_C}{\alpha + (1-\alpha)|C|}\right). \tag{2}$$

Consequently, the payoff to player i in clique C is

$$u_i = \frac{1}{2}\frac{|C|}{(\alpha+|C|)^2}\left(b_i + \frac{\alpha|C|\overline{b}_C}{\alpha + (1-\alpha)|C|}\right)^2. \tag{3}$$

Proof. From the first-order condition, we have

$$s_i = \frac{1}{|C|}\left(b_i - \alpha s_i + \alpha\sum_{j \in C} s_j\right) \quad \Longrightarrow \quad b_i = (|C| + \alpha)s_i - \alpha\sum_{j \in C} s_j.$$

Summing over all $i \in C$, we get

$$|C|\bar{b}_C = (|C| + \alpha - \alpha|C|)\sum_{j \in C} s_j \quad \implies \quad \sum_{j \in C} s_j = \frac{|C|\bar{b}_C}{\alpha + (1-\alpha)|C|}.$$

Substituting into the first-order condition and solving yields the result. □

One consequence of Lemma 3 is that welfare is higher in larger cliques only if spillovers are sufficiently strong. For any $\alpha < 1$, the denominator in (3) contains a higher power of $|C|$ than the numerator. Holding private incentives fixed, this means that utility declines if the clique becomes large enough. In the case with $\alpha = 1$, equation (3) gives

$$u_i = \frac{|C|}{2} \frac{\left(b_i + |C|\bar{b}_C\right)^2}{(1+|C|)^2}.$$

If $|C|$ gets larger without significantly affecting the average private incentive $\bar{b}_C$, this payoff increases: the first factor is proportional to $|C|$, and the second factor approaches $\bar{b}_C^2$.

Our first proposition characterizes conditions under which the empty graph is part of a pairwise stable outcome. Note that if G is empty, the unique equilibrium actions are $s_i = b_i$ for each player i. In the following results, we always assume players are ordered so that $b_1 \leq b_2 \leq \cdots \leq b_n$, and we write $\bar{b} = \frac{1}{n}\sum_{i=1}^n b_i$ for the average private incentive.

Proposition 1. The empty graph is part of a pairwise stable outcome if and only if $\frac{b_{i+1}}{b_i} \geq 2\alpha$ for each $i = 1, 2, \ldots, n-1$. Moreover, there exists a threshold $\underline{\alpha} \geq \frac{1}{2}$ such that, whenever $\alpha < \underline{\alpha}$, no nonempty graph is possible in a pairwise stable outcome. If the b_i are all distinct, then the threshold satisfies $\underline{\alpha} > \frac{1}{2}$. Otherwise, we have $\underline{\alpha} = \frac{1}{2}$.

Proof. See Appendix. □

The first part of Proposition 1 tells us that the empty graph is stable whenever the private incentives are sufficiently spaced out. How spaced out they need to be is increasing in the strength of spillovers. Moreover, if the spillover parameter α is small enough, then the empty graph is the only graph that can appear in a pairwise stable outcome. Our next result provides an analogous characterization of conditions under which the complete graph is part of a pairwise stable outcome.

Proposition 2. The following claims hold:

(a) There exists a pairwise stable outcome $(G, \mathbf{s})$ in which G is complete if

$$\frac{b_n}{b_1} \leq \frac{\alpha(1+n)}{(1-\alpha)(2\alpha+n)}. \tag{4}$$

(b) Moreover, if

$$\frac{b_n}{b_1} < \frac{2\alpha}{\alpha + (1-\alpha)(n-1)}, \tag{5}$$

then there is a unique pairwise stable outcome, and in it, G is complete.

(c) Conversely, if

$$b_n\left(2\alpha^2+n(1-2\alpha^2)\right)>b_1\alpha\left(4\alpha-1+2n(1-\alpha)\right),$$

then the complete graph is not part of any pairwise stable outcome.

Proof. See Appendix. □

The main message of Proposition 2 is that the complete graph becomes harder to sustain as private incentives b_i get more spread out: if the ratio $\frac{b_n}{b_1}$ is too high, there are pairwise stable outcomes with disconnected graphs, and the complete graph may not be part of any stable outcome. An important consequence of the first claim is that the complete graph is part of a pairwise stable outcome whenever α is sufficiently close to 1. If $\alpha = 1$, the complete graph is always part of a pairwise stable outcome. More generally, stronger spillovers encourage more connected graphs. The proof implies that claims (a) and (b) are tight whenever $b_1 = b_2 = \cdots = b_{n-1}$—if (5) fails, there is a pairwise stable outcome in which the first $n-1$ players form a clique, and player n is isolated. As n gets larger, or α gets smaller, the inequalities (4) and (5) become harder to satisfy, and outcomes with disconnected graphs become more likely.

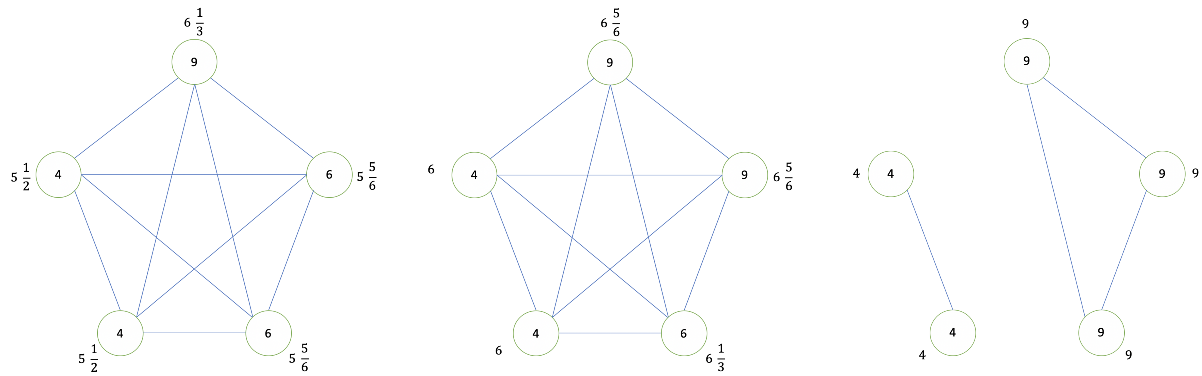
Note that even though these results are specific to complete graphs, the analysis readily extends to subsets of players, giving both necessary and sufficient conditions for cliques to form.

4.3 The importance of intermediate-ability types

Specializing the model to include just three productivity types allows us to transparently relate the model to the findings of Carrell et al. (2013). Suppose $\alpha = 1$, and private incentives take one of three values $b_i \in \{b_\ell, b_m, b_h\}$, satisfying the following inequalities: $2 < b_h/b_\ell < 4$, $b_h/b_m < 2$, and $b_m/b_\ell < 2$. We can interpret type b_ℓ as having low ability, type b_h as having high ability, and type b_m as having intermediate ability. Given an outcome $(G, \mathbf{s})$, we interpret the action s_i as the academic performance of cadet i, and links are friendships through which peer effects operate.

Since $\alpha = 1$, the complete graph is always part of a pairwise stable outcome. However, it should also be clear that without any middle types, an outcome with two isolated cliques—one of low types taking action $s = b_\ell$ and one of high types taking action $s = b_h$—is also pairwise stable. This outcome is clearly worse for low-type players as they take both lower actions and have fewer connections. Moreover, this outcome is the only one that survives a natural refinement. We call a pairwise stable outcome $(G, \mathbf{s})$ **uncoordinated** if there exists a sequence of graphs and action profiles $(G^{(0)}, \mathbf{s}^{(0)}, G^{(1)}, \mathbf{s}^{(1)}, \ldots)$, ending at $(G, \mathbf{s})$, in which

- $G^{(0)}$ is empty,
- $\mathbf{s}^{(k)}$ is a Nash equilibrium holding $G^{(k)}$ fixed, and
- we have $G^{(k+1)} = G^{(k)} + ij$ for some pair ij, and both $u_i(G^{k+1}, \mathbf{s}^{(k)}) \geq u_i(G^k, \mathbf{s}^{(k)})$ and $u_j(G^{k+1}, \mathbf{s}^{(k)}) \geq u_j(G^k, \mathbf{s}^{(k)})$ with at least one strict inequality.



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

Figure 2: 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.

In words, a pairwise stable outcome is uncoordinated if it is reachable through myopically beneficial link additions, starting from an empty graph and assuming that players reach a Nash equilibrium action profile following each new link.[22]

When both the complete graph and segregated cliques are pairwise stable, there is good reason to expect the latter outcome in practice—uncoordinated stable outcomes formalize this idea. However, if we include enough middle types, we can always eliminate the bad pairwise stable outcome.

Three illustrative squadrons Suppose there are 5 cadets, and we take $\{b_\ell, b_m, b_h\} = \{4, 6, 9\}$. 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 squadron, we replace a cadet of intermediate ability with one of high ability, and we are interested in how the actions and welfare of the low-ability cadets change. In the context of Carrell et al. (2013), the first two squadrons represent combinations that should occur frequently in the chance assignments of the first cohort, while the last one represents the designed groups.[23]

[22]This selection criterion implicitly assumes that players have no prior relationships at the start of the adjustment process. Given information on prior relationships, one could adapt this criterion to select an outcome reachable from the initial state.

[23]This is consistent with the facts reported in Carrell et al. (2013), who note that the protocol for designing squadrons, in order to group low-ability cadets with many high-ability peers, tended to exclude cadets of intermediate ability and place these into more homogeneous squadrons—we do not discuss these here.

In the first two squadrons, the unique pairwise stable outcome involves a complete graph. In squadron 1, the action vector is $\mathbf{s} = \left(5\frac{1}{2}, 5\frac{1}{2}, 5\frac{5}{6}, 5\frac{5}{6}, 6\frac{1}{3}\right)$, and in squadron 2, 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, and low-ability cadets benefit from this small change in group composition. What happens if we add another high-ability cadet? In squadron 3, the unique uncoordinated pairwise stable 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.

4.4 Discussion

The predicted outcome in the designed squadrons is fragmentation: a critical mass of high-ability cadets forms their own clique, which is consistent with the explanation that Carrell et al. (2013) give for the unintended consequences they observed. Surveying their participants subsequently, they found that treatment squadrons had significantly higher rates of ability homophily than control squadrons.[24] Indeed, even though the designed squadrons had more high-ability cadets (compared to a random squadron), the low-ability cadets in those squadrons were actually less likely to have high-ability cadets as study partners. This is strong evidence of a new force preventing the peer effects from materializing. Our theory explains this via the strategic forces arising from endogenous network formation.

Although the uncoordinated refinement is particularly compelling in this setting—cadets generally do not know one another beforehand—note that the complete graph is still part of a pairwise stable outcome in the "designed" squadron 3—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 if a policymaker undertakes a more coordinated effort to facilitate collaboration, it may be possible for stable friendship networks to mediate peer effects between low- and high-ability students.

5 Other applications

We briefly address two additional applications. We first study *status games*, in which payoffs incorporate social comparisons—this allows us to interpret stylized facts about clique formation. We then discuss how our analysis can provide a foundation for group matching models that assume a coarse notion of network formation in which agents choose cliques in which to participate. Prior work studying network games together with network formation focuses predominantly on the case of link-action complements and positive spillovers, and such models predict nested split graphs. In contrast, our main applications fall into cells in which stable graphs consist of ordered overlapping cliques.

[24] These calculations controlled for opportunities to link.

5.1 Status games and ordered cliques

A natural model of competitions for status features action–link complements and negative spillovers. For instance, when people care about their relative position in their social neighborhood, (i) having more friends who engage in conspicuous consumption creates stronger incentives to consume more, but at the same time (ii) those who consume conspicuously 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 externalities across neighbors (e.g., due to health consequences or crowding out more productive behaviors). 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 entails that such situations drive the formation of social cliques ordered according to both their popularity and their effort at the activity in question.

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 ranking among them in terms of their sociability and status-conferring behaviors.[25] 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 indeed that 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 the salience of status comparisons with more popular individuals, 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

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

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

[25] 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.

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,[26] but suffers a loss $s_j - s_i$ if neighbor j invests more effort. There are no other linking costs. 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 clearly falls into the class (1) of separable network games, with negative spillovers and weak links-action 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 maximum equilibrium 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 $\frac{k(k+1)}{2} \leq n$ (which is approximately $\sqrt{2n}$).

5.2 Foundations for group-matching models

Models of endogenous matching that go beyond standard pair matching frameworks often posit that individuals belong to a *group* of others. Externalities and strategic interactions then occur within or across groups—with the crucial feature that 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

[26] Note the graph here is unweighted.

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 the common action set S 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. Payoffs take the form

$$u_i(G, \mathbf{s}) = v(s_i, t_i) + \sum_{j \in G_i} g(s_i, s_j), \tag{6}$$

in which v and g are continuous. We further assume that players' have unique best responses, holding the graph and other players' actions fixed. Write $s_t^* = \arg\max_{s \in S} v(s, t)$ for the action that a type t player would take if isolated with no neighbors—this is the *privately optimal action.* Payoffs exhibit a *weak preference for conformity* if player i's optimal action always lies somewhere in between her privately optimal benchmark and the actions of her neighbors. That is, for $\hat{s} = \arg\max_{s_i \in S} u_i(G, s_i, s_{-i})$, we have

$$\min\{s_{t_i}^*, \min_{j \in G_i}\{s_j\}\} \leq \hat{s} \leq \max\{s_{t_i}^*, \max_{j \in G_i}\{s_j\}\}$$

for all i and G.

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

- $g\left(s_t^*, s_{t'}^*\right) \geq 0$ for any $t, t' \in T_k$ and any k.
- Either $g\left(s_t^*, s_{t'}^*\right) \leq 0$ or $g\left(s_{t'}^*, s_t^*\right) \leq 0$ with at least one strict inequality 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 3. Suppose a network game with network formation has payoffs of the form (6), exhibits a weak preference for conformity, and types form natural cliques. If the game exhibits either positive spillovers and action–link substitutes or negative spillovers and action–link complements, then there exists a pairwise stable outcome in which the network is exactly the partition into natural cliques.

Proof. We carry out the proof assuming positive spillovers and action–link substitutes—the other case is analogous. Since types form natural cliques, there is a partition $\{T_1, T_2, \ldots, T_K\}$ 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 $\underline{s}_k$ and $\overline{s}_k$ denote the lowest and highest values respectively of s_t^* for some type $t \in T_k$. Continuity together with Weak preference for conformity implies that there exists an equilibrium in actions with $s_i \in [\underline{s}_k, \overline{s}_k]$ for every player i with type $t_i \in T_k$. Given two such players i and j, we have

$$g(s_i, s_j) \geq g(s_i, \underline{s}_k) \geq g(\overline{s}_k, \underline{s}_k) \geq 0,$$

in which the first inequality follows from positive spillovers, and the second follows from action–link 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 $\underline{s}_\ell \geq \overline{s}_k$. For player i with type $t_i \in T_k$ and j with type $t_j \in T_\ell$ we have

$$g(s_i, s_j) \leq g(\underline{s}_k, s_j) \leq g(\underline{s}_k, \overline{s}_\ell) < 0,$$

so the players have no incentive to link. □

Under mild 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 that comes from group matching assumptions.

6 Existence

While pairwise stable outcomes exist in all of our applications, we have not yet addressed the general question of the existence of pairwise stable outcomes. There are two reasons why existence is non-trivial in our setting. First, the presence or absence of a link is a discrete event, so we cannot use standard arguments that rely on continuity. Second, pairwise stability requires the absence of profitable joint deviations to form new links. Nevertheless, there are natural sufficient conditions that ensure existence of pairwise stable outcomes. In what follows, we assume that players' action sets are complete lattices with order $\geq$.

Definition 5. A network game with network formation exhibits **strategic complements** if for any graph G, any $s_i' > s_i$, and any $s_{-i}' > s_{-i}$, we have

$$u_i(G, s_i', s_{-i}) \geq (>)\ u_i(G, s_i, s_{-i}) \quad \Longrightarrow \quad u_i(G, s_i', s_{-i}') \geq (>)\ u_i(G, s_i, s_{-i}').$$

The game exhibits **convexity in links** if for any profile $\mathbf{s}$, any graph G, any pair ij, and any collection of edges E, we have

$$\Delta_{ij} u_i(G, \mathbf{s}) \geq (>)\ 0 \quad \implies \quad \Delta_{ij} u_i(G + E, \mathbf{s}) \geq (>)\ 0.$$

A network game with network formation exhibits strategic complements if, holding the graph fixed, the underlying normal form game exhibits strategic complements. The definition imposes a single-crossing condition on players' strategies, which implies that best responses are weakly increasing in others' actions. The game exhibits convexity in links if, holding actions fixed, adding links to the network weakly increases players' incentives to form links. Note that in all of our examples, linking incentives are independent of G holding others' actions fixed, so this condition trivially holds.

To state our result, we also need to extend the notions of link-action complements/substitutes and positive/negative spillovers to arbitrary games.

Definition 6. A network game with network formation exhibits **link-action complements** if

$$\Delta_{ij} u_i(G, \mathbf{s}) \geq (>)\ 0 \quad \implies \quad \Delta_{ij} u_i(G, s_i', s_{-i}) \geq (>)\ 0)$$

whenever $s_i' > s_i$. The game exhibits **link-action substitutes** if the above inequality holds whenever $s_i' < s_i$.

Definition 7. A network game with network formation exhibits **positive spillovers** if

$$\Delta_{ij} u_i(G, \mathbf{s}) \geq (>)\ 0 \quad \implies \quad \Delta_{ij} u_i(G, s_j', s_{-j}) \geq (>)\ 0$$

whenever $s_j' > s_j$. The game exhibits **negative spillovers** if the above inequality holds whenever $s_j' < s_j$.

Proposition 4. Suppose a network game with network formation exhibits strategic complements and convexity in links. If either

(a) the game exhibits action-link complements and positive spillovers, or

(b) the game exhibits action-link substitutes and negative spillovers,

then there exists a pairwise stable outcome. Moreover, the set of graphs that occur in pairwise stable outcomes contains a maximal and minimal element.[27]

Proof. Since the game exhibits strategic complements, for any fixed G there exist minimal and maximal Nash equilibria of the induced normal form game—this follows from standard arguments using Tarski's fixed point theorem. Likewise, since the game exhibits convexity in links, for any fixed profile $\mathbf{s}$ there exist minimal and maximal pairwise stable graphs. To get the minimal graph, start from an empty graph and iteratively add links that pairs of players

[27] In fact, one can also show using convexity in links that the minimal graph is part of a pairwise Nash stable outcome.

jointly wish to form. Convexity in links implies that no player will later wish to remove a link that was added earlier, so we must eventually terminate at a stable graph. Similarly to get the maximal graph, start from a complete graph and iteratively delete links that one of the players wishes to remove.

We now define two maps $\overline{B}(G, \mathbf{s})$ and $\underline{B}(G, \mathbf{s})$ mapping outcomes to outcomes. Let $\overline{B}(G, \mathbf{s})$ return an outcome $(\overline{G}, \overline{\mathbf{s}})$ in which $\overline{G}$ is the maximal pairwise stable graph given $\mathbf{s}$, and $\overline{\mathbf{s}}$ is the maximal Nash equilibrium given G. Similarly, let $\underline{B}(G, \mathbf{s})$ return an outcome $(\underline{G}, \underline{\mathbf{s}})$ in which $\underline{G}$ is the minimal pairwise stable graph given $\mathbf{s}$, and $\underline{\mathbf{s}}$ is the minimal Nash equilibrium given G.

In case (a), positive spillovers and link-action complements imply that the graphs $\overline{G}$ in $\overline{B}(G, \mathbf{s})$ and $\underline{G}$ in $\underline{B}(G, \mathbf{s})$ are weakly increasing in $\mathbf{s}$—holding the rest of the graph fixed, higher $\mathbf{s}$ makes link ij more desirable to both player i and player j. Similarly, positive spillovers and link-action complements imply that the profiles $\overline{\mathbf{s}}$ in $\overline{B}(G, \mathbf{s})$ and $\underline{\mathbf{s}}$ in $\underline{B}(G, \mathbf{s})$ are weakly increasing in G. This means that both $\overline{B}$ and $\underline{B}$ are monotone maps with respect to the natural product order on $\mathcal{G} \times \mathcal{S}$, so Tarski's theorem implies minimal and maximal fixed points exist for both—the maximal fixed point of $\overline{B}$ is the maximal pairwise stable outcome, and the minimal fixed point of $\underline{B}$ is the minimal pairwise stable outcome.

Case (b) follows from similar reasoning after reversing the order on action profiles. Negative spillovers and link-action substitutes implies that the graphs in $\overline{B}(G, \mathbf{s})$ and $\underline{B}(G, \mathbf{s})$ are weakly decreasing in $\mathbf{s}$, and the profiles are weakly decreasing in G, so we can again apply Tarski's theorem. □

Proposition 4 only applies within two out of four cells for the class of games in Section 3.2. In general, we cannot ensure existence for the other two cells, as the following example illustrates. Suppose there are two players with a common action set $S = \{0, 1\}$ and payoffs

$$u_i(G, \mathbf{s}) = \begin{cases} s_i & \text{if } G \text{ is empty} \\ 2s_{-i} + \frac{2s_i s_{-i} - 1}{4} - s_i & \text{if } G \text{ is complete.} \end{cases}$$

For player i, the marginal value of a link to player $-i$ is

$$2(s_{-i} - s_i) + \frac{2s_i s_{-i} - 1}{4}.$$

This is increasing in s_{-i}, and for the relevant range of values it is decreasing in s_i—the game exhibits positive spillovers and link–action substitutes. Moreover, it should be clear that, if the two players are linked, the game exhibits strategic complements. Nevertheless, there is no pairwise stable outcome. In the empty graph, each player optimally takes action 1, and the marginal value of adding the link between them is $\frac{1}{4} > 0$, so they should form the link. In the complete graph, each player optimally takes action 0, and the marginal value of the link is then $-\frac{1}{4} < 0$, so they should each drop the link.

Even though existence in not always assured, the structural result in Corollary 1 greatly simplifies the process of searching for a stable outcome. As we have already seen in three applications, starting with a clique structure and checking whether it is stable often provides a simple way to establish existence.

We note that our existence result extends the main finding in Hellmann (2013), obtained in a setting with network formation only. The paper 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 convexity in links in Definition 5.

7 Discussion

7.1 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—rigid patterns across different layers can combine to form more realistic arrangements.

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 weekend ties, and homophily, driven by the work ties. This description ties into Simmel's account, subsequently developed by many scholars, of cross-cutting cleavages.

A second approach is to introduce noise. König et al. (2014) provide an example, describing a dynamic process in which agents either add or delete one link at a time, and the underlying incentives exhibit positive spillovers and action–link 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, and under suitable assumptions obtain a noisy version of our ordered cliques prediction.

7.2 Foundations for pairwise (Nash) stability

As presented, pairwise (Nash) stability is a static solution concept that entails the absence of particular individual and pairwise deviations. A key feature is that deviations in links and actions are treated separately: Players consider link deviations holding actions fixed and action deviations holding links fixed. Why not require robustness to simultaneous deviations in both actions and links?

There are two reasons. One is a pragmatic view of how link formation and action choice actually work: In practice, decisions over these dimensions often *are* considered separately. Occasionally, people must invest to form and maintain relationships (e.g., doing a costly favor, attending a social event)—these are the times at which linking costs are actually paid and people are prone to reconsider relationships. Opportunities to revise productive actions (e.g., investing in a certain kind of expertise at work) occur at other times. We therefore consider it plausible that individuals consider these revisions separately, taking the current state of play as otherwise fixed.

The second reason is methodological. Simultaneous deviations along multiple dimensions make it delicate to define how a counterparty should respond to a link offer. Should the recipient of an offer condition on other deviations by the sender? (E.g., "If Bob is offering me a link, he would also change other actions and links.") Should she contemplate subsequent changes in her own behavior? E.g., "In the network that is likely given Bob's link offer to me, I will want to drop certain other links.") Such considerations open the important but very complicated Pandora's box of farsighted stability. Following this logic to its natural conclusion requires that players are not only very sophisticated, but also omniscient (or at least in possession of a full theory) about how others respond to deviations. Allowing just one deviation at a time avoids this issue.

As this argument only applies to deviations that require another player's consent, one might still ask: Why not allow a broader set of unilateral deviations? This is precisely what pairwise Nash stability does. While we still consider pairwise stability a more appropriate solution concept for the first, practical, reason above, refining the solution concept to pairwise Nash stability has no impact on our structural results—since a pairwise Nash stable outcome is pairwise stable, any statement true of all pairwise stable outcomes is also true of all pairwise Nash stable outcomes. Requiring robustness to other unilateral deviations can only further refine the outcome set, and our main results continue to apply.

In the rest of this subsection, we present two dynamic models that provide foundations for our solution concepts, one for pairwise stability and one for pairwise Nash stability. These make explicit certain adjustment processes that imply our main stability conditions at equilibrium. Throughout the section, we restrict attention to finite action spaces and generic payoffs, so player i has a unique myopically optimal action s_i given $\mathbf{s}_{-i}$ and G.

7.2.1 A revision game

We first study a dynamic game that makes explicit the "occasional revisions" foundation for pairwise stability.[28] Players have revision opportunities arriving at random times, and myopic best response is optimal as long as discount rates are sufficiently high, or the time between revisions is sufficiently long.

[28]See Jackson and Watts (2002) for an early antecedent of stochastic revisions in a game of network formation and action choice.

Time t is continuous, all players observe the current state $(G^{(t)}, \mathbf{s}^{(t)})$, and players can change their actions and links only at random arrival times. Each player i has an independent Poisson clock with rate λ, which rings at times $\{\tau_k^i\}_{k=0}^{\infty}$. At each time τ_k^i, player i has an opportunity to revise her strategic action s_i. Additionally, each ordered pair of players ij has an independent Poisson clock with rate λ, which rings at times $\{\tau_k^{ij}\}_{k=0}^{\infty}$. At each time τ_k^{ij}, if $j \in G_i$, player i has the option to delete link ij, and if $j \notin G_i$, player i has the option to propose a link to player j. If i proposes a link to j, player j can either accept or reject at that instant. If player j accepts, we add ij to the graph, and otherwise the graph is unchanged. Players receive a constant flow payoff according to the current state (actions and links), and there is a common discount factor $\delta \in (0, 1)$. To complete the model, we specify an arbitrary initial condition $(G^{(0)}, \mathbf{s}^{(0)})$.

Proposition 5. Fix any $\delta < 1$. If $\lambda > 0$ is sufficiently small, the following statements hold:

(a) If there exists a subgame perfect Nash equilibrium of the revision game and an almost surely finite stopping time τ such that, in that equilibrium, $(G, \mathbf{s})$ is played on path at all $t \geq \tau$, then $(G, \mathbf{s})$ is pairwise stable.

(b) Conversely, if $(G, \mathbf{s})$ is pairwise stable, there exists a subgame perfect Nash equilibrium of the revision game, with initial condition $(G^{(0)}, \mathbf{s}^{(0)}) = (\mathbf{G}, \mathbf{s})$, in which $(G, \mathbf{s})$ is played at all times $t \geq 0$.

Proof. See Appendix. □

The main idea in the proof is that a sufficiently long delay until the next revision opportunity makes the immediate payoff implications of revising an action or link the dominant ones.

7.2.2 A two-stage game

We next present an explicit non-cooperative protocol for making joint deviations. This two-stage game provides a foundation for pairwise Nash stability.

Beginning from some outcome $(G, \mathbf{s})$, the deviation game proceeds in two stages. In the first stage, a player i is selected uniformly at random, and i is allowed to make any unilateral deviations she wishes—she can change her strategic action s_i and delete any subset of her links $S \subseteq G_i$. With probability $1 - \epsilon$, the game ends here. With probability ϵ, we move to the second stage in which i is allowed to propose a link to a single other player j. If i makes a proposal to j, player j chooses whether to accept or reject, and the game ends with payoffs determined by the final outcome $(G', \mathbf{s}')$. An outcome $(G, \mathbf{s})$ is an *equilibrium of the ϵ-PNS-game* if, starting at this outcome, we remain at the outcome $(G, \mathbf{s})$ in any subgame perfect Nash equilibirum of the deviation game.

Proposition 6. The outcome $(G, \mathbf{s})$ an equilibrium of the the ϵ-PNS-game for all sufficiently small ϵ if and only if it is pairwise Nash stable.

Proof. See Appendix. □

The deviation game captures an intuition that players can always change their own actions, or delete links, but an opportunity must arise in order to form a new link. If these are sufficiently infrequent, then it does not make sense to plan one's behavior in anticipation of such an opportunity.[29] Formally, when ϵ is small, options available in the second stage have no bearing on optimal decisions in the first.

8 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. Sound behavioral predictions and policy recommendations depend on taking these interactions into account, rather than studying each aspect separately. We offer a flexible formal framework that unites two types of models and solution concepts, pertaining to strategic actions and link choice. This unification enriches what we can capture, allowing us to make new and sharper predictions in important cases. We identify simple conditions that allow a stark characterization of equilibrium network structures and the behavior they support. Several widely studied applications fit within this framework, and we highlight new insights that emerge from applying our results.

One key point—for which our applications serve as a proof-of-concept—is that the structural predictions the theory offers greatly reduce the space of possible networks, as well as the actions they can support. This is crucial for the tractability both of numerical calculations and theoretical analyses of how equilibria and welfare depend on the environment.

Given the highly structured networks our theory predicts, the framework requires further elaboration to fit realistic network structures. There are two directions we have argued are promising. One introduces noise in linking decisions or incentives. We expect that qualitative insights about linking patterns are robust, but noise raises theoretical questions about how much predictions change relative to our benchmark, and econometric questions about what one can identify from an observed network. A second direction examines models that combine different types of relationships. This would allow "overlaying" the simple structures arising in our characterizations, and exploring the economic implications of interactions between different kinds of relationships.

References

Adler, P. A. and P. Adler (1995): "Dynamics of inclusion and exclusion in preadolescent cliques," *Social Psychology Quarterly*, 145–162.

Baccara, M. and L. Yariv (2013): "Homophily in peer groups," *American Economic Journal: Microeconomics*, 5, 69–96.

[29] The same result holds if we reverse the order of the stages—player i makes a link offer in the first stage, and with probability ϵ she can change her action or delete links in the second. While the order is unimportant from a technical perspective, we believe the order we use has a more natural interpretation.

Badev, A. (2021): "Nash Equilibria on (Un)stable Networks," *Econometrica*, 89, 1179–1206.

Baetz, O. (2015): "Social Activity and Network Formation," *Theoretical Economics*, 10, 315–340.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's who in Networks. Wanted: the Key Player," *Econometrica*, 74, 1403–1417.

Bandyopadhyay, S. and A. Cabrales (2020): "Pricing group membership," CESifo Working Paper.

Belhaj, M., S. Bervoets, and F. Deroïan (2016): "Efficient networks in games with local complementarities," *Theoretical Economics*, 11, 357–380.

Bloch, F. and M. O. Jackson (2006): "Definitions of equilibrium in network formation games," *International Journal of Game Theory*, 34, 305–318.

Bolletta, U. (2021): "A model of peer effects in school," *Mathematical Social Sciences*, 114, 1–10.

Bramoullé, Y. and R. Kranton (2007): "Public Goods in Networks," *Journal of Economic Theory*, 135, 478–494.

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

Cabrales, A., A. Calvó-Armengol, and Y. Zenou (2011): "Social Interactions and Spillovers," *Games and Economic Behavior*, 72, 339–360.

Carrell, S., B. Sacerdote, and J. West (2013): "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81, 855–882.

Chade, H. and J. Eeckhout (2018): "Matching information," *Theoretical Economics*, 13, 377–414.

Chandrasekhar, A. G. and M. O. Jackson (2021): "A network formation model based on subgraphs," *Available at SSRN 2660381*.

Davis, J. A. and S. Leinhardt (1967): "The structure of positive interpersonal relations in small groups." .

Galeotti, A. and S. Goyal (2009): "A Theory of Strategic Diffusion," *Rand Journal of Economics*, 40, 509–532.

Gest, S. D., A. J. Davidson, K. L. Rulison, J. Moody, and J. A. Welsh (2007): "Features of groups and status hierarchies in girls' and boys' early adolescent peer networks," *New Directions for Child and Adolescent Development*, 2007, 43–60.

Hellmann, T. (2013): "On the Existence and Uniqueness of Pairwise Stable Networks," *International Journal of Game Theory*, 42, 211–237.

——— (2020): "Pairwise Stable Networks in Homogeneous Societies with Weak Link Externalities," Forthcoming in European Journal of Operational Research.

Herskovic, B. and J. Ramos (2020): "Acquiring Information Through Peers," *American Economic Review*, 110, 2128–2152.

Hiller, T. (2017): "Peer Effects in Endogenous Networks," *Games and Economic Behavior*, 105, 349–367.

Homans, G. C. (1950): *The Human Group*, New York: Harcourt.

Immorlica, N., R. Kranton, M. Manea, and G. Stoddard (2017): "Social Status in Networks," *American Economic Journal: Microeconomics*, 9, 1–30.

Jackson, M. (2008): *Social and Economic Networks*, Princeton University Press.

Jackson, M. O. (2019): "The friendship paradox and systematic biases in perceptions and social norms," *Journal of Political Economy*, 127, 777–818.

Jackson, M. O. and A. Watts (2002): "On the formation of interaction networks in social coordination games," *Games and Economic Behavior*, 41, 265–291.

Joshi, S. and A. Mahmud (2016): "Network Formation Under Multiple Sources of Externalities," *Journal of Public Economic Theory*, 18, 148–167.

König, M., C. Tessone, and Y. Zenou (2014): "Nestedness in Networks: A Theoretical Model and Some Applications," *Theoretical Economics*, 9, 695–752.

Sadler, E. (2021): "Ordinal Centrality," *Journal of Political Economy*, forthcoming.

A Omitted proofs and details

A.1 Formalization of anonymity and separability conditions

We introduce the key payoff conditions. For this purpose, we fix a numerical representation $u_i(G, \mathbf{s})$ for each player's preferences over outcomes (which are taken to be complete and transitive), which has both ordinal and cardinal content (insofar as it captures the intensity of preferences for various links, or equivalently preferences over lotteries).

Definition 8. Let σ be a permutation of the agents and let G_σ and $\mathbf{s}_\sigma$ denote the graph and action profile after agents are relabeled according to σ.[30] Linking incentives are **anonymous** if incentives to form links do not depend on players' labels: player i strictly prefers to add link ij at $(G, \mathbf{s})$ if and only if agent $\sigma(i)$ strictly prefers to add link $\sigma(i)\sigma(j)$ at $(G_\sigma, \mathbf{s}_\sigma)$.

Our next definition, of separability, posits that the incremental values of links are separable.[31]

Definition 9. Linking incentives are **separable** if the following holds for every i, j, and $\mathbf{s}$. Player i's valuation of adding the link ij,

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

depends only on s_i and s_j, and not on any other actions or links.

In this section we establish a simple result:

Lemma 4. Linking incentives are separable and anonymous if and only if payoffs can be represented via the form (1), which we reproduce here for convenience:

$$u_i(G, \mathbf{s}) = v_i(\mathbf{s}) + \sum_{j \in G_i} g(s_i, s_j). \tag{7}$$

Proof. The "if" direction is trivial. For the "only if" direction, we suppose that linking incentives are anonymous and separable. Fix any $\mathbf{s}$. Let $v_i(\mathbf{s})$ be defined as $u_i(\varnothing, \mathbf{s})$, where $\varnothing$ is the empty graph. Now let the neighbors of i in G be enumerated as $j_1, j_2, \ldots, j_d$. Let G_k be the graph with edges $(i, j_{k'})$ for all $k' \leq k$, with G_0 understood to be the empty graph $\varnothing$. By separability of linking incentives, we can inductively show

$$u_i(G, \mathbf{s}) = v_i(\mathbf{s}) + \sum_{k=1}^{d} [u_i(G_k, \mathbf{s}) - u_i(G_{k-1}, \mathbf{s})].$$

Separability further implies that $u_i(G_k, \mathbf{s}) - u_i(G_{k-1}, \mathbf{s})$ is a function, say $g_k(s_i, s_j)$, only of s_i and s_j. Thus we write

$$u_i(G, \mathbf{s}) = v_i(\mathbf{s}) + \sum_{k=1}^{d} g_k(s_i, s_{j_k}).$$

[30] So that G_σ has link $\sigma(i)\sigma(j)$ if and only if G has link ij, and $s_{\sigma(i)} = s_i$.

[31] This definition could, of course, be restated in terms of choice behavior over lotteries for links.

Anonymity then requires that g_k does not depend on k. Thus we may write

$$u_i(G,\mathbf{s}) = v_i(\mathbf{s}) + \sum_{k=1}^{d} g(s_i, s_{j_k}),$$

which is the desired form. □

A.2 Proofs of results in Section 2

Proof of Lemma 1 Fixing an outcome $(G,\mathbf{s})$, define the binary relation $\succeq_{\text{in}}$ via $i \succeq_{\text{in}} j$ if there is no player k such that

$$\Delta_{ik}u_k(G,\mathbf{s}) < 0 \quad \text{and} \quad \Delta_{jk}u_k(G,\mathbf{s}) \geq 0.$$

That is, every player with a weak incentive to link with j also has a weak incentive to link with i. Because linking preferences are consistent, the relation $\succeq_{\text{in}}$ is complete—if we do not have $i \succeq_{\text{in}} j$, then there exists k with a weak incentive to link with j but not i, and if we do not have $j \succeq_{\text{in}} i$, then there exists ℓ with a weak incentive to link with i but not j, contradicting consistent link preferences. It is straightforward to check that this relation is also transitive. Similarly, the binary relation $\succeq_{\text{out}}$ defined via $i \succeq_{\text{out}} j$ if there is no player k such that

$$\Delta_{ik}u_i(G,\mathbf{s}) < 0 \quad \text{and} \quad \Delta_{jk}u_k(G,\mathbf{s}) \geq 0$$

is complete and transitive.

Suppose there exists i^* maximal under $\succeq_{\text{in}}$, and i_* minimal under $\succeq_{\text{in}}$, such that $i^* \succ_{\text{out}} i_*$. Fix any players i and j. If i_* wants to link with j, then i^* wants to link with j because $i^* \succeq_{\text{out}} i_*$, and i also wants to link with j by alignment. Similarly, if i^* does not want to link with j, then i_* and i do not want to link with j. This tells us that i^* is maximal under $\succeq_{\text{out}}$ and i_* is minimal under $\succeq_{\text{out}}$.

We now show that $i \succeq_{\text{out}} j$ whenever $i \succeq_{\text{in}} j$. If not, there exists k such that j wants to link with k but i does not. Since i^* is maximal under $\succeq_{\text{out}}$, we know i^* wants to link with k. We now have

$$i^* \succeq_{\text{in}} i \succeq_{\text{in}} j,$$

and both i^* and j want to link with k. Alignment now implies that i wants to link with k, a contradiction. A similar argument shows that if $i^* \preceq_{\text{out}} i_*$ for any i^* and i_* maximal and minimal respectively under $\succeq_{\text{in}}$, then $i \preceq_{\text{out}} j$ whenever $i \succeq_{\text{in}} j$. □

A.3 Proofs of results in Section 4

Proof of Proposition 1 Recall player i finds it strictly beneficial to link with j if and only if $\frac{s_i}{s_j} < 2\alpha$—the first claim follows. It is weakly beneficial only if $\frac{s_i}{s_j} \leq 2\alpha$, and this can never hold for both i and j if $\alpha < \frac{1}{2}$. If $\alpha = \frac{1}{2}$, it holds for both if and only if $s_i = s_j$—if $b_i = b_j$ for some pair ij, then i and J can optimally form a clique, so the empty graph is not uniquely stable. Going forward, we assume the b_i are distinct.

Suppose we G is not empty, and let S be the largest connected component. Write $\overline{s}$ and $\underline{s}$ for the maximal and minimal equilibrium actions among players in S. If the outcome is stable, we need $\frac{s_i}{s_j} \leq 2\alpha$ whenever i and j are linked, implying $\frac{\overline{s}}{\underline{s}} \leq (2\alpha)^n$. Using the first-order condition, we have

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

so for $i \in S$, we must have

$$\frac{b_i + \alpha d_i \underline{s}}{d_i + 1} \leq s_i \leq \frac{b_i + \alpha d_i \overline{s}}{d_i + 1} \leq \frac{b_i + \alpha d_i (2\alpha)^n \underline{s}}{d_i + 1}.$$

Within a connected component, there exist two players i and j with the same degree—if there are m players in a component, then there are $m - 1$ possible degree values, and the pigeonhole principle tells us two players have the same degree. Suppose $i, j \in S$ have $d_i = d_j = d$, and without loss of generality assume $b_i < b_j$. We then have

$$(d+1)s_i \leq b_i + \alpha d \overline{s} \leq b_i + \alpha d (2\alpha)^n \underline{s}, \quad \text{and} \quad (d+1)s_j \geq b_j + \alpha d \underline{s}.$$

Choosing $\alpha > \frac{1}{2}$, but sufficiently close to $\frac{1}{2}$, we can conclude that $(2\alpha)^n s_i < s_j$, which implies that i and j cannot connected—there exists a link that a player would like to sever. Hence, for α sufficiently close to $\frac{1}{2}$, the only pairwise stable outcome involves an empty graph. □

Proof of Proposition 2 To ensure that the complete graph is part of a pairwise stable outcome, we only need to check that player n, who takes the highest action in the complete graph equilibrium, does not want to delete her link to player 1, who takes the lowest action. Using Lemma 3, this means

$$2\alpha \geq \frac{s_n}{s_1} = \frac{b_n(\alpha + (1-\alpha)n) + \alpha n \overline{b}}{b_1(\alpha + (1-\alpha)n) + \alpha n \overline{b}}, \tag{8}$$

which is equivalent to

$$(b_n - 2\alpha b_1)(\alpha + (1-\alpha)n) \leq (2\alpha - 1)\alpha n \overline{b}. \tag{9}$$

This inequality is hardest to satisfy if we minimize the right hand side. To do this, we take $b_j = b_1$ for each $j \neq n$. Hence, the complete graph is part of a pairwise stable outcome if

$$(b_n - 2\alpha b_1)(\alpha + (1-\alpha)n) \leq \alpha(2\alpha - 1)(b_n + (n-1)b_1),$$

which is equivalent to

$$b_n(1-\alpha)\left(2 + \frac{n}{\alpha}\right) \leq b_1(1+n) \quad \Longrightarrow \quad \frac{b_n}{b_1} \leq \frac{\alpha(1+n)}{(1-\alpha)(2\alpha+n)}.$$

For the second claim, we need to show that no other graph can occur in a pairwise stable outcome. Recall the first-order condition

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

Notice that the best response to any $\mathbf{s}$ in which $s_i \leq b_n$ for each i is a vector $\mathbf{s}'$ in which $s_i' \leq b_n$ for each i, so we necessarily have $s_i \leq b_n$ in the unique Nash equilibrium on any fixed graph. Conversely, let $\underline{s}$ denote the lowest equilibrium action across all graphs G that are not complete. We necessarily have

$$\underline{s} \geq \frac{1}{d+1}\left(b_1 + \alpha d \underline{s}\right) \quad \Longrightarrow \quad \underline{s} \geq \frac{b_1}{\alpha + (1-\alpha)(d-1)}$$

for each $d = 0, 1, 2, \ldots, n-2$. The weakest bound occurs for $d = n-2$—we need not consider the case in which the player taking the lowest action has degree $n-1$ because Corollary 1 then implies the graph is complete. Consequently, the ratio of any two players' actions across all graphs that are not complete is bounded above by

$$\frac{b_n(\alpha + (1-\alpha)(n-1))}{b_1} < 2\alpha,$$

implying that all players have a strict incentive to link.

For the last claim, we return to equation (9) and *maximize* the right hand size by taking $b_2 = b_3 = \cdots = b_n$. We find that the complete graph cannot be pairwise stable if

$$\left(b_n - 2\alpha b_1\right)\left(\alpha + (1-\alpha)n\right) > \alpha(2\alpha - 1)\left((n-1)b_n + b_1\right),$$

which is equivalent to the stated condition. □

A.4 Proofs of results in Section 7.2: Foundations for stability

Proof of Proposition 5 First, suppose there is a subgame perfect Nash equilibrium and an almost surely finite stopping time τ such that, in that equilibrium, $(G, \mathbf{s})$ is played after τ, but this outcome is not pairwise stable. Since the outcome is not pairwise stable, there exists some player i with a profitable action or link deviation. Choose some $\tau_k^i > \tau$, so $(G, \mathbf{s})$ is the outcome at τ_k^i. If i has a profitable action deviation, she gets strictly more utility from some outcome $(G, (s_i', s_{-i}))$. If λ is small enough, then i's expected gain from this deviation is positive even if the outcome changes to the worst possible one for i at the next revision opportunity, since the action sets are finite, payoffs are bounded.) Similarly, if i has a profitable deviation along link ij, choose some $\tau_k^{ij} > \tau$. If $ij \in G$, then by analogous reasoning, for small enough λ, player i finds it profitable to delete link ij. If $ij \notin G$, player i similarly finds it profitable to propose a link to player j, who then finds it profitable to accept. The upper bound on λ necessarily depends on the deviation in question, but since there are finitely many possible configurations, and finitely many possible deviations, one can simply take a minimum over all such bounds.

Suppose conversely that the initial condition $(G^{(0)}, \mathbf{s}^{(0)}) = (G, \mathbf{s})$ is pairwise stable. Any deviation leads to an outcome delivering a strictly lower flow payoff. If λ is small enough, the cost of such a deviation outweigh any potential benefit in a later subgame. □

Proof of Proposition 6 If the outcome $(G, \mathbf{s})$ is an equilibrium of the ϵ-PNS-game for any $\epsilon > 0$, then it is pairwise Nash stable by definition. If $(G, \mathbf{s})$ is pairwise Nash stable, then conditional on reaching the second stage with the outcome $(G, \mathbf{s})$ intact, no further changes can occur in any subgame perfect Nash equilibrium. In the first stage, any deviation leads to a strictly worse outcome for player i, so any such deviation can only be profitable following another deviation in the second stage. Taking ϵ small enough, and using the boundedness of payoffs (due to the finiteness of the outcome space), ensures that ex ante player i has no incentive to deviate in the first stage. □