Chapter 11

Game-Theoretic Modeling of Network Formation

Chapter 6 showed the importance of understanding strategic network formation, highlighting a tension between social welfare and individual incentives to maintain relationships. As discussed in the introduction to that chapter, there are many settings where links are formed in a cognizant manner, especially in applications where the individual nodes are firms, organizations, or countries, which have explicit objectives when it comes to their relationships. And, even friendships and other more purely social relationships exhibit costs and benefits that influence which ones emerge and endure. This leads to a rich set of questions regarding the modeling of network formation.

- There are issues of how to define equilibrium and stability notions: Can players adjust many relationships at a time? Can players coordinate their choices?

- There are issues of sophistication: Are players farsighted or myopic? Do players take into account how the links that they form influence others? Do they make errors?

- There are issues of dynamics: Can players revise links over time? Are there evolutionary pressures on their choices? Are there random forces of opportunity that determine which relationships can be formed?

- There are issues of bargaining and transfers: Can players compensate others for the relationships that they do (or do not) maintain, either through negotiated payment or through favors? Is this bargained over at the time of network formation?

- There are issues of the formation of directed networks: How should we model network formation if links can be formed unilaterally? How does it depend on whether one or both players involved in a directed link benefit from its presence?
- There are issues associated with the strength of links: How do we model strength of links? What happens if players can allocate effort or resources to maintaining different links? How will the outcome depend on the context?

Given the diversity of questions above, there is no simple message emerging this chapter. Instead, this chapter examines foundational questions concerning network formation. There is, however, a pervasive question that ties things together: "Under what circumstances do incentives lead to the formation of efficient networks?"

11.1 Defining Stability and Equilibrium

In Chapters 1 and 6 we saw the concept of pairwise stability as a method of modeling network formation. Why should we use that concept rather than an explicit game? This was discussed briefly in Section 6.1, but let us re-examine that question while examining possible non-cooperative games of network formation and alternative notions of equilibrium and stability.

11.1.1 An Extensive Form Game of Network Formation

Aumann and Myerson [20] provided an early model of network formation. More specifically they were interested in the formation of a communication graph that served as a basis for a cooperative game (as discussed in Section 12.2). The formation game they examined extends to serve as a model of network formation and is described in our setting as follows.

Players move sequentially and propose links which are then accepted or rejected. The extensive form game is based on an ordering over all possible links, denoted $(i_1j_1, \ldots, i_Kj_K)$. When the link i_kj_k appears in the ordering, the pair of players i_kj_k decide on whether or not to form that link, knowing the decisions of all pairs coming before them and forecasting the play that will follow them. Player i_k moves first and says "yes" or "no," and then player j_k says "yes" or "no," and the link forms if both say "yes". A decision to form a link is binding and cannot be undone. However, if a pair i_kj_k decide not to form a link, but some other pair coming after them forms a link, then $i_k j_k$ are later allowed to reconsider their decision. This feature allows a player 1 to make a threat to 2 of the form "I will not form a link with 3 if you do not. But if you do form a link with 3, then I will also do so." The way in which this is captured is that the game moves through all the links a first time. If at least one link forms, then the game starts again with the same ordering, moving this time only through the links that have not yet been formed. The game continues to move through the remaining unformed links in order, until either all links are formed or there is a round such that all of the links that have not yet formed have been considered and no new links have formed.

This approach has the advantage of always having a pure strategy subgame perfect equilibrium.[1] Its main drawbacck is that the game can be very difficult to solve, even in very simple settings with only a few players. Moreover, the ordering of links can have a substantial impact on which networks emerge, and it is not so clear what a natural ordering is.

11.1.2 A Simultaneous Link-Announcement Game

Given the intractability of the sequential ordering and its inherent asymmetries, Myerson [473] suggests another game in the context of the formation of communication graphs, which also extends to the formation of networks. It is probably the most "natural" simultaneous move game of network formation. I will refer to it as the *link-announcement game*. Each player simultaneously announces the set of players with whom he or she wishes to be linked. The links that are formed are those such that both of the players involved in the link named each other.

More formally, the strategy space of player i is $S_i = 2^{N \setminus \{i\}}$.[2] If $s \in S_1 \times \cdot \times S_n$ is the profile of strategies played, then link ij forms if and only if both $j \in s_i$ and $i \in s_j$. The network that forms is

$$g(s) = \{ij | i \in s_j \text{ and } j \in s_i\}.$$

In modeling the networks that emerge from the link announcement game, we can use any of a variety of game theoretic solutions, such as Nash equilibrium.

The payoffs in the link formation game are described by a profile of utility functions, $u = (u_1, \ldots, u_n)$, which indicate the payoffs of each player as a function of the network.

[1] It is a finite game of perfect information, and so has at least one solution found by backward induction.

[2] 2^A is a notation for the set of all subsets of A, also known as the power set of A.

Figure 11.1.2. Both Networks are Nash Equilibria of the Link Announcement Game

A network $g \in G(N)$ is *Nash stable* if it results from a pure strategy Nash equilibrium of the link-announcement game, where player i's payoff as a function of the profile of strategies is $u_i(g(s))$.

This game is much easier to describe than the Aumann and Myerson extensive form, and it avoids inducing a priori asymmetries between the players or links. Arguably, any network that will be stable over time in the sense that no players would like to delete any links would have to be an equilibrium of this game. The main drawback of the game that it has too many Nash equilibria, including some which are easily seen to be unreasonable. In particular, $s_i = \emptyset$ for all i is *always* a Nash equilibrium, regardless of the payoffs. Each player refuses to link with any other player, because he or she correctly forecasts that the other players will do the same. This is seen most starkly in the dyadic case, as pictured in Figure 11.1.2. Here both networks are equilibria, although clearly the network where the link forms is the only reasonable one.

Thus, while the link formation game may at first seem to be a natural way to model network formation, it is not reasonable when using Nash equilibrium alone as a solution concept. Basically, Nash equilibrium allows players to refuse to form links, and thus effectively to "delete" links, but it does not capture the fact that it may be mutually advantageous for two players to form a new relationship. We need to move beyond

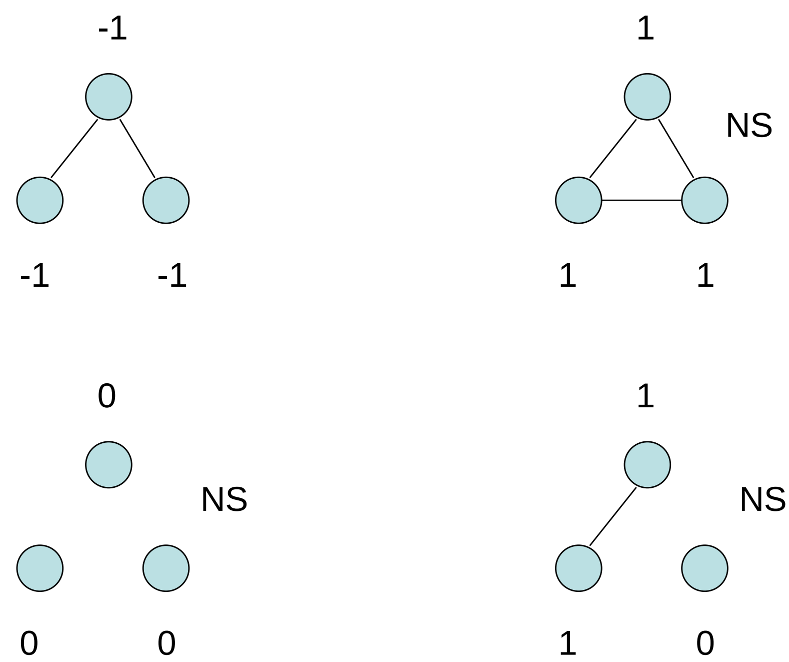

Figure 11.1.2. All Networks except Two-Link Networks are Nash Stable and all Strategies in the Link-Announcement Game are Undominated.

Nash equilibrium to capture this.

In the example pictured in Figure 11.1.2, it is a dominant strategy for each player to propose to link with the other player. This suggests that one way around the shortcoming of Nash equilibrium in modeling network formation might be to use a refinement of Nash equilibrium where players do not play weakly dominated strategies. However, a slight enrichment of the example in Figure 11.1.2 shows that this will not work. Consider a triad such that the empty network leads to a payoff of 0 for all players, a single link leads to a payoff of 1 for each of the linked players (and 0 for the other), a two-link network leads to a payoff of -1 for all players, and the complete network leads to a payoff of 1 for all players. This pictured in Figure 11.1.2.

In this example, all strategies in the link-announcement game are undominated. This means that the empty network is an outcome of a Nash equilibrium that only uses undominated strategies, where every player announces the empty set of players.[3]

In order to address the fact that it takes the consent of both players to form a link

[3]This is also a trembling hand perfect equilibrium. It is not a strict Nash equilibrium. However, requiring strictness in this game leads to very general existence problems, as outlined in Exercise 11.3.

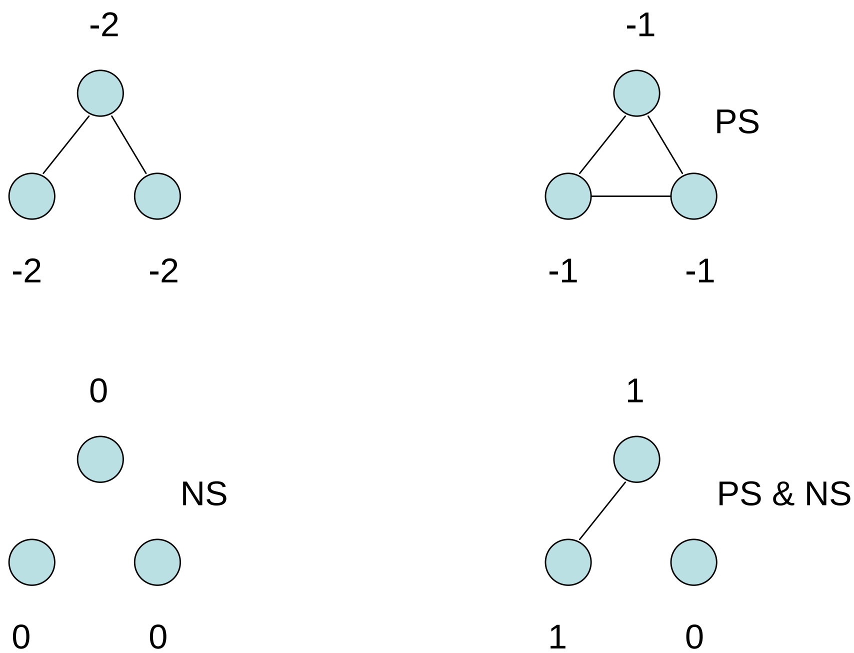

Figure 11.1.3. An Over-Connected Pairwise Stable Network. Payoffs on permutations of these networks are the permuted payoffs.

in an undirected network, one has to explicitly consider coordinated actions on the part of pairs of players. This forces one to move beyond Nash equilibrium, and standard refinements of it, and somehow coalitional considerations (at least for pairs of players) have to be considered. That is the reasoning behind pairwise stability.

11.1.3 Pairwise Nash Stability

Although pairwise stability overcomes the difficulties inherent in examining Nash equilibria of the link-announcement game, it restricts attention to changes of one link at a time. This can lead to over-connected networks being pairwise stable, even when some player would benefit from deleting multiple links at once, as is pictured in the example in Figure 11.1.3.

In the example in Figure 11.1.3, the "reasonable" network is the one that is both Nash stable and pairwise stable. This has led to a concept of pairwise Nash stable networks.

A network is *pairwise Nash stable* if it is Nash stable and pairwise stable.[4]

[4]This refinement was first discussed by Jackson and Wolinsky [343] and has been used in various

As people who have worked with game theoretic solution concepts are aware, given any equilibrium or stability concept one can find some setting where it makes a questionable prediction or is deficient in some way. This reflects the fact that as applications vary, so do the sorts of deviations that are feasible or salient. Can players communicate and coordinate their actions? Can they make multiple changes at once? Can more than two players coordinate at a time? Is the application one where players move simultaneously, or are we really just looking for stable points of some process where the timing might be sequential? Can players revise their actions? Although this means that finding a single solution concept that works well in all settings is a futile activity,[5] it does not mean that we should give up on studying variations of solutions, since different ones can be more appropriate and/or useful in different settings, and we should not give up modeling simply because a universal solution concept is not available.

A number of variations on stability concepts are discussed in the exercises of this chapter, but let me push a bit further to discuss a few of the basic considerations and some solutions that have been used to capture them.

11.1.4 Strong Stability

In some settings, players have open lines of communication and more than two players can coordinate their link formation decisions at the same time. Alternatives to pairwise stability and pairwise Nash stability that consider larger coalitions of players were first considered by Dutta and Mutuswami [200].[6] The following is a slight variation on Dutta and Mutuswami's definition, from Jackson and van den Nouweland [338]. It always selects from among the pairwise Nash stable networks.

A network $g' \in G$ *is obtainable from* $g \in G$ *via deviations by* $S \subset N$ if

(i) $ij \in g'$ and $ij \notin g$ implies $\{i, j\} \subset S$, and

(ii) $ij \in g$ and $ij \notin g'$ implies $\{i, j\} \cap S \neq \emptyset$.

---

studies (e.g., Goyal and Joshi [283] and Belleflamme and Bloch [48]). For more detailed studies of the relationships between Nash stability, pairwise Nash stability, and pairwise stability, as well as refinements of Nash equilibria that justify pairwise Nash stability for a wide class of settings, see Calvó-Armengol and Ilkilic [118], Gilles and Sarangi [266], Bloch and Jackson [75], and Ilkilic [323].

[5] This view is not universally held among game theorists.

[6] There was some also some early discussion of core-based allocations in the exchange network literature; e.g., by Bienenstock and Bonacich [58].

The above definition identifies changes in a network that can be made by a coalition S without the consent of any players outside of S. Part (i) requires that any new links only involve players in S, in line with the consent of both players being needed to add a link. Part (ii) requires that at least one player of any deleted link be in S, in line with the idea that either player in a link can unilaterally sever the relationship.

A network g is *strongly stable* with respect to a profile of utility functions $u = (u_1, \ldots, u_n)$ if for any $S \subset N$, g' that is obtainable from g via deviations by S, and $i \in S$ such that $u_i(g') > u_i(g)$, there exists $j \in S$ such that $u_j(g') < u_j(g)$.

The relationship between this definition and the definition of Dutta and Mutuswami [200] is examined in Exercise 11.5, and relates to whether or not a blocking coalition has to just have some members be strictly better off and others be weakly better off, as above, or have all members of a blocking coalition be strictly better off. The definition given here is consistent with pairwise stability, as the strongly stable networks are always a subset of pairwise stable networks, and in fact a subset of the pairwise Nash stable networks.

We see an example of the implications of strong stability in Figure 11.1.4. In that example, a one link network is pairwise Nash stable, as is the complete network. However, only the complete network is strongly stable.

Strongly stable networks are necessarily Pareto efficient, as one of the groups that can potentially deviate to form a better network is the society as a whole. Thus, if some network is Pareto dominated by another network, so that the second network is weakly better for all players and strictly better for some, then that will provide a viable deviation and so the dominated network will not be strongly stable. In addition to this efficiency property, strongly stable networks are immune to all sorts of coordinated deviations by players, and so are very robust. However, they only make sense as a predictive tool in situations where such coordination is feasible, and thus might be limited to situations where players have substantial knowledge about the opportunities for network formation and the payoffs, and can also readily communicate with each other. Also, while strongly stable networks are very robust and are Pareto efficient, there are many contexts where they fail to exist. The issue of existence of various sorts of stable networks is an important one to which I now turn, before returning to discuss other ways of modeling network formation.

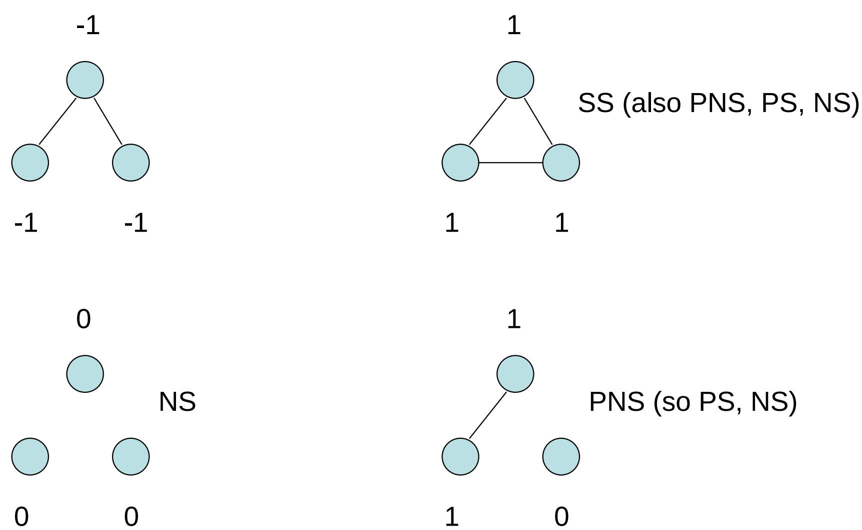

Figure 11.1.4. Strong Stability: An example with multiple pairwise stable, Nash stable, and pairwise Nash stable networks, but a unique strongly stable network. Permutations of these networks lead to permuted payoffs.

11.2 The Existence of Stable Networks

While the the link-announcement game always has an equilibrium, this is due to the fact that there is always a trivial equilibrium where no links form because no player expects any other player to be willing to form a link. Once we move to refinements for which the empty network is not always stable, such as pairwise stability, pairwise Nash stability, strong stability, or other such refinements, existence is not always guaranteed. Let us explore when stable networks exist.

11.2.1 Improving Paths, Dynamics, and Cycles

In studying the existence of various forms of stable networks, it is useful to consider some simple dynamics. The idea is to examine the sequences of networks that might emerge as players add or delete links to improve their payoffs. The resting points of these processes will be stable points, and so understanding these sequences helps in understanding when stable networks exist and what might happen when stable networks do not exist.

Let us say that two networks are *adjacent* if they differ by only one link. That is, g and g' are adjacent if either $g' = g + ij$ for some $ij \notin g$ or $g' = g - ij$ for some $ij \in g$.

A network g' *defeats* an adjacent network g if either

- $g' = g - ij$ and $u_i(g') > u_i(g)$, or if
- $g' = g + ij$ and $u_i(g') \geq u_i(g)$ and $u_j(g') \geq u_j(g)$, with at least one inequality holding strictly.

A network is pairwise stable if and only if it is not defeated by an (adjacent) network.

The following definition from Jackson and Watts [340] captures this notion of sequences of networks where each network defeats the previous one.

An *improving path* is a sequence of distinct networks $\{g_1, g_2, \ldots, g_K\}$, such that each network g_k with $k < K$ is adjacent to and defeated by the subsequent network g_{k+1}.

This usage of "path" refers to a sequence of networks and should not be confused with a path inside a network. The idea here is to examine the sequences of networks that can emerge as players add and delete links in a way that makes them better off. Clearly, the resting points of such a process are the pairwise stable networks. That is,

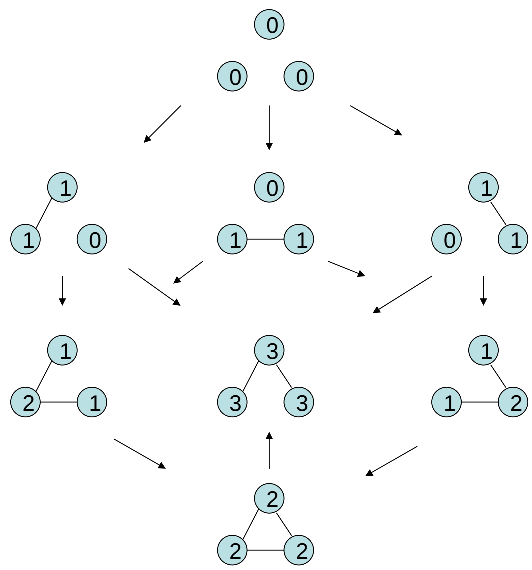

Figure 11.2.1. Improving Paths: The payoffs are listed in the nodes and the arrows point towards a network that defeats the one from which the arrow emanates. Following the arrows provides improving paths. There is a unique pairwise stable network.

a network is pairwise stable if and only if it has no improving paths emanating from it. Improving paths are illustrated in Figure 11.2.1.

The notion of improving paths is a myopic one, in that the agents involved in adding or deleting links are doing so without forecasting how their actions might affect the evolution of the process. This is a natural variation on best response dynamics, and has some experimental justification (e.g., see Pantz and Ziegelmeyer [497]), but nevertheless exhibits some forms of bounded rationality. I return to discuss farsighted network formation in Section **??**.

If there does not exist any pairwise stable network, then there must exist at least one *improving cycle* – that is, a sequence of adjacent networks $\{g_1, g_2, \ldots, g_K\}$ such that each defeats the previous one and such that $g_1 = g_K$. The possibility of cycles and non-existence of a pairwise stable network is illustrated in the following example from Jackson and Watts [340].

EXAMPLE **11.2.1** *Non-existence of a Pairwise Stable Network*

There are $n \geq 4$ players who obtain payoffs from trading with each other. The players have random endowments and the benefits from trading depend on the realization of these random endowments. The more players who are linked, the greater the gains from trade, but with diminishing marginal returns.

In particular, there is a cost of a link of $c = 5$ to each player involved in the link. The utility of being alone is 0. Not accounting for the cost of links, the benefits to each player in a dyad is 12, the benefits for being connected (directly or indirectly) to two other players is 16, and of being connected to three other individuals is 18.[7]

The resulting payoffs for several of the key network configurations are pictured in Figure 11.2.1.

[7] In terms of the economic background behind these payoffs, they can be derived as follows. There are two consumption goods and players each have a Cobb-Douglas utility function for the two goods of $u(x, y) = 96xy$, where x is the consumption of the first good and y is the consumption of the second good. A player's endowment is either (1,0) or (0,1), each with probability 1/2, and the realizations are independent across players. Players within each component trade to a Walrasian equilibrium within their component, regardless of the precise set of links in the component. For example, the networks $\{12, 23\}$ and $\{12, 23, 13\}$ lead to the same expected trades, but different costs of links. In a dyad there is a $\frac{1}{2}$ probability that one player has an endowment of (1,0) and the other has an endowment of (0,1). They then trade to the Walrasian allocation of $(\frac{1}{2}, \frac{1}{2})$ each and so their utility is 24 each. There is also a $\frac{1}{2}$ probability that the players have the same endowment and then there are no gains from trade and they each get a utility of 0. Expecting over these two situations leads to an expected utility of 12. Similar calculations for more players lead to the claimed payoffs.

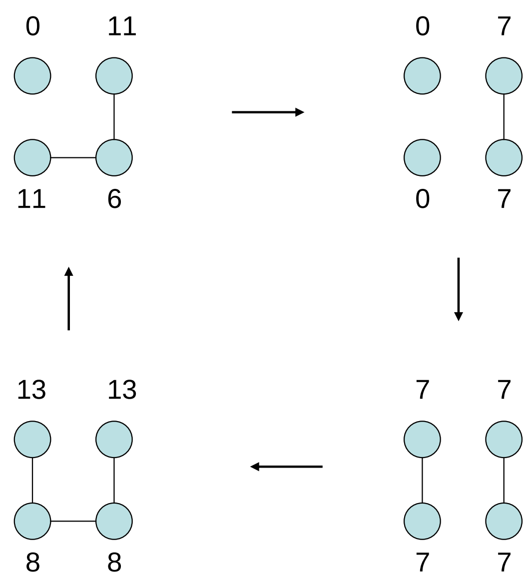

Figure 11.2.1. Nonexistsence of a Pairwise Stable Network: Payoffs pictured for one, two and three link networks and an improving cycle including the four networks. Networks with more than three links are defeated by networks with fewer links.

Any network with more than three links in this example is defeated by a network with fewer links, as some players will save the link cost by severing the link, and yet the full trading benefits are already realized with just three links. The critical aspect of the example is that two separate dyads gain by forming a link between them, expanding the network from two to four players. However, in any network of four players that has just three links, one of the players who has more than one link will save 5 utils in cost by severing a link and only lose 2 utils in trading benefits.

Clearly, the nonexistence of pairwise stable networks also implies the nonexistence of pairwise Nash stable networks. Moreover, pairwise Nash stable networks can fail to exist even when the sets of pairwise stable networks and Nash stable networks are both nonempty (see Exercise 11.8).[8]

While the above example shows that pairwise stable networks may not exist in some settings, there are settings where pairwise stable networks always exist. We have already seen several including distance-based generalizations of the connections model, the co-author model of Section 6.4, and a variety of market settings from Chapter 10.

A set of sufficient conditions for the existence of pairwise stable and pairwise Nash stable networks comes from ruling out improving cycles. Having no improving cycles also means that any dynamics that follow improving paths will find stable networks. The absence of improving cycles is related to the existence of what is known as what is known as an (ordinal) potential function.

An *ordinal potential function* for a society N with payoff functions $u = (u_1, \ldots, u_n)$ is a function $f : G(N) \to I\!R$ such that g' defeats g if and only if $f(g') > f(g)$ and g' and g are adjacent.[9]

For a society N, the payoff functions $u = (u_1, \ldots, u_n)$ *exhibit no indifference* if for any two adjacent networks, one defeats the other.

PROPOSITION **11.2.1** *[Jackson and Watts [339]] If a society (N, u) has an ordinal potential function then there are no improving cycles. Conversely, if a society is such*

[8]The existence of Nash stable networks is guaranteed by the observation that the empty network is always Nash stable. As argued above, that is not an interesting existence property as it simply derives from a coordination failure. The more interesting existence issue deals with pairwise stable and Nash stable networks.

[9]Ordinal potential functions were defined by Monderer and Shapley [451] for non-cooperative games, based on better responses. This definition is similar in spirit, but adapted to network formation settings and pairwise stability. For a more detailed look at the relationship between various sorts of potential functions and existence of equilibria in network formation games, see Gilles and Sarangi [270].

that payoffs exhibit no indifference, then there are no improving cycles only if there exists an ordinal potential function.

This result echoes results on ordinal potentials in non-cooperative games, and a corollary is that if society admits an ordinal potential function then there exists a pairwise stable network. This follows since any network that maximizes f must be undefeated.

Proof of Proposition 11.2.1: The proof that there are no improving cycles if a society has an ordinal potential function is straightforward and left to the reader. So, let us show that the converse holds when payoffs exhibit no indifference.

Suppose that there are no improving cycles and payoffs exhibit no indifference. Define $f(g)$ to be the number of networks g' such that there exists an improving path from g' to g.[10] We need to show that this is an ordinal potential function. Consider two adjacent networks g and g'. If g defeats g', then every network that has an improving path leading to g' also has an improving path leading to g. Moreover, g' has an improving path leading to g, but the reverse is not true, as otherwise there would be an improving cycle. Therefore $f(g) > f(g')$. Conversely, if $f(g) > f(g')$ and g' defeated g then we reach a contradiction by a similar argument. Thus, $f(g) > f(g')$ implies that g' does not defeat g, which by the no indifference condition implies that g defeats g'. Therefore, f is an ordinal potential function. ∎

The no indifference condition is needed for the proposition's conclusions, and is the subject of Exercise 11.10.

We can easily extend this analysis to the case of pairwise Nash stability, with some proper modifications to the definitions above. Let us say that the networks $g \neq g'$ are *weakly adjacent* if g' is either obtained from g by the addition of a single link or obtained by the deletion of some set of links such that there is some agent involved in all of the deleted links. If we then redefine "defeats," "improving path," "improving cycle," and "ordinal potential function" accordingly, Proposition 11.2.1 still holds, and the existence of an ordinal potential function with those definitions implies the existence of pairwise Nash stable networks.

The existence of an ordinal potential function is a demanding condition, and this emphasizes that the absence of improving cycles is also a demanding condition. Nevertheless, there exist situations where this is satisfied. In fact, there are always transfers that lead to payoffs for which there exists an ordinal potential function (see Section **??**).

[10] The idea behind this construction for finding ordinal potential functions is due to Milchtaich [443].

In some cases, simply examining the sum of all payoffs leads to an ordinal potential function. This works in the case of the Corominas-Bosch's [167] model of buyer-seller networks from Section **??** (see Exercise 11.11). It also works for the symmetric connections model when $\delta > c > n(\delta - \delta^{n-1})$ or when c is very large or small. More generally, in cases where potential functions exist, the analysis can be greatly simplified.

11.2.2 The Existence of Strongly Stable Networks

Let us now turn to examine the existence of strongly stable networks. Let us start by showing that strong stability demands that certain patterns in payoffs be present.

A profile of utility functions or payoff functions $u = (u_1, \ldots, u_n)$ is *anonymous* if for every permutation π on N (a one-to-one function mapping the set of agents N to N), it follows that $u_{\pi(i)}(g^\pi) = u_i(g)$, where $g^\pi = \{\{\pi(i), \pi(j)\}|ij \in g\}$ is the network obtained from g by permuting the positions of agents according to π.

Anonymity requires that payoffs depend only on players' positions within the network and not their labels. Payoff relevance is captured through the network structure and not through other innate characteristics of the players.

Recall the definition of component decomposability from Section 6.6.2: A profile of utility functions or payoff functions $u = (u_1, \ldots, u_n)$ is *component decomposable* if $u_i(g) = u_i(g')$ whenever $C_i(g) = C_i(g')$.

Component decomposability requires payoffs to players to depend only on the structure of their components and not to depend on the structure of other components. This allows for externalities within components, but precludes externalities across components. It holds in some settings, such as those where payoffs only depend on communication patterns within components, but not in ones where separate components interact with each other.

PROPOSITION **11.2.2** *[Jackson and van den Nouweland [338]] Consider a society with anonymous and component decomposable payoffs. If there exists a strongly stable network $g \in G(N)$ that is not connected, then all players must get an equal payoff.*

This is a variation on what is known as an "equal treatment" condition, which is implied in a wide variety of settings when requiring stability with respect to deviations by groups of agents. The proof is straightforward and only sketched here. The idea is that if all players do not get the same payoff, then there is a player in one component who gets less than the payoff of a player in another component. By replacing the higher payoff player j with the lower payoff player i, the payoffs to the other players in the new component of player i do not change, but i's payoff goes up (as implied by the anonymity and decomposability of payoffs when ensure that i's new payoff is j's old payoff), which is an improving deviation.

Proposition 11.2.2 shows that the existence of strongly stable networks imposes stringent requirements. For instance, in the symmetric connections model it implies that (for generic choices of parameters) the only networks that could possibly be strongly stable are networks with strong symmetry properties: those such that the cardinality of each extended neighborhood of every player is identical.

It is clear that if payoffs are equal across all players at every network, then strongly stable networks exist and coincide with the efficient networks. In that case, players' payoffs are perfectly aligned with society's total payoff. However, that demands that transfers be made across components in many contexts. Sufficient conditions for existence of strongly stable networks when there are no transfers across components require some definitions from the next chapter and are explored in Exercise 12.8.

11.3 Directed Networks

The modeling of network formation with directed networks differs from that with undirected networks, as links can be formed unilaterally, and the Nash equilibrium of a formation game becomes an appropriate modeling tool.

Clearly, whether a network is directed or undirected is not just a modeling choice, but instead depends on the application. Although many social and economic relationships involve some consent of both parties, there are some applications where links can be formed unilaterally. For example, one article can cite another without the consent of the first, and a web page can link to another without its consent. In those applications, one needs to adjust the network formation model to account for the unilateral nature of the formation process.[11]

In the case of directed networks, we can still write the payoffs as a function of the network that is formed, where now (fixing a society N) g is a directed network and

[11] A temptation is to mention things like phone calls or other sorts of broadcasting as falling into the directed case. However, those fall into a different category altogether. A phone call involves an asymmetry in the process since one person initiates the action, but it also requires that the both people be willing to hold a conversation, which is a costly activity. Access to many people and organizations is guarded.

$u_i(g)$ represents the utility to player i if g is the directed network that is formed. In the case where players can form a directed link unilaterally, one way to model network formation is to have each player list the set of directed links that he or she wishes to form (and the player can only list links from him or herself to another player), and then have the resulting network be the union of the listed links. This was suggested by Bala and Goyal [28]. More formally, we model this as follows.

A network g' is *obtainable from a network g by player i* if $g'_{kj} \neq g_{kj}$ implies that $k = i$.

Thus, a network g' is obtainable from a network g by player i if the only changes in the network involve links that are directed from i to other players.

A directed network g is *directed Nash stable* if $u_i(g) \geq u_i(g')$ for each i and all networks g' that are obtainable from network g by player i.

In cases where it is clear that we are discussing directed networks, I omit the "directed" from "directed Nash stability" and simply refer to a network as being Nash stable. Thus, a directed network g is Nash stable if and only if it is the outcome a Nash equilibrium of a game where the players simultaneously announce lists of directed links from themselves to other players and the network that forms is the union of those lists.

11.3.1 Two-Way Flow

There are several things to consider in a directed network in terms of how benefits accrue. For instance, in the case of a citation network, different values come from being cited as opposed to citing. That is similar in the case of web pages. Having links to other web sites can enhance the value of a web page and make it more attractive to visit, and being linked to by other sites makes a site easier to find and thus more likely to be visited. In this way, it might be that one side initiates a link, and yet both sides benefit from the link being present. As a first approximation of this, we can keep track of who forms the link, as that might involve specific costs (for instance space on a web page, time, etc.), but then allow the benefits of a link to be bilateral. This is what Bala and Goyal [28] term *two-way flow*.

11.3.2 Distance-Based Utility

To get some feeling for the formation of directed networks, let us start by considering a variation on the distance-based utility model from Section 6.3, but adjust this to allow for two-way flow and directeds.

Given a directed network g, let $\widehat{g}$ denote the undirected network obtained by allowing an (undirected) link to be present whenever there is a directed link present in g. That is, let $\widehat{g}_{ij} = \max(g_{ij}, g_{ji})$.

Recall that in the distance-based model, players get benefits from connections and indirect connections to other agents, where the value that they obtain from indirect connections is a decreasing function of the distance to the other player.

Let $b : \{1, \ldots, n-1\} \to I\!R$ denote the net benefit that a player gets from indirect connections as a function of the distance between the agents. The *distance-based utility model* is one where an agent's utility can be written as

$$u_i(g) = \sum_{j \neq i : j \in N^{n-1}(\widehat{g})} b(\ell_{ij}(\widehat{g})) - d_i(g)c,$$

where $\ell_{ij}(\widehat{g})$ is the shortest path length between i and j in the undirected network obtained from g and $d_i(g)$ is i's *outdegree*. Let $b(k) > b(k+1) > 0$ for any k and $c \geq 0$.

Again, this embodies the idea that a player sees higher benefits for being closer to other players. A special case of the distance-based utility model, analyzed by Bala and Goyal [28], is a directed adaptation of the symmetric connections model, where $b(k) = \delta^k$.

Proposition 6.3.1, characterizing the efficient networks in the undirected distance-based utility model, generalizes directly and shows that efficient networks in a directed version of the distance-based utility model share the same features as the symmetric connections model. The only difference is an adjustment that reflects the fact that only one player bears the cost of a link instead of two.

Let us say that a directed network g is a *directed star* if the associated undirected network $\widehat{g}$ is a star and if $g_{ij} = 1$ then $g_{ji} = 0$, so that links between two players only go in one direction.

PROPOSITION **11.3.1** *The efficient networks in the (directed version of the) distance-based utility model*

(i) consists of one directed link between each pair of players if $c < 2(b(1) - b(2))$,

(ii) is a directed star encompassing all nodes if $2(b(1)-b(2)) < c < 2b(1)+(n-2)b(2)$, *and*

(iii) is the empty network if $2b(1) + (n-2)b(2) < c$.

PROPOSITION **11.3.2** *Consider the directed version of the distance-based utility model.*

(i) If $c < b(1) - b(2)$, then the directed Nash stable networks are those that have one directed link between each pair of players.

(ii) If $b(1) - b(2) < c < b(1)$, then any directed star encompassing all nodes is directed Nash stable and for some parameters there are other directed Nash stable networks.

(iii) If $b(1) < c < b(1) + \frac{(n-2)}{2}b(2)$, then peripherally sponsored stars[12] are Nash stable and so are other networks (e.g., the empty network),

(iv) If $b(1) + \frac{(n-2)}{2}b(2) < c$, then only the empty network is directed Nash stable.

The proof is straightforward and the subject of Exercise 11.13.

Here we see very similar results to those of the nondirected case. With very high or very low costs to links, the efficient and stable networks coincide, while otherwise they may not. Again, efficient networks take the form of variations on stars or the complete network. The most interesting difference arises in the case of a peripherally-sponsored stars. Instability of stars in the undirected case can stem from the fact the hub of the star has to bear some costs and only sees direct benefits from connections and not any indirect benefits. With directed links, it is possible for only the outside players to direct the links, so that the hub does not have to bear any costs. Nevertheless, there are still inefficiencies, and most notably this arises since only one player bears the cost of a link while many players can benefit from its existence. Indeed, even when one can impose transfers, for a variety of settings there still exist conflicts between stability and efficiency. That is, variations on the results that we saw in Section **??** hold in the directed case, as explored by Dutta and Jackson [197].

11.3.3 One-Way Flow

While the two-way flow directed setting has much in common with the undirected setting, a "one-way flow" directed setting introduces some twists.

If we look at one extreme of the distance-based model, then a simple and intuitive characterization of both efficient and stable networks emerges. In particular, consider a one-way flow directed version of the symmetric connections model where we set $\delta = 1$. This is the benchmark where an arbitrarily distant connection provides the same benefit as a direct connection, and was analyzed by Bala and Goyal [28].

[12] This is a directed star where no link is formed by the center.

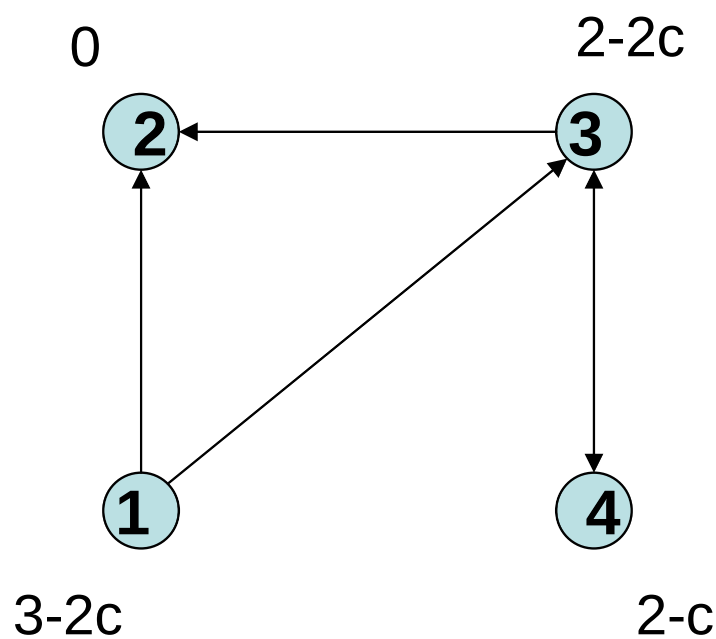

Figure 11.3.3. Payoffs in a One-Way Flow Model with no Decay.

In particular, let $R_i(g)$ denote the number of other players who can be reached from i via a *directed*-path in g. Then i's payoff is

$$u_i(g) = R_i(g) - cd_i(g) \tag{11.1}$$

where $d_i(g)$ is i's out degree.

While this model is clearly extreme since a player at a great distance is just as beneficial as a direct neighbor, it still provides some insight into the one-way flow setting. The payoffs are illustrated in Figure 11.3.3

In this setting, the characterizations of efficient networks and strict Nash stable networks are simple. First, we need a couple of definitions.

A network is an n-player *wheel* if it consists of n directed links and has a single directed cycle that involves n players. A wheel is illustrated in Figure 11.3.3.

A directed Nash stable network is *strictly Nash stable* if any change in the directed links from some player leads to a strictly lower payoff for that player.

PROPOSITION **11.3.3** *[Bala and Goyal [28]] The unique efficient network structure in a one-way flow model where there is no decay and payoffs are as in (11.1) is an n-player wheel if $c < n-1$ and an empty network if $c > n-1$. Moreover, if $c < 1$,*

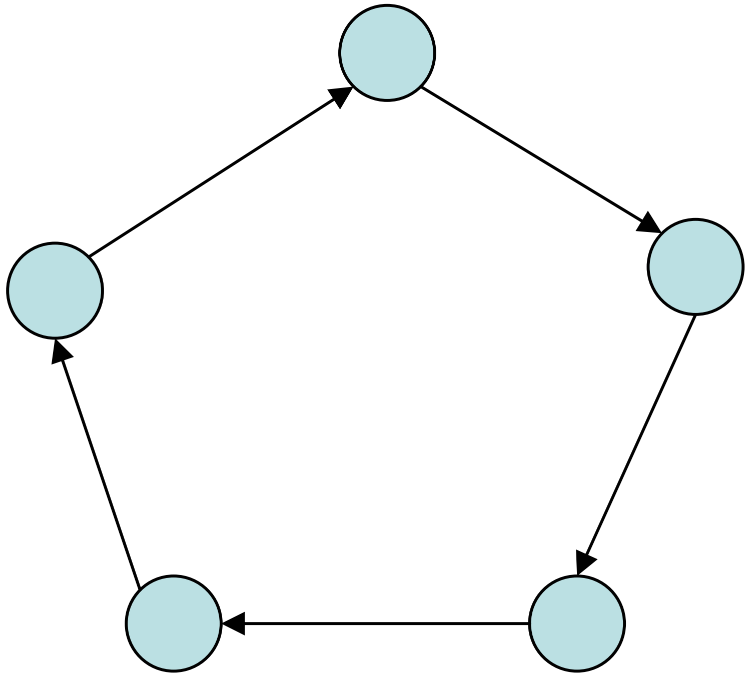

Figure 11.3.3. A Wheel on Five Nodes

then n-player wheels are the (only) strictly Nash stable networks, if $1 < c < n-1$ then n-player wheels and empty networks are the (only) strictly Nash stable networks; and if $c > n-1$, then the empty network is the unique strictly Nash stable network.

Proof of Proposition 11.3.3: First, let us show that a k player wheel is the unique total payoff maximizing network among those that are nonempty, involve at least k links and exactly k players have a link in or out. The cost is at least kc for any such network, which is the cost of a wheel. In order to have at least as high a payoff as a wheel, it must be that $R_i(g) = k-1$. Thus, there is a directed cycle containing all k players. If the network is not a wheel, and it contains a directed cycle with k players, it has more than k links, and so leads to a lower payoff than the k-player wheel. Thus, the only possibilities besides wheels for efficient networks are the empty network and networks that involve k players but have fewer than k links. In the latter case, it must have $k-1$ links to include k players in a component. Here thus must be a player i who has at least one link in but no links out and another player j who has at least one link out but no links in. Given that it is efficient for j to link to some k (and since j has no links in, only j benefits from that link), adding a link from i to j would increase payoffs by even more than the link from j to k does on the margin (since $R_i(g+ij) - R_i(g) \geq 1 + R_j(g) - R_j(g-jk)$ which follows since i reaches j and $R_i(g) = 0$), which is a contradiction.

Thus, different wheels (and combinations of wheels) and the empty network are the possible efficient networks. The remainder of the claim is straightforward, noting that if the value of a wheel with less than n players is positive, then the value of a wheel involving all players generates a higher per capita payoff.

Next, let us characterize the strictly Nash stable networks.

First, in the case where $c > n-1$ it follows directly that the only (strict) Nash network is the empty network, since a link can lead to a marginal payoff of at most $n-1-c$. Next, in a case where $1 < c < n-1$, the empty network is still a strict Nash equilibrium, as each link that a player adds will changes that player's payoff by an amount $1-c < 0$. In the case where $c < 1$ it is clear that the empty network is not Nash stable. The proof is then completed by showing that whenever $c < n-1$ any nonempty strictly Nash stable network must be a wheel involving all players, as it is clear that such a network is strictly Nash stable.

So, let $c < n-1$ and consider a nonempty strictly Nash stable network. First, note that all players have at out degree of at least one. Suppose not. There is at least one player j who strictly benefited from a link ij, since the network is nonempty. By duplicating that link, a player with no out links would also strictly benefit, which contradicts equilibrium. Next, note that each player must have at most one link coming in. Suppose to the contrary that players i and j both have links to player k. By deleting the link to k and adding a link to j (or keeping the link to j if i already has one), i's payoff can only increase as there is still a path to k (and hence to all other players reached through k) and i has not increased the number of links. Thus, i benefits weakly from such a change and so the network cannot have been a strictly Nash stable network, which is a contradiction. Hence we have a network such that every player has at least one link out and at most one link in, and hence every player has one link in and one link out. This must be a wheel including all players. ∎

The strict aspect of equilibrium is a useful device here as it really narrows down the set of stable networks dramatically. Moreover, in some experiments on this sort of model, Callendar and Plott [112] find evidence that strictness is a useful predictor of behavior. Part of the reason for the predictive power is the absence of any decay (since $\delta = 1$), which leads to lots of indifferences over which links form. More generally, when there is decay, many indifferences are naturally eliminated and a refinement to strict equilibrium does not make much difference.

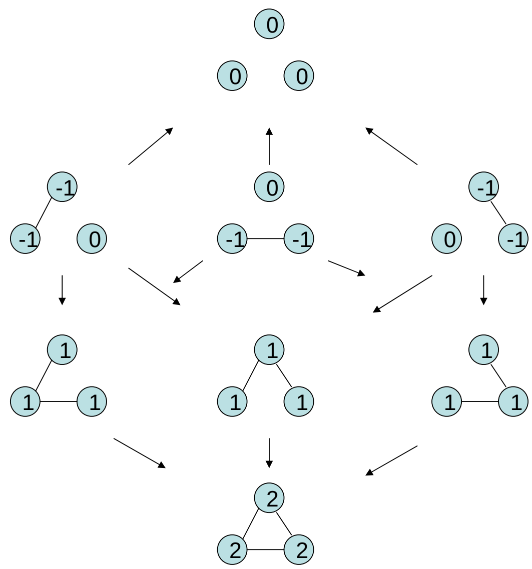

Figure 11.4. An Example with Two Pairwise (Nash) Stable Networks, where Improving Paths can get Stuck at the Empty Network.

11.4 Stochastic Strategic Models of Network Formation

Recall that in Section 6.3.2 we considered the following process for growing a network. At each point in time a link is randomly chosen, with equal weight on all links. If the link is not in the network, then the two players involved in the link have the choice to add it to the network and they add it if it makes each of them weakly better off in terms of payoffs and makes at least one of them strictly better off. If the link is already in the network, then either of the players involved in the link can choose to delete it, and it is deleted if that would increase the payoff for either player. If this process comes to rest then it will come to rest at a pairwise stable network. If there do not exist any pairwise stable networks, then it will keep cycling.

While such a process models the dynamics of network formation, it can get stuck at networks that are pairwise stable, but such that we would expect players to find their way around. To see this issue more starkly, consider a three-player society where the payoffs to different networks are as pictured in Figure 11.4.

For the payoffs pictured in Figure 11.4, the empty network and the complete network are both pairwise stable. However, a process simply following improving paths, as indicated below, can get stuck at the empty network. There are two different ways in which this process might get "unstuck". One variation would be to allow for trembles or some exogenous events that cause links to be added or deleted with some (small) probability ε. Once one link forms, then there is a good chance that another will be formed, and then the process would reach the complete network. Another variation is to consider "farsighted" players. Players might realize that if they add one link then other links would subsequently form. Thus, even though a single link would lead to negative payoffs, they might add it, anticipating that it will lead other links to form. These are both quite natural variations on the process described before, but from very different perspectives.[13] One simply introduces some randomness into the process, while the other relies on rational and forward-looking players. These also rely on different knowledge on the part of the players. In a farsighted process, players understand the incentives of other players and they forecast the subsequent evolution of the network, while the perturbed myopic process does not require knowledge on the part of the players other than whether a given link is beneficial on the margin. These are different types of arguments and thus might be more or less appropriate depending on the setting. Let us consider each in turn.

11.4.1 Random Improving Paths and Stochastic Stability*

Eexogenous randomness in the network formation process, so that links occasionally are added or deleted even though the benefits do not outweigh the costs, leads to a network formation process that can yield sharp predictions about which networks are likely to emerge. Such a variation on the improving path process was introduced by Jackson and Watts [**?**].

The process can be described starting at any network $g \in G(N)$. At each time $t \in \{1, 2, \ldots\}$ a link ij is randomly identified, with each link having an equal probability of being identified and with the randomness being independent across time.[14] Again, as in an improving path, if the link is not in the network and the players in question

[13] There are other perspectives, including that of strong stability. The arguments here are most pertinent to settings where such coordination between larger coalitions of players is not possible.

[14] The results described below extend to some more general processes where several links are identified at once, or which links are identified depend on the current network or the history of links that have been considered. However, this process is a useful one for the purposes of illustration.

would like to add the link (both weakly and at least one strictly), then the link is added, while if the link is already in the network then it is deleted if either of the two players strictly prefers to delete it. There is an added randomness to the process. With a probability of $1 - \varepsilon$ the intent of the players (to add a link, to delete a link, or to leave the network as it is) is carried out, and with probability $\varepsilon > 0$ the reverse occurs.

Thus, at each time some link is examined and with some probability the link is added, deleted, or ignored, depending on what the players would like to do under the concept of an improving path; and with some probability there is a "perturbation" and exactly the opposite occurs. Effectively, this process has small probabilities that some exogenous events happen, which might be errors on the part of players or some other interventions that break up beneficial relationships or introduce relationships that are not beneficial. There are many possible perturbations that could account for such randomness.

Given these random perturbations in the process, the process will now continue (with probability one) to have the network change indefinitely. Moreover, it will continue to visit each network over time. In fact, this process is now a finite state, aperiodic, irreducible Markov chain (recalling definitions from Section 4.5.8).[15] Thus, it has a steady-state distribution.

To get a feeling for this process, let us reconsider the example pictured in Figure 11.4.

If the process is at the empty network at some time, then it will only change to another network if there is an error. Thus, there is a ε chance that it will lead away from the empty network.

If the process is at a one-link network, then regardless of which link is identified, the players will wish to change the network, either adding a new link or deleting the new link. Thus, there is a $1-\varepsilon$ chance that it will change networks and only a ε chance that it will stay at a one link network. When it does change networks, it is twice as likely to lead to a two-link network as to lead to the empty network.

Once we get to a two-link network, then the players would choose to change the network only if the missing link is the one identified. In this case, there is a $1/3$ chance that the probability of leaving the network will be $1 - \varepsilon$ and a $2/3$ chance that the probability of leaving the network will be ε. In particular, there is a $(1-\varepsilon)/3$ chance of changing to the complete network, a $(2 - \varepsilon)/3$ chance of staying at the same network,

[15] Players behavior, and that of the system, depend only upon the current network and not on the history of how they got there.

and a $2\varepsilon/3$ chance of changing to a one-link network.

At the complete network, no player will wish to make any changes. So, the process will only change networks if an error occurs, and so changes to a two-link network with probability ε and stays put otherwise. Viewing this as a Markov chain, let the "state" of this system simply keep track of how many links the network has and consider the probability of transitioning from one "state" to another. The transition probabilities are described in the following matrix, where the ij-th entry is the probability that the network will change from a network with i links to one with j links.

$$\Pi(\varepsilon) = \begin{pmatrix} 1-\varepsilon & \varepsilon & 0 & 0 \\ \frac{1-\varepsilon}{3} & \varepsilon & \frac{2(1-\varepsilon)}{3} & 0 \\ 0 & \frac{2\varepsilon}{3} & \frac{2-\varepsilon}{3} & \frac{1-\varepsilon}{3} \\ 0 & 0 & \varepsilon & 1-\varepsilon \end{pmatrix}.$$

From this, it is easy to deduce the steady-state distribution of this process. It is a 1×4 vector μ such that $\mu\Pi = \mu$. This is the left-hand unit eigenvector, which in this case is described as follows.

$$\mu(\varepsilon) = \left(\frac{\varepsilon(1-\varepsilon)}{1+2\varepsilon}, \frac{3\varepsilon^2}{1+2\varepsilon}, \frac{3\varepsilon(1-\varepsilon)}{1+2\varepsilon}, \frac{(1-\varepsilon)^2}{1+2\varepsilon} \right).$$

Let us examine the properties of this process. As we let ε go to 0, $\mu(\varepsilon)$ tends to $(0, 0, 0, 1)$ and the time that the process spends in the complete network tends to 1 while the time that the process spends in any other network tends to 0.

To understand this process, note that for very small ε, once the process reaches the empty network, it stays there until an error occurs, and so it can stay there for a very long time as ε becomes small. In contrast, if it is either at a one or two link network, it will leave that state with very high probability. When it is at the complete network, it will stay there with very high probability. To see why the process spends almost all of its time in the complete network rather than the empty network when ε is small, note the following. If it is at the complete network, even if an error occurs it ends up at a two-link network. All two-link networks lead back to the complete network with very high probability. It actually takes an error to transition from a two-link network to a one link network. Thus, moving from the complete network to the empty network takes at least two errors to occur, which happens on the order of ε^2. In contrast, moving from the empty network to the complete network only takes one error. In particular, once a one-link network is reached, then there is a nontrivial probability of transitioning to a two link network and then to the complete network. Thus, the probability of transitioning from the empty network to the complete network is on the order of ε. As ε becomes small, the process is much more likely to transition from the empty network to the complete network than the other way around. Although the process can still stay at the empty network for many periods in a row upon reaching it, asymptotically it will spend much more time at the complete network.

Note that it is necessary to look at the limit of the $\mu(\varepsilon)$'s to find μ. If we examine the limiting improving path process directly without any errors, then that process is described by the transition matrix

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & \varepsilon & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

While $\mu = (0, 0, 0, 1)$ is a steady-state of that system, so is $(a, 0, 0, 1 - a)$ for any $a \in [0, 1]$. That is, the process without any mutations does not discriminate between the empty and complete networks. One needs the error process to really discover how stable various networks are to perturbations.

11.4.2 Stochastically Stable Networks*

With more than a few players, working directly with the perturbed improving path process $\Pi(\varepsilon)$ can be cumbersome, and so it is important to discover the set of stochastically stable networks more directly and generally. Here, we make use of a powerful theorem by Freidlin and Wentzell [239] that characterizes the steady-state distribution of Markov processes, which was adapted by Kandori, Mailath, and Rob [358] and Young [632] to understand how perturbed Markov chains behave as the probability of an error, ε, tends to 0. Jackson and Watts [340] show how such a general "stochastic stability" analysis can adapted to a network setting to derive insights about the evolution and dynamics of network formation. Let me outline some of the central tools and techniques.

Consider an improving path process with added errors as described above. Let $\mu(g, \varepsilon)$ be the steady-state probability that process is at network g when the process has error rate ε.

A network $g \in G(N)$ is *stochastically stable* if its steady-state probability is bounded below as the error rate, ε, tends to zero; that is, g is stochastically stable if $\mu(g, \varepsilon) \rightarrow_{\varepsilon} a > 0$.

When pairwise stable networks exist, any stochastically stable network must be pairwise stable. This is clear since the process will leave any non-pairwise stable network with a probability that is bounded below as ε goes to 0, while a pairwise stable network is only left if an error occurs.[16] When pairwise stable networks do not exist, the stochastically stable networks only include networks that sit on improving cycles of networks which are randomly visited over time. It could pick certain cycles and not others, as it might take many more errors to leave one cycle than another.

Thus, stochastic stability selects from among the pairwise stable networks, when they exist, and thus can provide a more refined prediction based on a sort of robustness argument. We already saw such a selection in the example above, where there were two pairwise stable networks and only one stochastically stable network. Stochastic stability identifies the most "robust" or easy to reach networks in a particular sense. The disadvantage of this approach is that the limit points of the dynamics can be difficult to identify in some applications. Nevertheless, there are many settings where it provides a meaningful refinement, as discussed below.

The characterization of stochastically stable networks follows from results characterizing the limiting properties of perturbed Markov chains, as I now outline.

Consider a Markov chain on a finite state space S with transition matrix Π. In the application to networks, the state space is the set of networks and the transition matrix is determined by the random identification of a link and then following an improving path (adding or deleting the link).

A set of mutations of Π is a set of transition matrices $\Pi(\varepsilon)$, one for each ε in a range $a > \varepsilon > 0$ for some $a > 0$, such that

(i) $\Pi(\varepsilon)$ is aperiodic and irreducible for each ε

(ii) $\Pi(\varepsilon)$ converges to Π as $\varepsilon \to 0$, and

(iii) $\Pi(\varepsilon)_{ss'} > 0$ implies that there exists $r \geq 0$ such that $0 < \lim_{\varepsilon \to 0} \frac{\Pi(\varepsilon)_{ss'}}{\varepsilon^r} < \infty$.

Part (i) ensures that the mutations add noise in a way such that any state can eventually be reached from any other state, and in an aperiodic way. Part (ii) ensures that for small ε the mutated matrix is close to the original matrix. The number r in (iii) is the *resistance* of the transition from state s to s', and roughly can be thought of as quantifying the level of error or mutation needed to get from state s to s'.

[16] It is possible for a (unique) stochastically stable network to be pairwise stable but not pairwise Nash stable. See exercise 11.14.

In the application to networks, the perturbations are found by including an ε error in the addition or deletion of the randomly identified link. Since this means that any network can lead to any adjacent network, as well as back to itself, the process satisfies (i), and (ii) is also clearly satisfied. To verify that (iii) is satisfied, first note that $\Pi(\varepsilon)_{gg'} > 0$ implies that g and g' are adjacent. There is a resistance of 0 in the case where g' defeats g (lies on an improving path with only one link), and so $\Pi_{gg'} > 0$ and setting $r = 0$ satisfies (iii). Otherwise, the transition only occurs if there is an error, and so $\Pi(\varepsilon)_{gg'}$ is simply ε divided by the number of links which could be identified by the process. Thus, $\Pi(\varepsilon)_{gg'}$ goes to 0 at the rate of ε and so $r = 1$ satisfies (iii).

Let me now state a theorem from Young [632] that characterizes the states that have positive probability under the limit of the steady state distributions of the mutations of the process.

Given any state s, an s-tree is a directed graph with a vertex for each state and a unique directed path leading from each state $s' \neq s$ to s. The *resistance* of s is the minimum across all s-trees of the summed resistance over directed edges in that tree.

THEOREM **11.4.1** *(Young [632]) Let Π be the transition matrix associated with a Markov chain on a finite state space with an associated set of mutations $\{\Pi(\varepsilon)\}$ and with corresponding (unique) stationary distributions $\{\mu(\varepsilon)\}$. Then the steady state distributions $\mu(\varepsilon)$ converge to a stationary distribution μ of Π. Moreover, a state s has positive probability under μ (and is thus stochastically stable) if and only if s has minimum resistance.*

From Theorem 11.4.1 it is easy to see that if a state s has positive probability under μ and there is an adjacent state s' that can be reached from s with no resistance, then the state s' will also have minimum resistance and thus will also have positive probability under μ. To see this, simply start with an s-tree with minimum resistance, and construct an s'-tree with at least as low a resistance as follows. Cut the directed link out from s' and form a directed link from s to s'. The new link has resistance 0, and so this tree has at least as low a resistance, and since s has minimum resistance, then s' must also.

By this reasoning, we can consider whole sets of states that can reach each other. The stochastically stable networks will either be pairwise stable networks or lie on cycles, where they can reach (and be reached by) other stochastically stable networks via an improving path. In terms of the results on stochastic stability of Markov chains, this is stated as follows.

The *recurrent communication classes* of Π, denoted $S_1, \ldots, S_J$, are disjoint subsets of states (not necessarily including all states) such that

- from each state there exists at least one path of zero resistance leading to some state in one of the recurrent communication classes,
- any state in a recurrent communication class can reach any other state in the same recurrent communication class by a path of zero resistance, and
- for any recurrent communication class S_j and states $s \in S_j$ and $s' \notin S_j$ such that $\Pi(\varepsilon)_{ss'} > 0$ for some ε, the resistance of the transition from s to s' is positive.

In the application to networks, the recurrent communication classes are either singletons consisting of a pairwise stable network, or a closed improving cycle, where closure refers to the third item above and means that there is no improving path leading out from a network in the cycle to a network that is not part of the cycle.

For two recurrent communication classes S_i and S_j, since $\Pi(\varepsilon)$ is irreducible for each ε, it follows that there is a sequence of states $s_1, \ldots, s_k$ with $s_1 \in S_i$ and $s_k \in S_j$ such that the resistance of transition from each consecutive state to the next in the sequence (e.g., from s_h to s_{h+1}) is defined by (iii) and finite. Let this be denoted by $r(s_h, s_{h+1})$. The resistance of transition from recurrent communication class S_i to recurrent communication class S_j is the minimum over all such sequences of $\sum_h r(s_h, s_{h+1})$, and is denoted $r(S_i, S_j)$.

Given a recurrent communication class S_i, an S_i-tree is a directed graph with a vertex for each communication class and a unique directed path leading from each recurrent communication class $S_j \neq S_i$ to S_i. The *stochastic potential* of a recurrent communication class S_i is then defined by finding an S_i-tree that minimizes the summed resistance over directed edges, and setting the stochastic potential equal to that summed resistance.

With these definitions in hand, we can relate resistance to stochastic stability.

THEOREM **11.4.2** *(Young [632]) Let Π be the transition matrix associated with a Markov chain on a finite state space with an associated set of mutations $\{\Pi(\varepsilon)\}$ and with corresponding (unique) stationary distributions $\{\mu(\varepsilon)\}$. Then the steady state distributions $\mu(\varepsilon)$ converge to a stationary distribution μ of Π, and a state s has positive probability under μ (and thus is stochastically stable) if and only if s is in a recurrent communication class of Π which achieves the minimal stochastic potential. This is equivalent to s having minimum resistance.*

This is similar to the previous theorem, except that it says that to identify the stochastically stable states, we need only work with the stochastic potential of the recurrent communication classes rather than keeping track of the resistance state-by-state. Thus, in identifying the stochastically stable networks, one needs only focus on pairwise stable networks and closed improving cycles of networks. This can substantially simplify the analysis. To get a feeling for this, and to see an example of how stochastic stability can refine the set of pairwise (Nash) stable networks, consider the co-author model from Section 6.4.

Recall that the payoff to player i in network g is

$$u_i(g) = \sum_{j:ij\in g} \left(\frac{1}{d_i(g)} + \frac{1}{d_j(g)} + \frac{1}{d_i(g)d_j(g)} \right)$$

for $d_i(g) > 0$, and $u_i(g) = 1$ if $d_i(g) = 0$.

When $n = 7$ there are 22 different pairwise (Nash) stable networks: the complete network and each network such that there are five completely connected players and a separate dyad. As Jackson and Watts [340] point out, only the complete network is stochastically stable. This is seen as follows. Each pairwise stable network other than the complete network has a resistance of 1 to the complete network. Indeed, it is easily checked that deleting the link in the dyad leads to a network that lies on an improving path to the complete network. Thus, when g is the complete network, we can construct a g-tree that has stochastic potential of 21, by pointing each other network directly to the complete network. In contrast, it takes several errors to get from the complete network to a network that lies on an improving path to some other pairwise stable network. If just only link in the complete network is severed, then the only improving path leads back to the complete network. Thus the resistance from the complete network to any other pairwise (Nash) stable network is more than 1. Constructing a g-tree for any other network leads to a stochastic potential of more than 21.[17]

Interestingly, the complete network is Pareto dominated by any of the other pairwise (Nash) stable networks, yet is the unique stochastically stable network, which offers a further illustration of the tension between efficiency and stability.

In some cases, there are weakenings and generalizations of stochastic stability that can be easier to work with, as shown by Tercieux and Vannetelbosch [589].

[17] There are no improving cycles, so the only recurrent communication classes are the ones that each consist of a different pairwise stable network.

11.4.3 Stochastic Stability Coupled with Behavior

It is possible to extend the apparatus of stochastic stability to include other considerations. For instance, Jackson and Watts [341] examine a graphical game in a setting where strategies co-evolve with the network. That is, players are choose between two actions in a coordination game where their payoffs depend on the play of their neighbors. At the same time as choosing their actions, they can also decide on adding or deleting links. With costs to links, they prefer to be linked to other players with whom they coordinate their play. Their payoffs are thus affected both by whom they link to and what strategies they play. Analyzing these things together provides for interplay between the network structure and play of the game. This can lead to stochastically stable outcomes that differ from the behavior that predicted when simply fixing the network structure (players who play more efficient actions can be more attractive), and can also lead to different predictions in terms of network structure than having fixed play. It can also lead to differences in the speed of convergence of play, as the evolving network structure can lead to the diffusion of certain types of play in a more rapid fashion than with a static network. The analysis ends up being sensitive to the details of the setting, but shows the importance of analyzing the co-evolution of behavior and network structure.[18]

11.5 Farsighted Network Formation

A very different perspective on network formation from that of random errors and myopic behavior is one where players are forward-looking and make no errors. This applies in very different circumstances, where players have a good idea of the setting and the relevant incentives that various players have to form and sever links. This has been explored from various vantage points, some of which involve explicit variations on network formation games (see Dutta, Ghosal and Ray [**?**]) as well as other approaches that capture farsightedness directly through variations on improving paths (e.g., Page, Wooders, and Kamat [494]). Let us examine this latter approach in more detail.

In the definition of improving path changes from one network to the next are improving for the players involved, but without anticipating the subsequent changes that

[18] That analysis has been extended to directed networks by Goyal and Vega-Redondo [288], Hojman and Szeidl [314], to anti-coordination games by Bramoullé et al [94], and to settings with geography by Droste, Gilles and Johnson [193]. See also Skyrms and Pemantle [563] for a reinforcement based evolutionary analysis of games played on networks.

will occur along the path. In contrast, the idea of a farsighted improving path captures the notion that the players anticipate the further changes along the path and compare the ending network and the current one.

Let us say that a network g' is *improving* for S relative to g if it is weakly preferred by all players in S to g, with strict preference holding for at least one player in S.

Next, consider a sequence of networks $g_1, \ldots, g_K$, and a corresponding sequence $S_1, \ldots, S_{K-1}$, such that g_{k+1} is obtainable from g_k via deviations by S_k. Such a sequence is a *farsighted improving path* if, for each k, the ending network g_K improving for S_k relative to g_k.

In the case where consecutive networks in the sequence are required to be adjacent, then this is a farsighted analogue of an improving path, while more generally it allows for large coalitional deviations.

Now we can say that a network g is *farsightedly pairwise stable* if there is no farsighted improving path from g to some other network g', such that each pair of consecutive networks along the sequence are adjacent.

Similarly, we can say that a network g is *farsightedly strongly stable* if there is no farsighted improving path from g to some other network g'.

These are both demanding requirements. They refine the set of pairwise stable and strongly stable networks, respectively. That is, any network that is farsightedly pairwise stable is necessarily pairwise stable and a network that is farsightedly strongly stable is necessarily strongly stable. These definitions require that networks be immune to both immediate deviations and farsighted sequences of deviations of arbitrary length.

In the example in Figure 11.1.4 the complete network is farsightedly strongly stable and thus farsightedly pairwise stable. But more generally, such definitions are difficult to satisfy. An aspect of the definitions that is too strong is that it does not require that a farsighted improving path end at a network that is stable itself. That is, if players are really farsighted, then they would not follow some farsighted improving path unless they really anticipated that the endpoint is justified as the stopping point of the process. Such a definition, however, becomes circular as it requires the endpoint to be farsightedly stable. This sort of existence problem is nicely handled through a concept developed by [148] and adapted to the network setting by Page, Wooders and Kamat [494].

The idea is based on a self-consistent set-based definition. In particular, a set of networks is said to be consistent if all deviations away from the network are expected to lead (in a farsighted manner) back to some network in the set, and one that is not improving for the original deviating coalition.

More formally, a set $A \subset G(N)$ is *consistent* if for each $g \in A$, and g' obtainable from g via deviations by some $S \subset N$, either $g' \in A$ and g' is not improving for S, or there exists a farsighted improving path from g' to some $g'' \in A$ such that g'' is not improving for S.[19]

It is easily checked that a union of consistent sets is consistent. Thus, Chwe suggests examining the largest consistent set, which he shows is always nonempty.

The idea of a consistent set A is that any network in the set is justified as being "stable" as follows. Consider $g \in A$, and g' obtainable from g via deviations by some $S \subset N$. If g' is improving for S, then there must be some expectation that discourages S from deviating to g'. In particular, there must exist a farsighted improving path moving away from g' that S anticipates will be followed and will lead to some $g'' \in A$ such that g'' is not improving for S. The presence of g'' in A implies that this is also a justifiable resting point, and so S can expect the process to stay there. This anticipation then deters the original deviation. If every deviation away from g can be deterred in this way, then it is a viable resting point, and the set is consistent in that its various elements are used in justifying each other as resting points.

To see how the largest consistent set can make different predictions from other solutions, let us re-examine a variation on the example from Figure 6.2.2, which involved a bargaining network. Allowing for coalitional deviations, we find that no network is strongly stable. The difficulty is that the efficient network is defeated by a network where a link is added. This in turn is defeated by the addition of another link, but then all players would be better off moving back to the efficient network.

The largest consistent set relative to the payoffs in Figure 11.5 includes the efficient network and two other networks. Without going through the full set of calculations needed to verify that this is the largest consistent set, let us examine a few of the key deviations to check that the set is consistent. The reasoning is a bit subtle, but holds together as follows. Suppose that we start at the efficient network. In order to check that this is part of a consistent set, we need to check that if some group deviates to change the network, that they could then anticipate some farsighted improving path that leads away from the deviation and to another network in the consistent set that would not be improving. A possible problematic deviation is players 1 and 2 threatening

[19] The definitions here differ slightly from Chwe's definitions, as here the definition of improving only requires weak improvement for some players and strict for at least one, while Chwe's definition requires improvement for all players.

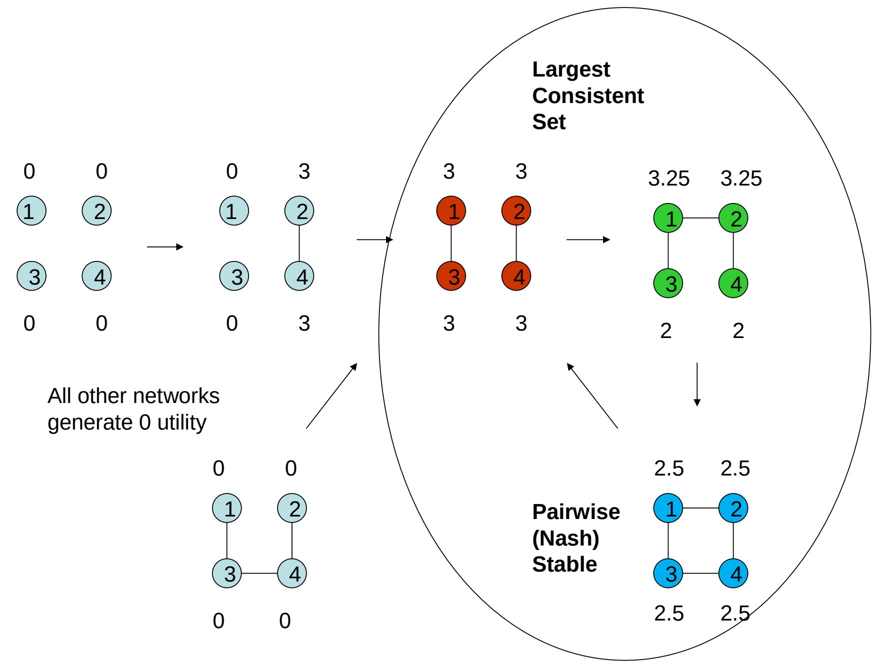

Figure 11.5. No Network is Strongly Stable, Pairwise (Nash) Stable Networks are Inefficient, while the Largest Consistent Set is Efficient

to add the top link, which would lead them to a higher payoff of 3.25 each. To check the condition for this to be a consistent set, we need to find a farsighted improving path away from this network leading to another network in the consistent set that is not improving for 1 and 2 relative to their original payoffs of 3. Indeed, players 3 and 4 could add a link, which would be improving for players 3 and 4 and lead to another network in the consistent set which offers a lower payoff to players 1 and 2 (2.5 each). Thus, this possible anticipated continuation reasoning deters 1 and 2 from adding a link in the first place, and thus is one way to justify the original efficient network being in the set. Next, let us consider the network with the three links where players 1 and 2 get a payoff of 3.25. Suppose that players 3 and 4 deviate to add a link resulting in a network where all players get 2.5. In order for the network with three links to be in the consistent set, we need to check that there is a farsighted improving path that leads to a consistent network with a payoff for players 3 and 4 that deters them from making this initial change. Here, there is a farsighted improving path which goes to the three link network at the bottom left of the figure, which then leads to payoffs of 0 for all players and then continuing on a farsighted improving path back to the original three-link network at the top right where players 3 and 4 have a payoff of 2 each. Thus, again, we find a subsequent farsighted improving path which can deter the original deviation. Finally, we need to check that the network at the bottom left is in the consistent set. An obvious deviation is to the efficient network. From that network there is a farsighted improving path which leads to the three link network at the top right, which is in the consistent set an leads to a lower payoff for players 3 and 4, thus deterring their original deviation.

This example shows the reasoning behind the consistent set, and why it is a set-based notion. Various networks in the set are sustained because of anticipated deviations leading to other consistent networks. The consistency refers to the fact that various networks in the set are used to sustain others, and the reasoning has to be fully consistent. Considering the largest such set implies that nothing outside of the set could be sustained via such reasoning.

PROPOSITION **11.5.1** *[Chwe [148]] Consider any N and profile of preferences. There exists a unique largest consistent set (so that every consistent set is a subset of it and it is consistent), and this set is nonempty.*

The ideas behind the proof are relatively straightforward. First, it is easily checked that the union of consistent sets is consistent, as there are fewer potential deviations to worry about that move outside of the set and more things that can deter those deviations. Thus, given the finite setting there will exist a largest consistent set simply by taking the union of all consistent sets. Showing that the largest consistent set is nonempty is the harder part of the proof. It relies on showing that a sequence of networks such that there is a farsighted improving path from any network to any subsequent network in the sequence, must be finite in length. This follows in our setting since an infinite path must repeat some networks, which cannot be improving relative to themselves.

The basic ideas behind the largest consistent set are not particular to the specific notions of "improvement" or "deviation" that we have been working with. Indeed, Chwe's original definitions do not mention networks at all. It is a quite versatile idea, that can be adapted to other settings, including networks with multiple relations between pairs or groups of players, directions, and other sorts of considerations. Such extensions are explored by Page, Wooders and Kamat [494] and Page and Wooders [493].[20]

[20] See Dutta, Ghosal, and Ray [**?**] and Herings, Mauleon, and Vannetelbosch [310] for other approaches to defining farsighted stability.

11.6 Transfers and Network Formation

As discussed in Chapter 6, the tension between efficiency and stability in network formation stems from externalities. Players do not take into account the indirect impact that their link-formation decisions have on others. At least in some contexts, this can be rectified if transfers are allowed so that players can pay each other for forming or severing links, or if players can bargain over how the value of relationships is allocated when forming links. Such transfers are applicable in many settings where links represent economic or business relationships, and transactions are occurring as part of the relationship. If the relationship is more advantageous for one party, then that can be contracted upon at the time of the relationship formation. One not only sees such agreements in purely economic relationships, but also in social ones. For instance, we even see transactions in some marriages in the form of dowries. Even with no wealth or goods exchanged, there can be an allocation of tasks or favors that is either implicitly or explicitly agreed to in order to maintain a relationship.

11.6.1 Forming Network Relationships and Bargaining

The fact that transfers can affect the network that forms means that it can be important to model the use of transfers as part of the network formation process. Such a process was first investigated by Currarini and Morelli [172] by examining a sequential network formation game that is described as follows.[21] The game is defined for cases where the utilities are component decomposable (recalling the definition from Section 6.6.2), so that the payoffs to a given component depend only on that component's configuration and not on the remainder of the network. Players are ordered exogenously, labeled in the order of their moves so that player 1 moves first, then player 2 and so forth. At his or her turn, player i announces the set of players with whom he or she is willing to be linked ($S_i \subset N \setminus \{i\}$), and a payoff demand $v_i \in I\!R$ – which is interpreted as the net payoff that the player wishes to get. The outcome of the game is then as follows. First one examines the network that could potentially form by including the links such that both players involved in the link announced each other. That is, the potential network that might form is $g(S)$ where $ij \in g(S)$ if and only if $j \in S_i$ and $i \in S_j$. This is not the final network, as one has to check to see whether the payoff demands that the agents made can all be satisfied. The network that is eventually formed is

[21] See Mutuswami and Winter [471] for elaboration on a similar model.

determined by checking which components of $g(S)$ are actually feasible in terms of the demands submitted. That is, if g' is a component of g, then g' is actually formed if $\sum_{i \in N(g')} v_i \leq \sum_i u_i(g)$, and otherwise *none* of the links in g' are formed.

The presumption is that if a component forms, then players in that component get payoffs of the v_i's that they demanded. Exactly how transfers are made, or what needs to be done to convert the initial utilities into this final allocation of payoffs is not specified and might involve some complicated transfers that travel some distance in the network.

Currarini and Morelli then show the following result, for a class of settings where payoffs satisfy a condition that they name size monotonicity.

Payoffs $u = (u_1, \ldots, u_n)$ satisfies *size monotonicity* if $\sum_i u_i(g) > \sum_i u_i(g - kj)$ for every g and every bridge $kj \in g$ (such that $g - kj$ has more components than g).

Proposition **11.6.1** *[Currarini and Morelli [172]] If payoffs satisfy size monotonicity, then every (subgame perfect) equilibrium of the Currarini and Morelli bargaining and network formation game leads to an efficient network.*

The intuition behind the result is as follows. It is helpful to consider a simple dyad to see the idea. Suppose that if the dyad is formed, player 1 will get a utility of 2 while player 2 will get a utility of -1. This is a beneficial relationship for player 1, but a costly one for player 2. There is a total utility of 1 to be had, and so it is efficient for the link to form; however, without any reallocation of value, player 2 would not be willing to form the relationship. Given that player 1 moves first, the equilibrium here is easy to see: player one states $S_1 = \{2\}$ and sets $v_1 = 1$; and then 2 responds with $S_2 = \{1\}$ and $v_2 = 0$.[22] This is the only subgame perfect equilibrium outcome. Effectively, player 1 pays player 2 a unit of utility in order to form the link. If the player roles were reversed and player 2 moved first, then player 2 would demand $v_2 = 1$ and leave player 1 with no value. The important aspect is that the players are now able to ensure that each gets enough value to form the efficient network. When there are more than two players the analysis becomes a bit more complicated, as the first movers have to forecast how much utility they can extract from the network, how much will be left, and what options the later players will have in terms of forming links; but the basic idea is that they will correctly forecast exactly how much they can extract

[22] In terms of the full specification of how 2 behaves, he or she with $S_2 = \{1\}$ and $v_2 = 1 - v_1$ whenever $v_1 \leq 1$ and otherwise says $S_2 = \emptyset$.

in equilibrium by maximizing with respect to the foreseeable equilibrium strategies of the subsequent players.[23]

This game has features that are important in the result, but are also a bit artificial. In particular, the fact that players move in a fixed, forecastable order and exactly once is important in backward induction and getting efficiency. The idea behind efficiency comes from the fact that suggesting links leading to an efficient network maximizes the payoff that each player is able to extract at his or her turn, given the previous demands of the other players and given how the remaining players will be forced to react. It is clear that it can provide some players with big advantages or disadvantages in terms of the network and payoffs that eventually emerge, and that this hinges on the fixed sequential ordering.

11.6.2 A Network Formation Game with Transfers

As an alternative to the above game, Bloch and Jackson [74] suggest a simultaneous move game where players can directly offer transfers to each other in order to form links. The motivation behind this is not that a simultaneous move game is how networks are formed. To the contrary, the idea is that if one models a richer setting where players can go back and forth and bargain, it will have to be that in the end they come to a point where none of them would want to change their proposed transfers or links. Thus the resting point of a more open and dynamic process will have to reach an equilibrium point where no player would gain from changing his or her action given the actions of the others. This is modeled as follows.

Let each player i announce a vector $t_i = (t_{i1}, \ldots, t_{in}) \in I\!R^n$ such that $t_{ii} = 0$ and with the interpretation that t_{ij} is the amount that i is willing to transfer to j to form a link, or, if this is a negative number, the amount that i requests in order to form the link.

A link ij is formed if and only if $t_{ij} + t_{ji} \geq 0$. In equilibrium this will hold with equality.

Let $g(t) = \{ij | t_{ij} + t_{ji} \geq 0\}$ denote the network that forms. Player i's payoff is then

$$\pi_i(t) = u_i(g(t)) - \sum_{j:ij \in g(t)} t_{ij}.$$

To solve this game, one can use Nash equilibrium; however, one then runs into the

[23] To see how this works in some richer examples, see Jackson [?] and the proof is provided in Currarini and Morelli [172].

same difficulties that are faced with Nash equilibrium in the basic network formation game. Tor instance, if a player expects all other players to demand enormous amounts to form a link (that is, to state very negative t_{ij}'s), then it is in his or her interest to do the same. Thus, the empty network is always an equilibrium, regardless of how attractive other networks are. Thus, one runs into the problem that nobody forms any links since they all correctly forecast that no one else will form a link.

To deal with this issue, Bloch and Jackson [74] adapt pairwise Nash equilibrium to this setting.

A *pairwise transfer equilibrium* is a profile of vectors of proposed transfers $(t_1, \dots, t_n)$ such that t is a Nash equilibrium, that is

$$\pi_i(t) \geq \pi_i(\widehat{t}_i, t_{-i})$$

for all i and $\widehat{t}_i \in I\!R^n$ (such that $\widehat{t}_{ii} = 0$); and, for any $ij \notin g(t)$

$$u_i(g(t) + ij) + u_j(g(t) + ij) \leq u_i(g(t)) + u_j(g(t)).$$

Note that this latter requirement is equivalent to requiring that there does not exist any $\widehat{t}_{ij}$ and $\widehat{t}_{ji}$ such that $\widehat{t}_{ij} + \widehat{t}_{ji} \geq 0$ and such that both i and j would be weakly better off with these new announcements and the addition of the link between them with at least one strictly better off. Thus, this is a parallel definition to pairwise Nash equilibrium, but allowing for transfers between the players.[24]

PROPOSITION **11.6.2** *[Bloch and Jackson [74]] In the distance-based utility model, for each efficient network, there exists a pairwise transfer equilibrium which results in that network and balanced transfers (so that the total transfers sum to 0).*

Bloch and Jackson do not prove this directly, but instead as part of a more general characterization which provides the necessary and sufficient conditions for a network to be supportable as an equilibrium of this transfer game. The direct proof in the case of the distance-based utility model is quite intuitive as it deals mainly with the star, and works by showing that the peripheral players can offer sufficient transfers to the center player to sustain the star as an equilibrium whenever it is efficient. That proof is left as Exercise 11.15.

The pairwise transfer equilibria in the transfer game above do not necessarily always include efficient networks. There are various reasons for this. One is that there

[24] For a notion of equilibrium incorporating transfers for the case of directed network formation see Johari, Mannor and Tsitsiklis **??**.

can be indirect externalities, so that the efficient network involves relationships that affect players who are not directly involved. A player might like to pay a neighbor to undertake more or fewer other relationships, but cannot as the types of transfers described above only affect the given link in question. For example, a player might benefit from having friends with more contacts, or else from having friends who are less distracted. Also, it could be that players would like to subsidize links that are far away in the network, so that they need to make transfers to players who are not even their neighbors. Furthermore, it might be that players are hurt by (distant) links that others wish to form. Here one would like to pay other players not to form relationships. For example, this is the case in some research and development settings, where a firm would like to be able to pay other firms not to collaborate with each other. Bloch and Jackson [74] also consider two other variations of such transfer games, one where players can offer to subsidize links that they are not involved with, and also can make transfers contingent on the network. They provide characterizations as to which types of externalities can be overcome by which sorts of transfers.

11.7 Weighted Network Formation

Most of the strategic network formation literature has looked at discrete linking decisions. That is, relationships are modeled as either being present or not, or either weak or strong, but without much richer choices. In many contexts decisions are much richer: for instance, we decide how much time to spend with different friends and how much effort to devote to various collaborations. Allowing for richer choices leads to new insights into how relationships form and what incentives players have to maintain efficient relationships. This is illustrated in the following model due to Rogers [537].

There is a finite society of $n \geq 3$ players who form a weighted and directed network. Each player has a budget of time that he or she can spend with other players. Let the budget for player i be denoted $B_i > 0$.

A feasible strategy for player i is a vector $(g_{i1}, \ldots, g_{in})$ such that $\sum_{j \neq i} g_{ij} = B_i$, $g_{ij} \geq 0$ for all $j \neq i$, and $g_{ii} = 0$. The interpretation is that g_{ij} is the amount of time that i spends with j. This is a directed network and these need not be reciprocal.

Each player i has some natural intrinsic base utility $v_i > 0$ that would be his or her payoff in the absence of any network interactions. In addition to that natural utility, the player benefits from other players' payoffs in an amount that depends on the time that is spent with other players times their payoffs. So spending time with a "happier" player leads to greater utility, all else held equal. However, there is a diminishing return to the time spent with any given other player.

In particular, in Rogers' [537] model, the payoff to player i is

$$u_i = v_i + \sum_{j \neq i} f(g_{ij}) u_j \tag{11.2}$$

where f is a nonnegative and continuously differentiable function such that $f(0) = 0$ and $\lim_{x \to 0} f'(x) = \infty$.

Here payoffs are self-referential, as the payoff to a given player depends on the payoffs to others, which in turn depend on the given player's payoff. This relates back to the sorts of eigenvector-based centrality measures we have discussed earlier, which had a similar self-referential formulation. So, this can also be thought of as an network formation problem in the face of endogenous centrality measures. The last condition on the derivative of f ensures that $g_{ij} > 0$ for all i and $j \neq i$, which makes the model easier to work with.

Letting $f(g)$ denote the $n \times n$ matrix with ij-th entry $f(g_{ij})$, we can write

$$u(g) = v + f(g)u(g),$$

where $u(g)$ and v are $n \times 1$ column vectors. This has the solution

$$u(g) = (I - f(g))^{-1} v = A(g)v,$$

whenever $A(g) = (I - f(g))^{-1}$ is well-defined.

In this setting a natural notion of equilibrium is simply a Nash equilibrium, where each player i is choosing $g_i = (g_{i1}, \dots, g_{in})$ to maximize $u_i(g)$.

The following is a strengthening of results by Rogers [537].

PROPOSITION **11.7.1** *[Rogers [537]] Suppose that* $\max_i f(B_i)$ *is small enough so that* $A(g) = (I - f(g))^{-1}$ *is well-defined, continuous, and nonnegative for all feasible* g *(and described by* $A(g) = \sum_p f(g)^p$*).*

- *All Nash equilibrium are interior (*$B_i > g_{ij} > 0$ *for all* i *and* $j \neq i$*).*
- *Any best response for a player* i *to a feasible and interior* g_{-i} *(and thus a Nash equilibrium strategy) is such that for each* j *and* h

$$f'(g_{ij})u_j(g) = f'(g_{ih})u_h(g). \tag{11.3}$$

- *If for each i and feasible and interior g_{-i} there is a unique g_i that satisfies (11.3), then g is a Nash equilibrium if and only if it is feasible, interior, and satisfies (11.3) for each i, j and h.*

- *If for each i and feasible and interior g_{-i} there is a unique g_i that satisfies (11.3), then the network strategy g_i that maximizes $u_i(g_i, g_{-i})$ given a feasible and interior g_{-i} also maximizes $u_k(g_i, g_{-i})$ given g_{-i} for each k, and so g_i maximizes the total sum of utilities $\sum_j u_j(g_i, g_{-i})$ given g_{-i}.*

Proof of Proposition 11.7.1: First, given that the limit, as x goes to 0, of the derivative of $f(x)$ is infinite, and that $u_j(g) \geq v_j > 0$ for each j, by (11.2) it follows that every Nash equilibrium (and every maximizer of the total sum of utilities) is such that $g_{ij} > 0$ for all i and $j \neq i$.

Consider any player i. Given that A is well-defined, given interior strategies of other players, it follows that regardless of i's strategy, all entries of A are strictly positive.[25]

In order to maximize u_i it must be that $\frac{\partial u_i}{\partial g_{ij}} = \frac{\partial u_i}{\partial g_{ih}}$ for each j and h other than i. For any k we can write

$$u_k = \sum_{\ell} A_{k\ell}(g) v_\ell.$$

Therefore

$$\frac{\partial u_k}{\partial g_{ij}} = \sum_{\ell} \frac{\partial A_{kl}(g)}{\partial g_{ij}} v_\ell.$$

To develop an expression for $\frac{\partial A_{kl}(g)}{\partial g_{ij}}$, we follow Rogers [537], who shows that differentiating $AA^{-1} = I$ leads to $\frac{\partial A(g)}{\partial g_{ij}} A(g)^{-1} = -A(g) \frac{\partial A(g)^{-1}}{\partial g_{ij}}$, and so $\frac{\partial A(g)}{\partial g_{\ell k}} = -A(g) \frac{\partial f(g)}{\partial g_{ij}} A(g)$. Therefore,

$$\frac{\partial A_{kl}(g)}{\partial g_{ij}} = f'(g_{ij}) A_{ki}(g) A_{j\ell}(g). \tag{11.4}$$

Substituting from (11.4), it follows that

$$\frac{\partial u_k}{\partial g_{ij}} = \sum_{\ell} f'(g_{ij}) A_{ki}(g) A_{j\ell}(g) v_\ell = f'(g_{ij}) A_{ki}(g) \sum_{\ell} A_{j\ell}(g) v_\ell = f'(g_{ij}) A_{ki}(g) u_j(g). \tag{11.5}$$

[25] We can write $A(g) = \sum_p f(g)^p$. Given that $g_{kj} > 0$ for all $k \neq i$ and $j \neq k$, it follows that for any i and j that the ij-th entry of $f(g)^p$ is positive for large enough p. To see this, recall that $f(g)^p$ will be positive if there is a directed walk of length p from i to j. Given that there are at least two other players, there is a path of some length from each other player to every player. Regardless of whom i connects to, i will also reach all other players.

Setting $k = i$ implies that $\frac{\partial u_i}{\partial g_{ij}} = f'(g_{ij})A_{ii}(g)u_j(g)$. This, and the facts that $A_{ii} > 0$ and $u_j(g) \geq v_j > 0$, implies that for every $j \neq i \neq h$:

$$f'(g_{ij})u_j(g) = f'(g_{ih})u_h(g).$$

Given that $A_{ki}(g) > 0$, (11.5) implies that the same condition characterizes the maximization of $u_k(g)$. The claims in the proposition follow directly. ∎

An important implication of the Proposition is that (when best responses are unique) any network that maximizes the total sum of utilities will be a Nash equilibrium network. The choice to maximize a given player's utility is the same choice as a society would make to maximize overall welfare.

What is special about this setting that leads to the congruence between efficiency and stability here, in contrast with the more general conflict between stability and efficiency that we have seen? There are several things which are important, and understanding them helps us to understand this conflict more generally. First, the problem faced by any given player in the Rogers' model is to allocate a given budget of "time" or "effort" on different relationships. Thus, the problem is solely one of allocating the budget across different relationships, rather than deciding on how much or many relationships to have in total. That is, generally we can think of a network formation problem faced by a player as having two main components: the total quantity of effort or relationships to maintain, and how to distribute that across the different relationships. The first part of the problem is missing here and is generally the problematic one. Usually the inefficiency of network formation stems from the fact that a given player does not properly account (with respect to social welfare) for the fact that his or her relationships also generate additional costs or benefits to others beyond his or her own private benefit. So the player either under or over invests in the total amount of relationships relative to what is socially valuable. The decision faced by players in the Rogers' model is solely allocative: Given the fixed budget of weight, how should a player spread it around? Here the player wants to spread it in a way that maximizes his or her payoff, and that would also be the same way that would maximize the indirect payoffs, since in this model indirect payoffs come through a given player's utility. That is, a player gets utility from his or her neighbors' utility, and so whatever makes them happy also makes him or her happy. This points to the second thing which is special about the model. A player gets indirect utility precisely through increases in neighbors' utility. To see how this is special, consider a network that has player 1 connected to both 2 and 3, who are each connected to player 4. In the Rogers model, the benefits that player 1 obtains from the indirect connection to player 4 come from both the utility that player 2 gets from being connected to 4 (independently of whether 3 is connected) and the value that player 3 gets from being connected to player 4 (independently of whether 2 is connected). In many contexts, it might be the that the marginal benefit to player 1 of having a second indirect path to player 4 is lower than the marginal benefit of having the first path. Finally, there is also symmetry in that all players get the same direct or indirect utility from any given connection, and they do not have any heterogeneity in which connections they would prefer.

While this model has special features, and they are responsible for a congruence between equilibrium networks and efficient networks, the model still provides a nice benchmark in terms of understanding the tradeoffs that players face in deciding how to allocate their time or effort across different relationships. There are many variations on the model that one might consider, including some by Rogers [537], a model by Brueckner [102] where effort translates into the probability that a link forms, and a model by Bloch and Dutta [71] where players do not face a budget constraint.

11.8 Agent-Based Modeling

When modeling network formation (or behavior on networks, or some combination of the two), a difficulty faced is that quite complex networks and/or patterns of behavior can emerge from fairly simple specifications, especially when even minimal sorts of heterogeneity (geography, age, costs, preference types, etc.) are introduced. Although there are many insights that one can derive analytically, there are some things that cannot be seen so directly. In many cases, it is more expedient to examine the behavior of large computer-simulated societies.

Such analyses have become more extensively used in the literature and are often referred to as *agent-based modeling*. These techniques can be very useful for a variety of purposes. As mentioned above, they can be used to analyze systems where equilibrium or dynamics cannot be determined analytically. They can also be used as tools to illustrate systems, or for exploratory analyses that help form hypotheses and conjectures. Such techniques are also useful in empirical analyses, for generating distributions of behaviors that would emerge under a model, which can then be compared to or fitted to observed data.

As with any form of analysis, there are important considerations in terms of how sensitive or robust the conclusions are to the specification. In agent-based modeling there are also issues of how many simulations to run, how long to run them for, how large a society to consider and so forth. As there are already a number of good sources on this subject, I will not discuss that here.[26]

11.9 Exercises

EXERCISE **11.1** *Nash Stability and Pairwise Stability.*

Provide an example of a society of individuals and utility functions such that the set of Nash stable networks is a strict subset of the set of pairwise stable networks. Provide an example where the reverse is true.

EXERCISE **11.2** *Pairwise Nash Stability*

Provide an example of a society of individuals and utility functions such that the set of pairwise Nash stable networks is a strict subset of the set of pairwise stable networks and also a strict subset of the set of Nash stable networks.

EXERCISE **11.3** *Strict Nash Equilibria in the Link-Announcement Game and Nonexistence*

Consider a potential dyad. Suppose that the payoff to having the link is negative for each player, while the payoff to not having the link is zero for each. Show that there is no strict Nash equilibrium of the link announcement game. Recall that a strict Nash equilibrium is a pure strategy Nash equilibrium where the actions played are the unique best responses to each other.

EXERCISE **11.4** *Strongly Stable Networks and the Connections Model.*

Consider the symmetric connections model (see Section **??**) with $\delta < c$ and $n \geq 4$. Identify a network that is pairwise Nash stable but not strongly stable for some choice of parameters. Find an example of a strongly stable network that is not efficient for a case where $n \geq 5$.

EXERCISE **11.5** *Deviations and Strongly Stable Networks.*

[26] Some starting references on this subject include: Axelrod [21], [22], Bonabeau [**?**], Brately, Fox, and Schrage [97], Epstein and Axtell [**?**], Gilbert and Troitzsch [262], Grimm and Railsback [296], Tesfatsion [592], and Tesfatsion and Judd [593].

The following definition follows one in Dutta and Mutuswami [200]. A network g is *strongly stable** with respect to if for any $S \subset N$ and g' that is obtainable from g via deviations by S, there exists $j \in S$ such that $u_j(g') \leq u_j(g)$.

Find an example of a network that is strongly stable* but not pairwise stable and hence not strongly stable.

EXERCISE **11.6** *Existence of Pairwise Stable Networks.*

Consider any N and profile of utility functions, one for each player. Show that either there exists a pairwise stable network or a closed cycle (as defined in Section **??**).

EXERCISE **11.7** *Improving Paths for Pairwise Nash Stability.*

Develop a definition of "improving path* that allows pairs of agents adding one link, or a single agent deleting multiple links and relate it to the existence of pairwise Nash stable networks. Provide an example where all improving paths* are part of cycles even though there exists a pairwise stable network.

EXERCISE **11.8** *Existence of Pairwise Stable and Pairwise Nash Stable Networks.*

Find an example where payoffs are anonymous,[27] and there exists a pairwise stable network, but there does not exist a pairwise Nash stable network. Does there exist such an example with $n = 3$ and where isolated players get a payoff of 0?

EXERCISE **11.9** *Simultaneous Stability.*

Consider the following variation on strong stability. A network g is *simultaneously stable* if for any $S \subset N$ such that $|S| \leq 2$, g' that is obtainable from g via deviations by S, and $i \in S$ such that $u_i(g') > u_i(g)$, there exists $j \in S$ such that $u_j(g') < u_j(g)$.

Thus, a network is simultaneously stable if no single player strictly prefers to delete some set of his or her links, and no two players would each weakly benefit (with at least one benefiting strictly) by deleting some of their links and/or adding a link between

[27] A profile of utility functions is anonymous if for every π that is a permutation on N (a one-to-one function mapping the set of agents N to N), it follows that $u_{\pi(i)}(g^\pi) = u_i(g)$, where $g^\pi = \{\{\pi(i), \pi(j)\} | ij \in g\}$ is the network obtained from g by permuting the positions of agents according to π.

them. This is a stronger requirement than Pairwise Nash stability, but weaker than strong stability because only coalitions of two players are considered.

Consider a setting where players come with "types" in some finite set Θ, and let player i's type be denoted θ_i. Suppose that payoffs are as follows. Let $s_i(g)$ be the number of players of i's same type to whom i is linked (so $\theta_i = \theta_j$ and $ij \in g$), and $o_i(g)$ be the number of players of types other than i's type whom i is linked to. Suppose that payoffs are as follows

- $u_i(g) = o_i(g) + 2s_i(g)$ if $o_i(g) + s_i(g) \leq d_i$, and
- $u_i(g) = 0$ if $o_i(g) + s_i(g) > d_i$,

where $d_i \geq 1$ is a capacity of links that i can maintain. So, players benefit more from links to their own types.

Show that the pairwise Nash networks are those such that: (i) no player exceeds his or her capacity and (ii) there is at most one player with fewer links than his or her capacity. Characterize the set of simultaneously stable networks. Show that if there are at least two players who have the same type, then the set of simultaneous stable networks is a strict subset of the set of pairwise Nash networks.

Exercise **11.10** *Ordinal Potential Functions and the Absence of Indifference*

Provide an example where the payoffs exhibit indifference and there are no improving cycles, but there does not exist an ordinal potential function (3 players will suffice).

Exercise **11.11** *An Ordinal Potential Function for a Buyer-Seller Network.*

Consider the Corominas-Bosch model from Section **??**. Show that the sum of all payoffs is an ordinal potential function.

Exercise **11.12** *Existence of Strongly Stable Networks and Top Convexity.*

Payoffs are *top-convex* if $\max_{g \in G(N)} \frac{\sum_{i \in N} u_i(g)}{|N|} \geq \max_{g \in G(S)} \frac{\sum_{i \in S} u_i(g)}{|S|}$ for all $S \subset N$.[28]

Suppose that payoffs are component decomposable and are such that any two players in the same component get the same payoff. Show that the set of strongly stable networks is nonempty if and only if payoffs are top-convex.

[28] This definition and result are due to Jackson and van den Nouweland [338].

Show that if we reallocate utility so that players within a component get an equal split of the total utility within a component, then payoffs in the symmetric connections model are top-convex.

EXERCISE **11.13** *Proof of Proposition 11.3.2 - Directed Nash Stable Networks in a Distance-Based Model.*

Prove Proposition 11.3.2.

EXERCISE **11.14** *A Stochastically Stable Network that is not Pairwise Nash Stable.*

Provide an example of a stochastically stable network that is pairwise stable but not pairwise Nash stable. Describe a variation on the random process that would instead select from Pairwise Nash Stable networks.

EXERCISE **11.15** *Proof of Proposition* **??***.*

Prove Proposition **??**.

EXERCISE **11.16** *Complementarities in Link Efforts with Convexities.*

Consider the following variation of a model by Bloch and Dutta [71].

Each agent has a unit of effort to allocate on different relationships. In particular, agent i chooses a vector of efforts, where $x_{ij} \in [0,1]$ is the effort that i invests on a relationship with agent j and where $\sum_{j \neq i} x_{ij} = 1$. The "strength" of overall relation between i and j is then $s_{ij} = \phi(x_{ij}) + \phi(x_{ji})$, where ϕ is an increasing function.

The payoff to i as a function of s (the matrix of s_{ij}'s) is $u_i(s) = \sum_{k \neq i} v_{ik}(s)$ where $v_{ik}(s)$ is determined as follows. Let P_{ik} be the set of the potential paths between i and k that could occur in any network. For a path $p = i_1 i_2, i_2 i_3, \ldots, i_{m-1} i_m$ let $v(p,s) = s_{i_1 i_2} \times s_{i_2 i_3} \times \cdots \times s_{i_{m-1} i_m}$. Then $v_{ik} = \max_{p \in P_{ik}} v(p,s)$.

Show that if ϕ is a strictly convex function then any Nash equilibrium choice of x_{ij}'s is such that, for each i, $x_{ij} = 1$ for some j and $x_{ik} = 0$ for all $k \neq j$. What are the equilibrium configurations for a three agent system when ϕ is strictly convex?

EXERCISE **11.17** *Schelling's Tipping and Segregation Models: An Agent-Based Computation Exercise.*

Consider the following simple model of segregation due to Schelling [549], [550]. A finite set of agents live on a line segment.[29] Agents are either red or blue. Agents have preferences over their neighborhoods: an agent prefers to be in a neighborhood where at least half of his or her neighbors are of the same color. Other than this agents have no further preference. An agent considers his or her neighborhood to consist of the k nearest agents to his left plus the k nearest agents to his right. (For agents near the end of the segment, simply wrap the segment around to form a circle when defining closest neighbors.) At any point in time each agent can be labeled as either "content" or "discontent," depending on whether at least fifty percent of their neighbors are of their color.

Schelling describes the following algorithm

- Start with n agents randomly positioned on a line segment (without any two agents at the same point) and with their colors randomly assigned with independent and equal probability of red or blue.
- Identify the "discontent" agents and label this set $D(0)$.
- Starting with the leftmost "discontent" agent, move that agent to a point where the agent is content, and do so in a manner so that the agent needs to leapfrog the fewest other agents (and break ties by moving to the left).
- Next, if the second agent who was discontent is now content (due to the change induced by the first intially discontent agent's move) then leave that secon initially discontent agent in place. Otherwise, move that agent to a point where he or she will be content (again leapfrogging as few agents as possible and moving to the left in the case of a tie).
- Iterate on this process until all of the agents in $D(0)$ have been considered.
- Now begin the whole process over again, identifying a new set of discontent agents $D(1)$, and moving discontent agents as described above.
- Iterate on this process until all agents are content.

Write a program to run this algorithm when $k = 4$ and $n = 100$. Run this program 100 times. For each run, record how many changes in color there are as one moves

[29] There are two dimensional versions of the model, which are known as "Schelling's Checkerboard Model".

along the segment from left to right at the starting configuration and at the ending configuration. This gives an idea of how many different segmented groups of agents of the same color there are on the line. Also record the average fraction of neighbors of the same color at both the starting and ending configurations.

Chapter 12

Allocation Rules, Networks, and Cooperative Games

Throughout this book we have seen different sorts of predictions of how powerful or central different agents are, as well as how influential they are, what terms of trade they might get in an exchange, and what utility they might end up with in a game played on a network. These predictions were derived using a variety of tools, ranging from specific measures of power and centrality to models of the spread of information, bargaining on networks, or strategic interaction on a network. Beyond these tools that are directed at specific applications, there is a more general perspective that builds on properties of how the aggregate payoff behaves as a function of the network and then deduces allocations of payoffs from those properties. It is an offshoot of cooperative game theory, which examines how productive value is split among the members of a society based on the relative contributions of different coalitions of players. This approach has both normative (how the total payoff should be split) and positive (how value is split) sides to it.

In this chapter I examine how tools of cooperative game theory have been extended and adapted to network settings. I begin with a brief background discussion of cooperative game theory and then turn to the extension of various methods and concepts to network settings. As we shall see, the concepts provide a nice basis for making predictions about the outcomes of multilateral bargaining on networks, as well as more generally for analyzing the power or influence of various players in a network. We will also see that there are different ways to extend and adapt cooperative tools to network settings, and that what is the "right" extension will depend the context.

12.1 Cooperative Game Theory

Cooperative game theory starts with a society of players just as a network setting does, but instead of thinking of how the players might be structured in terms of social networks, the foundation is built on simpler structures. In particular, the primitives are how the players might be grouped into subsets or “coalitions”.

There are several branches of the theory and an important distinction is made between “transferable utility” and “non-transferable utility (NTU)” games. In transferable utility games payoffs can be freely reallocated among players (so that payoffs are transferable), while in non-transferable utility games the payoffs can only come in given configurations. While both of these branches have natural cousins in the network setting, I focus on the more developed branch of transferable utility.[1]

Cooperative game theory provides a prediction (or prescription, depending on the interpretation) of how the total value generated in a society will (or should) be split among its members. It takes into account the relative values that every possible subset of players could generate, and then based on certain properties of how the allocation of value reacts (or should react) to these values. One interpretation is that the resulting allocation is a prediction of what the outcome will be if the players bargain over how to allocate the total value generated in the society, and when the values of various coalitions of players represent threat points of what they can earn by seceding from the society. As multilateral bargaining is difficult to model non-cooperatively, cooperative techniques can be quite useful in this regard. Another interpretation is that the resulting allocation is how a society should allocate its value, based on some principles.

[1]The terminology of “cooperative” and “non-cooperative games is perhaps no longer as useful as it was when first defined. The idea of a cooperative game began as an offshoot of a non-cooperative game. The value that could be generated by a group of players was what they could guarantee themselves by coordinating their strategies in the non-cooperative game. (See Luce and Raiffa [**?**] for a nice overview of these ideas.) That is, the value of a given coalition of players was derived by looking at the maximum value (in terms of the sum of their payoffs) that they could get across their choices of strategies when the remaining players react by collectively choosing their strategies minimize the coalition’s payoff. The idea of a value of a group of agents extends far beyond the original cooperation in a non-cooperative game derivation, and so the terminology is no longer so pertinent. Also, in many applications there is a well-defined value generated by a group of agents without requiring them to “cooperate,” especially when the theory is applied normatively. Nevertheless, the terminology remains from its historic roots.

12.1.1 Transferable Utility (TU) Cooperative Games

The society of players is denoted $N = \{1, \ldots, n\}$.

The productive values of different coalitions are are captured via a *characteristic function*, $w : 2^N \to I\!R$, where the value of a coalition $S \subset N$ is denoted $w(S)$. Let us normalize a characteristic function so that $w(\emptyset) = 0$.

Together (N, w) are referred to as a *transferable utility game* or a *TU game*. The set of all such games on a society N is denoted $W(N)$.

The term "transferable utility" refers to the idea that the value of a coalition can be transferred its members.

It helps to keep a few examples in mind.

Example **12.1.1** *Divide the Dollar.*

Consider a legislature that operates by majority rule and has a budget to split among its members. Normalize the value of the budget to 1. A proposal of how to split the budget can be passed if it receives the votes of a majority of the legislature's members. Thus, any coalition of at least $\frac{n}{2}$ members, and only such coalitions, can generate a value of 1.

This is represented by a TU game such that $w(S) = 1$ if $|S| \geq \frac{n}{2}$ and $w(S) = 0$ otherwise.

Example **12.1.2** *Simple Games.*

The divide-the-dollar game is an example of a more general class of games such that there are "winning" coalitions that can generate a value of 1. The divide the dollar setting is one where those coalitions are ones that contain a majority of the players. More generally, one can consider other possible rules for which coalitions can generate value.

A *simple game* is a TU game such that

- $w(S) \in \{0, 1\}$,
- $w(S) = 1$ and $S \subset S'$ implies that $w(S') = 1$,
- $w(N) = 1$,
- $w(S) = 1$ implies that $w(N \setminus S) = 0$.

Thus, a simple game specifies which coalitions generate a value of 1, is such that larger coalitions are at least as powerful as smaller coalitions, and is such that there cannot exist disjoint coalitions that each generate a value of 1 at the same time.

EXAMPLE **12.1.3** *A Quota Game with Veto Players.*

Another class of interesting TU games is a subset of the simple games where a specific player is needed in order to generate value. That is, a coalition is worthless unless it contains some specific player.

Let there be a quota $q \geq \frac{n}{2}$ and a set of veto players $C \subset N$ such that $w(S) = 1$ if and only if $|S| \geq q$ and $C \subset S$.

Thus, a coalition generates a value of 1 if and only if it meets the size quota and all of the veto players are included. An example of such a setting is the United Nations security council which has fifteen members and five veto players (the five permanent members: China, France, Russia, the U.S.A., and the U.K.). A resolution passes only if it receives "yes" votes of at least two thirds of its members and none of the five veto players vote "no".[2]

12.1.2 Allocating the Value

The values of different coalitions form the foundation for the analysis of a cooperative game, as they tell us the value there is to be split among the players and also the extent to which different groups of players are responsible for the generation of the value. The heart of the analysis is then how the value of the society is (or should be) allocated among its members. This is captured by an imputation.

An *imputation* is a function $\phi : W(N) \rightarrow I\!R^n$ such that $\sum_i \phi_i(w) = w(N)$.

An imputation thus indicates how much of the value generated by the full society is allocated to each player. Generally it is presumed that the grand coalition of the full society generates the maximum possible value.[3]

[2] The actual rules of the U.N. Security Council are a bit more complicated than a simple game as countries are allowed to abstain, and they sometimes do. In addition, different voting rules are used depending on the issue involved, but the basic structure is built on a quota game with veto players.

[3] There is a specification that examines more general settings where the value generated depends on how a society is partitioned and such that it might be efficient to have partitions other than the one where all players are grouped together. Such games are called games in partition function form, and can also be seen as special cases of the network setting discussed below.

An imputation can model different things. It might capture the result of a bargaining process or it might be a normative analysis of how the value should be allocated. It might also be a measure of the relative power of different members of the society.

A prominent imputation rule is the Shapley Value, introduced by Shapley [**?**]. It has a number of interesting properties, and can be seen as being based on the relative marginal contributions of players towards productive value.

12.1.3 The Shapley Value

The Shapley Value is an imputation defined by

$$\phi_i^{SV}(w) = \sum_{S \subset N \setminus \{i\}} (w(S \cup \{i\}) - w(S)) \left(\frac{\#S!(n - \#S - 1)!}{n!} \right).$$

A standard interpretation of the Shapley Value is as follows. Uniformly at random, choose an ordering of the players and let it be $\{i_1, i_2, \ldots, i_n\}$. Now, consider building the society up by adding one player at a time in this order. A player gets the marginal contribution that he or she makes to the society when added to the players who preceded him or her. So, a player i whose place in the order follows a coalition S gets $w(S \cup \{i\}) - w(S)$. There are $\#S!(n - \#S - 1)!$ such orderings, and averaging over all such orderings leads to the Shapley value.

In the divide-the-dollar game the Shapley Value allocates $1/n$ to each player given the full symmetry of the game. This is true of most any imputation. If instead we examine a simple majority game with a single veto player, then we see asymmetries and begin to get an impression of how the Shapley value operates.

Example **12.1.4** *A Majority Game with One Veto Player.*

Consider a quota game with an three players and a single veto player, player 1. In particular suppose that $w(S) = 1$ if $|S| \geq 2$ and $1 \in S$, and otherwise $w(S) = 0$.

The Shapley value of this game can be calculated as follows. There are 6 possible orderings of the players: (1,2,3), (1,3,2), etc. Player 1 contributes a marginal value of 1 in 4 of these 6 possible orderings (whenever 1 is not the first player included). The other two players each contribute a marginal value of 1 in just 1 of the 6 possible orderings (the ordering where player 1 comes first and he or she comes second). Thus, the Shapley Value of this game is

$$\phi^{SV}(w) = \left(\frac{2}{3}, \frac{1}{6}, \frac{1}{6} \right).$$

Player 1 gets a larger value than the other two players because of the asymmetric roles in this game. Players 2 and 3 get some value due to the fact that they do contribute productive value in that player 1 cannot generate any value without at least one of them.

12.1.4 The Core

Instead of having an imputation, such as the Shapley Value, which makes a unique prediction for each cooperative game, we might instead make a set of predictions based on some principles. The most prominent such predictions are based on the "core." The idea is that the allocation of the value of the whole society must be such that no coalition could secede and improve each member's payoff by allocating the value it generates alone to its members.

The *core* of a TU cooperative game is the set of all allocations $x \in I\!R^n$ such that

- $\sum_{i \in N} x_i = w(N)$, and
- $\sum_{i \in S} x_i \geq w(S)$ for all $S \subset N$.

When the core is nonempty, it makes powerful predictions, as an allocation in the core cannot be blocked by any coalition.

The core is sometimes empty as seen, for instance, in the divide-the-dollar game as follows. Consider any allocation x such that $\sum_{i \in N} x_i = w(N)$. There must exist i such that $x_i > 0$. Thus, it must be that $\sum_{j \neq i} x_j < 1$, while $w(\{j : j \neq i\}) = 1$. Thus, this cannot satisfy the second requirement in the definition of the core, as the coalition of players other than i are not getting the value that they would obtain by excluding player i. The core of this game is empty.

This shows the inherent instability of majority rule, and also previews the difference between predictions made based on an imputation rule such as the Shapley Value which in the divide-the-dollar game gives an equal allocation to each player, and the core. To see further differences, let us re-examine a game with a veto player.

EXAMPLE **12.1.5** *The Core in a Majority Game with One Veto Player.*

Reconsider the quota game from Example 12.1.4 with three players and a single veto player, player 1, and recall that the allocation under the Shapley Value was $\phi^{SV}(w) = \left(\frac{2}{3}, \frac{1}{6}, \frac{1}{6}\right)$.

The core in this game consists of a single allocation: (1,0,0). To see this, note that a core allocation must satisfy $x_1 + x_2 \geq 1$, $x_1 + x_3 \geq 1$, $\sum_i x_i = 1$, and $x_i \geq 0$ (given that $w(\{i\}) = 0$) for each i.

This example shows us that the core and Shapley Value are capturing different things. The Shapley value is not necessarily in the core. The core is built on ensuring that no coalition could block the allocation with a better allocation for its members, while the Shapley Value is derived from calculations on relative contributions.

The Shapley Value always lies in the core in some cases, including a subclass of games called convex games. Thus, that class of games is such that the core is nonempty.

A TU game w is *convex* if

$$w(S \cup \{i\}) - w(S) \leq w(S' \cup \{i\}) - w(S'),$$

whenever $S \subset S'$ and $i \notin S'$.

The convexity here refers to the fact that the marginal contribution of a player (weakly) increases as the size of the coalition he or she is added to is increased. If the value of a coalition only depends on the number of members it has, then this requires that that value be a convex function of the number of players. In such games, the grand coalition generates enough value to allocate in a way such that each coalition is getting at least its value, and so the core is nonempty, and, in fact, contains the Shapley value.

PROPOSITION **12.1.1** *If a TU game is convex, then the Shapley Value of the game is in the core of the game.*

The proof is relatively straightforward and the subject of Exercise 12.1.

With a brief introduction to cooperative game theory in hand,[4] let us now begin to bring network structures into play.

12.2 Communication Games

Myerson [473] introduced an interesting subclass of cooperative games that are called communication games.[5]

[4]For a bit more background on cooperative game theory see Myerson [474] and Osborne and Rubinstein [489].

[5]Myerson referred to the network as a "cooperation structure" and such games are also referred to as games with cooperation structures.

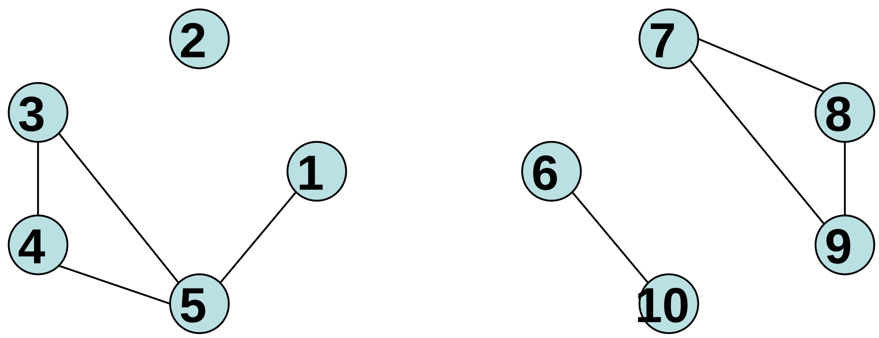

Figure 12.2. *A Communication Network*

Given a network (N, g), recall that $g|_S$ is the subnetwork of g restricted to the nodes in $S \subset N$ and $\Pi(g|_S)$ is the partition of S generated by the components of g restricted to S (as defined in Section **??**).

Myerson's definition begins with a convex TU cooperative game $(N, w) \in W(N)$, and augments this by a network $g \in G(N)$ that describes who can communicate with whom.

The *communication game* (N, w, g) induces a cooperative game $(N, \widehat{w}_g)$ such that

$$\widehat{w}_g(S) = \sum_{C \in \Pi(g|_S)} w(C).$$

The idea is that coalitions can only function to the extent that they can communicate. To see how this works, consider the network pictured in Figure 12.2.

Here, the coalition $\{1, 4, 5\}$ can function because the players are path-connected to each other, each lying in the same component, and thus $\widehat{w}_g(\{1, 4, 5\}) = w(\{1, 4, 5\})$.

In contrast, the coalition $\{1, 3, 4\}$ can only partially function and

$$\widehat{w}_g(\{1, 3, 4\}) = w(\{1\}) + w(\{3, 4\}) = w(\{3, 4\}).$$

Here, even though 1, 3, and 4, are path-connected in g, player 1 cannot communicate with 3 or 4 without 5 being present.

The value of coalition $\{1, 2, 6, 7\}$ is 0, since they cannot communicate at all in g.

12.2.1 The Myerson Value

The Shapley value has a natural extension to communication games.[6]

Myerson [473] defined an allocation rule for a communication game (N, w, g), namely

$$\psi^{MV}(w, g) = \phi^{SV}(\widehat{w}_g)$$

Although one can view a communication game as a specific form of a cooperative game (basically, the induced $(N, \widehat{w}_g)$), it also works the other way around. In the case where g is the complete network, a communication game reduces to a cooperative game in that $\widehat{w}_g = w$. Moreover, as one varies the network structure, one can see how the allocation of value varies. That is, as the network structure is varied, $\widehat{w}_g$ varies and so does the allocation, even though the underlying cooperative game remains fixed.

To see this, consider the divide the dollar game with three players. In that case th Shapley value allocates 1/3 to each player. That would be the allocation under the Myerson value if g is the complete network. However, if instead the network had only one link, then the two agents involved in the link would each get a value of 1/2. The most interesting case occurs when there is a two-link game. This now looks like a cooperative game where the middle player is a veto player, since without that player a coalition cannot function. Thus, the middle player gets 2/3 of the value, while the end players each get 1/6. This is pictured in Figure 12.2.1

While the communication games introduced by Myerson [473] bring networks into the context of cooperative game theory, they stop short of allowing one to fully analyze the allocation of the values in network settings. The difficulty lies in the fact that the actual value that a society generates is still based on a characteristic function, and it is mainly the allocation of value that is affected by the network structure, rather than the overall productive possibilities. To see the issue, consider a society of $N = \{1, 2, 3\}$, and any underlying cooperative game. A network of two links, say $g = \{12, 23\}$, and the complete network, $g' = \{12, 23, 13\}$, both lead to the same overall productive value in a communication game since they allow all three agents to communicate. While the Myerson Value allocates value differently to the players in the game, it still requires that both networks have the same value to allocate. In most any productive situation,

[6] For more of an overview of the literature on communication games and other allocation rules see Slikker and van den Nouweland [?].

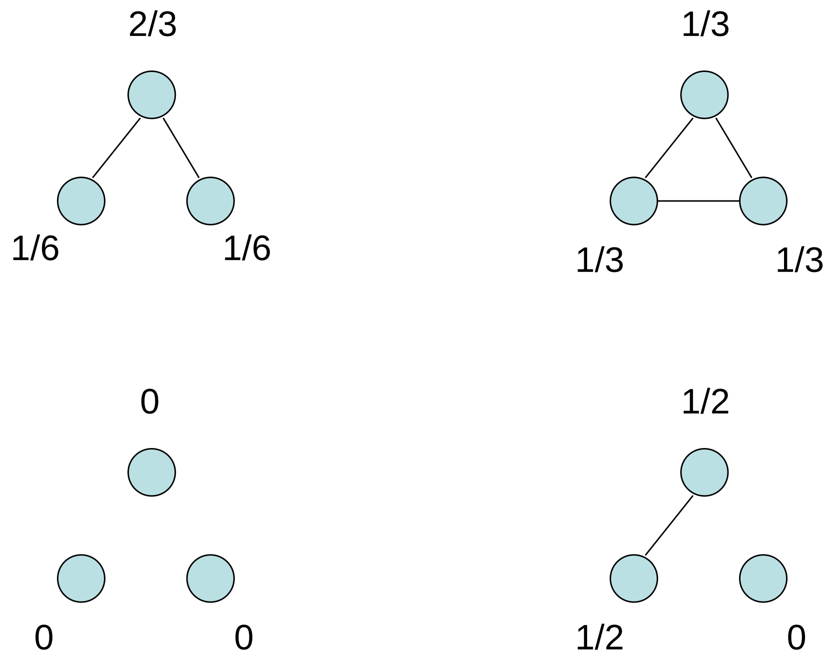

Figure 12.2.1. *The Myerson Value in a Communication Game: a 3-Player Divide-the-Dollar Game.*

including the structure of a firm or any sort of organization, links involve some costs and, more generally, network structure will affect productivity. Dealing with this issue requires a richer setting where the productive value is not based on a cooperative game, but is instead directly dependent on the network in place, as discussed next.

12.3 Networks and Allocation Rules

Jackson and Wolinsky [343] proposed a richer model than that of communication games, where the value that a society generates depends explicitly and directly on the network structure. This has as special cases both cooperative games and communication games, but uses networks as the primitive. It is described as follows.

Throughout, let a society N be given.

12.3.1 Value Functions

The productive value of a society is now determined directly by the network structure, and is captured via a value function.

A *value function* is a function $v : G(N) \to I\!R$.

Let us normalize the value of the empty network to be zero, so that $v(\emptyset) = 0$.

The set of all possible value functions for a society N is denoted $\mathcal{V}(N)$.

Note that any profile of utility functions $u = (u_1, \ldots, u_n)$ generates a value function defined by $v(g) = \sum_i u_i(g)$. Thus, the utility-based models of network formation that we have discussed, such as the connections model, distance-based utility models, the co-author model, and so forth, give rise to distinct value functions.

A prominent class of value functions is the set of component additive ones.

A value function v is *component additive* if $\sum_{h \in C(g)} v(h) = v(g)$.

Component additivity is a condition that rules out externalities across components, but still allows externalities within components. That is, the value of a given component does not depend on how other components are structured. It is quite natural in some contexts, for instance social interactions, but not in situations where different components interact with each other. If the value function is derived as the sum of a component decomposable profile of utility functions, then it will clearly be component additive.

Another prominent subclass of value functions is the set of anonymous ones.

Given a permutation of players π (a bijection from N to N) and any $g \in G(N)$, let $g^\pi = \{\{\pi(i), \pi(j)\} | ij \in g\}$.

Thus, g^π is a network that shares the same architecture as g but with the players relabeled according to π.

A value function is *anonymous* if $v(g^\pi) = v(g)$ for any permutation of the set of players π.

Anonymity says that the value of a network depends only on the structure of the network and not the labels of the players who occupy various positions. It requires that the critical productive determinant be social structure, and abstracts away from personal productive differences between individuals.

12.3.2 Allocation Rules

The analog of an imputation in the network setting is an allocation rule. It is a richer object because it not only depends on the value function, but also on the network structure that is in place.

An *allocation rule* is a function $Y : G(N) \times \mathcal{V}(N) \to I\!R^n$ such that $\sum_i Y_i(g, v) = v(g)$ for all v and g.

The definition of an allocation rule has a balance condition, $\sum_i Y_i(g,v) = v(g)$, built into it.

The idea of an allocation rule is to analyze how the total productive value or utility of a society ends up being allocated. It is the analog of an imputation, but now in the context of a network setting instead of a cooperative game. In many of the previous chapters, a profile of utility functions was taken as given. However, bargaining might occur about the terms of trade, or even about favors within a friendship or an allocation of chores. It could also be that there are taxes or subsidies imposed. An allocation rule keeps track of how the total value ends up being allocated after such a process.

An allocation rule depends on both g and v. This is important as it can then take account of a player i's role in productive value beyond the specific network in place. For instance, consider a network $g = \{12, 23\}$ in a situation where the value generated is 1 ($v(g) = 1$). Player 2's allocation might depend heavily on the values of other networks. For instance, if $v(\{12, 23, 13\}) = 0 = v(\{13\})$, then 2 is essential to the network and may receive a large allocation. If, on the other hand $v(g') = 1$ for all networks, then 2's role is not particularly special. This information can be relevant, especially in bargaining situations, which is why the allocation rule is allowed to depend on it.

12.3.3 Some Properties of Allocation Rules

There are some properties of allocation rules that are useful in studying extensions of the Shapley Value and Myerson Value to the network setting.

An allocation rule Y is *component balanced* if $\sum_{i\in S} Y_i(g,v) = v(g|_S)$ for each component additive v, $g \in G$, and $S \in \Pi(g)$.

Component balance requires that the value of a component of a network be allocated to the members of that component in cases where the value of the component is independent of how other components are organized. This would tend to arise naturally. It also is a condition that an intervening government would like to respect if it wishes to avoid secession by components of the network.

Given a permutation $\pi : N \to N$, let v^π be defined by $v^\pi(g) = v(g^{\pi^{-1}})$ for each $g \in G$ (recalling the definition of g^π from Section **??**). This is just value function obtained when agents' names are relabeled through π.

An allocation rule Y is *anonymous* if for any $v \in \mathcal{V}$, $g \in G$, and permutation of the set of players π, $Y_{\pi(i)}(g^\pi, v^\pi) = Y_i(g,v)$.

Anonymity of an allocation rule requires that if players are relabeled, then the allocation change accordingly.

12.3.4 Egalitarian Allocation Rules

An egalitarian allocation rule spreads value equitably among the members of a society. This can be done in different ways. One can simply spread value completely equally, or one might instead spread value the value of a component back to the members of that component.

The *egalitarian allocation rule* Y^e is defined by $Y_i^e(g,v) = \frac{v(g)}{n}$.

The egalitarian allocation rule has nice properties. Any efficient network will be pairwise (Nash) stable, and in fact strongly stable, if payoffs are given by the egalitarian rule since it maximizes the payoffs of all players. Moreover, there are no improving cycles, as we can see by setting $f(g) = \frac{v(g)}{n}$ and applying Proposition 11.2.1.

Despite these virtues, the egalitarian rule fails to satisfy component balance in component additive settings where not all components generate the same value. That is, there are many situations where it will makes transfers across components even though there are no externalities across the components. While this might be attractive from a normative perspective, it not as natural from a positive perspective (for instance, as the prediction of the outcome of a bargaining process), especially when the value function is component additive. Moreover, it can fail a basic stability property, as it could give incentives for components to secede, as they might end up being taxed as a whole. A natural variation on the egalitarian rule only equalizes allocations within components.

The *component-wise egalitarian allocation rule*, denoted Y^{ce}, is defined as follows. For a component additive v and network g, Y^{ce} is such that for any $h \in C(g)$ and each $i \in N(h)$

$$Y_i^{ce}(g,v) = \frac{v(h)}{\#N(h)}.$$

For a value function v that is not component additive, $Y^{ce}(g,v) = Y^e(g,v)$ for all g.

The component-wise egalitarian allocation rule only differs from the egalitarian rule in situations where it is clear that one can attribute a value to each component separately from the organization of the rest of the network; that is, in the situation of component additive value functions. Otherwise, there is no obvious value to attribute to a component, and then this allocation rule coincides with the egalitarian rule.

While the component-wise egalitarian rule is not quite as nice as the egalitarian rule in terms of all efficient networks being pairwise Nash stable and strongly stable, it still has some nice stability properties. For instance, when payoffs are given by the component-wise egalitarian allocation rule, there always exists a pairwise (Nash) stable network, and one can be found by a simple algorithm as outlined by Jackson [?] for the case of a component additive v (as otherwise it is the same as the egalitarian rule).

- Find a component h that maximizes the payoff $Y_i^{ce}(h,v)$ over i and h, and if there is more than one such component then choose the one that has the most agents.

- Follow the same policy on the remaining population $N \setminus N(h)$, and iterate.

The collection of resulting components forms the network.[7] While this identifies a pairwise Nash stable network, it does not always find a strongly stable network, but it does always find networks that are nearly strongly stable (see Exercise 12.7).

In many contexts the component-wise egalitarian allocation rule is such that efficient networks are pairwise stable (see Exercise 12.9).

Beyond egalitarian type allocation rules, there are allocation rules based on very different approach, namely Shapley value-style marginal contribution calculations.

12.3.5 The Myerson Value in Network Settings

The Shapley and Myerson Values have a natural extension to the context of networks, as shown by Jackson and Wolinsky [343]. That allocation rule is expressed as follows.

$$Y_i^{MV}(g,v) = \sum_{S \subset N \setminus \{i\}} (v(g|_{S \cup i}) - v(g|_S)) \left(\frac{\#S!(n - \#S - 1)!}{n!} \right).$$

The Myerson Value in this full network setting again allocates value using Shapley Value-style calculations, now based on how the value changes as the players comprising the network are changed.

The Myerson Value has some nice properties that distinguish it from other rules, as I now discuss.

[7] This follows a argument similar to one used by Banerjee, Konishi and Sönmez [?] to establish existence of core-stable coalition structures in a class of coalition formation games called "hedonic games."

12.3.6 Equal Bargaining Power, Fairness, and the Myerson Value

An allocation rule satisfies *equal bargaining power* if for any component additive v and $g \in G(N)$

$$Y_i(g,v) - Y_i(g-ij,v) = Y_j(g,v) - Y_j(g-ij,v),$$

for any link ij.

Equal bargaining power is a variation on a condition called "fairness" by Myerson. The fairness name can be thought of as viewing the property from a normative perspective, while the equal bargaining name looks at the property from a more positive viewpoint. Note that equal bargaining power does *not* require that players split the marginal value of a link. It just requires that they equally benefit or suffer from its addition. It is possible (and generally the case) that $Y_i(g,v) - Y_i(g-ij,v) + Y_j(g,v) - Y_j(g-ij,v) \neq v(g) - v(g-ij)$, so that the marginal value of a link is not allocated only to the two involved players.

At first sight, equal bargaining power seems like a natural condition. Why shouldn't two players involved in a relationship each gain or suffer equally from the addition of that relationship? (There will be an answer given to this shortly.) The is a powerful condition that is not satisfied by many rules including egalitarian-style rules. In fact, equal bargaining power in conjunction with component balance uniquely ties down the allocation rule. The following proposition from Jackson and Wolinsky [343] is a direct extension of Myerson's [473] result from the communication game setting to the network setting.

PROPOSITION **12.3.1** *[Myerson [473], Jackson and Wolinsky [343]] Y satisfies component balance and equal bargaining power if and only if $Y(g,v) = Y^{MV}(g,v)$ for all $g \in G$ and any component additive v.*

The proof follows the logic of Myerson's proof, but adapted to a network setting. One can show that there is a unique rule that satisfies equal bargaining power and component balance when v is component additive, and that the Myerson value satisfies these conditions.

Without providing the proof here, let me illustrate the ideas behind why these two properties uniquely tie down the allocation. Start by considering a one link component. In that case, if the link were deleted, both players would get a 0 payoff (by component balance) and the "normalization" that isolated nodes generate value 0.[8] Equal bargaining power then implies that the players get the same allocation, and component balance requires that the two players split the entire value of a single link, and so

$$Y_i(\{ij\}, v) = Y_j(\{ij\}, v) = \frac{v(\{ij\})}{2}.$$

Next, consider a two link component $h = \{ij, jk\}$. Component balance requires that

$$Y_i(h, v) + Y_j(h, v) + Y_k(h, v) = v(h). \tag{12.1}$$

Equal bargaining power requires that

$$Y_i(h, v) - Y_i(\{jk\}, v) = Y_j(h, v) - Y_j(\{jk\}, v),$$

and by component balance $Y_i(\{jk\}, v) = 0$, and so this implies that

$$Y_i(h, v) = Y_j(h, v) - \frac{v(\{jk\})}{2}. \tag{12.2}$$

Similarly,

$$Y_k(h, v) = Y_j(h, v) - \frac{v(\{ij\})}{2}. \tag{12.3}$$

Then from (12.1), (12.2), and (12.3) it follows that

$$3Y_j(h, v) - \frac{v(\{ij\})}{2} - \frac{v(\{jk\})}{2} = v(h),$$

or

$$Y_j(h, v) = \frac{v(h)}{3} + \frac{v(\{ij\})}{6} + \frac{v(\{jk\})}{6}.$$

Then (12.2) and (12.3) imply that

$$Y_i(h, v) = \frac{v(h)}{3} + \frac{v(\{ij\})}{6} - \frac{v(\{jk\})}{3},$$

and

$$Y_k(h, v) = \frac{v(h)}{3} - \frac{v(\{ij\})}{3} + \frac{v(\{jk\})}{6}.$$

These three expressions provide the Myerson Value for a component of the form $h = \{ij, jk\}$.

We see that the allocation for one and two link networks are unique under component balance and equal bargaining power. The idea is that we can derive the two

[8] This is implied by $v() = 0$ and component balance. We now see that this is more than a normalization, as it requires that all players generate the same value when isolated.

link network from two different one link networks, and so we have two different conditions tying down the allocation. Together with component balance, we then have three conditions tying down three allocations. As we continue to examine larger and larger components, an iterative logic can be used to determine the allocation uniquely at each step. The proof shows that there is at most one rule satisfying these conditions and so builds on this logic, but it does not work by explicitly deriving the allocation as in the exposition above.

There are also weighted versions of the Shapley and Myerson values, where the bargaining power is not equal but instead involves asymmetries among the players and some players receive systematically larger shares than other agents. Dutta and Mutuswami [200] extend the characterization to allow for weighted bargaining power and show that one obtains a version of a weighted Shapley (Myerson) Value.

12.3.7 Pairwise Stable Networks under the Myerson Value

A nice feature of the Myerson value is that pairwise stable networks always exist under it. In fact, the Myerson Value has an ordinal potential function, which then allows us to apply Proposition 11.2.1 to conclude that there are no improving cycles and that pairwise stable networks exist.

PROPOSITION **12.3.2** *[Jackson [**?**]] There exists a pairwise stable network relative to the Myerson Value allocation rule Y^{MV} for every value function v. Moreover, following improving paths relative to the Myerson value and any value function, and starting from any network, eventually leads to a pairwise stable network, and there are no improving cycles under the Myerson Value.*

Proof of Proposition 12.3.2: This is a corollary to Proposition 11.2.1. Let

$$f(g) = \sum_{S \subset N} v(g|_S) \left(\frac{(\#S-1)!(n-\#S)!}{n!} \right).$$

Then $Y_i^{MV}(g,v) - Y_i^{MV}(g-ij,v) = f(g) - f(g-ij)$, and so f is an ordinal potential function as required in Proposition 11.2.1. ∎

Although the Myerson Value leads to nice stability properties in terms of the existence of pairwise stable networks and the absence of cycles, it does not guarantee that the stable networks are even Pareto efficient. In fact, it can lead to systematic over-connection. This is illustrated by the following example and detailed in Exercise 12.6.

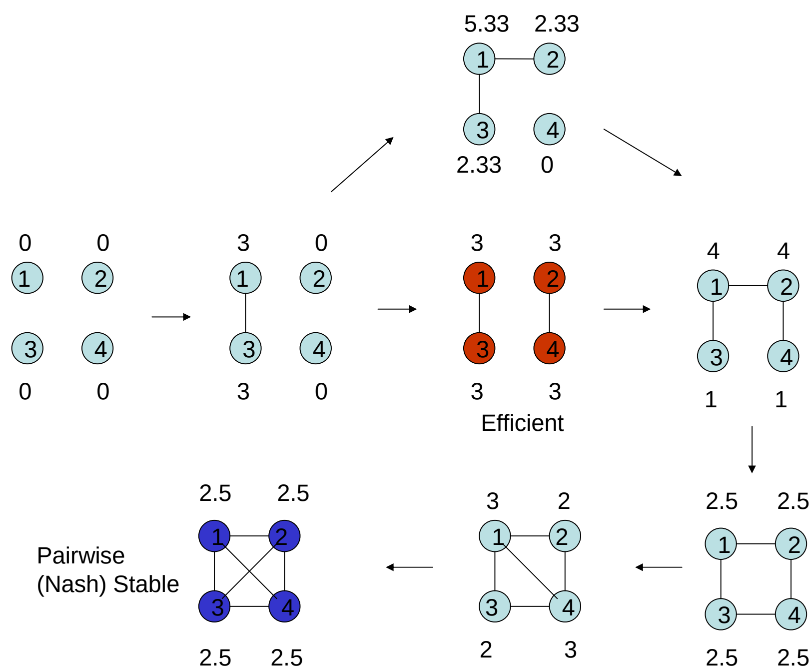

Figure 12.3.7. *Over-Connection and Pareto Inefficiency of Pairwise Nash Stable Networks under the Myerson Value: Dyads have Value 6 and Other Nonempty Components have Value 10.*

In Figure 12.3.7, we see the incentives for players to over-connect under the Myerson Value. In this example a dyad generates a value of 6 and any other nonempty component generates a value of 10. It is easy to check that the unique pairwise stable network is the complete network (which is verified via the improving paths pictured in the figure and a few that are not pictured), but this network is Pareto dominated by the efficient network. The reason for the inefficiency is that by having more links, there are more orderings under which a given player is important in contributing to the network. Thus, a player wishes to have more connections as they lead to increased bargaining power (as reflected in the Shapley value calculations).

12.4 Allocations Rules when Networks are Formed

If the network is something which can be formed at the players' discretion,[9] then one can argue that the Myerson Value it is correct neither from a normative standpoint, nor a positive standpoint, especially if the allocation rule is partly determined by the bargaining of players during the formation process. In particular, values of all networks, and not just sub-networks, should play some role in determining the allocation, as they are all viable alternatives. The Myerson Value takes into account subnetworks of a given network when calculating its value, but not other networks. To get a better feeling for this, let us examine some examples from Jackson [331].

Example **12.4.1** *A Criticism of the Myerson Value*

There is a three person society. Consider the two different value functions pictured below.

One value function v is $v(\{12\}) = v(\{23\}) = v(\{12, 23\}) = 1$, while $v(g) = 0$ for all other networks.

The other value function v' is such that all nonempty networks generate the same value of 1.

The Myerson Value assigns the same allocation to the agents in the network $g = \{1, 2, 3\}$ regardless of which of the two value functions is in place. That is,

$$Y^{MV}(\{12, 23\}, v) = Y^{MV}(\{12, 23\}, v') = \left(\frac{1}{6}, \frac{2}{3}, \frac{1}{6}\right).$$

[9] Actually, if the network is something that is fixed and cannot be altered, then it is not clear why one would pay attention to marginal contributions to the network, or why just subnetworks would be important. So, the criticisms here apply more broadly.

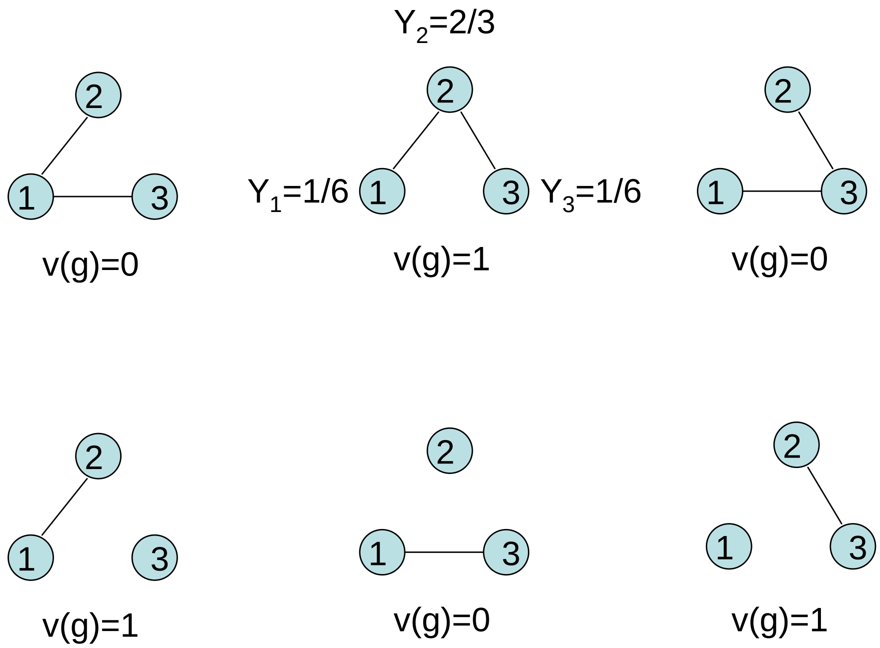

Figure 12.4. *The Myerson Value on v.*

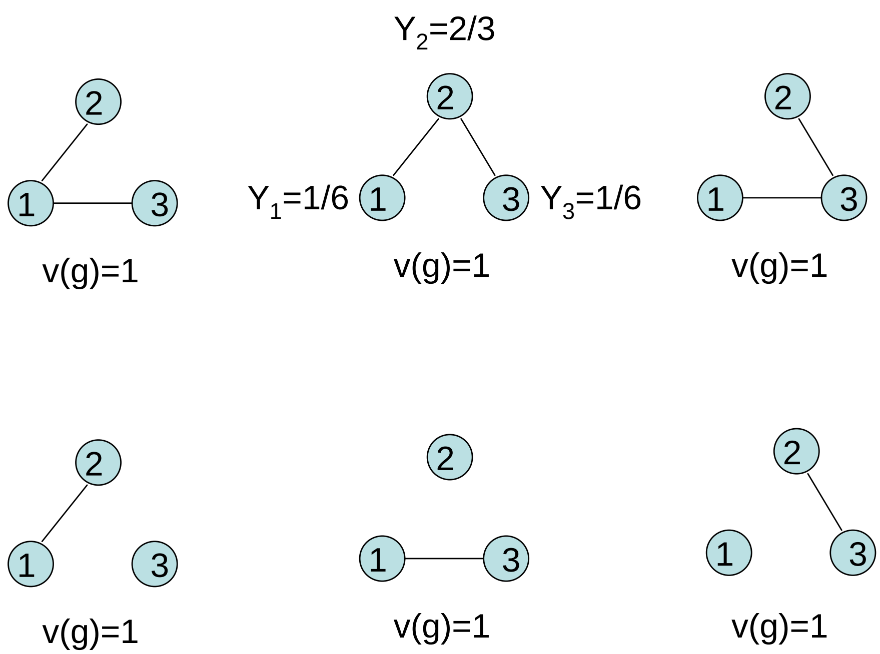

Figure 12.4. *The Myerson Value on v'.*

Player 2 is rewarded for being the central player in the network. Although the network is asymmetric, under the value function v', player 2 is not special in any way.

Let us consider the two main perspectives that might be taken. First, it could be that the network is something that can be adjusted and the allocation needs to take into account the fact that agents could rearrange themselves, or it might be that the allocation is the result of a bargaining process. From this perspective, the fact that player 2 is essential to generating value under v and not special under v', leads to very different disagreement points which should be reflected in the bargaining, but the Myerson value does not account for this. Second, it could be that the network is fixed and changing the network can only be relevant from normative perspective. If that were the case, then why should the allocation rule take into account subnetworks but not others? Basically, the criticism here is that the Myerson value takes into account how the value changes with respect to some but not all networks, and thus does not fully account for the roles of different players in generating value. This issue manifests itself here, but not in the cooperative game setting because the cooperative game setting generally views the grand coalition as forming, and so all other possible coalitional configurations are subsets. In a network setting, the efficient (or stable) networks will generally not be fully connected and thus the alternative networks include more than subnetworks.

A related issue can be seen when we examine the conditions that characterize the Myerson value. The next example from Jackson [331] shows some shortcomings of the equal bargaining power condition.[10]

Example **12.4.2** *A Criticism of Equal Bargaining Power*

Let $v(\{12\}) = v(\{23\}) = 1$ and $v(g) = 0$ for all other networks. Thus, single link networks that include player 2 result in a value of 1, and other networks result in a value of 0.

Any allocation rule, including the Myerson Value, that satisfies equal bargaining power and allocates 0 to players on the empty network will result in $Y_1(\{12\}, v) = Y_2(\{12\}, v)$, as in Figure 12.4.

While there might be situations where the value will be split evenly despite the fact that player 2 is essential for generating value but neither of the other players is, *requiring* that player 2 get the same allocation as the other player in a link is quite

[10] For a discussion of shortcomings of the component balance condition, and other examples, see Jackson [331].

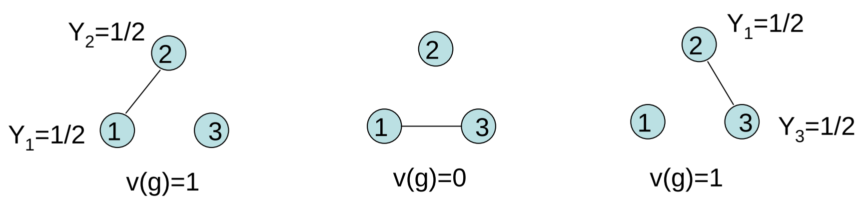

Figure 12.4. *A Critique of Equal Bargaining Power - Player 2 is required for to generate any value while other players are not, but the allocation is necessarily the same for all players under equal bargaining power.*

strong and should not be expected of a bargaining process. Indeed, this would be inconsistent with the sorts of outcomes we see in the exchange experiments of Cook and Emerson [161] as discussed in Section **??** or of Charness, Corominas-Bosch, and Frechette [135] as discussed in Section **??**. Even from a normative point of view, it is not entirely obvious that one should require that the allocation be symmetric in the dyad.

12.4.1 Defining Allocation Rules from Network Formation Possibilities

In order to account for the outside options that players and groups of players have available in constructing a network, Jackson [331] suggests the following alternative method of deriving allocation rules.

From any value function over networks we can define an associated cooperative game by[11]

$$w_v(S) = \max_{g \in G(S)} v(g).$$

The idea here is that $w_v(S)$ captures the value of a coalition S by measuring the maximal possible value that they could generate by forming a network among themselves when they are the only members of the society. This is a measurement of their threat value, or alternatively what they could generate for society without the help of any other agents.[12]

With these measures in hand, we can then allocate value based on Shapley value style calculations, now based on this auxiliary cooperative game that keeps track of the productive value of different groups of players.

So, in terms of allocating value, if g is efficient, then set

$$Y(g,v) = \phi^{SV}(w_v), \tag{12.4}$$

which is equivalently written as

$$Y_i(g,v) = \sum_{S \subset N \setminus \{i\}} (w_v(S \cup \{i\}) - w_v(S)) \left(\frac{\#S!(n - \#S - 1)!}{n!} \right).$$

[11] There is a slight abuse of notation here since v is defined on networks on N while g is a network on S, where S is a subset of N. This is easily translated to be a network on N where players outside of S are disconnected.

[12] I say "a measurement," as in cases where there are externalities across components of a network, it could be that the value generated by S depends on how other players are organized. This threat point is unambiguous if v is component additive, but is less clear otherwise.

Although this might appear to be similar to the Myerson value given its Shapley value-style calculations, it is a very different allocation rule. We see the differences immediately by noting that it gives different allocations for the two value functions given in Example 12.4.1. There it provides the same allocations under v as the Myerson value, but leads to a completely egalitarian allocation under v', while the Myerson value does not show any difference when the value function changes. It also leads to a higher allocation for player 2 in Example 12.4.2, in contrast to the Myerson value. This allocation rule violates both equal bargaining power and component balance, and is in turn characterized by conditions that are violated by the Myerson Value, as shown in Jackson [331].[13]

This way of defining allocation rules only ties down the allocation when an efficient network is chosen. There are many ways to then define the allocation on other networks. Jackson [331] suggests one possibility of simply adjusting allocations to be proportional to the allocation that would be obtained on an efficient network. That is, if some inefficient network generates 2/3 of the value of an efficient network, then each player would get 2/3 of the allocation that he or she would obtain under an efficient network. There are other ways to do this as well.[14]

12.4.2 The Core in Network Settings

Once we view the network as a flexible or changeable entity and see the associated cooperative game, then we can make more use of the cooperative game theoretic tool box. For example, there is a natural definition of the core in network settings.

A network-allocation pair $g \in G(N)$ and $y \in \mathbb{R}^n$ is in the *core* relative to (N, v) if

- $\sum_i y_i \leq v(g)$, and
- $\sum_{i \in S} y_i \geq w_v(S) = \max_{g \in G(S)} v(g)$ for all $S \subset N$.

The core includes the specification of both a network and an allocation of its value. The requirement is that no coalition could deviate, form a network on its own, and

[13] Jackson called this the player-based flexible network allocation rule. He also examined "link-based" variations on such a rule. There, one allocates value to links based on their importance in providing value, and then players get value from their links. Such an idea has roots in the communication game setting as studied by Meessen [441] and Borm, Owen, and Tijs [**?**].

[14] See Navarro [477] for further discussion and other allocation rules.

generate a higher value than what they are being allocated.[15]

So, analogously to its role in cooperative game theory, the core concept captures allocations which are stable to deviations from various groups, and can lead to allocations that differ from those derived from Shapley Value style calculations.

An allocation rule Y is *core consistent* if for any v such that the core is nonempty, there exists at least one g such that $(g, Y(g, v))$ is in the core.

While the Myerson value is not always in the core, and thus not core consistent, there are allocation rules which are core consistent. The nucleolus (of Schmeidler [**?**]) is an imputation defined on cooperative games that is core consistent relative to cooperative games. There is a natural analog of the nucleolus in a network setting, termed the networkolus by Jackson [331], that is core consistent in network settings.

Let $B(g, v) = \{y \in I\!R^n | \sum_i y_i = v(g)\}$ be the balanced allocations for g under v.

Let $e_S(y) = \sum_{i \in S} y_i - w_v(S)$ be the excess allocated to coalition S at an allocation y relative to their threat value under v, and let $e(y)$ denote the vector with entries indexed by a list of the nonempty S's, $S \subset N$.

Given an efficient g, let $Y^N(g, v) = y$ be the unique allocation such that $e(y)$ leximin dominates $e(y')$ for all $y' \in B(g, v)$.[16]

The networkolus examines how much various coalitions are getting relative to their threat values. In the case where the core is nonempty, so that there is some allocation that gives each coalition at least its threat value, then the networkolus equilibrates (to the extent possible) the excess value given to each coalition. More generally, even when the core is empty, it provides an allocation, and although in that case some of the excesses will be negative, the networkolus still minimizes the amount by which any coalition falls below its value.

[15] Again, this definition makes the most sense for component additive value functions, as otherwise the threat value generated by S in forming its network could depend on how the other players are organized. This is an issue that has resulted in various core definitions in cooperative settings, and appears here as well.

[16] A vector e leximin dominates a vector e' if there is some scalar x such that for any $x' < x$, e and e' have the same number of entries with value x', while e has fewer entries with value x. The calculation of the networkolus can be a difficult task without some underlying structure on v, but it can be shown that this is well-defined through a straightforward extension of results on the nucleolus.

12.5 Concluding Remarks

As we have seen here, tools from cooperative game theory can be adapted to provide insight into how the value of a network might be allocated among the players in a society and how this depends on the network in place and the values generated by alternative networks. T here are different perspectives that one might take, making this either a question of how value should be allocated or of how an allocation results from some process. Although we can adapt concepts from cooperative game theory to network settings, there are issues that arise in the network setting that lead to new questions regarding how value should be allocated. This is still a largely unexplored area.

12.6 Exercises

Exercise **12.1** *Convex TU Games.*

Prove Proposition 12.1.1.

Exercise **12.2** *A Convex 3-Player TU Game.*

Consider a three-player TU game where $w(\{1,2,3\}) = 2$ and $w(\{2,3\}) = 1$ while $w(S) = 0$ for all other S. Find the core allocations and Shapley value of the game.

Exercise **12.3** *The Core in an Exchange Network.*

Consider the society described in the left-hand network in Figure 10.3.1, where singletons are worthless, a coalition of any two "B" players is worth 8, a coalition of A with a B is worth 24, a coalition of three players is worth the same as the maximal value across its subsets of size 2, and the grand coalition is worth 32. Show that the unique core allocation is 4 for each B player and 20 for the A player. Find the Shapley Value for this game. Show that it differs from the core allocation.

Exercise **12.4** *Additivity of the Myerson Value.*

Let a value function v be such that $v(g) = v_1(g)+v_2(g)$ for two other value functions v_1 and v_2 and every $g \in G(N)$. Show that

$$Y^{MV}(v, g) = Y^{MV}(v_1, g) + Y^{MV}(v_2, g).$$

Exercise **12.5** *The Myerson Value in the Symmetric Connections Model*

Consider a star network comprising all players in the symmetric connections model. Find the Myerson Value allocation.

Consider a three player society and show that there exists a range of δ and c such that a star is efficient but only the complete network is pairwise stable under the Myerson value allocation rule.

Exercise **12.6** *Over-connection under the Myerson Value.*

Consider a value function v such that $v(g) = b(g) - c\sum_i d_i(g)$ where b represents benefits and $c > 0$ is a cost of maintaining a link.

$b(g)$ is *monotone* if

- $b(g') \leq b(g)$ if $g' \subset g$, and
- $b(\{ij\}) > 0$ for any ij.

The following is a special case of a result from Jackson [329].

Proposition **12.6.1** *Let $n \geq 4$, and consider an anonymous and monotone benefit function b for which there is some efficient network g^* relative to b which is symmetric and not the complete network. There exists $\overline{c} > 0$ such that for any $c < \overline{c}$, any pairwise stable network relative to Y^{MV} and the value function $v(g) = b(g) - c\sum_i d_i(g)$ is Pareto dominated by some subnetwork.*

Prove Proposition 12.6.1.

Hint: first show that if $ij \notin g$ then $Y_i(g+ij, b) - Y_i(g, b) \geq b(\{ij\}) > 0$, and conclude that the complete network is the unique pairwise stable network under b. Then apply the additivity of the Myerson Value from Exercise 12.4 and work with small costs.

Exercise **12.7** *Possible Non-Existence of Strongly Stable Networks under the Component-Wise Egalitarian Rule.*

Show that there exists a component additive value function for which there is no strongly stable network under the component-wise egalitarian rule. Show that if we weaken strong stability to say that a network is stable if there is no deviation by a coalition of agents that is *strictly* improving for *all* members of the deviating coalition, then there will exist a strongly stable network under the component-wise egalitarian rule for any v.

EXERCISE **12.8** *The Existence of Strongly Stable Networks under the Component-wise Egalitarian Allocation Rule.**

The following condition and result are from Jackson and van den Nouweland [338].

A value function is *top-convex* if $\max_{g\in G(N)} \frac{v(g)}{|N|} \geq \max_{g\in G(S)} \frac{v(g)}{|S|}$ for all $S \subset N$.

Show that if the value function is component additive, then under this condition, the per capita value of each component of an efficient network is equal, and is at least as high as the per capita value of any component of any network.

Suppose that payoffs are governed by the component-wise egalitarian allocation rule and consider an anonymous and component additive value function. Show that the set of strongly stable networks is nonempty if and only if the value function is top convex. Moreover, show that in that case, the strongly stable networks are the efficient ones.

EXERCISE **12.9** *The Component-Wise Egalitarian Allocation Rule and Bridges.*

A pair of a network and a component additive value function, (g, v), is *bridge-monotonic* if,

$$v(C_i(g))/\#C_i(g) \geq \max\left[v(g^1)/\#N(g^1), v(g^2)/\#N(g^2)\right]$$

for every bridge ij in g such that $v(g) \geq v(g - ij)$, where $\#C_i(g)$ is the number of players in i's component of g, and g^1 and g^2 are the components of g bridged by ij.

Prove the following proposition.

PROPOSITION **12.6.2** *[Jackson and Wolinsky [343]] If g is efficient relative to a component additive v, then g is pairwise stable for Y^{ce} relative to v if and only if (g, v) is bridge-monotonic.*

EXERCISE **12.10** *The Shapley Value in the Connections Model.*

Consider the symmetric connections model with parameters such that a star is the unique efficient network structure. Let v be defined by $v(g) = \sum_i u_i(g)$, where u_i is from the symmetric connections model. Compare the Myerson value allocation for a star to the Shapley Value of w_v, as defined in (12.4), for a star.

EXERCISE **12.11** *Anonymity and the Shapley Value in Network Settings.*

Let v be an anonymous value function. Show that for any efficient network, the Shapley Value of w_v, as defined in (12.4), results in the same allocation as the egalitarian allocation rule.

Exercise **12.12** *The Monotonic Cover of a Value Function*

Given a value function v, Jackson [331] defines its *monotonic cover* $\widehat{v}$ by

$$\widehat{v}(g) = \max_{g' \subset g} v(g').$$

Consider an efficient g and a component additive v. Define an allocation rule Y so that the allocation at g, v is described by

$$Y_i(g, v) = \sum_{S \subset N \setminus \{i\}} (\widehat{v}(g^{S \cup i}) - \widehat{v}(g^S)) \left(\frac{\#S!(n - \#S - 1)!}{n!} \right).$$

where g^S is the complete network on the nodes S (viewed as network on N). Show that this is the same allocation as the Shapley Value of w_v, as defined in (12.4).

Exercise **12.13** *The Core in Example 12.4.1*

Determine the core networks and allocations (as defined in Section 12.4.2) under the value functions in Example 12.4.1.

Exercise **12.14** *The Networkolus and the Core.*

Consider a value function v for a three player society such that $v(\{12\}) = v(\{23\}) = 1$, $v(\{12, 23\}) = w$, and $v(g) = 0$ for all other networks, where $w > 0$.

Find the Myerson value, the player-based flexible network allocation, the networkolus, and all of the core allocations.