Competition and Networks of Collaboration*†

Nikita Roketskiy‡
New York University

January, 2012

Abstract

I develop a model of collaboration between competitors. In this model, agents collaborate in pairs, and a structure of collaboration is represented by a network. Agents' payoffs consist of two terms. The first term, which I call "performance," increases in the number of the agent's connections. The second term, a "prize" in a competition, increases in the agent's relative position in the distribution of performance. In contrast with most of the network-formation literature, I assume that the agents are forward looking, and I use von Neumann-Morgenstern stable sets as a solution for the network formation process. First, I find a necessary and sufficient condition for the stability of the efficient network. Second, I find a set of networks that are stable whenever the efficient network is not. These networks consist of two mutually disconnected complete components. I interpret this pattern as a group competition between "insiders" and "outsiders." The larger of the two groups, namely the group of insiders, gains an advantage in the competition by staying disconnected from the outsiders. Finally, I extend my model to accommodate heterogeneity in agents' abilities and show that a high ability does not necessarily translate into a strong position in the competition.

*I am very grateful to my advisors, Alessandro Lizzeri and Debraj Ray, for their guidance, support and encouragement. I thank (in alphabetical order) Heski Bar-Isaac, Joyee Deb, Guillaume Fréchette, Sergei Izmalkov, Sergei Kovbasyuk, John Leahy, Stefania Minardi, Leonardo Pejsachowicz, Michael Richter, Ariel Rubinstein, Burkhard Schipper, Ennio Stacchetti, Kevin Thom, and all participants of seminars at NYU for their insightful comments. All errors are my own.

†Please, see `https://files.nyu.edu/nvr213/public/papers.html` for the up-to-date version of the paper.

‡Department of Economics, NYU, 19 West 4th Street, New York, NY, 10012, tel: 917-385-4802, email: `nroketskiy@nyu.edu`, web: `https://files.nyu.edu/nvr213/public/`.

1 Introduction

We often observe collaboration between direct competitors. For instance, firms that compete in the market for a final product often collaborate at the R&D stage. Similarly, co-workers who compete for a promotion, usually collaborate with their rivals in the course of the competition. Agents in these situation face a dilemma: If they collaborate, they become stronger competitors, but they also strengthen their rivals' positions. Nevertheless, we often see agents choosing extensive collaboration over autarkic competition.

This paper studies the following questions about such collaboration. Under what conditions do competitors collaborate at the efficient level? If those conditions fail to hold, what are stable patterns of collaboration? How do these patterns depend on the exogenous strength of the competitors?

I address these questions with a model in which an endogenous structure of collaboration is represented by a network. I start with a basic setup that features homogeneous agents and simple bilateral collaboration. Any two agents in this model can collaborate with each other, and each agent can collaborate with as many agents as he wants. Naturally, in this setup, a link represents a collaboration between two agents. A link between two agents exists so long as both of them approve it. If one of the two agents decides to break a link, the collaboration ceases.

Agents in my model produce an output and seek collaboration in order to be more productive. Each agent's output increases with the number of direct partners (or, equivalently, links) he has. In addition, agents compete with each other for a prize. I model the competition as a tournament in which the prize is split equally between the agents with maximal output. An agent's payoff consists of two terms. The first term is a portion of the prize an agent wins, and the second term is the output an agent produces.

I assume that agents are forward-looking and capable of cooperation. Whenever an agent decides which action to take, he examines the ultimate outcomes of these actions rather than their immediate effects. I also assume that agents are immune to collective-action problems and failure of coordination: If a coalition of agents can credibly coordinate their actions for their mutual benefit, they do so. The vast majority of papers on network formation assume that agents are myopic (see Jackson and Wolinsky (1996), Dutta and Mutuswami (1997), Jackson (2008), and many others). In this paper, I depart from this tradition. Forward-looking agents fit better for my purposes since I study small networks with extensive communication between agents. It is natural, for example, to assume that, prior to signing a bilateral agreement, firms carefully consider the long-run effect it will have on their current partners.[1]

[1]Such coordination of actions between allied firms is very common. For example, on August 16, 2011, HTC, a Google partner on the market for handsets, filed a lawsuit against Apple, accusing the firm of infingement of HTC patents. This happened a day after Google announced plans to acquire Motorola Mobility. Both Google and HTC foresaw that an alliance of Apple and Microsoft

I use von Neumann-Morgenstern stable sets (see von Neumann and Morgenstern (1944)) to define a solution for this model.A von Neumann-Morgenstern stable set of networks must satisfy two properties: external and internal stability. Intuitively, internal stability guarantees that agents that are connected in a stable networks do not attempt to switch to another stable network, and external stability guarantees that agents that are connected in an unstable network would rearrange their links into a stable network. The notion of von Neumann-Morgenstern stable sets is contextual: Agents consider a particular network to be stable because it satisfies external and internal stability with respect to itself and other elements of a stable set.[2]

This setup speaks to a variety of applications, such as networks of bilateral agreements between firms, collaboration between employees, networks of organized crime, racial and ethnic discrimination in small societies, and others. Although these applications are very different from each other, they all have several common features that play a central role in my model. Agents in these applications compete with each other, and their performance in the competition depends on their collaboration with rivals.

For instance, let us consider some firms that operate in the same market for a final product. They have an opportunity to collaborate at the R&D stage, and, as the data show, firms often use this opportunity extensively.[3] Under what conditions do firms collaborate at the efficient level? How does competitiveness of the market influence collaboration? How does collaboration affect the market structure? My model can be used to tackle all these questions.

Although there is an extensive literature on firms' collaboration, my model offers new insights. It differs from the existing literature in two respects. First, I characterize conditions under which efficient outcomes *cannot* be stable. Second, my model makes sharper predictions about the structure of stable collaboration networks. The vast majority of papers on R&D collaboration consider firms to be myopic (see Goyal and Moraga-Gonzalez (2001), Goyal and Joshi (2003), and Marinucci and Vergote (2011)). These papers find that efficient outcomes are *always* stable, but there are some inefficient outcomes that can be stable as well. In an efficient outcome, a firm has to collaborate with many others. If a firm is myopic, it can be trapped in a situation in which, for example, it has no connections. If there are increasing returns to connections, a myopic firm cannot escape this trap by creating new connections

would try to hinder the acquisition of Motorola Mobility, and HTC's lawsuits are seen by some as an attempt to protect Google's acquisition (see Robin Kwong and Joseph Menn, (2011, August 16), "HTC files fresh lawsuit against Apple." *Financial Times*).

[2]Formally, von Neumann-Morgenstern stable sets are fixed points of so-called blocking correspondence.

[3]For instance, Hagedoorn (2002) reports that the amount of inter-firm collaboration in R&D almost tripled in the period from 1980 to 1998. He also reports that joint ventures constituted roughly half of R&D partnerships in 1980, but by 1998, this share had dropped to roughly 10%. He argues, that firms' preferences shifted from joint ventures towards contractual partnerships such as sharing agreements.

one by one. Put differently, these models are driven by the lack of firms' foresight: A firm fails to coordinate its actions with its "future self." Dutta et al. (1998) avoid this problem by looking at CPNE of a network-formation game. They find results similar to mine, but their framework is very limited—they consider a game with only three firms.

The main results of my paper are as follows. I find necessary and sufficient conditions for efficient outcomes to be stable. To obtain this condition, I look at the collection of networks in which winners are connected only to other winners. Let us call this collection $\mathcal{C}$.[4] An efficient network is stable if and only if it maximizes the utility of winners in the set $\mathcal{C}$. This criterion resembles a *union mentality*. Winners are concerned only with their own payoffs, just as unions are concerned only with the well-being of union members. Winners do not internalize the effect they impose on others, just as unions do not internalize the effect they impose on nonunion employees.

I also find stable networks of collaboration when this condition does not hold. A set of networks that maximizes winners' payoffs in the set $\mathcal{C}$ is always stable. These networks exhibit a special property: Each network consists of two mutually disconnected groups, both of which feature full collaboration. One group necessarily constitutes a majority and, hence, is a group of winners. Members of the larger group secure their superiority in the competition by refusing to collaborate with members of the smaller group. I call the larger group "insiders" and the smaller one "outsiders." We can interpret this rule as a norm, that favors the majority and is enforced by it.

These results are coherent with properties of networks of collaboration that we observe in practice. To illustrate this, let us go back to the example with R&D collaboration. Theory suggests that if the stakes in the competition are high enough, the efficient network of collaboration, in which firms sign all available collaboration agreements, is not stable. Moreover, there are stable networks in which firms that dominate the market—i.e., "insider" firms—refuse to collaborate with "outsider" firms. This strategy allows the "insiders" to maintain their dominant position in the market. Such configurations are observed in practice. For example, Bekkers et al. (2002), in an overview of the market for GSM infrastructure and terminals, report that an alliance of five firms led by Motorola dominated this market in the 1990s. Their total market share in 1996 was above 85%. Bekkers et al. (2002) argue that bilateral cross-licensing agreements played a central role in this alliance. Its participants made their portfolios of GSM-essential patents available to their allies and unavailable to the rest of the firms. This strategy is predicted by my results, and it has proven to be very successful in capturing almost all of the GSM market.

More generally, my model provides two important insights on patent wars. First, if stakes in the competition are high, patent wars are inevitable. Second, the possibility of patent war forces firms to form alliances.

To obtain more insights, I extend this model to the case of both heterogeneous agents and more-complex collaboration relationships. I assume, that agents differ

[4] Notice, that an efficient network belongs to this collection of networks.

in their ability to convert collaboration into output. Using this setup, I find stable networks with empirically relevant properties. As in the basic model, these stable networks feature two mutually disconnected groups of unequal size. This time, however, I find that, first, the smaller of the two groups features tighter collaboration, and, second, a high ability of an agent does not necessarily transform into a strong position in the competition.

Both results stem from the following observation. A high-ability member of the larger group can be easily replaced by an outsider with a low ability. Indeed, links that they can both generate are the same from the others' point of view. The high-ability agent is a strong competitor and a potential threat to members of the larger group. Hence, he is less welcome in the larger group unless he pretends to have a low ability by restricting his output. In the stable networks that I find, high-ability agents follow this strategy of restricting their own output in exchange for membership in the larger group.

I do not study collaboration in competition in its full generality. Instead, I focus on one aspect of it that was *not* examined previously in the literature. Cooperative and forward-looking agents may collectively manipulate their connections to gain an advantage in a competition. Such behavior results in structural holes in networks of collaboration. Such networks are inefficient since they favor some agents more than others. This aspect is best explained with an example.

Consider three agents, A, B and C. In my model, the efficient network for these three agents is a triangle (complete network). Imagine that agent A deviates from the efficient network by dropping his link with agent C. In the resulting network, agent B has two links, and agents A and C each have one link. The absence of a link between agent A and agent C imposes a positive externality on agent B, since he becomes the only winner in the competition. Now imagine that agent B reciprocates to agent A's action and also drops his link with agent C. Let us compare the resulting network with only one link, $A \leftrightarrow B$[5], to the efficient network $\{A \leftrightarrow B, A \leftrightarrow C, B \leftrightarrow C\}$. On the one hand, both A and B bear some losses when they delete their links with C. On the other hand, in the network with one link, both A and B get half of the prize, whereas in the efficient network, they get only one third each. For some parameters, the benefits outweigh the losses, and, therefore, it is profitable for the coalition $\{A, B\}$ to drop their links with C.

One may object that such a deviation by agents A and B does not lead to a stable network because there might be further deviations from it. For instance, agents A and C both find it profitable to revive their missing link: Agent A becomes the only winner, and agent C gains a link. There are, indeed, deviations from the network with only one link, $A \leftrightarrow B$, but I argue that they are not credible. Suppose, that all three agents perceive networks $\{A \leftrightarrow B\}$, $\{A \leftrightarrow C\}$, and $\{B \leftrightarrow C\}$ as stable. Then, the deviation by agents A and C will be followed by agents B and C, who will drop their

[5] A link between two agents, A and B, is denoted by $A \leftrightarrow B$. All notations are introduced in Section 2.

links with A and connect with each other. This counter-deviation will lead us back into a stable set of networks. This observation supports agents' beliefs about stability of networks $\{A \leftrightarrow B\}$, $\{A \leftrightarrow C\}$, and $\{B \leftrightarrow C\}$.

This idea may be extended to a setup with an arbitrary number of agents. Let us start with a complete network. A group of agents that constitutes a majority can unilaterally terminate all its collaboration with the rest of the agents. This action leads to losses in output caused by the absence of some links. Notice, that this loss is asymmetric across members of the larger and the smaller groups. Suppose that the sizes of the groups are k_1 and k_2 such that $k_1 > k_2$. The loss that an agent incurs is determined by the number of links he loses, which is k_2 for a typical member of the larger group and k_1 for a member of the smaller one. The difference of losses in output gives members of the larger group an advantage in the competition and allows them to keep the prize within their group. Notice that members of the smaller group are powerless against such a tactics by the larger group since one side can unilaterally severe all the links that connect the two groups.

The rest of the papers is organized as follows. I set up the model in Section 2. In Section 3, I introduce the solution concept for my model and discuss its properties. The main results are presented in Section 4. I extend the model to accommodate heterogeneity in agents' talents in Sections 5. Related literature is discussed in Section 6. Finally, Section 7 concludes. All proofs are in the Appendix.

2 Framework

There are N identical agents in my model. Each agent produces an output and competes against the others in a tournament. The agents who produce the largest output win the tournament.

A key ingredient of my model is collaboration between competitors. The more the agents collaborate, the higher their outputs are. I assume that collaboration takes place in pairs. Although collaboration in groups is not allowed as a primitive of the model, it can be represented as a set of the collaborations in pairs. I do not put any restriction on the number of partners agents can have.

I represent a structure of collaboration relationships by a network. Naturally, nodes represent agents and links—collaboration. Some additional definitions related to networks are discussed in Section 2.1.

The agents set up their links with each other prior to producing the output and competing in the tournament. I assume that the network of collaboration is fully observable by all agents at each step of this process. An agent is free to choose his partners, but each link requires the consent of his counterpart. An existing link, however, can be unilaterally removed by any of the two participants.

I do not assume any particular protocol that agents have to follow in setting up their links and, instead, I use a cooperative approach.I define which deviations agents can pursue from a current network. After that, I define stable networks as ones that are immune to such deviations. I discuss this approach in detail in Section 3.

Once all links have been established, each agent produces an output. I assume that an agent's output is increasing in the number of his links: The output of an agent who collaborates with x partners, is given by $f(x)$. In addition, I assume that each new partner strictly improves the performance of the agent—i.e., $f(x)$ is strictly increasing.

Finally, once all agents have produced their output, they compete with each other in the tournament, and those with the maximal output are announced as the winners. The winners are awarded a prize, which they split equally among themselves. I denote the size of the prize by g and the number of winners by w. Thus, each winner gets $\frac{g}{w}$.

I assume that the collaboration is beneficial for the agents, even after I control for its effect on the competition outcome. This means that even an agent who has no chance of winning prefers to have as many links with others as possible. In particular, I assume that an agent receives his output as a part of his payoff.[6] This assumption holds in many applications; the compensation schemes for employees, for example, usually take into account both their absolute and relative performance. In this example, the part of the compensation that depends on the employee's absolute performance is determined by, among many other things, the extent of the employee's collaboration with others.

Notice that both the division of the prize and the agents' outputs are determined by a current network. Suppose that an agent i has x_i links in the network. By x_{-i}, I denote the vector that contains the numbers of links of all other agents. Then, the agent i's payoff is given by

$$u_i(x_i, x_{-i}) = \begin{cases} f(x_i) + \frac{g}{w} & \text{, if } x_i \geq x_j \text{ for all } j \in N \\ f(x_i) & \text{, otherwise.} \end{cases}$$

This payoff consists of two parts discussed above: the share of the prize $\frac{g}{w}$ that the agent receives if he wins the tournament, and the agent's output $f(x_i)$.

I do not allow monetary transfers between the agents in this model. In general, the possibility of transfers has a stark effect on the outcome of the network-formation process (see, for instance, Dutta and Mutuswami (1997)). In particular, if transfers are allowed and no additional constraints are imposed, efficiency can be achieved. However, if I require component-neutrality of transfers, as Dutta and Mutuswami (1997) do, my results still hold.

[6]The assumption that agent's payoff is linear in his output is not very restrictive, since output plays only an ordinal role in the tournament. I could have assumed that this part of agent's payoff is an increasing function of his output, and such assumption would not change the results.

2.1 Networks

As already mentioned, the structure of the collaboration is modeled using a network, in which nodes represent the agents, and links—the collaborations.

The set N of agents is the set of nodes (or vertexes). Let Λ be a set of all subsets of N of size two. An element $a \leftrightarrow b \in \Lambda$ represents the link (or the edge) between agents a and b. Nodes a and b are called the ends of the edge $a \leftrightarrow b$. A network ψ is a collection of links between agents—i.e., a subset of Λ. Therefore, the set 2^{Λ} is the collection of all networks on N.

Let $\psi \in 2^{\Lambda}$. With some abuse of notations, I say that $\psi(a,b) = 1$ if $a \leftrightarrow b \in \psi$, and $\psi(a,b) = 0$ otherwise. Also, let $M \subset N$. By $E_M(\psi)$, I denote the set of edges that have the elements of the set M as their ends (if $M = N$, I drop the index and write $E(\psi)$). Also, by $I_M(\psi)$, I denote the set of edges that have *only* the elements of the set M as their ends. Finally, $O_M(\psi) \equiv E_M(\psi) \backslash I_M(\psi)$.

Any set operation on networks are taken with respect to the set of edges. For example, for $\psi, \gamma \in 2^{\Lambda}$, $\psi \cap \gamma$ is the network that satisfies $E(\psi \cap \gamma) = E(\psi) \cap E(\gamma)$.

A path between two nodes, a and b, is a collection of edges of the form $\{a_0 \leftrightarrow a_1, a_1 \leftrightarrow a_2, \ldots, a_{k-1} \leftrightarrow a_k\}$ such that $a_0 = a, a_k = b$ and $a_i \neq a_j$ for any $i \neq j$. If such a path exists in ψ, I say that a and b are connected in ψ. The set of nodes M is connected in ψ if any $a, b \in M$ are connected in ψ. The maximal (in the set-inclusion sense) connected set of nodes M is called a component. In this case, with a slight abuse of notations, I call $E_M(\psi)$ a component as well. Component M is regular if, for any $a, b \in M : |E_a(\Psi)| = |E_b(M)|$, and is complete if for any $a \in M : |E_a(\psi) = |M| - 1$.

By $M(\psi)$, I denote the set of nodes with the maximal number of edges in ψ:

$$M(\psi) = \arg\max_{k \in N} \{|E_k(\psi)|\}.$$

Given these notations, we can rewrite agent i's utility in network ψ as

$$U_i(\psi) = f(|E_i(\psi)|) + \mathbf{1}_{i \in M(\psi)} \frac{g}{|M(\psi)|}.$$

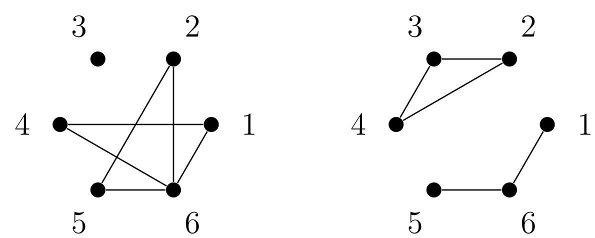
Throughout the paper, I refer to the tournament winners as "insiders" and to the rest as "outsiders." The meaning of these terms will become more clear once I present the results.

3 Stability

In this section, I define the solution concept that I use throughout the paper. Before proceeding to the formal definitions, let me say, that I follow the approach of cooperative game theory. I use the stable sets of von Neumann and Morgenstern (1944) combined with the notion of farsighted agents first introduced by Harsanyi (1974). A similar approach is used in the literature on networks, two-sided matching, and coalition formation (see Ray and Vohra (1997), Diamantoudi and Xue (2007), Herings et al. (2009) and others). Although this modeling choice might seem to be ad hoc at this stage, I invite the reader to follow me to the end of this section. After introducing the definitions, I discuss the strengths and limitations of this solution in detail.

The solution that I use for my model is a stable set of networks, which is defined to be immune to *credible* deviations by coalitions of agents. Before introducing the formal definition of a stable set, I must describe what these coalitions *can* do and what they *would like* to do.

Previously, I mentioned that a link between two agents can appear only if they both agree to it, but that an existing link can be broken even if only one of the two partners drops it. This rule naturally extends to the coalitions of agents. Let us look, for example, at networks ψ_1 and ψ_2 in Figure 1. What coalition of agents can enforce



(a) Network ψ_1; (b) Network ψ_2.

Figure 1

the transition from network ψ_1 to the network ψ_2? Observe that links $2 \leftrightarrow 3$, $3 \leftrightarrow 4$, and $2 \leftrightarrow 4$ do not exist in network ψ_1, but do exist in network ψ_2. This means that the participation of agents 2, 3, and 4 is necessary for this transition. Also, observe that links $1 \leftrightarrow 4$, $4 \leftrightarrow 6$, $2 \leftrightarrow 5$, and $2 \leftrightarrow 6$ must be removed in order to reach the network ψ_2. Agents 2 and 4 can do that as well, so we can conclude that coalition $\{2, 3, 4\}$ can enforce the transition from the network ψ_1 to the network ψ_2.

The following definition generalizes this principle.

Definition 1. *A coalition S can enforce a transition from a network ψ to γ or*

$$\psi \xrightarrow{S} \gamma$$

if for all $i, j \in N, i \neq j$ the following holds:

(i) $\gamma(i,j) > \psi(i,j) \implies \{i,j\} \subset S$*;*

(ii) $\gamma(i,j) < \psi(i,j) \implies \{i,j\} \cap S \neq \emptyset$*.*

Observe that there is an asymmetry in the process of creation and removal of links. In particular, $\psi \xrightarrow{S} \gamma$ does *not* imply that coalition S can reverse this process—i.e.,

$\gamma \overset{S}{\to} \psi$. If we go back to the example in Figure 1, we can see that $\psi_1 \overset{\{2,3,4\}}{\to} \psi_2$ and $\psi_2 \overset{\{1,2,4,5,6\}}{\to} \psi_1$.

The previous definition states only what agents can do, but it does not take into account their incentives. The traditional approach says that the transition will take place if all active participants are better off. It is captured by the notion of *blocking*: If transition from γ to ψ can be enforced by some coalition of agents that benefit from the transition, network ψ blocks network γ. Depending on which coalitions are allowed to act, there are notions of setwise or pairwise (myopic) blocking (see Jackson (2008), Goyal (2007), Roth and Sotomayor (1992)).

Definition 2. *A network ψ setwise blocks γ or*

$$\psi \rhd \gamma$$

if there exists a coalition S such, that

(i) $\gamma \overset{S}{\to} \psi$

(ii) $\mathbf{U}_S(\psi) > \mathbf{U}_S(\psi)$.[7]

In addition, ψ pairwise blocks γ (or $\psi \rhd_p \gamma$) if $|S| = 2$.

Usually, a stable network is defined as one that is not blocked by any other network. I depart from this approach for several reasons. First, according to this definition, agents cannot plan their actions ahead. For instance, there can be a situation in which an agent prefers to have a certain number of links or no links at all.[8] If he is unconnected, it follows from this definition, that he does not want to create the first link because he can not foresee that he will have a chance to create the second one.

Second, agents evaluate their actions as if a network that results from a transition is always stable. This requirement is excessively strong, and stable outcomes often do not exist. In particular, in my model, I can always find reasonable parameters for which the set of unblocked networks (or an abstract core defined with respect to $\rhd$) is empty.

Proposition 1. *Suppose that $N > 2$. There is always g and $f(\cdot)$ such that no setwise unblocked networks exist.*

This issue is not just a technical detail. The fact that all possible networks are blocked raises a concern that some blocking networks are not credible (because they themselves are blocked as well). It is possible to weaken the definition of a stable network by ruling out non-credible blocks and to still get reasonable predictions.

[7]For two vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} > \mathbf{y}$ if $x_i > y_i$ for all i.

[8]It happens, for example, when links are complementary and costly.

Suppose that there is a sequence of transitions that is enforced by a sequence of coalitions:

$$\psi_1 \overset{S_1}{\to} \psi_2 \overset{S_2}{\to} ... \overset{S_K}{\to} \psi_{K+1}.$$

I do not require that each step be an immediate improvement for the acting coalition. Instead, all acting coalitions must prefer the ultimate outcome of this transition to the networks in which they are asked to act. For example, the coalition S_i must prefer the network ψ_K to the network ψ_i. Otherwise, if this property is violated for some coalition S_i, the transition will stop at step i and the network ψ_K will never be reached. One might argue, that if coalition S_i refuses to act, there might be some other coalition $\widehat{S}_i$ that can step in and proceed with the transition. It is indeed true, and I am going to account for it by looking at *all* possible transition paths.

Let us formally define the blocking for farsighted agents.

Definition 3. *A network ψ setwise farsightedly blocks γ or*

$$\psi \succ \gamma$$

if there exists a sequence $\{(S_i, \phi_i)\}_{i=1}^K$, $\forall i = 1, ..., K : S_i \subset N$ and $\phi_i \in 2^\Lambda$ such, that

(i) $\gamma = \phi_1 \overset{S_1}{\to} \phi_2 \overset{S_2}{\to} \dots \overset{S_K}{\to} \psi$

(ii) $\mathbf{U}_{S_k}(\psi) > \mathbf{U}_{S_k}(\phi_k)$ *for all* $k \leq K$.

A sequence $\{(S_i, \phi_i)\}_{i=1}^K$ that satisfies properties (i) and (ii) is called γ-to-ψ path.

This definition is widely used in the literature on coalition formation with forward-looking agents. Ray (2007) discusses this notion in detail. The main difference is that I use the notion of enforceability that is tailored for network-formation problems.

Notice that setwise farsighted blocking is weaker[9] than setwise blocking; hence, Proposition 1 also applies to it. Therefore, instead of looking at unblocked networks, I look at networks that are unblocked only by stable networks, and I define the former to be members of the latter. Put differently, a stable set of networks consists of *all* networks that are unblocked by other networks in this set.

Such sets are fully characterized by internal and external stability.

Definition 4. *A set of networks $\mathcal{R}$ is von Neumann-Morgenstern farsightedly stable, if it satisfies the following two conditions:*

(IS) for any $\psi, \gamma \in \mathcal{R} : \psi \not\succ \gamma$;

(ES) for any $\gamma \notin \mathcal{R}$ *there exist* $\psi \in \mathcal{R} : \psi \succ \gamma$.

[9] Formally, $\rhd \subset \succ$.

Let $A(\cdot)$ be a function that for a given set of networks $\mathcal{X}$ returns a set $A(\mathcal{X})$ of all networks that are unblocked by any network in $\mathcal{X}$. Then, $\mathcal{R}$ is von Neumann-Morgenstern farsightedly stable if and only if

$$\mathcal{R} = A(\mathcal{R}).$$

It is implicitly assumed in the definition that the agents believe that a set $\mathcal{R}$ consists of stable networks. This belief, however, is reinforced by the two properties of internal and external stability. In contrast with the core, the stability of an element of the stable set relies on other elements of the same set. It is also implicitly assumed that agents are optimistic: If they find themselves in the network outside of a stable set, it is enough for them to have only one path back into the set to believe that this set is actually stable. All agents that are active in the transition believe that there are going to be no deviations along the path and that they will definitely reach the stable set.

In general, there exist numerous myopically profitable deviations from any stable set $\mathcal{R}$. The definition of a stable set requires that, for any deviation, there exists a path back to the stable set. However, there is no explicit requirement that the original deviators are worse off once the new stable networks is reached. One might think that agents could use deviations to switch from a less desirable stable network to a more desirable one. It turns out that this is not the case. Such a property of stable sets follows from the combination of internal and external stability. It is called *consistency* and it was introduced by Chwe (1994).

Suppose that a coalition of players deviates from a stable network to some network outside of the stable set. Then, there always exists a path back to the stable set such, that at least one of the deviators is worse off compared to the original stable network. Since there is always at least one agents who is punished, the deviations will not take place.

Chwe (1994) states, that stable sets are immune to one-step deviations by a coalition of players. This result can be easily extended to a sequential deviation as well.

Proposition 2 (Chwe, 1994)**.** *Let $\mathcal{R}$ be a stable set, and $\rho \in \mathcal{R}$ be a stable network. Take any $\psi \succ \rho$, and any ρ-to-ψ path $\{(S_k, \gamma_k)\}_{k=1}^{K}$. For any stable network $\widehat{\rho} \in \mathcal{R}$ such that $\widehat{\rho} \succ \psi$, there exists an agent $i \in \bigcup_{k=1}^{K} S_k$ such that*

$$U_i(\rho) \geq U_i(\widehat{\rho}).$$

As already mentioned, most of the literature on network formation in general and on collaboration in particular uses pairwise-stability notion a la Jackson and Wolinsky (1996). Although pairwise stability is a reasonable solution for large networks in which the effect of a single agent on the network is negligible (or, put it differently, agents are network-takers), it is unequal to the task of understanding the environments in which a single agent has much control over the network structure. This is the main reason why my approach is very different from the traditional one.

Herings et al. (2009) study a similar stability notion. They, however, look at pairwise farsighted blocking and they replace internal stability with the weaker condition. Although this weakening guarantees existence, it also makes their stability less discriminative: Some networks that can never be stable in my framework are stable in theirs. Similarly, one can use the adaptation of Chwe's largest consistent set, which is always non-empty, to solve a network-formation problem, but as in case of Herings et al. (2009), the largest consistent set has little predictive power.

For the rest of the paper, for conciseness, I often write "block" to mean "setwise farsightedly block" and "stable" to mean "von Neumann-Morgenstern farsightedly stable". Also, I refer to networks from a given stable set as stable. However, one should keep in mind that these networks are stable *with respect to* the corresponding stable set. Finally, one should not confuse stable networks with elements of an abstract core.

4 Results

There are two main questions to be answered. Under what conditions are efficient networks stable? If efficient networks are not stable, then what networks are?

To answer these questions, I must first characterize efficient networks. In my setup, the complete network Λ is the unique efficient network. Intuitively, it is easy to see: The distribution of the prize in the tournament does not have an effect on aggregate welfare, since the sum of the prizes is constant and equal to g. Therefore, an efficient network is one that has the maximal aggregate output. Recall that any link strictly improves the outputs of the two agents that are connected by it; thus, the complete network is the unique efficient network.

Proposition 3. *The complete network Λ is the unique efficient network.*

Let us start with the first question. What are the necessary and sufficient conditions for the unique efficient network Λ to belong to a stable set? The following theorem provides these conditions.

Theorem 1. *There exists a stable set $\mathcal{R}$, and a network $\rho \in \mathcal{R}$ such that $M(\rho) = N$ if and only if*

$$N \in \underset{\frac{N}{2} < m \leq N}{\arg\max} \left\{ f(m-1) + \frac{g}{m} \right\}. \tag{1}$$

Theorem 1 does not mention the efficient network Λ explicitly. Notice, however, that the efficient network Λ is symmetric; hence, $M(\Lambda) = N$. To obtain the desired characterization, one has to combine this observation with Theorem 1.

Corollary 2. *The efficient network* Λ *belongs to some stable set if and only if (1) holds or, equivalently, if and only if*

$$\frac{g}{N} \leq \min_{\frac{N}{2} < m \leq N} \left\{ \frac{m}{N-m} \left(f(N-1) - f(m-1) \right) \right\}$$

This result is very different from the analogs obtained in the literature (see, for example, Goyal (2007) or Marinucci and Vergote (2011)). Previous literature argues that the efficient outcome is *always* stable, and, in addition, some inefficient outcomes can also be stable. In contrast with that, Corollary 2 says that the efficient network *cannot* be stable if condition (1) is not satisfied.

This difference in results is due to the assumption that the agents are forward-looking and are able to coordinate with their peers when creating their connections. For a moment, I focus on forces that drive the result in Corollary 2 since they have not been discussed in the existing literature.

In my model, any agent in any network benefits from creating an extra link. Let us look at an arbitrary agents who is not collaborating with some other agents—i.e., the number of his links $x < N - 1$. An extra link for this agent is always profitable since he gets an extra payoff of $f(x+1) - f(x) > 0$, and his chances of winning the tournament weakly increase.

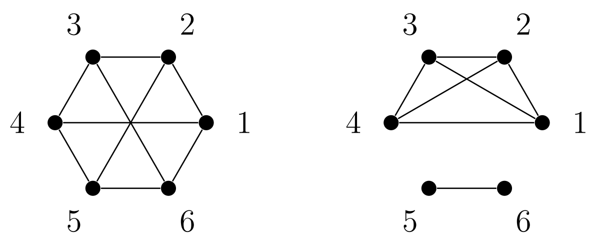
This agent has two options: to create the missing link or to leave things as they are. As I mentioned earlier, if the agent chooses the first option, his payoffs strictly increase. However, if he chooses the second option, or alternatively if he terminates one of his links, he imposes a *positive externality* on the rest of his competitors. Indeed, if these two agents do not collaborate, their performance in the tournament is lower than it could be. As a result, their competitors' chances of winning are higher. The price that the agent who refuses to collaborate pays for imposing this externality is the opportunity cost.

Such an externality is common not only for winner-takes-all tournaments but also for other forms of competition. Intuitively, if one competitor is slowed down, others are better off. The literature on sabotage in tournaments shows that, when a marginal costs of slowing an opponent down is lower than marginal costs of increasing own performance, sabotage is present in equilibrium (see Lazear (1989), Chen (2003), Konrad (2000)). The positive externality discussed above plays a special role in these models: Konrad (2000) shows that it dissolves the effect of sabotage when the number of participants is large.

In my paper this externality also plays a central role, but its effect is opposite to what Konrad (2000) finds. Agents internalize the effect of the externality and collectively invest in sabotaging their opponents. Note, that sabotage in my model is very costly: an agent that terminates a link is *always* worse off, thus sabotage is individually suboptimal, and standard models predict that agents do not engage in sabotage in equilibrium. On the contrary, I find that there is still room for sabotage through termination of collaboration in this case. since this sabotage is valuable for active saboteurs: The benefits of the positive externality, if it is properly exploited, can outweigh the total opportunity costs of all agents involved. One way to exploit the externality is for agents to coordinate their missing links.

To illustrate how coordination helps to exploit the externality, let us look at two networks in Figure 2. In the regular network ψ_r, all agents have three links each; thus they each get a payoff of $f(3) + \frac{g}{6}$. Any agent in this network misses two links, but the externality that this agent imposes on others is canceled out by the externality others impose on him.

Observe, however, that agents 1,2,3, and 4 can coordinate their missing links and use the externality they impose on each other for their own good. In particular, they can delete their links to agents 5 and 6 and create missing links with each other. In the resulting network, ψ_n, each of them still has three links, but their payoff, $f(3)+\frac{g}{4}$, is higher than in ψ_r since agents 5 and 6 do not get their share of the prize anymore.



(a) Regular network ψ_r; (b) Nonregular network ψ_n;

Figure 2: Externality in regular and nonregular networks.

In this case, the agents who coordinate their actions to exploit the externality have the same number of links at the beginning and at the end. This does not have to be the case—a group of agents might be willing to lose some links if, in return, they get a high enough share of the prize.

This effect drives the results in Theorem 1 and Corollary 2. In particular, if the condition (1) does not hold, it means that there exists $m \in (\frac{N}{2}, N)$, such that

$$f(m-1) + \frac{g}{m} > f(N-1) + \frac{g}{N}.$$

This inequality guarantees that one can find a group of m agents who are willing to disconnect themselves from the rest in order to be the only winners in the competition.

Needless to say, the existence of such a deviation from the efficient network alone does not guarantee that the efficient network cannot be stable. To prove the result, one must show that such deviation is credible—i.e., is not blocked by other credible networks.

Once m agents disconnected themselves from others, they will not act on returning back to the efficient network. Formally speaking, the efficient network does not block current network. In order for the efficient network to be an element of a stable set, there must be another network, let us call it ψ, that blocks the current one (otherwise, the deviation discussed above is credible and, hence, the efficient network is not stable).

In order for networks ψ to block the current network, some of the current m winners must be offered an increase in payoffs if they switch to network ψ. This means that network ψ either has very many or very few links. In either case, agents that do not win the tournament in network ψ will be able to enforce the transition to the efficient network. Given that, some of the agents who are essential for the transition from the current network into network ψ will not pursue it: They know that this transition will be followed by another one, and they will end up in the efficient network. Formally, this means that $\Lambda \succ \psi$ and the internal stability is violated. Intuitively, this means that agents cannot be credibly promised to stay in network ψ. The following example illustrates this point.

Example 1. Suppose that there are three agents: 1,2, and 3. Also, suppose that the prize in the tournament, g, satisfies

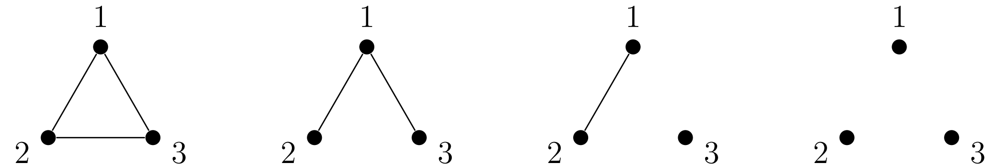
$$g > 6(f(2) - f(1)).$$

Clearly, in this case, condition (1) is violated, so Corollary 2 tell us that the efficient network cannot be stable. Let us establish it without using the corollary. I will show that, in this case, the unique stable set consists of three networks, each consisting of a single link: $\mathcal{R} = \{\{1 \leftrightarrow 2\}, \{2 \leftrightarrow 3\}, \{1 \leftrightarrow 3\}\}$.

Now consider Figure 3, which depicts all possible networks up to a permutation of the agents. First, notice that for any network except ψ_2, there are two agents who strictly prefer some network in $\mathcal{R}$ to the original network. In particular, agents 1 and 2 prefer network $\{1 \leftrightarrow 2\}$ over both Λ and ψ_3, and agents 2 and 3 prefer network $\{2 \leftrightarrow 3\}$ over ψ_1. Since these pairs are sufficient to enforce these transitions, $\mathcal{R}$ is externally stable.

The internal stability of $\mathcal{R}$ is also easy to see. As all connected agents are indifferent between their partners, they will not participate in any transition between the networks in $\mathcal{R}$. So far, I have established that the set $\mathcal{R}$ is stable. The question is if there are any other stable sets. The answer is "no," and I prove it by contradiction.

Assume that there is another stable set $\mathcal{P}$. This stable set must contain the network ψ_1 since it is the only network (up to a permutation of agents) that blocks the network ψ_2. However, the network ψ_1 does not block the efficient network Λ; hence, the latter should also belong to the stable set. The desired contradiction arises because agents 2 and 3 can enforce the transition from ψ_1 to Λ and, moreover, they both strictly prefer Λ over ψ. This means that $\Lambda \succ \psi_1$ and the internal stability of the set $\mathcal{P}$ is violated.



(a) Network Λ (b) Network ψ_1 (c) Network ψ_2 (d) Network ψ_3

Figure 3: Example 1

Intuitively, although the network ψ_1 blocks the network ψ_2, this block is not credible. One of the two agents who enforce the transition—namely, agent 1—foresees that after he collaborates with agent 3, the latter will also collaborate with agent 2. Moreover, agent 1 thinks of the collaboration between agents 2 and 3 (i.e., of $\Lambda \succ \psi_1$) as being credible since the other network ψ_1, which is (wrongly) assumed to be stable, does not block Λ.

The novel tension between stability and efficiency presented in Corollary 2 is particularly striking if one tries to approach this model with the traditional stability concept—i.e., pairwise stability. If agents are myopic (i.e., always act towards immediate increase in their payoffs as in Jackson and Wolinsky (1996)), the pairwise stability unambiguously predicts that the efficient network is the unique stable one (see, for example, Goyal and Joshi (2003) or Goyal (2007)).

Theorem 1 offers more than just the necessary and sufficient conditions for the stability of the efficient outcome. It also establishes that if the efficient network is not stable, then any stable network features ex post inequality in agents' payoffs. Even though agents are identical, if condition (1) does not hold, it is impossible that they all have the same payoff in a stable outcome. Some agents are necessarily discriminated against by others.

What are the implications of Theorem 1 for the examples we discussed in the Introduction? Let us look at firms that compete with each other in the market. Often there is a room for collaboration between such firms. Hagedoorn (2002) reports that there has been a sharp increase in collaboration between competing firms in the period of 1980–1998; this collaboration mostly took a form of sharing agreements. One example of such a sharing agreement, that fits our model well, is cross-licensing. Cross-licensing agreements allow participants to use each other's patent portfolios without paying royalties. Signing such an agreement is virtually costless. The firms that have free access to their rivals' patents gain a competitive advantage over firms that have to pay royalties for using the same patents. This advantage, however, does not come for free: A typical cross-licensing agreement also allows a firm's counterparts to access its patents and, hence, makes them stronger rivals.

In such markets, we often observe patent wars. In a course of a patent war two groups of firms "attack" each other by filing patent infringement lawsuits. These groups feature tight collaboration that strengthens their positions against the opponents. Cross-licensing agreements between the allies are one of the instruments of such a collaboration and are often used in patent wars[10] .

Applied to this environment, Theorem 1 predicts that if the stakes in the competition are high (i.e., the surplus that firms can extract from operating in this market is large), patent wars between alliances of firms are unavoidable. Groups (or alliances) of firms manipulate the structure of the collaboration network in order to hinder the performance of their competitors. The resulting network of collaboration is nonregular, and firms that sign more cross-licensing agreements gain a larger share of the market.

Another application of Theorem 1 is discrimination. McAdams (1995) argues that race discrimination in the U.S. exists to maintain the gap in social status between whites and blacks. People who value being wealthier than others, may sacrifice some mutually beneficial connections in order to maintain the gap in wealth between discriminating and discriminated groups.

Theorem 1, in this case, states that if the concern for social status is significant, discrimination necessarily arises in the stable network of collaboration. It is important to note that this theory predicts discrimination that arises in a society of identical agents. This prediction is consistent with the observation that discrimination happens along economically irrelevant markers such as skin color or ethnicity.

My result is similar to one obtained in Pęski and Szentes (2011), as the discrimination arises in the population of identical agents (that may only differ in payoff-irrelevant characteristics) and is supported by a social norm[11]. This social norm prescribes that a group of agents must discriminate against the other group even if discrimination is not individually optimal. If an agent refuses to discriminate, he gets ostracized (i.e., others start to discriminate against this agents). I discuss the social norm that enforce discrimination in detail after I present next result (Theorem 3).

The main difference between Pęski and Szentes's spontaneous discrimination and result in Theorem 1 is the underlying mechanism. Spontaneous discrimination arises as a contagious failure of coordination: If there is a small amount of discrimination present, fear of being ostracized makes everyone starts following this social norm. In particular, the equilibrium with no discrimination Pareto-dominates equilibria with discrimination. In contrast with that, in my model, discrimination arises because it is collectively optimal: A majority always discriminates against a minority to ensure a difference in social status.

Of course, one has to be cautious when applying Theorem 1 to the environments discussed above because my framework is extremely stylized. In order to obtain more-

[10]A recent example is a struggle between the alliance of Microsoft and Apple on one side and Google on the other side; these two alliances were engaged in a patent war over the market for smartphones throughout 2011.

[11]There are numerous other explanations for race discrimination, such as taste-based discrimination as in Becker (1971), or statistical discrimination as in Phelps (1972) and Arrow (1973).

accurate predictions, the framework has to be modified to accommodate the specifics of a particular environment.

Although Theorem 1 hints at some properties of stable sets, it does not tell us what networks are stable. This brings me to the second question of this section: If the efficient network is not stable, then what networks are? The next theorem provides the answer to this question.

Theorem 3. *Let,*

$$\mathcal{M}^* = \arg\max_{\frac{N}{2}<m\leq N} \left\{ f(m-1) + \frac{g}{m} \right\}, \text{ and} \tag{2}$$

$$\mathcal{G}(m) = \left\{ \psi \in 2^{\Lambda} \mid \exists M \subset N : |M| = m, \psi(i,j) = \begin{cases} 0, & \text{if } |\{i,j\} \cap M| = 1 \\ 1, & \text{otherwise} \end{cases} \right\} \tag{3}$$

The set of networks $\mathcal{G}^* = \bigcup_{m\in\mathcal{M}^*} \mathcal{G}(m)$ *always exists and is stable.*

I construct a stable set of networks for any set of the parameters. This construction consists of several steps. First, I find the maximizer for the problem that was introduced in Theorem 1. For simplicity, let us assume that the maximizer is unique—i.e.,

$$m^* = \arg\max_{\frac{N}{2}<m\leq N} \left\{ f(m-1) + \frac{g}{m} \right\}.$$

Take the set M that consists of any m^* agents, and look at the network in which the set M and its complement $N\backslash M$ form two complete components—the agents in M are connected to all other agents in M and are not connected to the agents outside of M (similarly, the agents in $N\backslash M$ are connected only to all other agents in $N\backslash M$). Let us denote this network by γ. The set of networks $\mathcal{G}^*$ obtained from γ by all permutations of the agents is stable. The maximizer m^* determines the size of the group of winners in any network that belongs to the constructed stable set.

Note that such a set always exists since the maximization problem (2) is finite. The example of the set $\mathcal{G}^*$ for $N = 4$ and $m^* = 3$ is given in Figure 4.

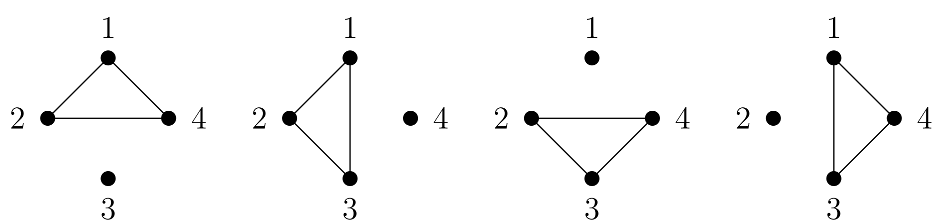


Figure 4: Set of networks $\mathcal{G}^*$ for $N = 4$ and $m^* = 3$

Observe that in any network in the stable set $\mathcal{G}^*$, the set of winners (or insiders) does not have any links with the set of losers (or outsiders). The insiders do not connect with the outsiders because, otherwise, the former expose themselves to the danger of being overthrown by the latter. In this set of stable outcomes, the competition is shifted to the group level. I do not allow for group competition in the fundamentals of the model, and yet, the agents organize endogenous groups and act as if these groups compete against each other.

The criterion for the size of the insiders' group resembles the so-called "union mentality." The size of the insiders' group is set to maximize the utility of its typical member, and, at the same time, neither the well-being of the outsiders nor any efficiency considerations are taken into account.

As I discussed above, a group of insiders coordinates on which links they sacrifice to eliminate their competitors. I notice that the number of links sacrificed is, in general, excessive. For example, for $N = 6$ and $m^* = 4$, the stable network is shown in Figure 5a. There exists a network in which the same set of agents wins the tournament and, at the same time, each agent has a strictly higher payoff (see Figure 5d)—and yet, this network is not stable. In this stable set, the winners can maintain their status only if they isolate themselves completely from the outsiders. In such a case, there is nothing the outsiders can do to destabilize the current network.

The subgroup of winners, in principle, may attempt to collaborate with outsiders and obtain an extra payoff. The stable set is constructed in such a way, that any attempt of this kind will be followed by excluding the deviators and replacing them with outsiders. This threat is credible since any deviation from the stable network decreases the payoffs of the agents that did participate in the deviation.

Let us go back to the example with firms signing cross-licensing agreements. Theorem 3 describes the structure of a particular stable set in which a group of agents (or, in this case, firms) forms a component that dominates the competition. Insiders' dominant position is enforced through the absence of links to the outsiders.

Bekkers et al. (2002) provide empirical evidence from the GSM industry that fits the predictions of Theorem 3. According to their overview, by the time the GSM standard was established, numerous companies held patents that are essential in developing the products that comply with the standard. Five of them—Ericsson, Nokia, Siemens, Motorola, and Alcatel—actively participated in cross-licensing their patents with each other. The same five companies later dominated the market for GSM infrastructure and terminals, with a total market share of 85 percent in 1996. At the same time, three other companies, Phillips, Bull, and Telia, held roughly as many patents as Alcatel and were not able to convert them into a significant market share. Moreover, they performed worse than Ericsson and Siemens, which did not have large patent portfolios and, yet, were ranked the largest and the third-largest GSM companies in 1996.

Bekkers et al. (2002) argue that the companies that dominated the GSM market in the 1990s could maintain the control over the market because they where able to shift the competition from the individual level to the group level. On the individual level, none of them had a competitive advantage over their rivals. However, they could gain an advantage on the group level by tailoring the size of their alliance.

As mentioned earlier the result in Theorem 3 can be applied to a problem of discrimination. This result combined with Proposition 2 explains the social norms that enforce discrimination. Agents belong to two mutually exclusive groups—majority and minority—and are allowed to collaborate only with members of their own group. If some member of a majority deviates from this norm and collaborates with an outsider, he gets expelled from majority.

The same as Pęski and Szentes (2011), my model does not explain why skin color is used as a criterion for discrimination. However, Theorem 3 shows that, whatever this criterion is, it must correspond to a certain majority-minority ratio. This can explain why in some countries there is a discrimination on the basis of other attributes such as ethnic origin, religious views or caste membership.

The set of networks $\mathcal{G}^*$ is not a unique stable set. However, this set is not arbitrary: It is crucial for understanding the driving forces behind Theorem 1. Consider the following example that provides an intuition for Theorem 1 and Theorem 3.

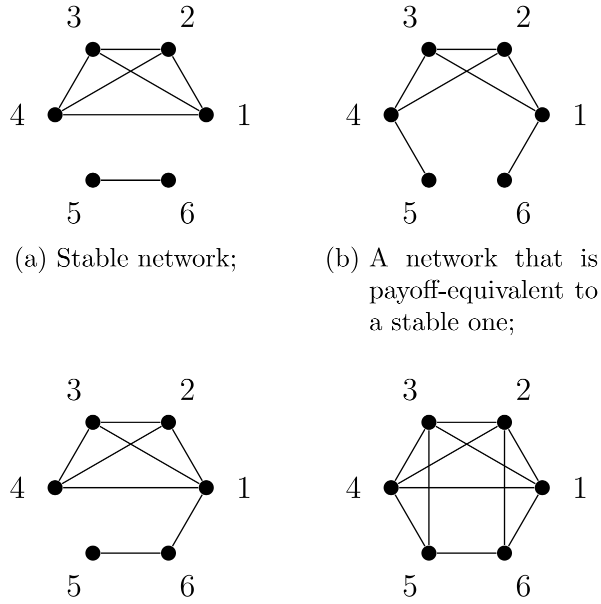
4.1 Six-Agent Example

Assume that $N = 6$ and $m^* = 4$. Theorem 3 tells us that there exists a stable set $\mathcal{G}^*$, which consists of the networks with two complete components of sizes 4 and 2. One such network is shown in Figure 5a.

The set $\mathcal{G}^*$ has several properties that are worth illustrating. First, it is internally stable (which is one of the requirements for stability). Second, in any network in set $\mathcal{G}^*$, the insiders are not connected to the outsiders. Third, set $\mathcal{G}^*$ is consistent. I discuss the intuition behind these properties in detail in the context of this example.

Let us start with internal stability. Set $\mathcal{G}^*$ is internally stable because any network in $\mathcal{G}^*$ consists of two complete components. Take any agent, say agent 1. In the network in Figure 5a, agent 1's payoff is $f(3) + \frac{g}{4}$. If agent 1 were in the smaller component, as agent 6 is in the current network, his payoff would be $f(1)$. These are the only two payoffs any agent receives in any network in $\mathcal{G}^*$. Since agents 1, 2, 3, and 4 already receive the maximum possible payoff, they would not want to switch to any other network in $\mathcal{G}^*$. The only agents that would prefer to switch to some other network in $\mathcal{G}^*$ are 5 and 6. However, the only action they can take is dropping the one link they have and switching to the network in Figure 6b. As a result of this action, the payoffs of agents 1, 2, 3, and 4 do not change; thus, agents 5 and 6 are stuck in one of the two networks in Figure. 6. They both strictly prefer the stable network in Figure 6a to the one in Figure 6b.

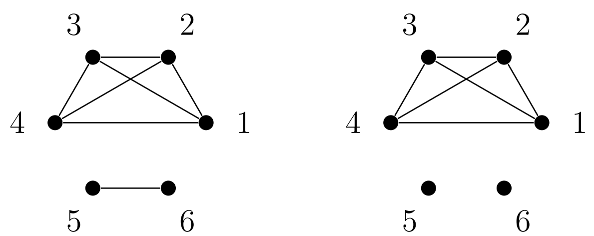
Each network in $\mathcal{G}^*$ features the group of insiders who control the distribution of the prize through enforcing a particular structure of the collaboration. For each agent, there are many sets of partners with whom he can take over the tournament. Intuitively, for internal stability to hold, agents must be indifferent between their



(a) Stable network; (b) A network that is payoff-equivalent to a stable one; (c) Deviation by agents 1 and 6; (d) The network that strictly Pareto dominates the stable one.

Figure 5: Stable network and deviations from it.

partners.



(a) Stable network; (b) The only possible deviation by agents who do not win the tournament.

Figure 6: Internal stability

The second property of $\mathcal{G}^*$ is that in any network in this set, the insiders do not collaborate with the outsiders. There is only one reason why some links are missing in any stable network, not only in the networks in $\mathcal{G}^*$. The irregularity of a network favors some agents who create it by assigning a share of the prize that is larger than

$\frac{g}{N}$ to these agents. Put differently, a group of agents weaken their rivals by not collaborating with them.

If the the prize is assigned to the agents with the highest performance, a minimal difference in performance is sufficient for the purpose of dominating the tournament. This minimal difference can be achieved by terminating fewer than $\frac{N}{4}$ links. Notice that for every network in the set $\mathcal{G}^*$, there are $m^*(N - m^*) \geq \frac{N}{4}$ missing links. For instance, in our current example, it is enough to drop one link to ensure that $m^* = 4$ agents win the tournament. However, in each network in the set $\mathcal{G}^*$, there are eight missing links.

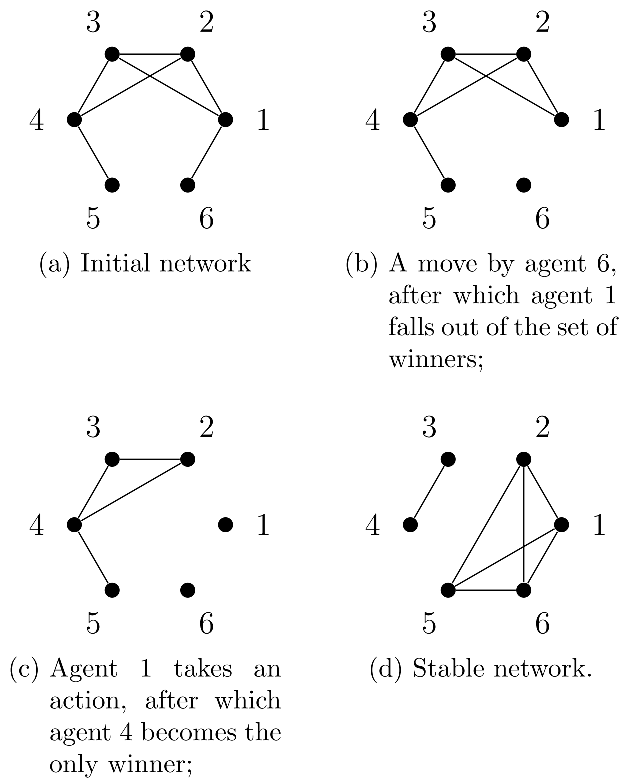
This extra inefficiency, which comes from terminating more links than necessary for ensuring the dominant position in the tournament, is the price for (internal) stability. To illustrate this, let me look at the network in Figure 5b. This network is payoff-equivalent to the stable one in Figure 5a since agents 1, 2, 3, and 4 have three links each, and agents 5 and 6 have one link each in both networks. The only difference between the two is that agents that win the tournament are connected to ones that do not. The links between insiders and outsiders create an opportunity for the the latter to become winners in the tournament. For example, agent 6 can start a transition to a stable network, in which he is one of the tournament winners. In order to do this, agent 6 has to drop his link to agent 1 (see Figure 7b). Once this is done, agent 1 no longer wins the tournament, and his payoff is low enough for him to participate in the transition. If agent 1 drops all his links, then agent 2 also drops out of the set of winners. At this point (see Figure 7c), there are five agents whose payoff is below $f(3) + \frac{g}{4}$. Four of them (agents 5, 6, 1, and 2) can form a new set of winners in the network in Figure 7d.

The same result holds even for the networks, that strictly Pareto dominate stable ones. Even though all six agents would myopically agree to switch from the stable network to the one in Figure 5d, they could not keep this deviation stable in the long run because agents 5 and 6 would be tempted to act further and to force a network that favors them the most.

Intuitively, links between insiders and outsiders stand in the way of the stability of large sets (set $\mathcal{G}^*$ consists of $C_N^{m^*}$ networks) because these links provide the outsiders with a tool to rearrange the structure of a network in their favor. By terminating these links, outsiders "invite" insiders to renegotiate their connections and their membership in the larger group.

As Proposition 2 states, any stable set is consistent: For any deviation from the stable set, there exists a credible counterdeviation back to the stable set, such that some of the agents who originally deviated prefer the initial network to the current one. This property ensures, that deviations can be punished in a consistent way.

To illustrate this, suppose that agents 1 and 6 deviate from the stable network in Figure 5a to the network in Figure 5c. If the latter network were stable, both agents would be better off, since both agents would get an extra link, and agent 1 would still be the winner in the tournament. It is clear, however, that the resulting



(a) Initial network

(b) A move by agent 6, after which agent 1 falls out of the set of winners;

(c) Agent 1 takes an action, after which agent 4 becomes the only winner;

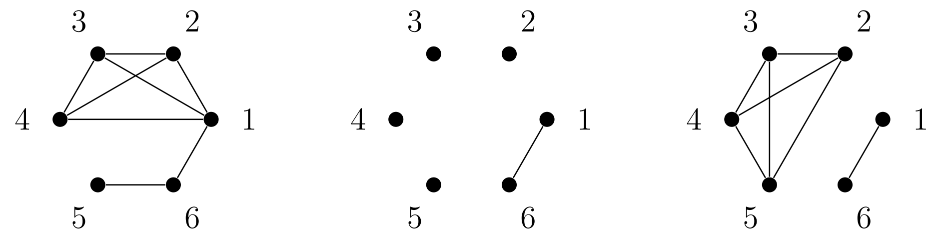
(d) Stable network.

Figure 7: Links between insiders and outsiders.

network is not stable. Moreover, there is a counterdeviation such, that agent 1 loses all but one of his links and drops out of the set of winners. Since, in the network in Figure 8a, agent 1 is the only winner in the tournament, agents 2, 3, 4, and 5 have low enough payoffs to act on switching to a stable network. These agents could drop all their links, leaving agents 1 and 6 in isolation (see Fig. 8b), and relink with each other later on. As a result, agents 2, 3, 4, and 5 would be the new set of insiders (see Fig. 8c). As agent 1 prefers the initial network in Figure 8a to the in Figure 8c, he would never attempt the considered deviation in the first place. Since all agents are the same, any insider that deviates from the stable network and “hurts” other insiders can be effectively replaced with some outsider. Such punishment is credible because any outsider is happy to take a place of an agent who participated in the deviation.

Consistency of a stable set does not guarantee that all deviators are punished. However, even if one of the participants in the deviation is credibly threatened, this particular deviation cannot take place. Since this result holds for *all* deviations, none of them is credible.

Although I do not provide the full characterization of stable sets, I point to some interesting properties that are common to all of them. Theorem 1 shows that, if a



(a) Deviation by agents 1 and 6; (b) Agents 1 and 6 in isolation. (c) Stable network.

Figure 8: A deviation by agents 1 and 6 and a path back into the stable set.

prize in the competition is above certain threashold, in all stable networks there are agents who are discriminated. No network with perfect equality of payoffs across all agents can be a part of a stable set.

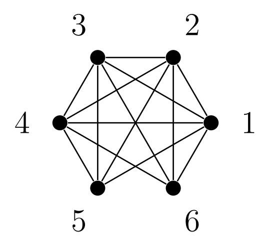


Figure 9: A stable singleton set.

For example, the singleton set in Figure 9 is always stable. In this network, agents 5 and 6 do not win the tournament, and their payoffs are strictly smaller than the ones of agents 1, 2, 3, and 4.

5 Heterogeneous Agents

When agents are identical, in the stable networks that I find, the group of winners is selected arbitrarily. A natural question, that arises, is how the agents split into insiders and outsiders if some of them become more productive than the others. Is it the case that the most (or least) productive agents become the most likely to be discriminated against? What happens to the payoff inequality across agents? How does it compare to the payoff inequality in the efficient outcome?

To answer these questions, I study an extended model that features the heterogeneous agents and various intensities of collaboration between them. In particular, I study the case in which the agents are heterogeneous in their ability to transform collaboration into output. I start by introducing various intensities of collaboration. After that, I proceed to the model with heterogeneous agents'.

I assume that an agreement between two agents i and j is characterized by two numbers x_{ij} and x_{ji} between 0 and 1. These numbers represent the intensity of the collaboration. In particular, x_{ij} represents the intensity of "help" that agent j provides to agent i. An agent i, after all agreements are finalized, receives an amount of "help" that is equal to

$$x_i = \sum_{j \neq i} x_{ij}.$$

The aggregate output (or performance) of an agent i in this case is $f(x_i)$.

The agents' payoffs consist of two parts. As in the symmetric model, all agents get their outputs as payoffs. In addition to this, agents benefit from outperforming their competitors. An agent i gets a prize g with probability

$$p_i = \frac{f(x_i)}{\sum_j f(x_j)}.$$

Agent i's aggregate payoff in this game is

$$u_i(x_i, x_{-i}) = f(x_i) + g\frac{f(x_i)}{\sum_j f(x_j)}.$$

The agreements in this setup are more complex than those in the symmetric model. In particular, it is possible that agent i lowers the performance of agent j without necessarily changing his own performance: Switching from the an agreement (x_{ij}, x_{ji}) to an agreement $(x_{ij}, \widehat{x}_{ji})$, where $\widehat{x}_{ji} < x_{ji}$, will achieve this effect.

A network ψ in this model is not just a collection of pairs from the set N. Instead, a network is a function that, for each *ordered* pair of agents, returns the intensity of the *directed* link: $\phi : N \times N \to [0, 1]$. Using our previous notation, if two agents, i and j, are connected by a link (x_{ij}, x_{ji}) in network ϕ, then $\phi(i, j) = x_{ij}$ and $\phi(j, i) = x_{ji}$.

The stability concept that was used for the basic model carries over to the extended model without any modifications.

I find that the result of Theorem 3 also holds in this framework. The only difference is that the size of the group of insiders is defined using the new payoff function. Similarly to the basic model, I can define

$$\mathcal{M}^* = \arg \max_{\frac{N}{2} < m \leq N} \left\{ f(m-1) + \frac{g/m}{1 + \frac{(N-m)f(N-m-1)}{mf(m-1)}} \right\} \quad (4)$$

and

$$\mathcal{G}^* = \bigcup_{m \in \mathcal{M}^*} \left\{ \gamma \mid \exists M \subset N : |M| = m, \gamma(i,j) = \begin{cases} 0, & \text{if } |\{i,j\} \cap M| = 1 \\ 1, & \text{otherwise} \end{cases} \right\}. \quad (5)$$

The set $\mathcal{G}^*$ is stable. Formally, this observation is a part of a more general result, which is stated later in this section (see Proposition 4 on page 29).

In the stable set $\mathcal{G}^*$, the internal collaboration within groups of winners and outsiders is at the maximal level, and the collaboration between groups is absent. This observation supports the statement that the absence of links between the two groups of agents is crucial for the stability of this configuration. Indeed, in the current extension of the model, agents can fine-tune their collaboration much more than in the basic model, and yet, they choose not to collaborate with some of their potential partners.

There are several dimensions of agents' heterogeneity relevant for this model. In this section I assume that agents are heterogeneous in productivity: By α_i, I denote agent i's ability to convert the incoming links into output. In addition, I assume that the agent's technology is linear in collaboration intensity:

$$f_i(x_i) = \alpha_i x_i.$$

The effect of another important dimension of heterogeneity—i.e., heterogeneity in the ability of agents to influence each others' performance in the competition through collaboration—remains an open question.

I impose an additional assumption on the distribution of the agents' abilities. I only look at the situations in which the difference between the most and the least talented agents is moderate.

Assumption 1.

$$\frac{\max\{\alpha_1, ..., \alpha_N\}}{\min\{\alpha_1, ..., \alpha_N\}} \leq \frac{(N+1)(N-1)}{(N-3)(N-5)}$$

Theorem 3 points to a stable set in which all networks consist of two complete components. If the above assumption is violated, the agents in the smaller component might prefer to split further into two new components. In the current model every agent who has link obtains a portion of the prize. If there is an agent with a very high ability, others might want to completely isolate him as he is too competitive.

I find stable sets of networks that resemble ones described in Theorem 3. Networks in the new stable sets feature two disconnected, tightly intralinked groups of agents. Note that the term "group of winners" is not applicable in this framework because any agent with nonzero output gets a certain share of the prize. Moreover, since agents are heterogeneous in their abilities, an agent with just a few links could, in principle, win a larger share of the prize than an agent with many links. Nevertheless, in the sets that I find, the larger group still can be called the group of insiders based on the following observation. Any agent strictly prefers to be in the larger group than in the smaller group because the former one has more opportunities for collaboration.

In the model with homogeneous agents, the stable set is chosen from among the sets that consist of networks with two complete components. In the maximization problem (4), m parameterizes the collection of such sets.

My current setup requires me to look at all sets of networks that consist of two components. I cannot restrict my attention only on complete components, because, in this model, the participants' preferences are much less aligned than in the case of homogeneous agents. To find the stable set, I have to take two steps: First, I characterize the collection of sets that are potential candidate sets, and then, I provide the conditions under which a candidate set is stable.

I define a candidate set for each size $m \in \left(\frac{N}{2}, N\right)$ of the majority group. I denote the candidate set by $\mathcal{R}_m$. Such sets are constructed to satisfy the following criteria.

Let us look at the set of all networks that have a complete component of size $(N-m)$:

$$\mathcal{D}_m = \left\{ \psi \mid \exists M \subset N : |M| = m \text{ and } \forall i \in N \backslash M : \psi(i,j) = \begin{cases} 1, \text{ if } j \in N \backslash M \\ 0, \text{ otherwise} \end{cases} \right\}$$

Define a set of functions that point to a complete component (CC) in each of these networks:

$$\mathcal{C}_m = \left\{ c : \mathcal{D}_m \to 2^N \mid \forall \psi \in \mathcal{D}_m : c(\psi) \text{ form a CC in } \psi \text{ and } |c(\psi)| = N - m \right\}.$$

Definition 5. *Fix $\frac{N}{2} \le m < N$. A candidate pair for m is a set of networks $\mathcal{R}_m \subset \mathcal{D}_m$ and a function $c_m \in \mathcal{C}_m$ such that*

(i) for any set $M \subset N : |M| = m$ there exists at least one network $\rho \in \mathcal{R}_m$ such that $c_m(\rho) = N \backslash M$;

(ii) for any agent $i \in N$ and any networks $\psi, \phi \in \mathcal{R}_m$ such that $i \notin c_m(\psi) \cup c_m(\phi)$:

$$U_i(\psi) = U_i(\phi);$$

(iii) for any network $\rho \in \mathcal{R}_m$ there exists no network ψ such that

$$\psi \in \{\phi \in \mathcal{D}_m \mid \exists \widehat{c}_m \in \mathcal{C}_m : c_m(\rho) = \widehat{c}_m(\phi)\}$$

and

$$\mathbf{U}_{N \backslash c_m(\rho)}(\psi) > \mathbf{U}_{N \backslash c_m(\rho)}(\rho).$$

Lemma 1. *A candidate pair $(\mathcal{R}_m, c_m)$ exists for any $m : \frac{N}{2} < m < N$.*

For each candidate pair $(\mathcal{R}_m, c_m)$, I define a vector of the maximum payoffs that agents receive in networks inside $\mathcal{R}_m$: $\overline{\mathbf{U}}(\mathcal{R}_m)$. In particular, agent i receives his maximum payoff in $\mathcal{R}_m$ when he belongs to the larger of the two groups. Take some $\rho \in \mathcal{R}_m : i \notin c_m(\rho)$; then, $\overline{U}_i(\mathcal{R}_m) = U_i(\rho)$.

Now, I can proceed to finding the stable sets. The properties above are enough to show that any candidate set is internally stable. All I have to do now is find a candidate set that is also externally stable. Take a network from outside of the candidate set. If there are agents that are well-connected and, at the same time, unhappy (relative to the networks in the candidate set), they can initiate the transition into the candidate set. If it is true for any network outside of the candidate set, the latter is externally stable.

Proposition 4. *Take a candidate pair $(\mathcal{R}_k, c_k)$. If there exists no network $\psi \notin \mathcal{R}_k$ such that*

(i) $M = \{i \in N \mid U_i(\psi) \geq \overline{U}_i(\mathcal{R}_k)\} \neq \emptyset$ *and*

(ii) $N \backslash M \subset K = \{i \mid \psi(i,j) = \psi(j,i) = \mathbf{1}_{j \in K}\}$,

then $\mathcal{R}_k$ is stable.

Let us obtain more intuition about this proposition by looking at the situation when it is violated. Take some candidate set $\mathcal{R}_k$. Suppose that there is a network $\psi \notin \mathcal{R}_k$, such that there is a set of agents M, and every agent in that set prefers ψ over networks in $\mathcal{R}_k$. Also, the rest of the agents form a complete component and play the role of outsiders in this situation. This is so because whatever they do—and they cannot do much since their aggregate performance is already at the maximum level—they cannot threaten the agents in M. At the same time, since the agents in M like the current network more than any potentially stable network, they will never make a first step to switch to a network in $\mathcal{R}_k$. As a result, the set $\mathcal{R}_k$ fails to be externally stable.

It turns out that this condition is also sufficient for external stability. In any network ψ that fails to satisfy conditions (i) and (ii) of Proposition 4, there exists a set of agents who are willing and able to initiate transition to a network in $\mathcal{R}_k$. In particular, the agents in ψ who prefer this network to any network in $\mathcal{R}_k$ are connected to agents who do not.

Notice that the internal stability of a candidate pair $(\mathcal{R}_k, c_k)$ implies that agent do not use their abilities to the full extent. The higher the ability of the agent is, the more he has to sacrifice in return for membership in a dominant group. This property of a candidate pair is reflected in the following proposition.

Proposition 5. *For any candidate pair $(\mathcal{R}_k, c_k)$ take two agents, i and j, such that $\alpha_i > \alpha_j$. Let $\rho \in \mathcal{R}_k$ be the network in which both agents are in the majority group: $i, j \notin c_k(\rho)$. Then,*

$$\frac{U_i(\Lambda)}{U_j(\Lambda)} > \frac{U_i(\rho)}{U_j(\rho)} = \frac{\overline{U}_i(\mathcal{R}_k)}{\overline{U}_j(\mathcal{R}_k)}.$$

The inequality of payoffs within the dominant group is lower in any candidate set, than in the efficient network. The payoffs' distribution gets "squeezed" because talented agents represent a higher threat to the other participants, but, at the same time, can be easily replaced by less-talented ones. If an agent with a high ability wants to stay with the dominant group, he must compete modestly. Since this proposition applies to all candidate sets, it also applies to the stable ones.

In the following section, I present a simple three-agent example that illustrates the definition of candidate pairs and Propositions 4 and 5.

5.1 Three-Agent Example

Suppose that there are three agents with abilities $\alpha_1 > \alpha_2 > \alpha_3$. As before, the complete network is the unique efficient one. In this network, the payoff of agent i is

$$U_i(\Lambda) = 2\alpha_i + g\frac{\alpha_i}{\sum_j \alpha_j}.$$

In this section I find the stable set of networks for this three-agent model to illustrate Propositions 4 and 5.

Since $N = 3$, there is a unique candidate pair—$(\mathcal{R}_2, c_2)$. The set $\mathcal{R}_2$ consists of the three networks that can be found in Figure 10. The definition of the candidate pair requires that agents in the larger group are indifferent between their partners. This requirement gives us three conditions:

$$\begin{aligned}
x_3\left(1+\frac{g}{x_2\alpha_2+x_3\alpha_3}\right) &= y_3\left(1+\frac{g}{y_1\alpha_1+y_3\alpha_3}\right)\\
x_2\left(1+\frac{g}{x_2\alpha_2+x_3\alpha_3}\right) &= z_2\left(1+\frac{g}{z_1\alpha_1+z_2\alpha_2}\right)\\
y_1\left(1+\frac{g}{y_1\alpha_1+y_3\alpha_3}\right) &= z_1\left(1+\frac{g}{z_1\alpha_1+z_2\alpha_2}\right)
\end{aligned}$$

Take an arbitrary solution to this system of equations $(x_2, x_3, y_1, y_3, z_1, z_2)$. Suppose, for example, that both $x_2 < 1$ and $x_3 < 1$. Then, I can find a network $\widehat{\psi}$ that has the same set of links (possibly with different intensities)—i.e., $\mathcal{N}(\psi_1) = \mathcal{N}(\widehat{\psi})$[12] such that $\widehat{\psi}(2,3) = \frac{x_2}{\max\{x_2,x_3\}}$ and $\widehat{\psi}(3,2) = \frac{x_3}{\max\{x_2,x_3\}}$. Notice, that both $\widehat{\psi}(2,3)$ and $\widehat{\psi}(3,2)$ are weakly less than 1, hence are feasible, and both agents 2 and 3 strictly prefer $\widehat{\psi}$ to ψ_1. The definition of a candidate pair requires that such a network $\widehat{\psi}$ does not exist, so it must be that one of the agents in each of the three discriminating pairs is assigned a link with intensity 1. In particular, $x_3 = y_3 = z_2 = 1$. The rest of the solution is

[12] I define $\mathcal{N}(\cdot)$ in the following way: $\mathcal{N} \circ \psi(i,j) = 1$ if $\max\{\psi(i,j), \psi(j,i)\} > 0$.

$z_1 = \frac{\alpha_2}{\alpha_1} \leq 1$, $y_1 = \frac{\alpha_2}{\alpha_1} x_2 \leq 1$ where $x_2 \leq 1$ solves[13]

$$x_2 \left(1 + \frac{g}{x_2\alpha_2 + \alpha_3}\right) = 1 + \frac{g}{2\alpha_2}. \tag{6}$$

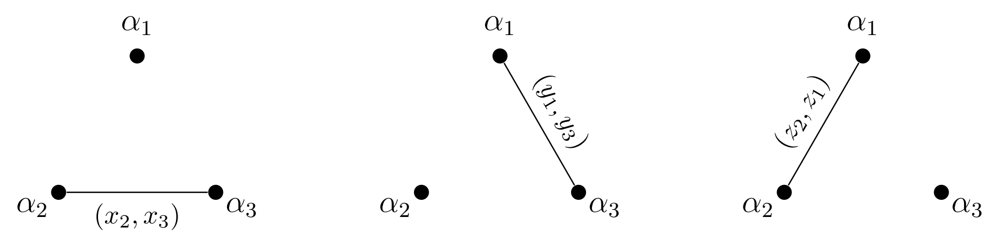


Figure 10: Networks ψ_1, ψ_2 and ψ_3 (from the left to the right)

I show that if

$$g > \frac{2\alpha_2(2x_2 - 1)(2\alpha_2 x_2 + \alpha_3)}{\alpha_3}. \tag{7}$$

any network outside of $\mathcal{R}_2$ is blocked by one of the networks in $\mathcal{R}_2$. This inequality can be derived from the conditions in Proposition 4.

I start with networks that have only one link. Take a network $\phi_1 \notin \mathcal{R}_2$ such that it has only one link (see Fig. 11). There is a network $\psi \in \mathcal{R}_2$ such that $\mathcal{N}(\phi_1) = \mathcal{N}(\psi)$. Either agent i or j prefers ψ to ϕ_1: Suppose that it is i. Then, agent i can break the current arrangement in ϕ_1, and after that, agent j will be willing to create a new network ψ. To summarize, for any $\phi \notin \mathcal{R}_2$ such that $\mathcal{N}(\phi)$ has only one link, there always exists $\psi \in \mathcal{R}_2 : \psi \succ \phi$.

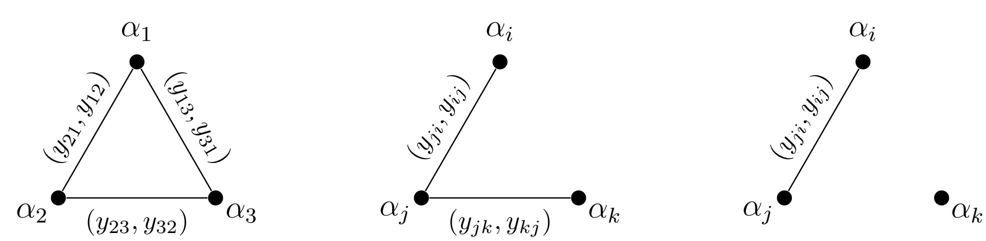


Figure 11: Networks ϕ_3, ϕ_2 and ϕ_1 (from left to right)

[13]Notice, that $\frac{d}{dx_2}F(x_2) \geq 0$, $F(0) = 0$ and $F(1) = 1 + \frac{g}{\alpha_2+\alpha_3} > 1 + \frac{g}{2\alpha_2}$, where

$$F(x_2) = x_2 + \frac{gx_2}{x_2\alpha_2 + \alpha_3},$$

is a continuous function on $[0, 1]$; hence, the solution to the equation (6) exists and is less than 1.

Take a network ϕ_3 (Fig. 11) and ask the following question: Under what condition on g, for any triple $(y_1, y_2, y_3) \in (0, 2]^3$, does there exist an $i \in \{1, 2, 3\}$ such that agent i's payoff in at least one of the three networks ψ_1, ψ_2 or ψ_3 is larger than in ϕ_3? Formally, I need to find a condition on g such that at least one of the three following inequalities holds for any triple $(y_1, y_2, y_3) \in (0, 2]^3$:

$$\begin{aligned}
\alpha_1 y_1 \left(1 + \frac{g}{\sum\limits_i \alpha_i y_i}\right) &< \alpha_2 \left(1 + \frac{g}{2\alpha_2}\right) \\
\alpha_2 y_2 \left(1 + \frac{g}{\sum\limits_i \alpha_i y_i}\right) &< \alpha_2 \left(1 + \frac{g}{2\alpha_2}\right) \\
\alpha_3 y_3 \left(1 + \frac{g}{\sum\limits_i \alpha_i y_i}\right) &< \frac{\alpha_3}{x_2} \left(1 + \frac{g}{2\alpha_2}\right).
\end{aligned}$$

From these three inequalities, it follows that the condition on g is

$$g > \max_{(y_1,y_2,y_3)\in(0,2]^3} \left\{ \frac{\min\left\{\frac{\alpha_1}{\alpha_2} y_1, y_2, x_2 y_3\right\} - 1}{\frac{1}{2\alpha_2} - \frac{\min\left\{\frac{\alpha_1}{\alpha_2} y_1, y_2, x_2 y_3\right\}}{\sum\limits_i \alpha_i y_i}} \right\} = \frac{2\alpha_2(2x_2 - 1)(2\alpha_2 x_2 + \alpha_3)}{\alpha_3}.$$

Under condition (7), in any network ϕ that either has a set of links $\mathcal{N}(\phi_3)$ or $\mathcal{N}(\phi_2)$, one can find an agent, that prefers one of the three networks in $\mathcal{R}_2$ to ϕ.

Finally, I show that for any network $\phi \notin \mathcal{R}_2$, I can find $\psi \in \mathcal{R}$ such, that $\psi \succ \phi$. If $\mathcal{N}(\phi)$ has only one link, I am done. But suppose that $\mathcal{N}(\phi)$ has more than one link. By the argument, under condition (7), I can find an agent i that prefers some element of $\mathcal{R}_2$ to ϕ. Let me look at the link between the other two agents, j and k, in ϕ. Delete all links of agent i in ϕ obtaining a new network ϕ_{jk}. Now, there can be two cases: ϕ_{jk} either belongs to $\mathcal{R}_2$ or does not. Suppose that $\phi_{jk} \notin \mathcal{R}_2$. Then, either agent j or k prefers one of the elements in $\mathcal{R}_2$ to ϕ_{jk}. Without loss of generality, suppose that it is j. Then, j can break the agreement with k in ϕ_{jk}, and, after that, arrange the agreement ψ with i such that $\psi \in \mathcal{R}_2$, which means that $\psi \succ \phi$. Now, suppose that $\phi_{jk} \in \mathcal{R}$. Then, either there exists a second agent in ϕ—say j, that prefers an element in $\mathcal{R}_2$ to ϕ—or agent i has non-trivial links to both agents j and k. In the first scenario, agents i and j simultaneously break all existing agreements and then link with each other in $\psi \in \mathcal{R}_2$. In the second scenario, agent i breaks just one link to agent j, switching to network ϕ_{ij}, which both agents i and j enjoy less than an element in $\mathcal{R}_2$. After that, agents i and j simultaneously break links with agent k and then relink with each other in $\psi \in \mathcal{R}_2$. In both scenarios, $\psi \succ \phi$.

Notice that finding one agent who prefers a stable network to an unstable one is enough for proving that the former is stable. This observation is true in general and it is at the heart of Proposition 4. An agent who is *indirectly* connected to many agents has certain power in renegotiating the structure of the network with his indirect connections. He does it by changing his direct connection, and therefore "upsetting" his neighbors. Once the payoffs of his neighbors is low, they are open for renegotiation and can "upset" their neighbors as well. This process continues until there are enough agents to form a dominant group.

If condition (7) holds, the set $\mathcal{R}_2$ satisfies both internal and external stability conditions and, hence, is von Neumann-Morgenstern stable. There are several new properties of this set that were absent in the symmetric model. Condition (7) depends only on α_2 and α_3. The closed-form solution for the threshold for g is

$$\frac{g}{\alpha_3} > \frac{12 - \left(1 + 2\frac{\alpha_2}{\alpha_3}\right) + 3\sqrt{\left(1 + 2\frac{\alpha_2}{\alpha_3}\right)^2 + 16}}{4}$$

I can set $\alpha_3 = 1$ as a normalization. Observe, that the threshold is an increasing function of α_2. Indeed, when the inequality increases (in this case, the measure of inequality is α_2/α_1), it becomes more costly in terms of absolute output to sustain equal payoffs across ψ_1, ψ_2 and ψ_3. Therefore the collusive outcome can be sustained only if the size of the tournament prize also goes up.

Also, note that the threshold does not depend on α_1. The coalition of agents 2 and 3 have enough power to exclude agent 1 from the competition for the prize. Agent 1 has no choice but to keep his own performance down to appear not too strong a competitor to agents 2 and 3. When linked, for example, to agent 3, agent 1 mimics agent 2's productivity. No matter how high his true productivity is, he can compensate for it by the low intensity of his links—i.e., by low x_{1i}.

Proposition 5 states that the inequality between agents in a dominant group is lower than inequality between the same agents in the efficient outcome. To illustrate this proposition, let us look at the payoffs of agents 1 and 2: These payoffs are equal. Indeed, agent 1 has to mimic less productive agent 2 in exchange for the membership in the dominant group.

Finally, Assumption 1 does not have any effect in this example. It is so because the smaller group consists of one agents and therefore it cannot split into two new groups.

6 Related Literature

In this section I discuss the literature that is related to the current paper. I start with general literature on network-formation problems and then proceed to the applications.

6.1 Network Formation

There are several approaches to solving network-formation problems in the literature. The vast majority of the literature employs pairwise stability notion developed by Jackson and Wolinsky (1996). There are, however, the alternatives to pairwise stability. In this section, I will only discuss the solutions that assume that a link between two agents requires their mutual consent.

One can approach the network-formation problem as a noncooperative game in which agents sequentially create and terminate links. The seminal paper on network-formation, Aumann and Myerson (1988), takes this approach. The authors assume that there is a deterministic rule which defines active players. Agents can only create, but not terminate links, and the game ends when no pair agents wants to create a missing link.

Another example of noncooperative approach is Dutta et al. (2005). They study the infinite horizon dynamic network-formation game endowed with Markov Perfect Equilibrium. Agents are farsighted, therefore each agent, when evaluating certain strategy, takes into account the effect of his actions on the future play. They establish the conditions, under which the efficient outcome is stable. In my model I also assume that agents are farsighted, but I allow the network structure to settle before the payoffs are realized.

Both Aumann and Myerson (1988) and Dutta et al. (2005) (and virtually any other dynamic noncooperative model of network formation) assume a certain protocol of agents' moves in the game, and the equilibria are, in general, sensitive to this protocol. The advantage of my approach is that no particular protocol is assumed and and agents consider a wide variety of possible actions they take. Of course, the downside is that I only consider stable outcomes and my model can say very little about the dynamic transition process between networks.

Another alternative is to look at the cooperative solutions. The current paper takes this path, and there are several paper that are related to mine. In particular, my stability notion is related to one introduced by Ray and Vohra (1997) and later by Diamantoudi and Xue (2007). Both Ray and Vohra (1997) and Diamantoudi and Xue (2007) study the model of coalition formation with farsighted agents. I modify their blocking relation to make it applicable to my network formation model. Ray and Vohra (1997) and Greenberg (1990) provide a discussion on the beliefs of agents that can support the stable outcome.

Solution concept, that I use is also closely related to one developed by Herings et al. (2009). The difference is twofold. When they define the blocking relation for networks, they allow weak improvement for all but one agent. I exclude this possibility, because it can lead to large number of cycles, in particular between networks that are in my set $\mathcal{G}^*$. As a result the dominance relation becomes to rich, which in its turn can lead to existence problems. Also, instead of using von Neumann-Morgenstern stable sets, they use what they call pairwise farsighted stability. Their notion is defined in such a way, that existence is automatically guaranteed[14]. Their solution in general lacks the internal stability property. The stable set that is defined in my paper is always pairwise farsightedly stable. The reverse is not true.

Chwe (1994) introduces a solution, which he calls largest consistent set. This solution is defined using farsighted blocking relation that is identical to the one I use in this paper. Any stable set of mine is a subset of Chwe's largest consistent set. In particular, the consistency of stable set (see Proposition 2) follows from this observation.

Finally, one can use a hybrid solution that combines the features of cooperative and noncooperative approaches. An example of a hybrid solution is Konishi and Ray (2003). The authors develop a solution for a coalition-formation problem. In principle, such solution can be easily adapted to a network framework. One important advantage of Konishi and Ray's solution is that it describes the process of transition between situations of the game (i.e., coalitional structures, or, in the case of this paper, networks). The downside of this approach is that it requires the specification of the protocol[15] and the set of stable outcomes is sensitive to this specification.

6.2 Applications: Collaboration between Firms

My paper is closely related to Goyal and Joshi (2003) and Marinucci and Vergote (2011). Both papers develop a model of competition in R&D. In their model firms can use research joint ventures to save on costs of R&D. They find, that under certain conditions, asymmetric networks of cooperation is possible in the equilibrium. The main difference between my results and Marinucci and Vergote (2011) is that they can not exclude the complete network from the equilibrium: there is always an equilibrium that admits efficient outcome. Also, the results by Marinucci and Vergote hold only under very strong condition for the payoff functions, that are due to the way they model the competition in R&D.

There exists an extensive literature on collaboration between firms that looks at the coalitions of firms rather than bilateral agreements between them. The survey of these works can be found in Bloch (2002) or Ray (2007). The predictions obtained in this literature significantly differ from ones obtained in network-formation models discussed above. In particular the grand coalition (which is the analog of the complete network) is usually not stable because there exist a smaller coalition that prefers to reduce the amount of collaboration in exchange for a greater market power. These results are obtained under assumption that participation in a coalition is exclusive. The results I obtain are similar but I do not use this assumption: In my paper, groups

[14] They define two properties that a pairwise farsightedly stable set must satisfy and then say that the solution is the minimal (in the set inclusion sense) set that satisfies these two properties. Also they note that grand set satisfies these two properties, hence the existence follows from the finite number of sets.

[15] The protocol specifies the sequence of active agents or coalitions.

are *endogenously* exclusive. Therefore my paper can be viewed a bridge between coalition- and network-formation literature that settles the difference in the results.

Also, my paper contributes to the program announced in Salop and Scheffman (1983). In their paper, Salop and Scheffman (1983) state, that firms can capture the market by increasing the costs of production for their rivals. In the companion paper, Salop and Scheffman (1987) describe various strategies that are available to firms to raise the costs for their competitors. They notice that some of those strategies can be more effective than predatory pricing. The mechanism, that is described in my model, can be used by a coalition of firms to gain control over the market. The coalition of the firms does not need to engage in predatory pricing to implement this mechanism. In fact, prices are likely to rise, since some of the firms will be forced to leave the market, or to include the royalty fees in their prices.

6.3 Applications: Discrimination and Segregation

My paper is related to papers on discrimination and segregation. In this section, I focus on models that treat race (and other characteristics such as ethnicity, religion, caste, etc.) as payoff-irrelevant.

McAdams (1995) argues that race discrimination is a tool that is used by a white majority to maintain their superior social status. My paper is complementary to McAdams (1995) since it provides a formal model of discrimination that is driven by status concerns. In particular, I derive a social norm that supports individual incentives to discriminate against a minority.

Although similar social norm is derived in Pęski and Szentes (2011), the underlying mechanism is different. In Pęski and Szentes's model, discrimination is a result of coordination failure, whereas, in my model, the discrimination, on the opposite, is a collective strategy. Note that Theorem 3 predicts a complete segregation between majority and minority. A similar result is obtained in Eeckhout (2006), but the framework of that paper is very different: Eeckhout considers a marriage market in which matched agents play prisoners' dilemma repeatedly. Segregation makes punishment for unmatched agents more severe, hence improves cooperation between matched agents.

7 Conclusion

This paper studies networks of collaboration between competitors. Prior to competition, agents set up collaboration relationships with each other. Once a network of the collaboration is settled, each agent produces an output that depends on the number of the agent's partners in the network. The competition is modeled as a tournament, and, therefore, the agents with the highest output win the competition and split the prize among themselves. In my model, I assume that the agents value their connections not only because the latter improve their chances of winning the tournament, but also because the agents value their output per se. This implies that the efficient outcome is the complete network. Moreover, each link in the network is bilaterally optimal: Any two unconnected agents find it optimal to connect with each other.

I assume that the agents are farsighted and can coordinate their actions with each other. I use von Neumann-Morgenstern stable sets to characterize stable outcomes in my model. My framework works best for small networks with an abundance of communication between agents. It can be applied to such problems as patent wars between alliances of firms, collaboration networks in promotion tournaments, the internal structure of organized crime, etc.

Although my model does not capture all the the details of these applications, it provides the important insights that are common to all of these environments. First, it characterizes the necessary and sufficient conditions for the stability of the efficient outcome. Second, it highlights stable networks that have empirically relevant properties. These networks feature two complete components. Such a structure can be interpreted as a competition between two groups: the group of "insiders" that dominates the competition and the group of "outsiders" that is forced to lose. These groups are represented by the two components.

This result is particularly interesting because group competition is not a primitive of the model. Nevertheless, the stable networks that I find are organized in a fashion that resembles a competition on the group level.

I depart from the traditional approach of the network-formation literature in my assumption that agents are farsighted. This allows me to obtain novel predictions about stable outcomes. In particular, in contrast with the existing literature, I find the conditions under which the efficient outcome *cannot* be stable.

Finally, I see some worthwhile directions for future research. First, it would be interesting to investigate if there can be more than two components in a stable network. The stable networks that I find feature two complete components. The smaller component does not receive any of the prize, so its members do not have any incentives to split further, into two components. However, if we assume that the agents in the smaller component get a significant share of the prize, the smaller component may split into two.

Second, it would be interesting to study a model that accommodates the agents that are heterogeneous in their effect on others. In the model that I study, the agents differ from each other in their ability to convert the collaboration into output. Put differently, the agents are either good or bad "performers." A natural question, then, is: What if the agents are either good or bad "collaborators"?

Proofs

Proof of Proposition 1

First, observe that $\psi \cup \{ij\} \triangleright \psi$ for all $i \neq j$ and $\psi : \psi(i,j) = 0$. This means that any network, except for Λ is blocked. Suppose that $f(N-1) + \frac{g}{N} < f(m-1) + \frac{g}{m}$ for some $m \in \left(\frac{N}{2}, N\right)$. Take a coalition $S = \{1, 2, ..., m\}$, and let us look at a network ψ such that

$$\psi(i,j) = \begin{cases} 0 & \text{, if } |\{i,j\} \cap S| = 1, \\ 1 & \text{, otherwise.} \end{cases}$$

Then, $\psi \triangleright \Lambda$, and hence the core of $(2^{\Lambda}, \triangleright)$ is empty.

Lemma A1

Lemma A 1. *If $\psi \succ \gamma$, then there exist the sequences $\{(S_i, \gamma_i)\}_{i=1}^{K}$, $\forall i = 1, ..., K$: $S_i \subset N$ and $\gamma_i \in 2^{\Lambda}$ such, that $\gamma = \gamma_1 \overset{S_1}{\to} \gamma_2 \overset{S_2}{\to} \dots \overset{S_K}{\to} \psi$ and $\psi \succ \gamma_k$ for all $k \leq K$*

Proof. By definition there exist a b-sequence of networks such, that $U_{S_k}(\psi) \gg U_{S_k}(\gamma_k)$ for all $k \leq K$. If we take the tail of that sequence starting from some k we get a new sequence that is a b-sequence for pair ψ and γ_k. □

Lemma A2

Lemma A 2. *If*

$$\underset{m > \frac{N}{2}}{\arg\max} \{f(m-1) + g(m)\} \cap N = \emptyset$$

complete network can not be stable: $\{\Lambda\}$ is never stable.

Proof. Let $m^* = \underset{m > \frac{N}{2}}{\arg\max} \{f(m-1) + g(m)\}$. By contradiction, assume, that $\{\Lambda\}$ is a stable set. By external stability $\Lambda \succ \psi$ for all $\psi \in 2^{\Lambda}$. Let $\mathcal{G}$ be a set of networks, in which agents $1, ..., m^*$ form a complete component. Take an arbitrary $\psi \in \mathcal{G}$. By Lemma A1, there exists a b-sequence $\psi = \psi_1 \overset{S_1}{\to} \dots \overset{S_K}{\to} \Lambda$. Let ψ_i be the latest element in the sequence that lies in $\mathcal{G}$. Then, $\psi_{i+1} \cap \mathcal{G} = \emptyset$ and hence $S_i \cap \{1, ..., m^*\} \neq \emptyset$. Also, for any $i \in \{1, ..., m^*\}$

$$U_i(\Lambda) = f(N-1) + g(N) < f(m^*-1) + g(m^*) = U_i(\psi).$$

However, if (ψ_i, S_i) is a part of a b-sequence, that ends in Λ, $U_{S_i}(\Lambda) \gg U_{S_i}(\psi)$, hence the contradiction. □

Proof of Proposition 3

Take some network ψ. The aggregate utilitarian welfare of this network is

$$W(\psi) = \sum_{i \in N} U_i(\psi) = g + \sum_{i \in N} f(x_i)$$

Since $f(\cdot)$ is strictly increasing, $W(\Lambda) > W(\psi)$ for any $\psi \neq \Lambda$.

Proof of Theorem 1

Assume by contradiction that $\Lambda \in \mathcal{R}$, where $\mathcal{R}$ is a stable set. Take $\psi \in \mathcal{G}^*$. I know, that $\Lambda \not\succ \psi$ (see Lemma A2). Also, it is easy to see, that $\psi \succ \Lambda$. Since $\mathcal{R}$ is stable, there exist $\gamma \in \mathcal{R}$ such that $\gamma \succ \psi$.

I am going to show that $\Lambda \succ \gamma$, which leads to the contradiction, since by assumption both of those networks are in $\mathcal{R}$.

Define the functions $\underline{U}(x) = \min_{i \in N}\{U_i(x)\}$ and $\overline{U}(x) = \max_{i \in N}\{U_i(x)\}$. Note, that $\overline{U}(\Lambda) = \underline{U}(\Lambda)$.

Since I am proving that $\Lambda \succ \gamma$, I need to find a sequence $\gamma \rightarrow \gamma_1 \rightarrow ... \rightarrow \Lambda$ that satisfies the definition of dominance relation $\succ$. I am going to come up with recursive definition of this sequence, i.e. $\gamma_{i+1} = D(\gamma_i)$. Before introducing function $D(\cdot)$, I have to define the domain of it.

By $M(\psi) = \{i \mid i \in \arg\max_{k \in N} |E_k(\psi)|\}$ I denote the set of nodes in ψ that have maximum number of edges. Let

$$\Omega = \{\psi \in 2^\Lambda \mid \psi \neq \psi' \sqcup \psi'', \text{ such that } N(\psi') \subset M(\psi) \text{ and } \psi' \text{ is a component}\}$$

With Ω in hands, I define $D : \Omega \rightarrow 2^\Lambda$. Take $\psi \in \Omega$. $D(\psi)$ is obtained from ψ by performing two operations:

(i) delete all edges in $\psi \backslash E_{M(\psi)}(\psi)$;

(ii) delete one arbitrary edge in $O_{M(\psi)}(\psi)$.

Note that $(\psi \backslash E_{M(\psi)}(\psi)) \cap O_{M(\psi)}(\psi) = \emptyset$, hence operation (ii) is well defined.

The following lemma makes sure that if I start from an element in Ω, the sequence that is defined by recursive rule $\gamma_{i+1} = D(\gamma_i)$ does not go outside of the domain of $D(\cdot)$.

Lemma 2. $D(\Omega) \subset \Omega \cup \{\chi \in 2^\Lambda \mid |E(\chi)| = 1\}$

Proof. Take $\psi \in \Omega : |E_{M(\psi)}| = 2$. By definition of Ω I know that $M(\psi)$ must be a singleton in this case. After applying procedure (i) of $D(\cdot)$ I obtain $\hat{\psi}$, that has only two edges: $|E(\hat{\psi})| = 2$, hence $D(\psi) \in \{\chi \in 2^\Lambda \mid |E(\chi)| = 1\}$.

Now, take $\psi \in \Omega : |E_{M(\psi)}| > 2$. After applying procedure (i) of $D(\cdot)$ I obtain $\hat{\psi} \in \Omega$, that satisfies $\hat{\psi} \backslash E_{M(\hat{\psi})}(\hat{\psi}) = \emptyset$.

If $|M(\psi)| = 1$, it is clear that $D(\psi) \in \Omega$. Suppose, that $|M(\psi)| > 1$. By definition of Ω there exist $i \in M(\psi) : E_i(\psi) \cap O_{M(\psi)}(\psi) \neq \emptyset$. Take $j \in M(\psi)\backslash i$. Again by definition of Ω, either i and j belong to the same component of ψ or there exist $k \in N\backslash M(\psi)$, such that j and k belong to the same component. Procedure (ii) of $D(\cdot)$ deletes edge that belongs to $E_i(\psi) \cap O_{M(\psi)}(\psi)$. Now, the components that were disjoint with i in ψ are untouched by procedure (ii), hence they satisfy the definition of Ω. Also note that $M(D(\psi)) = M(\psi)\backslash i$, and i is still connected to elements in $M(\psi)$ that shared the same component with i in ψ. Hence I can not find a component in $D(\psi)$, that consist of only elements of $M(D(\psi))$, i.e. $D(\psi) \in \Omega$. □

Finally, I can define the sequence $\{\gamma_i\}_{i \in K}$, that satisfies the definition of dominance. Start with $\gamma_0 \in \Omega$, and apply the recursive rule $\gamma_{i+1} = D(\gamma_i)$ until $D(\gamma_{K-2}) \in \{\chi \in 2^\Lambda \mid |E(\chi)| = 1\}$. Observe, that $|E(\gamma_{K-2})| = 2$ and $|M(\gamma_{K-2})| = 1$. Pick an arbitrary regular network $\gamma_{K-1} \supset \gamma_{K-2}$ of degree 2[16]. The last element of the sequence is $\gamma_K = \Lambda$.

If $\gamma \in \Omega$, then I set $\gamma_0 = \gamma$. Suppose $\gamma \notin \Omega$. Then, there exist $\hat{M} \subset M(\gamma)$ that forms a component. I observe that $|M(\gamma)| \leq \frac{N}{2}$, since by original assumption $\gamma \succ \psi$. Take an arbitrary agent $i \in N\backslash M(\gamma)$ and obtain network γ_0 by connecting this agent to all nodes in $N\backslash M(\gamma)$. Obviously, $M(\gamma_0) = i$ and $\gamma_0 \in \Omega$.

I observe, that under this algorithm, when switching from γ_i to γ_{i+1} a link is deleted only if one of the nodes is in $N\backslash M(\gamma_i)$, and link is created if both of the nodes are in $N\backslash M(\gamma_i)$. Also, for any network $\psi \neq \Lambda$, $U_{N\backslash M(\psi)}(\psi) < \overline{U}(\Lambda)$.

I have shown that $\Lambda \succ \gamma$ by constructing the sequence $\gamma \to \gamma_0 \to ... \to \gamma_K = \Lambda$, that satisfies definition of setwise dominance, which contradicts my assumption that $\gamma \in \mathcal{R}$.

Finally, as I just showed, $\Lambda \notin \mathcal{R}$, hence $\exists \psi \in \mathcal{R} : \psi \succ \Lambda$. Take a network $\lambda : M(\lambda) = N$. Then for any $\psi \succ \Lambda$, I have $\psi \succ \lambda$, hence $\lambda \notin \mathcal{R}$.

Proof of Theorem 3

I have to prove that $\mathcal{G}^*$ is both internally and externally stable.

I start with internal stability. Take two networks $\psi, \gamma \in \mathcal{G}^*$. Internal stability requires that $\psi \not\succ \gamma$. Suppose, by contradiction that $\psi \succ \gamma$. Then, there exist a sequence $\gamma = \gamma_1 \to ... \to \gamma_K = \psi$, that satisfies the definition of dominance. Suppose, that all networks γ_i with $i \leq k$ are such, that $E_{M(\gamma)}(\gamma_i) = E_{M(\gamma)}(\gamma)$, and γ_{k+1} is the first network in the sequence that violates this property. Then, $\gamma_k \to \gamma_{k+1}$

[16] It is easy to see that it always exist. For example I can use the following procedure: two remaining edges in γ_{K-2} are between agent i and say l and m. Starting with agent l connect all agents but agent i in one line that ends with agent m. The resulting network γ_{K-1} is a cycle.

violates the definition of dominance since $U_{M(\gamma)}(\gamma_k) = U_{M(\gamma)}(\gamma) \leq U_{M(\gamma)}(\psi)$, hence the contradiction.

For external stability, I need to show, that for any $\psi \notin \mathcal{G}^*$ there exist $\gamma \in \mathcal{G}^*$, such that $\gamma \succ \psi$. Since the set $\mathcal{G}^*$ consist of networks that are equivalent up to permutation of agents, for any $\gamma, \gamma' \in \mathcal{G}^*$, I have $\overline{U}(\gamma) = \overline{U}(\gamma')$. With some abuse of notation I will say that $\overline{U}(\mathcal{G}^*) = \overline{U}(\gamma)$ for some $\gamma \in \mathcal{G}^*$.

Take an arbitrary $\psi \notin \mathcal{G}^*$. Suppose that $U_{M(\psi)}(\psi) \leq \overline{U}(\mathcal{G}^*)$. Pick an arbitrary set M of agents such that $|M| = m^*$ and modify ψ by dropping all edges in $E_M(\psi)$. Let us call the resulting network ψ_2. Network ψ_3 is obtained from network ψ_2 by connecting agents in the set M into a complete component. Finally network $\psi_4 \in \mathcal{G}^*$ is obtained from ψ_3 by connecting agents in the set $N \backslash M$ into a complete component. It is easy to see that in transitions $\psi_1 \to \psi_2 \to \psi_3$ only agents from the set M are involved and their payoff in the final network is larger than along the way: $U_M(\psi_4) = \overline{U}(\mathcal{G}^*)$. The rest of the agents are only involved in the last step of transition $\psi_3 \to \psi_4$ and since $N \backslash M \cap M(\psi_3) = \emptyset$ and each agent in $N \backslash M$ gets weakly more links in ψ_4 than in ψ_3, the definition of dominance is met by the sequence $\psi = \psi_1 \to ... \to \psi_4$.

Now suppose, that $U_{M(\psi)}(\psi) > \overline{U}(\mathcal{G}^*)$. I have to go through three cases: (i) $|M(\psi)| < \frac{N}{2}$, (ii) $|M(\psi)| = \frac{N}{2}$ and (iii) $|M(\psi)| > \frac{N}{2}$.

In case (i) when $|M(\psi)| < \frac{N}{2}$, pick an arbitrary set $M \subset (N \backslash M(\psi)) : |M(\psi)| < |M| \leq m^*$. Obtain network ψ_2 by simultaneously dropping all edges in $O_M(\psi)$ and adding all missing edges between elements in M. In the resulting network $M(\psi_2) = M$. Also, $U_M(\psi_2) \leq \overline{U}(\mathcal{G}^*)$ with equality only in case of $|M| = m^*$. If $|M| < m^*$, pick an arbitrary set $\hat{M} \supset M : |\hat{M}| = m^*$. Obtain network ψ_3 by creating a complete component that consists of $\hat{M}$ and finally obtain ψ_4 by adding all missing links between agents in $N \backslash \hat{M}$. As before, I can see, that $\mathcal{G}^* \ni \psi_4 \succ \psi$, since we've constructed proper sequence $\psi = \psi_1 \to ... \to \psi_4$.

In case of $|M(\psi)| = \frac{N}{2}$, if $M(\psi)$ forms a union of components, take agents $N \backslash M(\psi)$ and obtain network ψ_2 by duplicating components formed by $M(\psi)$, using agents $N \backslash M(\psi)$. The resulting network has a property that $M(\psi_2) = N$ and that $\overline{U}(\psi_2) = \underline{U}(\psi_2) < \overline{U}(\mathcal{G}^*)$. In the next step, obtain network ψ_3 from network ψ_2 by selecting the set $M \supset (N \backslash M(\psi)) : |M| = m^*$, simultaneously dropping all edges in $O_M(\psi_2)$ and adding all missing edges between elements in M. The resulting network ψ_3 consists of complete component of size m^* and some other components. Finally obtain network $\psi_4 \in \mathcal{G}^*$ by creating all missing links between agents in $N \backslash M$. Obtained sequence $\psi = \psi_1 \to ... \to \psi_4$ satisfies the definition of dominance, hence $\mathcal{G}^* \ni \psi_4 \succ \psi$.

If $|M(\psi)| = \frac{N}{2}$ and there exist an element in $M(\psi)$ that is connected to $N \backslash M(\psi)$. Drop one link between $M(\psi)$ and some $i \in N \backslash M(\psi)$ and obtain network ψ'. Observe that $|M(\psi')| = |M(\psi)| - 1 < \frac{N}{2}$, hence I can use the argument above (see case (i)) to construct the sequence $\psi' = \psi_1 \to ... \to \psi_4$ with one modification, that when set M is picked it must be that $i \in M$. This will guarantee that I can add ψ to the beginning of this sequence and still satisfy the definition of dominance.

Finally, if $|M(\psi)| > \frac{N}{2}$ and $U_{M(\psi)}(\psi) > \overline{U}(\mathcal{G}^*)$, it must be the case that each element in $M(\psi)$ has weakly more links then $|M(\psi)|$, i.e. each agent in $M(\psi)$ is connected to some agent in $N\backslash M(\psi)$. I already discussed the case when $|M(\psi)| = N$, so I only need to go through the case when $|M(\psi)| < N$. Pick $\hat{M} \subset M(\psi) : |\hat{M}| = \frac{|M(\psi)|}{2}$ if $|M(\psi)|$ is even and $|\hat{M}| = \frac{|M(\psi)|+1}{2}$ if $|M(\psi)|$ is odd. Break the links that connect $\hat{M}$ to $N\backslash M(\psi)$ and obtain network ψ'. Clearly $M(\psi') = M(\psi)\backslash\hat{M}$ and $|M(\psi')| < \frac{N}{2}$. Now I can use the argument for case (i) with the modification that the set M must be such that $N\backslash M(\psi) \subset M$. This condition will guarantee that I can add network ψ to the beginning of the sequence $\psi' \to ... \to \psi_4 \in \mathcal{G}^*$ in a way that goes along with the definition of dominance.

Proof of Proposition 2

The external stability of $\mathcal{R}$ guarantees the existence of $\widehat{\rho} \in \mathcal{R}$. Fix a b-sequence $\{(S_k, \gamma_k)\}_{k=1}^{K}$, that supports $\psi \succ \rho$. By contradiction assume, that for any agent $i \in \bigcup_{k=1}^{K} S_k$, I have $U_i(\rho) < U_i(\widehat{\rho})$. Then, $\widehat{\rho} \succ \rho$, which contradicts an assumption that $\mathcal{R}$ is stable.

Proof of Lemma 1

Let us look at the set of networks $\mathcal{R}_m$ each of which consists of two groups, that are not connected between each other. The sizes of the groups are m and $N - m$, where $m > \frac{N}{2}$. For a network $\psi \in \mathcal{R}_m$, I am going to denote a larger group by $M(\psi)$.

Suppose I have N agents ordered by their productivity: $\alpha_1 > \alpha_2 > ... > \alpha_N$.

For the internal stability of $\mathcal{R}_m$, I require that for any two networks $\psi, \phi \in \mathcal{R}_m$, and for any agent $i \in M(\psi) \cap M(\phi)$ I have, that $U_i(\psi) = U_i(\phi)$. Also, I require, that in any network $\psi \in \mathcal{R}_m$, the agent $i_{(1)}(M(\psi)) = \arg \min_{i \in M(\psi)} \{\alpha_i\}$, should have the highest possible intensity of connections, i.e. $M - 1$.

Given these two properties, I conclude that for any two networks $\psi, \phi \in \mathcal{R}_M$ such that $i_{(1)}(M(\psi)) = i_{(1)}(M(\phi))$, the aggregate output in ψ is the same as in ϕ. The set

$$\Psi_k = \{\psi \in \mathcal{R}_m \mid i_{(1)}(M(\psi)) = k\}$$

is the set of networks in which the agent k is the one with the lowest productivity within the majority group. Then, I am going to denote the aggregate output in those networks by Y_k. It is easy to see, that any time agent i participates in the majority in some networks in Ψ_k, the intensity of his links is always the same. I am going to denote it by x_i^k.

Take two networks $\psi, \phi \in \Psi_k$, such that $k \notin M(\psi)\Delta M(\phi) = \{i, j\}$. Since the aggregate output in ψ is the same as in ϕ, I have that

$$\alpha_i \left(x_i^k - (N - M - 1)\right) = \alpha_j \left(x_j^k - (N - M - 1)\right) \tag{8}$$

Take a network $\psi \in \Psi_k$, such that agent 1 belongs to the majority in $\psi : 1 \in M(\psi)$. The aggregate output in ψ is

$$Y_k = \sum_{i \in M(\psi) \setminus k} \alpha_i \left(x_i^k - (N - M - 1) \right) + \alpha_k (M - 1) + \sum_{i \in N \setminus k} \alpha_i (N - M - 1),$$

but as I noted, $\alpha_i \left(x_i^k - (N - M - 1) \right)$ is constant across i, so

$$Y_k = (M - 1) \left(\alpha_1 \left(x_1^k - (N - M - 1) \right) + \alpha_k \right) + \sum_{i \in N \setminus k} \alpha_i (N - M - 1), \qquad (9)$$

If I know the aggregate outputs for all networks, i.e. Y_k, for $k = m, ..., N$, I can solve for x_1^k using the equations (9), and after that I can solve for the rest of x_i^k using equations (8).

All I need to find is Y_k for $k = m, ..., N$. First I reduce this problem to the problem of two unknowns. Take a network $\psi \in \Psi_k$ such, that agent $m \in M(\psi)$. Then it must be that

$$\alpha_m x_m^k \left(1 + \frac{g}{Y_k} \right) = \alpha_m (m - 1) \left(1 + \frac{g}{Y_m} \right), \qquad (10)$$

so once I know Y_m I can solve for all $Y_k, k > M$.

To come up with the equation for Y_m, observe that

$$\alpha_i x_i^N \left(1 + \frac{g}{Y_N} \right) = \alpha_i x_i^m \left(1 + \frac{g}{Y_m} \right).$$

The aggregate output for networks in Ψ_m is

$$Y_m = \sum_{1 \leq i \leq m-1} \alpha_i x_i^m + \alpha_m (m - 1) + \sum_{m+1 \leq i \leq N} \alpha_i (N - m - 1).$$

By substituting x_i^m with x_i^N from the previous equation I obtain

$$\frac{\left(Y_m - \alpha_m (2m - N) - \sum_{i \in N} \alpha_i (N - m - 1) \right)}{\left(Y_N - \alpha_N (2m - N) - \sum_{i \in N} \alpha_i (N - m - 1) \right)} = \frac{\left(1 + \frac{g}{Y_N} \right)}{\left(1 + \frac{g}{Y_m} \right)}$$

This is an extra equation, that solves for Y_m.

Proof of Proposition 4

First observe that $\mathcal{R}_k$ is internally stable by construction. I only need to prove that $\mathcal{R}_k$ satisfies external stability. With some abuse of notation, by $\overline{U}_i$ I denote a maximum payoff of agent i in the set $\mathcal{R}_k$:

$$\overline{U}_i = \max_{\psi \in \mathcal{R}_k} \{U_i(\psi)\}.$$

Take an arbitrary $\psi \notin \mathcal{R}_k$ and look at the set of agents that would prefer to switch to one of the networks in $\mathcal{R}_k$ rather than stay in ψ:

$$A(\psi) = \{i \in N \mid U_i(\psi) < \overline{U}_i\}$$

If $|A(\psi)| \geq k$, then clearly I can find $\rho \in \mathcal{R}_k$, such that $\rho \succ \psi$. Suppose, $|A(\psi)| < k$.

I am going to set up an induction argument at this point. Assume, that there is a network ψ_l such, that $|A(\psi_l)| < k$. I will show, that there is a network ψ_{l+1}, such that:

(i) $A(\psi_l) \cup j \subset A(\psi_{l+1})$, for some $j \in N$,

(ii) $\psi_l = \psi^0 \overset{S_0}{\to} \psi^1 \overset{S_1}{\to} ... \overset{S_n}{\to} \psi^{n+1} = \psi_{l+1}$ such, that

  (a) $\forall i = 1, ..., n : \overline{U}_{S_i} > U_{S_i}(\psi^i)$, and

  (b) $\bigcup_{i=1}^{n} S_i \subset A(\psi_{l+1})$.

If this statement holds, then I can construct a strictly increasing sequence $A(\psi) \subset A(\psi_1) \subset ... \subset A(\psi_L)$, such that $|A(\psi_L)| = k$. Note that the conditions that I used guarantee, that the sequence of corresponding networks is a b-sequence for a $\rho \succ \psi$, where $\rho \in \mathcal{R}_k$, and $U_{A(\psi_L)}(\rho) = \overline{U}_{A(\psi_L)}$.

Let us get back to proving the induction. I claim that I can find ψ_{l+1} that satisfies the conditions above. Let us look at $\widehat{\psi}_l$ such, that

(a) $\psi_l \overset{A(\psi)}{\to} \widehat{\psi}_l$,

(b) $A(\psi_l)$ form a component in $\widehat{\psi}_l$,

(c) $x_{A(\psi_l)}(\widehat{\psi}_l) = \alpha x_{A(\psi_l)}(\rho)$, where $\rho \in \mathcal{R}_k : U_{A(\psi)}(\rho) = \overline{U}_{A(\psi)}$.

Observe, that the condition (c) guarantees, that for any two agents $i, j \in A(\psi)$

$$\frac{U_i(\widehat{\psi}_l)}{U_j(\widehat{\psi}_l)} = \frac{\overline{U}_i}{\overline{U}_j}$$

and hence there exists $\alpha_0 \in \left[0, \frac{|A(\psi_l)|-1}{k-1}\right]$ such, that $A(\psi_l) = A(\widehat{\psi}_l(\alpha_0))$[17]. Starting from α_0, increase α gradually until $A(\widehat{\psi}_l(\alpha)) \cap (N \backslash A(\psi_l)) \neq \emptyset$ or one of the agents

[17] I abuse notatations slightly. The function $\widehat{\psi}_l(\alpha)$ maps $\alpha \in \left[0, \frac{|A(\psi_l)|-1}{k-1}\right]$ into network described by conditions above

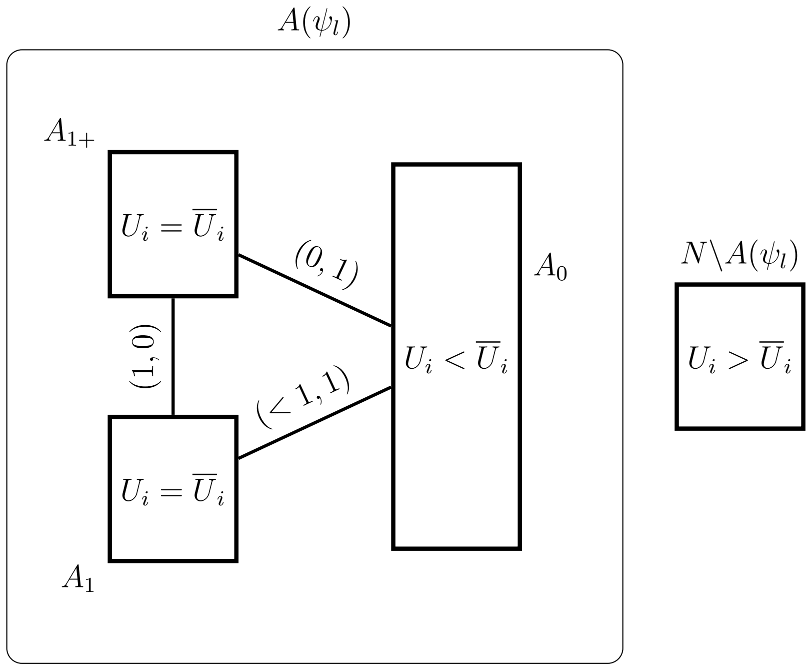


Figure 12: The structure of network ϕ_l^0.

in $A(\psi_l)$ reaches his capacity in links. In principle, it could also be the case, that $U_{A(\psi_l)}(\widehat{\psi}_l(\alpha)) = \overline{U}_{A(\psi_l)}$, but that would be a violation of conditions of the proposition.

If $A(\widehat{\psi}_l(\alpha)) \cap (N \backslash A(\psi_l)) \neq \emptyset$, I am done. If not, then I keep increasing α for all agents that did not reach the capacity constraint yet. Again I stop either when $A(\widehat{\psi}_l(\alpha)) \cap (N \backslash A(\psi_l)) \neq \emptyset$, in which case I am done, or when the agents that did not yet reach the capacity constraint have a payoff $\overline{U}_i$. Suppose the latter happens at network ϕ_l^o.

I am going to show that it is possible to raise the aggregate output of the set of agents $A(\psi_l)$ up to a point, when some agent $i \in N \backslash A(\psi_l)$ have a payoff equal to $\overline{U}_i$. Moreover, it is possible to raise it in a reversible way, i.e. such that the coalition $A(\psi_l)$ can go back to a network ϕ_l^0 only utilizing agents whose current payoff is lower than $\overline{U}$.

First, let us look at the properties of ϕ_l^0 (see Fig. 12). Notice, that agents in A_0 have $|A(\psi_l)|$ incoming connections each. Also, Agents in A_1 have strictly more than $|A_0|$ connections.

If I set $x_i = |A(\psi_l)|$ for all $i \in A_1$. In this case, the payoffs of all agents in A_{2+} become strictly less than $\overline{U}$. I can proportionally increase the intensity for agents in A_{1+} until either their payoffs become equal to $\overline{U}$ again, or one of the agents in $N \backslash A(\psi_l)$ gets payoff $\underline{U}$. Suppose the former happens. If the former happens I repeat the same procedure again, i.e. set $x_i = |A(\psi_l)|$ for all $i \in A_1$.

I continue until I hit the network in which there exists $i \in N \backslash A(\psi_l)$ such that

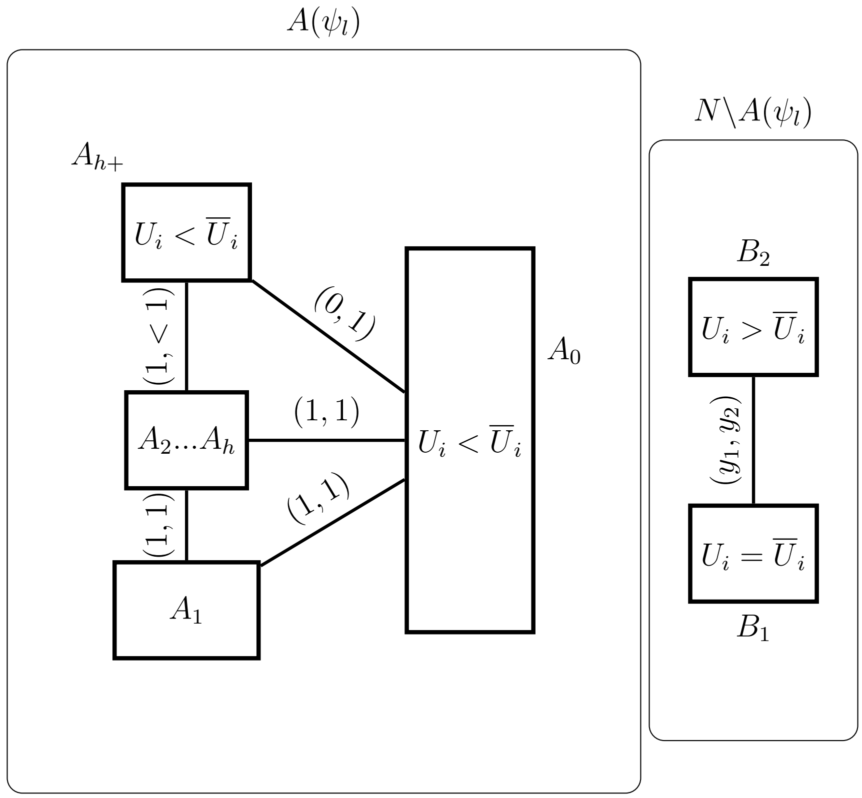


Figure 13: The structure of network ψ_l^h.

$U_i = \overline{U}_i$. Suppose it happens at step h, i.e. the network ϕ_l^h.

Since $|A_0| > |N\backslash A(\psi_l)|$, there exists network $\widehat{\phi}_l^h$ such, that agents in B_1 have the same intensity of incoming links, but all of them come from A_0, and the rest of the links are exactly the same as in ϕ_l^h.

There exists a sequence of networks, that starts in ϕ_l^h and leads back to ϕ_l^0, and only agents whose payoffs are lower than $\overline{U}$ are active along this sequence. Indeed, agents in A_0 can break link to agents in A_0, which guarantees, that latter have payoffs below $\overline{U}$. After that, agents in $A_1 \cup A_0$ break their links with agents in A_2, and simultaneously link with each other. After this step, agents in A_2 are the ones with the low payoff. Agents proceed in the same fashion, working their way up along the sequence of A_i until they reach the top. This algorithm applies for any transition from ϕ_l^j to ϕ_l^{j-1}.

Notice now, that for any network ψ_l^j I can find a corresponding network $\widehat{\phi}_l^j$, in which agents in B_1 have their links coming from A_0 instead of B_2.

Once I reach $\widehat{\phi}_l^0$, all agents in $A(\psi_l)$ delete all their link. As a result all agents in $A(\psi_l) \cup B_1$ have no links.

Proof of Proposition 5

First, note that

$$\frac{U_i(\Lambda)}{U_j(\Lambda)} = \frac{\alpha_i}{\alpha_j},$$

and

$$\frac{U_i(\psi)}{U_j(\psi)} = \frac{\alpha_i x_i^N}{\alpha_j x_j^N}.$$

By equation (8) I obtain that

$$\alpha_i x_i^N = \alpha_j x_j^N - (\alpha_j - \alpha_i)(N - M - 1),$$

hence,

$$\frac{U_i(\psi)}{U_j(\psi)} - \frac{U_i(\Lambda)}{U_j(\Lambda)} = \left(\frac{x_j^N - (N - M - 1)}{x_j^N} \right) \left(1 - \frac{\alpha_i}{\alpha_j} \right) < 0.$$

References

Arrow, Kenneth J., "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973.

Aumann, Robert and Roger Myerson, "Endogenous Formation of Links Between Players and Coalitions: An Application of the Shapley Value," in A. Roth, ed., *The Shapley Value*, Cambridge University Press, 1988.

Becker, Gary S., *The Economics of Discrimination*, University of Chicago Press, 1971.

Bekkers, Rudi, Geert Duysters, and Bart Verspagen, "Intellectual property rights, strategic technology agreements and market structure: The case of GSM," *Research Policy*, 2002, *31* (7), 1141 – 1161.

Bloch, Francis, "Coalitions and Networks in Industrial Organization," *Manchester School*, 2002, *70* (1), 36–55.

Chen, Kong-Pin, "Sabotage in Promotion Tournaments," *Journal of Law, Economics, and Organization*, 2003, *19* (1), 119–140.

Chwe, Michael Suk-Young, "Farsighted Coalitional Stability," *Journal of Economic Theory*, 1994, *63* (2), 299 – 325.

Diamantoudi, Effrosyni and Licun Xue, "Coalitions, agreements and efficiency," *Journal of Economic Theory*, 2007, *136* (1), 105 – 125.

Dutta, Bhaskar and Suresh Mutuswami, "Stable Networks," *Journal of Economic Theory*, 1997, *76* (2), 322 – 344.

_ , Anne van den Nouweland, and Stef Tijs, "Link formation in cooperative situations," *International Journal of Game Theory*, 1998, *27*, 245–256. 10.1007/s001820050070.

_ , Sayantan Ghosal, and Debraj Ray, "Farsighted network formation," *Journal of Economic Theory*, 2005, *122* (2), 143–164.

Eeckhout, Jan, "Minorities and Endogenous Segregation," *Review of Economic Studies*, 01 2006, *73* (1), 31–53.

Goyal, Sanjeev, *Connections : An Introduction to the Economics of Networks*, Princeton University Press, Princeton, 2007.

_ and Jose Luis Moraga-Gonzalez, "R&D Networks," *RAND Journal of Economics*, Winter 2001, *32* (4), 686–707.

– **and Sumit Joshi**, "Networks of collaboration in oligopoly," *Games and Economic Behavior*, April 2003, *43* (1), 57–85.

Greenberg, Joseph, *The Theory of Social Situations: An Alternative Game-Theoretic Approach*, Cambridge University Press, 1990.

Hagedoorn, John, "Inter-firm R&D partnerships: an overview of major trends and patterns since 1960," *Research Policy*, 2002, *31* (4), 477 – 492.

Harsanyi, John C., "An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition," *Management Science*, 1974, *20* (11), pp. 1472–1495.

Herings, P. Jean-Jacques, Ana Mauleon, and Vincent Vannetelbosch, "Farsightedly stable networks," *Games and Economic Behavior*, 2009, *67* (2), 526–541.

Jackson, Matthew O., *Social and Economic Networks*, Princeton, NJ, USA: Princeton University Press, 2008.

– **and Asher Wolinsky**, "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 1996, *71* (1), 44 – 74.

Konishi, Hideo and Debraj Ray, "Coalition formation as a dynamic process," *Journal of Economic Theory*, May 2003, *110* (1), 1–41.

Konrad, Kai A., "Sabotage in rent-seeking contests," *Journal of Law, Economics, and Organization*, 2000, *16* (1), 155–165.

Lazear, Edward P., "Pay Equality and Industrial Politics," *The Journal of Political Economy*, 1989, *97* (3), 561–580.

Marinucci, Marco and Wouter Vergote, "Endogenous Network Formation in Patent Contests and Its Role as a Barrier to Entry," *forthcoming in Journal of Industrial Economics*, 2011.

McAdams, Richard H., "Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination," *Harvard Law Review*, 1995, *108* (5), pp. 1003–1084.

Pęski, Marcin and Balázs Szentes, "Spontaneous Discrimination," *mimeo*, 2011.

Phelps, Edmund S, "The Statistical Theory of Racism and Sexism," *American Economic Review*, September 1972, *62* (4), 659–61.

Ray, Debraj, *A Game-Theoretic Perspective on Coalition Formation*, Oxford University Press, October 2007.

_ **and Rajiv Vohra**, "Equilibrium Binding Agreements," *Journal of Economic Theory*, 1997, *73* (1), 30 – 78.

Roth, Alvin E. and Marilda A. Sotomayor, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis,* Econometric Society Monographs, Cambridge University Press, 1992.

Salop, Steven C. and David T. Scheffman, "Raising Rivals' Costs," *The American Economic Review*, 1983, *73* (2), 267–271.

_ **and** _ , "Cost-Raising Strategies," *The Journal of Industrial Economics*, 1987, *36* (1), pp. 19–34.

von Neumann, John and Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944.