Exclusion and cooperation in networks

Ule, A.

Citation for published version (APA):
Chapter 4

Cooperation in finitely repeated network dilemma games

Modeling an interactive situation as a non-cooperative game is perhaps the most common approach in game theory. The standard assumption is that players are rational and have a perfect foresight. If they are not perfectly informed about the situation or about the past behavior of other players they still know all the relevant chances. In addition, preferences and information of each player are common knowledge to all of them. These assumptions facilitate a relatively straightforward analysis that often provides simple and exact results. The conventional task is to find the equilibria of the game. An equilibrium is any collection of actions, one for each player, such that it is in each player’s self interest to play her action if all other players play theirs.

A well known example is that of a finitely repeated prisoner’s dilemma game which, with rational and perfectly informed players, has a unique equilibrium. In this equilibrium all players mutually defect in all periods even though they know that mutual cooperation would have been better for each of them. In this chapter we contrast this result by showing that cooperation among such players may yet be achieved when they can choose their partners endogenously. We do this by exploring the subgame-perfect equilibria of finitely repeated network dilemma games. Our focus is on the effect of linking constraints and linking costs for cooperation. Our intuition is that competition for partners, induced by scarcity or costs of links, may induce rational players to cooperate. This is confirmed by our results.

In the following section we motivate our approach, outline the results and discuss our contribution in view of the related literature. In section 2 we introduce and formally define the network dilemma games with linking constraints. The analysis of these games is separated into two sections: section 3 deals with mutual link formation and section 4 with unilateral link formation. In section 5
we formally define the network dilemma games with linking costs and analyze them for both models of link formation. A discussion of our results is given in section 6, where we also suggest some topics for future investigation. We summarize the chapter in section 7.

4.1 Introduction

The prisoner’s dilemma game is a paradigm used to study the failure to achieve cooperation in groups of rational and informed players. The finitely repeated prisoner’s dilemma, the multi-player prisoner’s dilemma and the prisoner’s dilemma played on a fixed network are examples of games where the unique equilibrium consists of strategies that prescribe defection in each of the periods even though mutual cooperation benefits each of the players. Cooperative equilibria exist only in infinitely repeated prisoner’s dilemma games or in those with random ending (see the Folk theorem and its extensions by Friedman, 1971, Aumann and Shapley, 1994, and Fudenberg and Maskin, 1986).

To achieve cooperation in the finitely repeated prisoner’s dilemma the assumptions about players and their environment must be relaxed (see Kreps et al., 1982 for relaxation about information and Neyman (1985) for relaxations about rationality). In this chapter we investigate whether cooperation can be achieved in equilibrium under standard assumptions, if we change the game by permitting the players to choose their partners.

We consider network dilemma games with either mutual or unilateral link formation. Traditional equilibrium concepts from non-cooperative game theory, such as Nash equilibrium and subgame-perfection, can straightforwardly be applied to network formation games with unilateral linking (see Bala and Goyal, 2000). However, these concepts may be too weak for network dilemma games with mutual linking. In particular, in such games a large number of unintuitive Nash equilibria exist simply because of mis-coordination where pairs of players want to establish a link but neither proposes it, because neither expects the other player to do so. In our investigation of repeated network dilemma games with mutual linking this multiplicity of Nash equilibria further translates into the existence of unintuitive subgame-perfect equilibria. To avoid the problem of multiplicity of equilibria in network formation games the requirement of Nash equilibrium is usually supplemented with that of network stability. This approach has been formalized by the concept of a pairwise-stable equilibrium (Goyal and Joshi, 2003) and its variants (Calvó-Armengol, 2004, and Goyal and Vega-Redondo, 2004). However, these equilibrium concepts are defined only for network formation games with exogenous linking costs and benefits. They cannot be applied to network dilemma games, where benefits of links depend on actions taken by players in the prisoner’s dilemma. In this chapter we define the concept of a linking-proof equilibrium which may be seen as an extension of the concept of pairwise-stable equilibrium to network formation games with endogenous benefits. For repeated games with mutual linking we define a related concept of a linking-proof subgame-perfect equilibrium.
The analysis in this chapter is in two parts. In the first part we investigate network dilemma games with linking constraints. In our network dilemma games, as defined in the previous chapter, players are free to establish any number of links. In many real situations, however, agents are strictly constrained in the maximal number of connections they can maintain. One such example is the human social network: people select only a few of the other people with whom they keep regular contacts. The number of contacts is obviously negligible in comparison to the number of all other people but typically, and more crucially, it is small also compared to the number of people in some reference group, such as school class or job environment.

Constraining the number of links that players can propose or establish has a dramatic effect for cooperation in network dilemma games with low outside option. We say that a player's linking constraint is strict when she cannot simultaneously establish links to all other players. In this chapter we show that cooperation can be achieved in a (linking-proof) subgame-perfect equilibrium as long as at least two players have a strict linking constraint. This condition is relatively weak and not satisfied only in a few boundary cases, such as by the original network dilemma game without any linking constraints. We proceed to identify network dilemma games in which cooperation can be achieved using only simple and intuitive trigger strategies that threaten to punish any deviation with exclusion. Such a strategy profile exists only when there is a sufficient variety of (linking-proof) equilibrium networks, so that the group can effectively punish any player by removing some or all of her established links and still structure into an equilibrium network. We show that this is possible if linking constraints are relatively severe, that is, if each player's upper bound on the number of her neighbors is relatively small. The human social network is again a suitable example.

The intuition behind our results is as follows. When linking is strictly constrained and the outside option is low the players can structure their interaction into several different equilibrium networks. Because the outside option is low, each player prefers to establish as many links as possible. In some equilibrium networks all players establish their maximal number of links, but in others some players establish less or even no links. We could say that some equilibrium networks are efficient and others are inefficient. Naturally, in the final period of a finitely repeated network dilemma game all rational players defect. However, because their interaction in the final period can be structured into a variety of networks they can condition the structure on the actions taken in the previous periods. To see this, suppose that players begin by cooperating in an efficient network. If one of them defects the remaining players can punish her by structuring their final period interactions into an inefficient equilibrium network in which this player establishes relatively little or no links, earning only the low outside option. This player is thus punished by exclusion. When trigger strategies fail to sustain cooperation we may still use, slightly less simple, recursive trigger strategies. For this it is sufficient that at least one player can be excluded from a connected equilibrium network. This, in turn, is possible with almost any linking constraints.
The second part of the chapter focuses on network dilemma games with linking costs. Consider players endowed with fixed amounts of available time. They can distribute this time between their social interactions and some alternative beneficial activities. The opportunity cost of each interaction is the value of the most valuable of the foregone activities. A rational player decreases this cost by foregoing only the least valuable of the activities. However, the opportunity cost increases with the number of established interactions and the rational player eventually stops adding new interactions.

This motivates the following assumptions behind our network dilemma games with linking costs. We say that a player sponsors a link if she proposes it and it is established. Linking is not constrained but sponsored links are costly. For each player the marginal cost of a link increases with the number of sponsored links, so that each additional link is costlier than the previous one. The resulting cost function is then convex. As in other network dilemma games, the benefits of links are determined by the actions in the prisoner’s dilemma. The analysis of such games is facilitated by the analysis of network dilemma games with linking constraints. We characterize the following sufficient condition for cooperation in a (linking-proof) subgame-perfect equilibrium: each player must be willing to sponsor a few but not all links even if all players defect. Obviously, if all players cooperate then the benefit of each link is larger and more links may be sponsored. Hence, the networks among cooperators may be denser than those among defectors. We demonstrate this in the concluding example, where a subgame-perfect equilibrium of a two-period network dilemma game with linking costs is such that in the first period the complete network is established and all players cooperate, and in the final period a wheel network is established and all players defect.

Our results hold both for mutual and unilateral link formation models. This is interesting in view of the considerable differences between the two models. For example, if linking is mutual then any link can be removed unilaterally. A player is induced to cooperate through threats of exclusion. On the other hand, if linking is unilateral then punishment by removing a link is not always possible. A player in this model cooperates because this attracts free links, sponsored by the other players.

We can relate our results to the existing literature about subgame-perfect equilibria in finitely repeated games. Friedman (1985) and Benoit and Krishna (1985) have proven a limit Folk theorem for finitely repeated games when the set of static equilibrium payoffs has sufficient dimensionality. Within network dilemma games, a large variety of equilibrium networks implies high dimensionality of static equilibrium payoffs. It follows from Friedman’s result that this may be sufficient to sustain non-equilibrium outcomes, such as cooperation, in a subgame-perfect equilibrium. Furthermore, Smith (1995) has proven the limit Folk theorem for finitely repeated games under relaxed conditions on dimensionality of static equilibrium payoffs. Our characterization of when the recursive trigger strategies can sustain cooperation in network dilemma games is related to his condition of recursively distinct Nash payoffs. For a review of related results in a unified framework see Benoit and Krishna (2000).
We contribute to this literature by highlighting the possibility that the dimensionality of equilibrium payoffs can be increased by endogenizing the interaction structure. When the interaction structure is fixed then in some games, such as the prisoner's dilemma, only the static equilibrium outcomes can be sustained in a subgame-perfect equilibrium. We demonstrate that this may change if the interaction network is endogenous: outcomes other than static equilibria may be achieved in a subgame-perfect equilibrium of the network dilemma game under any of the two models of link formation. Furthermore, given that the concept of a subgame-perfect equilibrium may be too weak for games with mutual linking we show that, when linking is constrained, the above results hold also under the additional requirement of linking-proofness.

Our study is one of relatively few that consider situations where agents form the network and also determine its value through behavior in the network interactions. Droste et al. (2000), Jackson and Watts (2002), and Goyal and Vega-Redondo (2005) study games of coordination in endogenous networks, assuming either unilateral or mutual link formation. These games may be viewed as equivalent to our network dilemma games for situations where interaction over the network has a character of a coordination game instead of the prisoner's dilemma game. All three papers consider an adaptive dynamic process with myopic best-responding players, and study the long-run stability of different conventions in relation to the linking costs. Their somewhat surprising main conclusion is that high linking costs imply high earnings and efficiency.

Evolutionary dynamics is the preferred approach also in studies of social dilemmas on endogenous interaction structures. These studies assume that behavior spreads via imitation rather than the best response, ostensibly because defection is the unique best response in social dilemma situations. We review this literature in chapter 6, where we also show that cooperation can be sustained in network dilemma games even with myopic best-responding players, if they use limited forward looking. An equilibrium analysis of an infinitely repeated prisoner's dilemma game, played on an endogenous network by rational players with perfect foresight, is given in Galeotti and Meléndez-Jiménez (2004). In this chapter we consider finitely repeated network dilemma games and analyze whether cooperation can be sustained in a subgame-perfect equilibrium through strategic linking behavior.

We are aware of only two other studies that study strategic behavior in similar situations. Hirshleifer and Rasmusen (1989) study a finitely repeated n-player prisoner's dilemma game with the possibility of ostracism. Each period of the prisoner's dilemma is followed by a vote whether to ostracize any players from the group. Hirshleifer and Rasmusen show that cooperation can be achieved in the initial periods of a subgame-perfect equilibrium if ostracism in the last period of the game is costless for ostracizers and costly for the ostracized. They do not deal with individual exclusion but assume that each individual conforms to the outcome of the vote.

The other related result is about finitely repeated multiple prisoner's dilemma games with (high) outside option, analyzed in Hauk and Nagel (2001). Hauk and Nagel show that cooperation can be achieved in a Nash equilibrium of the
repeated game if link formation is unilateral, but not if link formation is mutual. They stop short of showing that cooperation can be achieved also in a subgame-perfect equilibrium of one of the repeated games. This can be deduced from results in our section 4.4.

This is how to read this chapter. For each class of games we first characterize the equilibrium networks of one-shot games and then the subgame-perfect equilibria of the repeated games. A number of proofs use lengthy combinatorics, mostly to deal with the existence of networks having specific properties. These proofs are not essential and we gather them in the appendix to the chapter. The proofs that carry important arguments are placed in the main sections. To support the intuition we supplement formal analysis with a sequence of examples. Numerical examples are often escorted also by graphical illustrations. A quick, but not comprehensive, overview of our results can be acquired by reading through descriptions of the models and following our analysis through the examples.

4.2 Network dilemma game with linking constraints

In the basic game players are free to propose and establish any number of links to any subset of other players. In this chapter we consider the games in which players may be constrained in the number of links they can propose.

Let $k_i$ be the integer denoting the maximal number of links that player $i$ can propose. We refer to $k = (k_1, \ldots, k_n)$ as the vector of linking constraints of players $N = \{1, \ldots, n\}$. Proposing and establishing links is costless. A player $i$ can propose a link to any other player in the group and can propose up to $k_i$ links.

Each player also chooses an action in a prisoner's dilemma game. Players cannot discriminate in their actions. One game is played over each established link. For simplicity (and with no loss of generality) we assume that the value of the outside option is $o = 0$. Hence, following classification from chapter 3 we say that outside option is low if $d > 0$, and high if $d < 0$.

Let $A_i = \{C, D\}$ be the set of actions and let $P_i(k_i) = \{p_i \in \{0, 1\}^n \mid p_{ii} = 0 \text{ and } \sum_{j \in N} p_{ij} \leq k_i\}$ be the set of linking choices of player $i$. The set of moves of player $i$ is denoted by $J_i(k_i) = A_i \times P_i(k_i)$. Let $J(k) = \times_{i \in N} J_i(k_i)$ and let $J_{-i}(k) = \times_{j \in N \setminus \{i\}} J_j(k_j)$ be the set of moves of player $i$'s opponents. When we refer to a game with linking constraints, move refers to a pair $(a_i, p_i) \in J_i(k_i)$ and profile of moves

---

1In section 4.6.2 we show that all results of this chapter apply also if players are free to choose different actions in interactions with different neighbours, as long as information about past action choices is complete.
refers to a profile \((a, p) \in J(k)\). For each profile of moves \((a, p)\) the payoff to player \(i\) is given by

\[
\pi_i(a, p) = \sum_{j \in L_i(p)} v(a_i, a_j) = \sum_{j \in N} v(a_i, a_j)g_{ij}(p).
\] (4.1)

In line with (3.2), we can also define the payoff to player \(i\), given network \(g\) and a profile of actions \(a\), by

\[
\pi_i(a, g) = \sum_{j \in L_i(g)} v(a_i, a_j) = \sum_{j \in N} v(a_i, a_j)g_{ij}.
\] (4.2)

We refer to the stage game \(T(k) = (N, J(k), \pi)\) as the network dilemma game with linking constraints, or shortly, the dilemma game.

We do not consider the possibility that players may randomize in their choices. Randomization can be analytically useful, e.g. to smooth the response functions, to make the action sets convex and, consequently, to ensure the existence of Nash equilibria. In our view there is no need for this in the analysis of network dilemma games, because at least one Nash equilibrium in pure strategies always exists. Including equilibria in mixed strategies would add little to our results but significantly complicate the notation and analysis. Furthermore, in repeated games the mixed strategies are difficult to enforce because deviations from mixed strategies cannot be detected. This is usually resolved either by assuming that each randomization process itself can be observed or that players condition their moves on the outcome of some public randomization mechanism. However, for the games in this chapter it seems more intuitive to assume that only pure strategies are used.

Move \((a_i^*, p_i^*) \in J_i(k)\) is a best response of player \(i\) to the profile of moves \((a_{-i}, p_{-i}) \in J_{-i}(k)\) if

\[
\pi_i(a_i^*, p_i^*, a_{-i}, p_{-i}) \geq \pi_i(a_i', p_i', a_{-i}, p_{-i})
\]

for any \((a_i', p_i') \in J_i(k)\). A Nash equilibrium of game \(\Gamma(k)\) is a profile of moves \((a^*, p^*) \in J(k)\) such that for each player \(i\),

\[
\pi_i(a^*, p^*) \geq \pi_i(a_i, p_i, a_{-i}^*, p_{-i}^*)
\] (4.3)

for any \((a_i, p_i) \in J_i(k)\). Network \(g\) is an equilibrium network if there exists a Nash equilibrium \((a^*, p^*)\) such that \(g = g(p^*)\).

The game obtained when a network dilemma game with linking constraints \(\Gamma(k)\) is repeated \(T\)-times is denoted by \(\Gamma_T(k) = (N, J(k), \pi, T)\). Recall that when discussing a repeated game \(\Gamma_T\) the constituent game \(\Gamma\) is the stage game and the equilibria of \(\Gamma\) are the static equilibria. We refer to the repeated game \(\Gamma_T(k)\) as the finitely repeated dilemma game with linking constraints or, shortly, the repeated dilemma game.

The profile of moves at \(t\) is denoted by \((a^t, p^t)\). The action of player \(i\) in period \(t\) is \(a_i^t\) and her linking choice at \(t\) is \(p_i^t\). The history at the end of time \(t\) is
the sequence of moves \( h^t = ((a^1, p^1), ..., (a^T, p^T)) \). The total payoff to player \( i \) at the end of the repeated dilemma game is determined by the terminal history \( h^T \) and given by

\[
\Pi_i(h^T) = \sum_{t=1}^{T} \pi_i(a^t, p^t). 
\]

A strategy for player \( i \) in the game \( \Gamma^T(k) \) is a function \( \sigma_i \) which selects, for any history of play, an element of \( J_i(k) \). We say that a strategy profile is semi-cooperative if at least one semi-cooperative relation (see section 3.1) is established in an outcome of at least one period. Otherwise, we say it is defective. We say that a strategy profile is cooperative if, in an outcome of at least one period, (i) each player has at least one neighbor and (ii) all players cooperate.

The following proposition describes some obvious properties of subgame-perfect equilibria of network dilemma games. It states that in any such equilibrium all players defect in the last period of the game. Therefore, a subgame-perfect equilibrium is cooperative only if there is an early period in which all players cooperate. The cooperative subgame-perfect equilibria that we describe in this chapter will be such that (i) all players cooperate during all initial periods, (ii) then sequentially or simultaneously turn to defection, until (iii) all players defect during one or more final periods. As this proposition shows, if such an equilibrium is possible for a network dilemma game repeated \( T \)-times, then one exists also when the game is repeated more than \( T \)-times.

**Proposition 4.1** If a cooperative subgame perfect equilibrium exists for \( \Gamma^{T'}(k) \) then there exists an integer \( \gamma \leq T' \) such that a cooperative subgame perfect equilibrium for \( \Gamma^T(k) \) exists if and only if \( T \geq \gamma \). In any subgame perfect equilibrium, (i) at least one relation is defective in each of the final \( \gamma - 1 \) periods, and (ii) all relations are defective in the final period.

**Proof.** Assume that a cooperative subgame perfect equilibrium (CSPE) exists for \( \Gamma^T(k) \). Obviously, there exists a smallest \( \gamma \) such that a CSPE exists for \( \Gamma^\gamma(k) \). In any one-shot dilemma game all relations are defective, hence \( \gamma \geq 2 \). For any \( T \) : if a CSPE \( \sigma \) exists for \( \Gamma^T(k) \) then one exists also for \( \Gamma^{T+1}(k) \); for example, ”play \( \sigma \) in the first \( T \) periods and play a static equilibrium in period \( T + 1 \).”

A subgame perfect equilibrium (SPE) of \( \Gamma^T(k) \) must select a SPE for any subgame \( \Gamma^t(k), t < T \) and any history. No CSPE exist for \( \Gamma^t(k) \) if \( t < \gamma \). All relations are defective in any Nash equilibrium of \( \Gamma(k) \). Hence, an equilibrium outcome path for the \( \Gamma^T(k) \) must be such that all relations are defective in the last period, and in none of the last \( \gamma - 1 \) periods are all relations cooperative.
4.3 Mutual link formation

In this section we consider the mutual link formation model. For a profile of linking choices \( p \) the network of established links \( g(p) \) is defined by \( g_{ij}(p) = \min\{p_{ij}, p_{ji}\} \).

For each network \( g \) and each player \( i \) let \( p_i(g) \) be the binary vector defined by \( p_{ij}(g) = g_{ij} \) for all \( j \). We say that player \( i \) supports network \( g \) when her linking choice is \( p_i(g) \). The profile \( p(g) = (p_1(g), \ldots, p_n(g)) \) is the minimal support for the network \( g \), in the sense that each link proposal in \( p(g) \) is necessary to establish \( g \).

To avoid trivialities we assume that \( c > 0 \), that is, the value of mutual cooperation is higher than the outside option. We also assume that \( 1 \leq k_i \leq n - 1 \) for each player \( i \).\(^2\) Let

\[ \mathcal{G}(k) = \{ g \mid l_i(g) \leq k_i \text{ for each } i \} \]

be the set of feasible networks. For reference we restate the Prisoner’s dilemma payoff matrix:

<table>
<thead>
<tr>
<th></th>
<th>( C )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C )</td>
<td>( c, c )</td>
<td>( e, f )</td>
</tr>
<tr>
<td>( D )</td>
<td>( f, e )</td>
<td>( d, d )</td>
</tr>
</tbody>
</table>

4.3.1 Stage game equilibria and linking-proof networks

The complete characterization of Nash equilibria for dilemma games is possible but cumbersome. The following proposition characterizes the actions and the networks that are established in a Nash equilibrium of a one-shot dilemma game. The set of equilibrium networks depends only on whether the outside option is low or high.

**Proposition 4.2** Consider the dilemma game \( \Gamma(k) \).

1. A player cooperates in a Nash equilibrium only if no links to her are proposed. Hence, in a Nash equilibrium all cooperative players are isolated.

2. If \( d > 0 \) then \((D, p(g))\) is a Nash equilibrium for any network \( g \in \mathcal{G}(k) \).

3. If \( d < 0 \) then the empty network is the unique equilibrium network.

\(^2\)Obviously, in a group with \( n \) players each of them can establish at most \( n - 1 \) links, hence it makes sense to assume that \( k_i \leq n - 1 \) for all \( i \). On the other hand, if a player cannot make any links (\( k_i = 0 \)) she will be isolated regardless of the linking behavior of other players. An isolated player does not affect the dynamics of play. We therefore assume \( k_i \geq 1 \) for all \( i \).
Proof. (1) Let \((a^*, p^*)\) be a Nash equilibrium (NE). There cannot be a pair of players \(i, j\) such that \(a_i^* = C\) and \(p_{ij}^* = 1\) as in this case \(i\) could strictly increase her payoff by switching to \(a_i = D\) and \(p_{ij} = 1\). There also cannot be a pair of players \(i, j\) such that \(a_j^* = C\) and \(p_{ij}^* = 1\) and \(p_{ji}^* = 0\) as in this case \(j\) could strictly increase her payoff choosing some \(p_j\) such that \(p_{ji} = 1\).

(2) For any \(g \in \mathcal{G}(k)\) the \(p(g)\) is the minimal support for \(g\) in the sense that all proposed links are reciprocated. Hence, no player can increase her payoff by unilaterally proposing additional links and can only (weakly) decrease her payoff if she removes some links. Obviously, no player can increase her payoff by switching to cooperation. The profile \((D, p(g))\) is therefore a NE.

(3) According to (1), if in NE players \(i\) and \(j\) are neighbors they should both defect. However, each could then increase her payoff by removing the link. Hence, in a NE there are no neighbors.

Statement (1) asserts that in a Nash equilibrium of a dilemma game all relations are defective. That is, in a Nash equilibrium players with at least one established link defect but the isolated players may cooperate. Consequently and as stated in (3), if the mutual defection payoff \(d\) is negative no links are established in a Nash equilibrium. For non-negative \(d\), however, statement (2) asserts that any feasible network is an equilibrium network. The dilemma game with \(d = 0\) is rather trivial. In the rest of the section we consider only games with \(d \neq 0\).

Consider a dilemma game with \(d > 0\). If all players defect each of them maximizes her payoff by establishing as many links as possible. It may thus be surprising that networks with relatively little or no links can be established in a Nash equilibrium. The multiplicity of equilibrium networks is a consequence of the mutual link formation assumption. A link between two players is established only if both propose it. If one of the players does not propose the link the other is indifferent between proposing the link or not. It may therefore happen that two players prefer to establish a mutual link but none proposes it because each knows that the other will not propose the link. Hence, in a Nash equilibrium there may exist a pair of players such that (i) each could beneficially establish at least one more link but (ii) none proposes a link to the other player. This appears especially unintuitive given that proposing links is costless (see e.g. the discussion in Dutta et al., 1998).

Several resolutions have been discussed in the related literature on network formation (see Goyal and Joshi, 2003, Calvó-Armengol, 2004, and Gilles and Sarangi, 2004). We follow the common approach in this literature and supplement the idea of the Nash equilibrium with a requirement of network stability. Roughly speaking, an equilibrium network is said to be stable if no pair of separated players exists such that, by adding the mutual link, each could strictly increase her payoff.

Our definition of network stability is inspired by the concept of pairwise-stability, introduced by Jackson and Wolinsky (1996). It is defined for network formation games with mutual linking, but with exogenously fixed cost and benefit rules. While it retains some flavor of the non-cooperative equilibrium,
pairwise-stability is analyzed with the tools of cooperative game theory.\(^3\) It is not immediately obvious how to appropriately generalize pairwise-stability to network formation games with endogenous benefit structure, and how to analyze such games with the tools of non-cooperative game theory. Our definition of linking-proofness below is chosen because it eliminates the unintuitive Nash equilibria in the dilemma games.

In a dilemma game the value of a link depends on the actions taken by the linked players. The dilemma game in this section is then a network formation game with mutual linking, no linking costs, and an endogenous benefit structure. The players are free to deviate by simultaneously changing their action, removing links, and adding a new link.

A profile of moves is a Nash equilibrium if no player wants to deviate given the moves of the other players. We say that an equilibrium profile of moves is linking-proof if no player wants to deviate even if there was a possibility to establish a new link. In other words, the only multilateral deviation considered is that of adding a new link. However, a player is free to deviate by adding a new link in parallel to any other unilateral changes to her move.

For a profile of linking choices \(p\) let \(p \oplus ij\) be another profile defined as

\[
(p \oplus ij)_{v'j'} = \begin{cases} 
1 & \text{if } i'j' = ij \\
\nu_{i'j'} & \text{otherwise.}
\end{cases}
\]

In words, \(p \oplus ij\) is a profile of linking choices obtained from \(p\) by assuming that player \(i\) proposes a link to \(j\) in addition to all the links proposed according to \(p\). If \(p_{ij} = 1\) then \(p \oplus ij = p\).

**Definition 4.1** The profile of moves \((a^*, p^*) \in J(k)\) is a linking-proof equilibrium (LP equilibrium) of game \(\Gamma(k)\) if (i) for each player \(i\) her move \((a^*_i, p^*_i)\) is a best response to \((a^*, p^*)\), and (ii) for each pair of players \(i, j\) either \((a^*_i, p^*_j)\) is a best response to \((a^*_i, p^* \oplus ji)\) or \((a^*_j, p^*_j)\) is a best response to

\(^3\)Below we give a slightly simplified definition of pairwise-stability (for the general definition and discussion see Jackson, 2004). The network formation game in the cooperative form is characterized by the set of feasible networks \(\mathcal{G}\) and by the allocation rule \(Y : \mathcal{G} \rightarrow \mathbb{R}^n\) which describes how the value generated by the network is shared among the players. The network \(g \in \mathcal{G}\) is said to be pairwise stable with respect to the allocation rule \(Y\) if

(i) for all \(ij \in g\), \(Y_i(g) \geq Y_i(g \oplus ij)\) and \(Y_j(g) \geq Y_j(g \oplus ij)\), and

(ii) for all \(ij \notin g\) such that \((g \oplus ij) \in \mathcal{G}\), \(Y_i(g \oplus ij) > Y_i(g)\) implies \(Y_j(g \oplus ij) < Y_j(g)\).

It is implicit in the definition of pairwise stability that the termination of a link can be done unilaterally but that the addition of a link requires mutual consent of both involved players. A network is pairwise-stable if no player can profit by unilaterally removing a link and no pair of players can simultaneously profitably deviate by adding a mutual link.

A few extensions of the original definition of pairwise-stability exist. For example, one may assume that adding a link is plausible if it makes both linked players strictly better off, or alternatively, if it makes both linked players weakly better off, or yet some intermediate concept (see Dutta and Mutuswami, 1997). The original definition does not consider deviations of one player by simultaneous removal of several links or by simultaneous removal of one link and addition of another link, which would bring the concept closer to the spirit of a non-cooperative equilibrium (see Calvó-Armengol, 2004, and Gilles and Sarangi, 2004).
Network $g^* \in G(k)$ is linking-proof if it is established in some linking-proof equilibrium.

Any linking-proof equilibrium is Nash equilibrium, due to (i). Furthermore, in a linking-proof equilibrium there is no pair of separated players such that each of them strictly prefers to change her move and establish the mutual link. In other words, in each pair of separated players at least one of them has no move that strictly improves her payoff and establishes the mutual link.

Supplementing an equilibrium concept with that of network stability is not the only route to study strategic form games of network formation with mutual linking. For example, a stronger equilibrium concept, such as coalition-proof Nash equilibrium (Bernheim et al., 1987) or Strong Nash equilibrium (Aumann, 1959), may also eliminate unintuitive Nash equilibria, as shown in Dutta and Mutuswami (1997), Dutta et al. (1998), Slikker and van den Nouweland (2001) and Jackson and van der Nouweland (2002). The problem with these equilibrium refinements is that they eliminate too many networks because they require stability against deviations by variously sized coalitions of players. Our concept of linking-proof equilibrium is much weaker: it supplements the Nash equilibrium only with the requirement that a pair of players is linked whenever it is possible and in their mutual interest, but does not permit any other coordinated actions.

We denote the set of linking-proof networks of the game $\Gamma(k)$ by $S(\Gamma(k))$. The following proposition asserts that, when $d > 0$, the set $S(\Gamma(k))$ contains all networks in which only the linking constraints prohibit any pair of separated players from adding a mutual link.

**Proposition 4.3** Consider dilemma game $\Gamma(k)$.

1. Let $d > 0$. Network $g^* \in G(k)$ is a linking-proof network if and only if there is no pair of separated players $i, j$ such that $l_i(g^*) < k_i$ and $l_j(g^*) < k_j$.

2. If $d < 0$ the empty network is the unique linking-proof network.

**Proof.** (1.a) Let $d > 0$. Consider $g^* \in S(\Gamma(k))$ and let $(a^*, p^*)$ be the LP equilibrium such that $g(p^*) = g^*$. Assume (A1): there is a pair of distinct players $i$ and $j$ such that $l_i(g^*) < k_i$ and $l_j(g^*) < k_j$.

Let $p^*_{ij} = 0$. Define $p'_i$ as follows: (a) if there is $j'$ such that $p^*_i \neq 1$ and $g^*_{i,j'} = 0$ then let $p'_i$ coincide with $p^*_i$ in all values except for setting $p'_{ij} = 1$ and $p'_{i,j'} = 0$ (if $i$ proposed a link which was not reciprocated propose instead the link to $j$).

(b) if there is no such $j'$ then let $p'_i$ coincide with $p^*_i$ in all values except for setting $p'_{ij} = 1$ (if $i$ did not propose the maximal number of links, hence propose also the link to $j$). Define $p'_j$ in a symmetric manner. As $p^* \in P(k)$ then $p'_i \in P_i(k_i)$ and $p'_j \in P_j(k_j)$.

Because $g^*_{ij} = 0$ it must be that either $p^*_{ij} = 0$ or $p^*_{ji} = 0$ or both. If $p^*_{ij} = 1$ then for player $i$, $(D, p'_i)$ is a better response to $(a^*, p^*)$ than is $(a^*_i, p^*_i)$, as thus the link $ij$ is established in addition to all other links of player $i$ and this
increases the payoff to player $i$ for at least $d > 0$. But this should not be possible in a LP equilibrium, hence $p_{ij}^* > 0$. In this case, however, $(D, p_i')$ is a better response to $(a^*, p^* \oplus ji)$ than is $(a_i^*, p_i^*)$.

Similarly we can show that $p_{ij}^* = 0$, in which case $(D, p'_i)$ is a better response to $(a^*, p^* \oplus ij)$ than is $(a^*_j, p_i^*)$. This, however, is again not possible in a LP equilibrium. We thus proved that (A1) cannot hold.

(1.b) Let $d > 0$. Consider network $g \in \mathcal{G}(k)$ such that $g_{ij} = 1$ for each pair of distinct players $i, j$ such that $l_i(g) < k_i$ and $l_j(g) < k_j$. We argue that $(D, p(g))$ must be a LP equilibrium. By Proposition 4.2 it is a Nash equilibrium which proves part (i) of the definition of LP equilibrium. We show next that part (ii) is also fulfilled.

Take a pair of distinct players $i$ and $j$. If $g_{ij} = 1$ then (ii) follows from (i). Assume therefore $g_{ij} = 0$, implying that either $l_i(g) = k_i$ or $l_j(g) = k_j$. W.l.o.g. let $l_i(g) = k_i$. Player $i$ cannot increase her payoff by removing or by moving some of her links. She also cannot add any new links. Hence $(D, p_i)$ is a best response to $(D, p \oplus ji)$, and (ii) is satisfied.

(2) Let $d < 0$. By Proposition 4.2 the empty network is the only equilibrium network and thus the only candidate to be linking-proof. Let $0_i$ be the vector denoting that player $i$ proposes no links, and let $0 = (0_1, \ldots, 0_n)$ be the corresponding profile. To see that $(D, 0)$ is a LP equilibrium, note that for each $i$, $(D, 0_i)$ is best response both to $(D, 0)$ and to $(D, 0 \oplus ij)$ for any $j \neq i$.

Let $d > 0$. If a pair of players in a linking-proof network are separated it must be that at least one of them has established her maximal number of links. The restriction to linking-proof equilibria may considerably reduce the set of networks established by an equilibrium profile and lead to existence problems. However, proposition 4.4 below states that a linking-proof network always exists. When only the empty network is established in a Nash equilibrium (i.e. when $d < 0$) this network is the unique linking-proof network. Several linking-proof networks may exist when $d > 0$. We demonstrate this in the following examples.

**Example 4.1 (no linking constraints)** Let $d > 0$ and let $k_i = n - 1$ for all $i$. In this game the linking is not constrained and players may establish any number of links. This constitutes our basic dilemma game $\Gamma$. Hence, if $g_{ij} = 0$ then $l_i < k_i$ and $l_j < k_j$ for any pair of distinct players $i$ and $j$ and $g$ is not linking-proof. The complete network is therefore the unique linking-proof network.

**Example 4.2 (uniform linking constraint)** Let $d > 0$ and let $n = 4$. Let $k = (2, 2, 2, 2)$. Any of the 64 possible networks can be established in a Nash equilibrium. However, only seven of them, illustrated in Figure 4.1, are linking-proof. Networks $g^1, g^{11}$ and $g^{111}$ constitute the three possible wheel networks in which each player links to two neighbors. In networks $g^1, g^2, g^3$ and $g^4$ one player is isolated and each of the other three players has two links. For $i = 1, 2, 3, 4$ the network $g^i$ forms when players $N \setminus \{i\}$ form a clique, thus leaving player $i$ without any links.
The following Proposition asserts that every dilemma game has a linking-proof equilibrium, and consequently, a linking-proof network.

**Proposition 4.4** There always exists a linking-proof equilibrium of the game $\Gamma(k)$.

**Proof.** Let $0_i$ be the vector denoting that player $i$ proposes no links, and let $0 = (0_1, ..., 0_n)$ be the corresponding profile. The profile $(D, 0)$ is LP equilibrium for dilemma games with $d < 0$. For $d < 0$ this is shown in Proposition 4.3. For $d = 0$ this is because if all players defect no player can increase her payoff either by switching to cooperation, or by removing, moving or adding links.

For $d > 0$ a LP network can be constructed with the following iterative procedure. Let $g^0$ be an empty network. For $x = 1, 2, ...$ let $g^x$ be the network obtained from $g^{x-1}$ by the addition of one link: take a pair of distinct players $i, j$ such that $g^x_{ij} = 0$, $l_i(g^{x-1}) < k_i$ and $l_j(g^{x-1}) < k_j$, and let $g^x_{ij} = 1$, and otherwise let $g^x_{ij} = g^{x-1}_{ij}$. Stop the procedure if there is no such pair of players.

This procedure stops after finite number of steps, because a new link is added in each step and there is a finite number of possible links. It follows from Proposition 4.3 that the resulting network is an LP equilibrium. $lacksquare$

### 4.3.2 Subgame perfect equilibrium of the repeated game

In the repeated game the players can condition their linking choices on the behavior of players in previous periods. More specifically, players may punish a deviation from the agreed outcome path by removing links to any player that deviated. For example, players may agree to cooperate and punish defectors by exclusion from the neighborhood of cooperative players.

Figure 4.1: Linking-proof networks of the dilemma game with $n = 4$ players, $k = (2, 2, 2, 2)$, and $d > 0$. Only established links are shown.
In a repeated game we may define exclusion as follows. Let $t > 2$. We say that player $i$ excludes player $j$ in period $t$ if $i$ and $j$ were neighbors at $t - 1$ and $i$ removes the link with $j$ in period $t$.

Two players may exclude each other simultaneously. For example, player $i$ may remove her link with player $j$ if she knows $j$ will also remove it. Recall that each player is indifferent between proposing a link or not if she knows that it will not be reciprocated. Two neighbors could therefore exclude each other even if the resulting network is such that each would prefer to keep the link established. We call an exclusion mutual if two players exclude each other but both miss at least one link in the resulting network.$^4$

On the other hand, a single player may be willing to exclude another regardless of the other player’s linking decision. It is possible, for example, to exclude a player and yet establish the maximal number of links. This can be done, for example, by removing the link with one player and establish a new one with another player, thus excluding the first and including the latter.$^5$ We say that player $i$ unilaterally excludes player $j$ if $i$ still establishes her maximal number of links in the resulting network. We call an exclusion unilateral if each excluding player establishes her maximal number of links in the ensuing network. There is no condition on the number of links that the excluded player establishes in the ensuing network. Unilateral exclusion is defined only for games with mutual link formation and is not related to unilateral link formation. See Figure 4.2 for examples of unilateral and mutual exclusion.

We begin with a few examples of the equilibria of repeated dilemma games. The first example illustrates that not every Nash equilibrium of the repeated dilemma game is subgame perfect. Recall that in order to distinguish between the Nash equilibria of the repeated game and those of the stage game we refer to the latter as the static equilibria.

**Example 4.3** Consider the dilemma game $\Gamma = (\{1,2\}, J, \pi)$ with two players and no linking constraints,$^6$ repeated twice. Let the outside option be low, $d > 0$.

Consider the following strategy $\sigma_i$ for player $i = 1,2$: "Propose the link and cooperate in the first period. In the second period defect, and propose the link if and only if the opponent cooperated in the first period."

The pair $(\sigma_1, \sigma_2)$ constitutes a Nash equilibrium of $\Gamma^2$ if $f < c + d$: given that player $i$ follows $\sigma_i$, the best response of player $j$ entails always proposing the link and defection in the last period; this given, defection in the first period earns her a total payoff of $f$, which is less than $c + d$ earned if she cooperates in the first period.

Assume now that player 2 defects in the first period nevertheless. In response, the player’s strategies require both players to defect, and that only player 2 proposes the link which, consequently, is not established. However, in the absence

---

$^4$We say that player $i$ is missing one or more links in network $g$ if she establishes less than her maximal number of links, that is, if $l_i(g) < k_i$.

$^5$If a player $i$ removes a link with player $j$ to establish a link with player $j'$ we say that $i$ relocates a link.

$^6$This game is a variant of the usual two-player prisoner’s dilemma game with outside option.
of commitment it is not obvious that in the second period player 1 will persist in playing $\sigma_1$. In fact, rather than to exercise the threat of removing the link, player 1 does better in the second period by reneging on her threat and propose the link anyway. If player 1 is rational she never excludes player 2. If player 2 is rational she anticipates this and defects in both periods. Realizing this, rational players would not even agree on playing $(\sigma_1, \sigma_2)$. The threat of unilateral exclusion, implicit in the strategy $\sigma_1$, is in this sense empty and the equilibrium $(\sigma_1, \sigma_2)$ is not subgame perfect.

The above example illustrates the general discussion in section 2.2.2 and motivates our focus on subgame perfect equilibria. For many dilemma games only trivial subgame-perfect equilibria exist, in the sense that a static equilibrium must be selected in each period. Recall assertion 2 of Lemma 2.1: if a stage game has a unique Nash equilibrium, the finitely repeated game has a unique subgame perfect equilibrium.

Example 4.4 (Exogenous network) The n-player prisoner’s dilemma game can be seen as a dilemma game without the linking choice. Rather, the network is fixed exogenously to be a complete network. Each player chooses an action and must play the game with all other players. The payoffs are given by (4.2), with $g$ being the complete network.

For each player defection strictly dominates cooperation. The stage game thus has a unique Nash equilibrium in which all players defect. Consequently, the finitely repeated n-player prisoner’s dilemma game has a unique subgame perfect equilibrium in which all players defect in each of the periods.
Assertion 3 of Lemma 2.1 states that if static equilibria are payoff-equivalent the set of the subgame perfect equilibria is characterized by all possible sequences of static equilibrium outcomes. Consequently, a player earns the same total payoff in each of the subgame perfect equilibria.

Example 4.5 (Endogenous network, high outside option) Consider any dilemma game and let \( d < 0 \). Proposition 4.2 states that the empty network is established in any static Nash equilibrium of this game. All its static equilibria are therefore payoff-equivalent because the payoff of each player in the empty network is 0, regardless of the profile of actions. By implication, each subgame perfect equilibrium of the repeated dilemma game selects a sequence of static equilibria. To summarize, whenever \( d < 0 \) the empty network is established along the outcome path of any subgame perfect equilibrium of the repeated dilemma game.

All the subgame-perfect equilibria that we constructed above have a common feature: they select a static equilibrium in each period. Statement (1) of Lemma 2.1 asserts that any profile of strategies that in each period, and regardless of the history, selects some static equilibrium is subgame perfect. For any repeated dilemma game a number of subgame perfect equilibria can therefore be constructed in this way, from sequences of static equilibria. However, in each static equilibrium all relations are defective. Hence, all subgame perfect equilibria constructed from sequences of static equilibria are defective.

No other subgame perfect equilibrium exist for finitely repeated dilemma games with high outside option, which is shown in the example above. In the following section we show that cooperative subgame-perfect equilibria may exist for dilemma games with low outside option.

4.3.3 Cooperation in subgame perfect equilibria of the repeated game

In section 4.3.3 we show that, for any repeated dilemma game with \( d > 0 \), cooperative relations may be sustained in a subgame perfect equilibrium via the threat of mutual exclusion. Mutual exclusion requires that both players simultaneously remove the link with each other. Punishment with mutual exclusion thus requires the participation of a punished player in her own punishment. Furthermore, the resulting network is not linking-proof. This may motivate the players, especially the punished one, to renege on their participation in the punishment. In section 4.3.3 we follow up by discussing the possibility to sustain cooperation via threats of unilateral exclusion. We show that this is possible for most vectors of linking constraints.

Cooperation through threats of mutual exclusion

Consider a dilemma game \( \Gamma(k) \) and let there be a network \( g^* \in \mathcal{G}(k) \) with no isolated players: \( l_i(g^*) \geq 1 \) for each player \( i \). For each \( i \) let \( g^i = g \ominus l_i \ominus \ldots \ominus n_i \)
be the network obtained from $g^*$ by removing all established links with player $i$: $l_i(g^t) = 0$. Obviously, $g^t \in \mathcal{G}(k)$ for each $i$. Note that $q^t$ is a Nash equilibrium only if (i) player $i$ proposes no links and (ii) no player proposes a link to $i$.

Let $d > 0$. It follows from Proposition 4.2 that $g^*$ and all $g^i$ can be established in the Nash equilibria $q^* = (D, p(g^*))$ and $q^i = (D, p(g^i))$. The payoff of player $i$ in each of these equilibria is proportional to the number of her neighbors. In particular,

$$\pi_i(q^*) = d \cdot l_i(g^*) > 0 \text{ and } \pi_i(q^i) = d \cdot l_i(g^i) = 0.$$  

Consider now the profile of moves $r^* = (C, p(g^*))$ where all players cooperate and establish the network $g^*$. The payoff of player $i$ in this profile is

$$\pi_i(r^*) = c \cdot l_i(g^*) > 0 = \pi_i(q^i).$$

Recall Definition 2.5, of a trigger strategy profile. Take two integers $T > t^* > 0$. Consider the trigger strategy profile $\sigma(r^*, q^*, \{q^i\}_{i \in N}, T, t^*)$, which can be described in words as follows:

"Cooperate in the early periods $1, \ldots, T - t^*$ and defect in the remaining periods. Establish network $g^*$ in each period. If in an early period some player $i$ defects, establish network $g^i$ and defect forever."

Each player $i$ earns strictly more in the equilibrium $q^*$ compared to equilibrium $q^i$. If player $i$ defects early, she earns $t^* \pi_i(q^i) = 0$ in the last $t^*$ periods. If there are no early defections she earns $t^* \pi_i(q^*)$. Hence, if she defects early she looses $t^* \pi_i(q^*)$ in the last $t^*$ periods. There exists a sufficiently large $t^*$ such that for each player this loss is larger than the benefit from any early deviation. If $t^*$ is such, then no player wants to defect in an early period and cooperation is sustained.

Proposition 4.5 confirms this conclusion: there exist $T, t^*$, such that $\sigma(r^*, q^*, \{q^i\}_{i \in N}, T, t^*)$ is a subgame perfect equilibrium of $\Gamma^T(k)$. By definition, $\sigma(r^*, q^*, \{q^i\}_{i \in N}, T, t^*)$ is cooperative as it induces the cooperative outcome $r^*$ in the early periods of the repeated game.

In the following we focus on the trigger strategy profiles where, similar to the profile outlined above, an early defection triggers a change of the network. For reference we use the notation

$$(a, g) = (a, p(g))$$

and define, for networks $g, g^*, g^1, \ldots, g^n \in \mathcal{G}(k)$, the following trigger strategy profile

$$\rho(g, g^*, \{g^i\}_{i \in N}, T, t^*) \equiv \sigma((C, g), (D, g^*), \{(D, g^i)\}_{i \in N}, T, t^*).$$

Proposition 4.5 Consider the dilemma game $\Gamma^T(k)$ with $d > 0$ such that there exists a network $g^* \in \mathcal{G}(k)$ with no isolated players. For each $i$ let $g^i$ be the network obtained from $g^*$ by removing all links with player $i$. For large enough $T$ there exists $t^*$ such that the trigger strategy profile $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ for $\Gamma^T(k)$ is fully cooperative and subgame perfect.
**Proof.** We have outlined the proof in the discussion above. Profiles of moves \((D, p(g^*))\) and \((D, p(g_1))\) are Nash equilibria. Furthermore, \(\pi_i(C, p(g^*)) > \pi_i(D, p(g^*)) > \pi_i(D, p(g_1))\). The rest of the proof follows from Theorem 2.1 in chapter 2.}

**Corollary 4.1** A cooperative trigger strategy profile for \(T_T^T(k)\) with \(d > 0\) exists for sufficiently large \(T\) if \(k_i > 1\) for each \(i\) and, if \(n\) is odd, \(k_j > 2\) for some \(j\).

**Proof.** If \(n\) is even, a network \(g^* \in G(k)\) with no isolated players may be established by linking pairs of players: for example, for each \(x = 1, ..., n/2\), players with indices \(2x - 1\) and \(2x\) are linked. If \(n\) is odd a network \(g^* \in G(k)\) with no isolated players may be established by linking two players to player \(j\) and link pairs of the remaining players: for example, if \(k_1 \geq 2\) players 1 and 2, players 1 and 3, and for each \(x = 2, ..., n/2\), players with indices \(2x\) and \(2x + 1\) are linked.

The trigger strategies of proposition 4.5 threaten to punish deviations in the early periods with mutual exclusion: after a player deviates from the equilibrium path all links with that players are mutually removed and she remains (mutually) isolated from the network throughout the remaining periods of the game.\(^7\)

**Example 4.6 (cooperation via mutual exclusion)** Consider again the dilemma game \(\Gamma = (\{1, 2\}, J, \pi)\) of Example 4.3, repeated twice. Let \(f < c + d\) so that, for each player, mutual cooperation in the first and defection in the second period is better than defection in the first and exclusion in the second period. Inspired by proposition 4.5 we consider the following strategy: 

"Propose the link and cooperate in the first period. In the second period defect, and propose the link if and only if both players cooperated in the first period."

This strategy differs from the strategy \(\sigma_i\) of Example 4.3 in that it requires second period link removal after deviation of any player, including own deviation. A deviation in the first period is punished by mutual, rather than unilateral exclusion of strategy \(\sigma_i\). The threat of exclusion is effective. Mutual exclusion is credible as no player can profit by unilateral abstention from excluding the other player. A pair of strategies described above is thus subgame perfect. It is important to note that punishment by mutual exclusion requires participation of both players.

The finitely repeated prisoner’s dilemma game without an outside option has a unique subgame perfect equilibrium in which both players defect in all periods. Example 4.6 demonstrates that the addition of a low outside option to the prisoner’s dilemma game is sufficient to obtain cooperative subgame perfect

\(^7\)Not all deviations need to be punished actively. Some deviations may decrease the payoff of the deviating player and need no subsequent punishment to be deterred. For example, if the play reaches a static equilibrium there is no deviation that increases the payoff of the deviating player. In this case proceeding, after some deviation, with the play as if no deviation occurred already constitutes an effective threat.
equilibria. Cooperation in these equilibria is sustained by the threat of mutual exclusion.

Mutual exclusion may not be the most intuitive form of punishment, because it requires participation of a punished player in her own punishment. Following her deviation the player is expected to remove all of her links and isolate herself throughout the remaining periods. Her isolation is an equilibrium because other players never again propose the link to her so that she is indifferent between proposing any links or not. However, there are several reasons why it may not be sensible to expect a player to contribute to her own punishment.

A player may be expected to participate in her own punishment only if it is in her own interest to do so, that is, if doing so constitutes the subgame perfect equilibrium across the remaining periods of the game. To make a threat credible it is sufficient that each player weakly prefers to participate in the punishment. This is the case with the threats in the trigger strategies of Proposition 4.5. Namely, knowing that she is being excluded by all other players the punished player is indifferent between participating in the punishment or not, hence she does participate in equilibrium. Similarly, knowing that the punished player will participate in the punishment all other players are indifferent between punishing or not, hence they punish in equilibrium.

However, the punished player never strictly prefers to remove a link in order to support her punishment. She will keep her links if even a smallest chance exists that the other players will not exclude her. Given that after the deviation all players defect, each player strictly prefers to have more links from having less. Proposing a maximal number of links thus never hurts. Yet, punishments with mutual exclusion, as described in Proposition 4.5, always requires that punishing players propose less than their maximal number of links and that the punished player proposes no links. If there is even a slight doubt about whether one of the parties will participate in the punishment, or about whether mutual participation is common knowledge, none of the parties will punish.

In the following section we discuss strategies with threats of unilateral exclusion. Such threats may not always be credible, as we demonstrated in Example 4.3. Nevertheless, we conclude that for almost all dilemma games with linking constraints it is possible to sustain cooperation in the subgame perfect equilibrium with threats of unilateral exclusion only. In such equilibria the punishing players are willing to exclude the punished player regardless of her linking behavior.

Cooperation through threats of unilateral exclusion

The ideas in this section may be outlined as follows. Consider an n-player dilemma game \( \Gamma(k) \) and let \( k \) be such that linking is strictly constrained, that is, \( k_i < n - 1 \) for each player \( i \). Each player must thus select at least one other player to whom she does not propose a link. If the dilemma game is repeated the choice of which player to exclude may be made contingent on player’s past behavior. For example, players may unilaterally exclude those other players who defected in the previous period. This in turn may be sufficient to discourage
period 1 2

equilibrium path

if pl. i defects in period 1

Figure 4.3: An illustration of a subgame perfect, cooperative and linking-proof strategy profile for the repeated game of example 4.7, with \( n = 4 \), \( k = (2, 2, 2, 2) \), \( d > 0 \) and \( f < c + d \).

early defections.

The following example demonstrates that unilateral exclusion can indeed form a credible threat when linking is constrained. The example also gives the sketch of the general analysis.

**Example 4.7 (cooperation via unilateral exclusion)** Consider the dilemma game of example 4.2 with \( d > 0 \), \( n = 4 \) and \( k = (2, 2, 2, 2) \) for each \( i \). Let the game be repeated twice and let \( f < c + d \). Consider the following strategy for player \( i \): "Cooperate in the first and defect in the second period. Establish the network \( g^1 \) in the first period. In the second period: establish \( g^1 \) if in the first period (i) no player defected, or (ii) two or more players defected; establish \( g^2 \) if another player \( j \neq i \) was the only player who has defected; make any linking choice otherwise (i.e. if \( i \) was the only player to have defected)."

If all players follow this strategy the wheel network \( g^1 \) forms in both periods, players cooperate in the first period and defect in the second period. If in the first period player \( i \) is the only to defect, she is unilaterally excluded by her neighbors and network \( g^1 \) is established in the second period. See Figure 4.3 for an illustration.

Network \( g^i \) is linking-proof. The opponents of player \( i \) are willing to establish \( g^i \) independently of \( i 's \) linking choice. The threat of unilateral exclusion is both credible and effective and the ensuing strategy profile is subgame perfect.

In the example above all threats with exclusion implement unilateral exclusion. A linking-proof network is established in each period along the equilibrium path, as well as in each period along any punishment path. The following proposition asserts that sequential formation of linking-proof networks implies that any exclusion is unilateral.
Proposition 4.6 Consider a repeated dilemma game $\Gamma^T(k)$ with $d > 0$, and an arbitrary $t \leq T$. If all networks which form along the history $h^t$ are linking-proof, then each exclusion during $h^t$ was unilateral.

Proof. Assume that, for some $\tau \leq t$, player $i$ excludes player $j$ between periods $\tau - 1$ and $\tau$ but none of them establish their maximal number of links in the ensuing network $g^\tau$. Then, $l_i(g^\tau) < k_i$ and $l_j(g^\tau) < k_j$, while $g_{ij} = 0$. Following Proposition 4.3 $g^\tau$ is not a LP network. Hence, $g^\tau$ is a LP network only if all exclusions between periods $\tau - 1$ and $\tau$ are unilateral. 

To assure that each exclusion is unilateral we thus concentrate on strategies such that, after any possible history, a linking-proof network is established in each period along the resulting outcome path.\(^8\)

Definition 4.2 Let $\sigma$ be a profile of strategies in the repeated dilemma game with linking constraints $\Gamma^T(k)$. Profile $\sigma$ is a linking-proof strategy profile if, for each period $t < T$ and each possible history $h^t \in H^t$, $g(\sigma^{t+1}(h^t))$ is a linking-proof network.

In what follows we restrict our interest to a subset of subgame perfect equilibria, by focusing on linking-proof subgame perfect profiles. Example 4.7 demonstrates that simple linking-proof subgame perfect equilibria can be constructed using trigger strategies. When trigger strategies fail to sustain cooperation, non-trigger strategies may be used.

We first characterize in Propositions 4.7 and 4.8 the games for which a linking-proof subgame perfect equilibrium can be constructed using simple trigger strategies. We then show in Theorem 4.1 that cooperative and linking-proof subgame perfect equilibria can be constructed for most dilemma games with $d > 0$, using recursive trigger strategies.

Trigger strategies

Consider a trigger strategy profile $\rho(g^*,g^*,\{g^i\}_{i \in N,T,t^*})$, defined by (4.5). Along the equilibrium path of $\rho(g^*,g^*,\{g^i\}_{i \in N,T,t^*})$ network $g^*$ is established in each period, all players cooperate during the early periods $1,...,T - t^*$ and defect during the remaining periods. Any deviation during the early periods triggers a change of the network: during the remaining periods all players defect and establish network $g^i$, where $i$ is one of the players which deviated. Profile $\rho(g^*,g^*,\{g^i\}_{i \in N,T,t^*})$ is linking-proof whenever each of the networks $g^*,g^1,...,g^n$ is linking-proof. In this case all threats are credible because in each period of each threat the players play a linking-proof equilibrium.

The following proposition asserts that a subgame-perfect and linking-proof trigger strategy profile exists if and only if there is a linking-proof network from which each player can be unilaterally excluded.

\(^8\)We do not assume that along the outcome paths a linking-proof outcome is chosen but only that a linking-proof network is established. Obviously, whenever a pairwise equilibrium is chosen all relations are defective.
Proposition 4.7 Consider an n-player dilemma game $\Gamma(k)$ with $d > 0$. Let $g^*, g^1, ..., g^n$ be linking-proof networks. There exist $T, t^*$, such that $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ is subgame-perfect for $\Gamma^T(k)$, if and only if for each $i$,

$$l_i(g^i) < l_i(g^*)$$

In this case, for any $T' \geq T$, $\rho(g^*, g^*, \{g^i\}_{i \in N}, T', t^*)$ is subgame perfect for $\Gamma^{T'}(k)$.

Proof. Let $d > 0$. Assume that $l_i(g^*) > l_i(g^i)$ for some $i$. If player $i$ defects in one of the early periods she thereby looses some of her neighbors. This decreases all her future per-period earnings,

$$\pi_i(D, p(g^i)) = d \cdot l_i(g^i) < d \cdot l_i(g^*) = \pi_i(D, p(g^*)) < c \cdot l_i(g^*) = \pi_i(C, p(g^*))$$

(4.6)

which makes this threat of unilateral exclusion effective for $t^*$ sufficiently small. If $l_i(g^*) > l_i(g^i)$ holds for each player $i$ then (4.6) holds for each $i$. The profile $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ then satisfies conditions of Theorem 2.1 to be subgame-perfect for all sufficiently high $T$ and $t^*$ such that $T > t^*$.

Assume now that $l_i(g^*) \leq l_i(g^i)$ for some $i$. Following $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ all players cooperate during periods $1, ..., T - t^*$ and defect in the remaining periods. In this case player $i$ earns $l_i(g^*)((T - t^*)c + t^*d)$. If, however, player $i$ defects already in period $T - t^*$, and follows $\rho$ otherwise, she earns

$$l_i(g^*)((T - t^* - 1)c + f) + l_i(g^i)t^*d > l_i(g^*)((T - t^*)c) + l_i(g^i)t^*d$$

$$\geq l_i(g^*)((T - t^*)c) + l_i(g^*)t^*d$$

thus strictly increasing her earning. The threat against her defection in period $T - t^*$ is not effective and $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ is not subgame-perfect.

When $d > 0$ the vector of linking constraints $k$ determines the set of linking-proof networks $\mathcal{S}(\Gamma(k))$. It would thus be of interest to know which vectors $k$ imply the conditions of Proposition 4.7. Providing a full characterization appears to be a difficult combinatorial task and is beyond the scope of this chapter. Instead, we give below a few partial characterizations and an interpretation of our results.

Proposition 4.8 Let $d > 0$. Given a sufficient number of periods, trigger strategies can be used to construct (i) cooperative, (ii) linking-proof, and (iii) subgame perfect equilibrium for a repeated dilemma game with linking constraints $k$, in any of the following cases:

1. $k_i = k$ for each $i$, where $2 \leq k \leq n - 2$,
2. $2 \leq k_i < \sqrt{2n - 9}$ for each $i$. 

53
In words, cooperation and linking proofness can be sustained (1) if linking constraints are uniform, or (2) if the set of all players is large relative to the maximal possible number of individual links. The proof of Proposition 4.8 follows a sequence of lemmas and is given in the appendix to this chapter.

The conditions derived in Proposition 4.8 exclude some interesting dilemma games. Consider, for example, the original network dilemma game with unconstrained linking, \( k = n - 1 \). In this game the complete network is the unique linking-proof network, as we demonstrated in Example 4.1. Using backwards induction we prove below that the unique subgame-perfect selects this equilibrium in each period. This implies that the complete network is established in each period and all players always defect.

All linking-proof subgame perfect equilibria are defective also when \( k_i = 1 \) for each \( i \) and \( n \) is even, because all linking-proof equilibria are payoff-equivalent. In contrast, if \( k_i = 1 \) for each player \( i \) but \( n \) is odd, linking-proof subgame perfect equilibrium can be constructed in which all relations are cooperative in early periods.

**Proposition 4.9** Consider an \( n \)-player dilemma game \( \Gamma(k) \) with \( d > 0 \) and \( k_i = k \) for each \( i \).

1. If \( k = n - 1 \), or if \( k = 1 \) and \( n \) is even, then each linking-proof subgame perfect equilibrium is defective.

2. If \( k = 1 \) and \( n \) is odd, there is a linking-proof subgame perfect equilibrium of \( \Gamma^T(k) \) with sufficiently large \( T \), such that one player is isolated and all players cooperate in the early periods.

**Proof.** (1) Consider the dilemma game without the linking constraints, that is, \( k_i = n - 1 \) for each player \( i \). A complete network \( g^c \) is the unique linking-proof equilibrium of this game, and the profile \( (D, p(g^c)) \) is the unique static equilibrium establishing \( g^c \). Any linking-proof and subgame perfect profile thus selects \( (D, p(g^c)) \) in the last period regardless of the history. By backwards induction, there is a unique such profile and it selects \( (D, p(g^c)) \) in each period.

A similar conclusion holds for the dilemma game \( \Gamma(k) \) with an even number of players \( n \) and \( k_i = 1 \) for each player \( i \). Each linking-proof network consists of isolated pairs, where each player is linked to one other player. In a linking-proof equilibrium all players defect. This implies that each player earns \( d \) in each linking-proof equilibrium profile. All pairwise-equilibria are thus payoff-equivalent. Any linking-proof and subgame perfect profile selects a linking-proof equilibrium in the last period. However, as these are payoff-equivalent, the penultimate period is treated just as the last one and a linking-proof equilibrium is again selected. Hence, by backwards induction a static linking-proof equilibrium must be selected in each period along the outcome path of any such strategy profile. Players defect in all periods along the equilibrium path of any linking-proof and subgame perfect profile.

(2) Construct a linking-proof network \( g \) as follows: link each odd \( i \) to \( i + 1 \) and leave \( n \) unlinked. The profile of the following strategies is linking-proof:
"Establish $g$ as long as there have been no deviations or if more than one player simultaneously deviated. Cooperate in periods 1, ..., $T - \gamma$ and defect in the remaining periods. If own neighbor is the only player to deviate then unilaterally exclude her and link to the previously isolated player. If isolated, propose link to any player whose neighbor defected."

Take a player $i$ with a neighbor and consider any period $r < T - \gamma$. Assume that all players $N \setminus \{i\}$ cooperate. If player $i$ defects, she gains $f$ immediately and 0 ever after. In contrast, if she cooperates she earns at least $c + \gamma d$. Punishment with exclusion is effective if $f < c + \gamma d$. The strategy profile is subgame perfect and cooperative if $T > \gamma > (f - c)/d$. ■

**Non-trigger strategies**

The trigger strategies above threaten to punish any early defection with unilaterally exclusion. If, however, some player cannot be unilaterally excluded the trigger strategies fail to achieve complete cooperation. Take, for example, a player who has the same number of neighbors in all possible linking-proof networks. This player must have the same number of neighbors in each period along the outcome path of any linking-proof profile and cannot be made to cooperate using the trigger strategies defined by (4.5).

**Example 4.8** Consider the dilemma game $\Gamma(2,1,1)$ with $n = 3$ players and vector of linking constraints $k = (2,1,1)$. Let $d > 0$. There exist only two linking-proof networks, $g^*$ and $g^1$, illustrated in Figure 4.4. Players 2 and 3 establish one link in each of these networks and earn the same per-period payoff in the two corresponding stable-equilibria. Since $l_2(g) = 1$ for each linking-proof network $g \in \{g^*, g^1\} \equiv S(\Gamma(2,1,1))$, conditions of Proposition 4.7 are not satisfied. Only player 1 can be punished effectively with unilateral exclusion. Therefore, there exists no cooperative linking-proof equilibrium trigger strategy (4.5) that is subgame-perfect.

It is interesting that it may nevertheless be possible to design strategy profiles that are linking-proof, subgame perfect, and cooperative, albeit not of the form (4.5). We describe the construction of such strategies below.
Trigger strategies threaten to retaliate any deviations with the repetition of a punishment static equilibrium. While such threats are most intuitive, and thus implemented in simple behavioral strategies of the Tit-for-tat or Grim type (see e.g. chapter 6), they may not be the most effective ones. Trigger strategies can not sustain outcomes which yield payoffs that are below the worst equilibrium payoffs. In contrast, Benoit and Krishna (1985) demonstrated that, if the so-called three phase punishment paths are used, any outcome that yields payoffs above the minimax levels may be sustained in early periods of a subgame perfect equilibrium. They proved a limit folk theorem for subgame perfect equilibria of finitely repeated games\(^9\), under the sufficient condition that each player has several distinct equilibrium payoffs and that the dimension of the set of feasible payoffs is \(n\).

The game is said to have distinct equilibrium payoffs if every player has two or more different equilibrium payoffs. This implies that every player has a strictly preferred equilibrium and can be punished by playing her non-preferred one. Conversely, if only one player has distinct equilibrium payoffs then only she can be punished in the last period of the repeated game. Nevertheless, other players may still be induced to play any of their actions, not by threats of punishment but by promises of rewards.

An informal outline of the idea can be given as follows. Consider a generic game \(F = (N, A, \pi)\) in which only player \(i\) has distinct equilibrium payoffs. Because she has distinct equilibrium payoffs player \(i\) can be made to play repeatedly any action: only if she complies her preferred equilibrium is played in the last periods. Let \(a_i, b_i \in A_i\) be two possible actions of player \(i\) and let \(\Gamma_{a_i}\) and \(\Gamma_{b_i}\) be games obtained from \(\Gamma\) by fixing the action of player \(i\) to either \(a_i\) or \(b_i\). Let \(j\) be a player who prefers if player \(i\) chooses \(a_i\) rather than \(b_i\), in the sense that \(j\) earns more in his worst equilibrium of \(\Gamma_{a_i}\) than in his worst equilibrium of \(\Gamma_{b_i}\). Player \(j\) may now be induced to play any of her actions if doing so is rewarded by player \(i\) playing \(a_i\). Player \(j\) may thus also be made to play any action repeatedly. Hence, both players may be induced to play any of their actions, given a sufficiently large number of periods. A third player, whose payoff crucially depends on the actions of player \(j\), may now in the same fashion be induced to play any of her actions, expecting reward from \(j\). Indeed, iterative rewarding may induce any feasible outcome in the early periods of the repeated game if the utilities of all players are appropriately interrelated.

See Smith (1995) for precise definitions and a thorough elaboration of this idea. Formally, the game is said to have recursively distinct equilibrium payoffs if there exists an ordering of the players 1, ..., \(n\) such that

(i) player 1 has at least two distinct equilibrium payoffs, and

(ii) for all \(i < n\), there exist strategy profiles \(a(i), b(i) \in A\) such that

\(^9\)The folk theorem was initially proven for infinitely repeated games. It states that any feasible and individually rational payoff vector can be obtained by a long-run undiscounted average of some subgame perfect equilibrium. See Friedman (1971) and Aumann and Shapley (1994) for its early formalizations.
- each player \(i+1, \ldots, n\) plays her best response in both \(a(i)\) and \(b(i)\),\(^{10}\) and

- the payoff to player \(i+1\) in \(a(i)\) is distinct from her payoff in \(b(i)\).

Smith (1995) shows that, if the stage game has recursively distinct equilibrium payoffs, any feasible and individually rational outcome may be approximated as an average payoff of the repeated game with sufficiently long horizon. Recursively distinct equilibrium payoffs are thus the necessary and sufficient condition for the limit folk theorem.

We cannot straightforwardly apply the limit folk theorem of Smith to dilemma games because of our additional restriction of linking-proofness. For instance, the strategy profiles designed in Smith (1995) punish an early deviation by selecting the outcome in which the deviating player earns her minimax payoff. In the dilemma game with low outside option the player attains her minimax payoff when she has no neighbors. A network in which player has no links may not be linking-proof, though. Indeed, there exist linking constraints such that some players establish a positive number of links in each linking-proof network.\(^{11}\)

Nevertheless, iterative rewarding can be used to design strategy profiles which are cooperative, subgame-perfect and linking-proof, in dilemma games with linking constraints that do not satisfy the conditions of Proposition 4.7. This is possible under two, albeit weak, conditions: there should exist a player \(i\) and two linking-proof networks \(g^*\) and \(g^i\) such that (i) player \(i\) has more neighbors in \(g^*\) than in \(g^i\), and (ii) \(g^*\) is connected.

The parallel between these conditions and that of the recursively distinct equilibrium payoffs is apparent: condition (i) implies that player \(i\) has distinct linking-proof equilibrium payoffs, while condition (ii) implies that payoffs of all players in some linking-proof network are interrelated. The proof of Theorem 4.1, below, is inspired by this parallel. To give the flavor of the proof we construct in the following example an instance of cooperative, subgame-perfect, and linking-proof strategy profile for the game \(\Gamma(2,1,1)\).

**Example 4.9 (4.8 continued)** Consider again the game \(\Gamma(2,1,1)\) with \(d > 0\), and linking-proof networks \(g^*\) and \(g^1\), shown by Figure 4.4. Let the game be repeated \(T\)-times and let \(t^0\) and \(t^1\) be integers such that \(0 < t^0 < t^1 < T\). See Figure 4.5 for the illustration of the following strategy profile:

- On the outcome path:
  - network \(g^*\) is established in all periods,
  - player 1 cooperates during periods \(\{1, \ldots, T - t^0\}\), and
  - players 2 and 3 cooperate during periods \(\{1, \ldots, T - t^1\}\).

- If player 1 defects early, network \(g^1\) is established and all players defect during the remaining periods.

\(^{10}\)In other words, for each player \(j \in \{i+1, \ldots, n\}\), \(a(i)_j\) is a best response to \(a(i)_{-j}\) and \(b(i)_j\) is a best response to \(b(i)_{-j}\), see chapter 2.

\(^{11}\)Take, for example, player 2 in the game \(\Gamma(2,1,1)\) of example 4.8.
• If player 2 or 3 defects early, network $g^*$ is kept and all players defect during the remaining periods.

In words, the threatened loss of links during $\{T - t^0 + 1, \ldots, T\}$ constitutes an incentive to player 1 to cooperate during $\{1, \ldots, T - t^0\}$. The incentive to players 2 and 3 to cooperate during $\{1, \ldots, T - t^1\}$ is the possibility to free ride on player 1 during $\{T - t^1 + 1, \ldots, T - t^0\}$. Players will conform to the strategy profile if one-period profit of any early defection is offset by the loss of profit in the resulting path. All threats are credible as they consist of a repetition of a static equilibrium.

More precisely, let $\Delta_0 = t^0$ and let $\Delta_1 = t^1 - t^0$. If players 2 and 3 follow the strategy profile they are rewarded in $\Delta_1$ periods by free riding on player 1, thus earning $\Delta_1 f$ instead of $\Delta_1 d$. The reward increases with its length and there certainly exists a positive integer $\Delta_1$ such that early deviation does not pay. Similarly, if player 1 deviates early she is punished by exclusion for at least $\Delta_0$ periods, earning 0 instead of earning at least $\Delta_0 d$. The punishment increases with its length and there exists a positive integer $\Delta_0$ such that early deviation does not pay. Find such $\Delta_0$ and $\Delta_1$, and set $t^0 = \Delta_0$, $t^1 = t^0 + \Delta_1$ and $T > t^1$. The resulting strategy profile is subgame perfect, linking-proof, and cooperative.

It is interesting to note that "intervals" $\Delta_0$ and $\Delta_1$ need not be very long. For example, with the prisoner’s dilemma game of Figure 2.1 it is sufficient to set $\Delta_0 = 2$ and $\Delta_1 = 1$. That is, if the game $\Gamma(2, 1, 1)$ is repeated at least 4 times the profile above induces all players to establish $g^*$ and cooperate during all but the last three periods.

In the example above only one iteration of rewarding is needed to sustain complete cooperation during early periods of the repeated game. More than one iteration may be necessary for dilemma games with general linking constraints. The definition of the recursive trigger profile below suggests that the number of necessary iterations is related to the structure of the initial network structure.¹²

Let $d > 0$ and let $g^*$ be a connected linking-proof network for $\Gamma(k)$. Define

$$N^0 = \{i \in N \mid l_i(g^i) < l_i(g^*) \text{ for some } g^i \in G(\Gamma(k))\}$$

to be the set of players who can be punished (for an early defection) by unilateral exclusion from $g^*$. Assume that $N^0$ is non-empty. Consider now the set $N^1$ of their neighbors,

$$N^1 = \left( \bigcup_{i \in N^0} L_i(g^*) \right) \setminus N^0.$$

Players in $N^1$ can be rewarded for cooperation with the opportunity to free ride on players in $N^0$. Now define recursively the sets $N^2, \ldots, N^m$ by

$$N^n = \left( \bigcup_{i \in N^{n-1}} L_i(g^*) \right) \setminus (N^{n-1} \cup N^{n-2})$$

¹²More precisely, the number of necessary iterations coincides with the maximal distance between a player who can be excluded from the network and a player who cannot be excluded.
Figure 4.5: An illustration of a subgame perfect, cooperative and linking-proof strategy profile for the repeated game of example 4.8, with $n = 3$, $k = (2,1,1)$, and $d > 0$.

and let $N^m$ be the last nonempty set in the sequence. By definition, each set $N^\eta$ consists of players whose shortest distance to a player from the set $N^0$ is $\eta$. Given that $g^*$ is connected, any player is at a finite distance from any other player. Each player thus belongs to some $N^\eta$. Hence, the family $\{N^\eta\}_{\eta=0}^m$ is a partition of the set of players $N$. For each player $i$ define $\eta(i)$ to be such that $i \in N^{\eta(i)}$.

**Definition 4.3** Consider an $n$-player dilemma game $\Gamma(k)$ with $d > 0$. Let $g^*$ be a connected linking-proof network and let $\{N^\eta\}_{\eta=0}^m$ be the corresponding partition of $N$. Assume that $N^0$ is non-empty. For each $i \in N^0$ let $g^i$ be a linking-proof network such that $l_i(g^i) < l_i(g^*)$. A recursive trigger strategy profile for $\Gamma^T(k)$, based upon $(g^*, \{g^i\}_{i \in N^0}, (t^\eta)_{\eta=0}^m)$, $0 < t^0 < ... < t^m < T$, is a strategy profile denoted by $\rho(g^*, g^*, \{g^i\}_{i \in N^0}, T, (t^\eta)_{\eta=0}^m)$ and given by

outcome path: establish $g^*$ in all periods, player $i$ cooperates during $\{1, ..., T - t^\eta(i)\}$ and defects otherwise,

threat for $N^0$: if a player $i \in N^0$ defects early, i.e. during $\{1, ..., T - t^0\}$, all players defect and establish $g^i$ during the remaining periods,

threat for $N^\eta$: if a player $i \notin N^0$ defects early, i.e. during $\{1, ..., T - t^\eta(i)\}$, all players defect and establish $g^*$ during the remaining periods,

Such set exists because the number of players is finite and all sets $L^\eta$ with $\eta < m$ are nonempty.
simultaneous deviations: if several players simultaneously defect early exercise
the threat for the deviating player with the lowest index.

The recursive trigger profile is clearly linking-proof. It is also cooperative
because all players cooperate in a connected network during \{1, ..., T - t_m\}. The
following theorem characterizes sufficient conditions for existence of a subgame
perfect recursive trigger profile.

**Theorem 4.1** Consider a dilemma game \( \Gamma(k) \) with \( d > 0. \) Let \( g^* \) be a con-
nected linking-proof network. If \( l_i(g^i) < l_i(g^*) \) for some player \( i \) and some
linking-proof network \( g^i \) then there exists a positive number \( T^* \) such that for
any integer \( T > T^* \) a subgame perfect recursive trigger profile for \( \Gamma^T(k) \) can be
constructed that is (i) subgame perfect, (ii) linking-proof, and (iii) cooperative,
selecting the outcome \( (C, g^*) \) during the periods \( 1, ..., T - T^* \).

**Proof.** The complete proof is given in the appendix to this chapter. However,
we provide here an outline of the proof.

All threats of the recursive trigger profile are credible, because they consist
of a repetition of a static equilibrium. We thus need to verify that there exist
positive integers \( t_0, ..., t_m \) such that all threats are effective.

From the perspective of a player \( i \), such that \( 0 < \eta(i) < m \), the outcome
path of a recursive trigger strategy profile \( \rho(g^*, g^*, \{g^i\}_{i \in N^\eta}, T, \{t^\eta\}_{\eta=0}) \) can be
divided into four parts as follows: (p1) during periods \( \{1, ..., T - t^\eta(i)\} \) player \( i \)
and all her neighbors cooperate; (p2) during periods \( \{T - t^\eta(i) + 1 + 1, ..., T - t^\eta(i)\} \)
her neighbors in \( N^\eta(i) \) defect, but player \( i \) and her other neighbors continue
to cooperate; (p3) during periods \( \{T - t^\eta(i) + 1, ..., T - t^\eta(i) - 1\} \) player \( i \) defects
and free-rides on her neighbors in \( N^\eta(i) \) who still cooperate; and finally (p4)
during periods \( \{T - t^\eta(i) - 1 + 1, ..., T\} \) player \( i \) and all her neighbors defect.

If player \( i \) defects during (p1) or (p2) then all her neighbors (and all other
players) defect in all subsequent periods, and \( i \) thus looses the payoff she would
otherwise have earned by free-riding on some of her neighbors during (p3). This
loss increases with the length of part (p3),

\[
\Delta^\eta(i) = (T - t^\eta(i) - 1) - (T - t^\eta(i)) = t^\eta(i) - t^\eta(i) - 1,
\]

in a linear manner. Hence, player \( i \) will not defect during (p1) or (p2) if the
length of part (p3) is so large that no unilateral deviation of player \( i \) in part
(p1) or (p2) exceeds the potential loss of payoff during part (p3). We show in
the appendix that, given a finite length of part (p2), a finite minimal length of
period (p3), \( \Delta^\eta(i) \), exists for each \( i \) such that the threat of punishment for any
her unilateral deviation is effective. No player in \( N^\eta \) can profit from deviating
during her part (p1) if

\[
\Delta^\eta \geq \max_{i \in N^\eta} \Delta^\eta(i).
\]

A player \( j \in N^m \) faces only the parts (p1), (p3) and (p4) but not part
(p2). Therefore, there exists the finite minimal length of period (p3), \( \Delta^\eta(j) \),
such that the threat of punishment for any her unilateral deviation is effective,
independently of $T$ and all $t^m$. No player in $N^m$ can profit from deviating during her part (p1) if (4.8) is satisfied for $\eta = m$.

Given that $\Delta^*(j)$ for each $j \in N^m$ is finite, there exists a finite $\Delta_m$ that satisfies 4.8 for $\eta = m$. Given a finite $\Delta_m$ there exists a finite $\Delta^*(i)$ for each $i \in N^{m-1}$, and thus a finite $\Delta_{m-1}$ that satisfies 4.8 for $\eta = m - 1$. Using this argument recursively, we prove in the appendix that a sequence $(\Delta_\eta)_{\eta=1}^m$ of finite integers exists such 4.8 is satisfied for each $\eta \geq 1$.

A player $i \in N^0$ faces only the parts (p1), (p2) and (p4) but not part (p3). If she deviates during (p1) or (p2) she is subsequently punished by exclusion during the remaining periods. This threat is effective if the number of periods during her part (p4), $\Delta_0 = t^0$ is sufficiently long. We prove in the appendix that, given a finite $\Delta_1$ a finite $\Delta_0$ exists such that no player $i \in N^0$ can profit from deviating during her parts (p1) or (p2).

Given this $\Delta_0 = t^0$ and the finite $(\Delta_\eta)_{\eta=1}^m$ that satisfy (4.8), the sequence $(t^\eta)_{\eta=0}^m$ can now be obtained by recursive application of (4.7), such that $\rho(g^*, g^*, \{g^*_i\} \in N^0, T, (t^\eta)_{\eta=0}^m)$ is subgame perfect for any $T > t^m$.

Conditions of Theorem 4.1 are much weaker than those of Proposition 4.8. Those of Proposition 4.8 require that a network is constructed such that each player can be punished by exclusion. In contrast, conditions of Theorem 4.1 require only a connected network such that at least one player can be punished by exclusion. Proposition 4.10 below shows that for a vast majority of linking constraints each connected linking-proof network satisfies conditions of Theorem 4.1.

Proposition 4.10 Consider an n-player dilemma game $\Gamma(k)$ with $d > 0$. If (i) $k_i \geq 2$ for each $i$ and (ii) $k_j \leq n - 2$ for some $j$ then connected linking-proof networks exist, and each satisfies conditions of Theorem 4.1.

Proof. Given in the appendix to this chapter.

The condition that network $g^*$ be connected is not necessary for the conclusions of Theorem 4.1. If $g^*$ is disconnected then the sufficient condition for existence of a subgame perfect recursive trigger profile is that at least one player from each connected component of $g^*$ belongs to $N^0$. That is, in each connected component of $g^*$ it must be possible to punish at least one player by exclusion. To construct a cooperative, linking-proof and subgame perfect profile one may then apply the construction of a strategy profile, outlined in the proof above, to each of the connected components.

4.4 Unilateral link formation

In this section we consider the unilateral link formation model. For a profile of linking choices $p$ the network of established links $g(p)$ is defined by $g_{ij}(p) = \max\{p_{ij}, p_{ji}\}$. Let

$$m_i(p) = \sum_{j \in N} p_{ij}$$
be the number of links proposed by player $i$. By the definition of the model, it is assumed that a player accepts any number of links proposed by other players, but can herself propose only a limited number of links. In particular, a player can establish more links than she can herself propose, $l_i(g(p)) \geq m_i(p)$, and may have any number of neighbors. Let

$$\mathcal{H}(k) = \{g(p) \mid m_i(p) \leq k_i \text{ for each } i\}$$  \hspace{1cm} (4.9)

be the set of feasible networks.

In this section we use directed graphs in illustrations of networks. Links are illustrated by directed arrows instead of undirected lines. The direction of an arrow describes which of the linked players proposed the link. It is important to remember that, in spite of this illustration, the interaction in the network is two-way. A proposal of only one player is needed to establish a link. However, once a link is established, both linked players earn their payoffs according to their actions in the prisoner’s dilemma game.

To avoid trivialities we assume that $c > 0$, that is, the value of cooperation is higher than the outside option. We also assume that $k_i \geq 1$ for at least one player, and for at least two players if the outside option is high.$^{14}$

### 4.4.1 Stage game equilibria and equilibrium networks

We say that a linking profile $p \in P(k)$ is maximal if, for each player $i$ either $l_i(g(p)) = n - 1$, or

$$m_i(p) = k_i \text{ and } p_{ij}p_{ji} = 0 \text{ for all } j.$$  

If a linking profile is maximal no player can unilaterally increase the number of her established links.

We say that linking profile $p \in P(k)$ is redundant if $p_{ij} = p_{ji}$ for all $i$ and $j$, that is, if each link is proposed by both players. The following proposition gives a complete characterization of Nash equilibria of game $\Gamma(k)$. Again, whether the outside option is low or high crucially determines the set of equilibrium networks.

**Proposition 4.11** In each Nash equilibrium $(a^*, p^*)$ of a game $\Gamma(k)$ all players defect, i.e. $a^* = D$.

1. Let $d < 0$. The profile $(D, p) \in J(k)$ is a Nash equilibrium of $\Gamma(k)$ if and only if $p$ is redundant.

2. Let $d > 0$. The profile $(D, p) \in J(k)$ is a Nash equilibrium of $\Gamma(k)$ if and only if $p$ is maximal.

$^{14}$The following are equilibria of a dilemma game with $d < 0$ and $k_j = 0$ for all but one player $i$. Two Nash equilibria exist; in both no links are proposed and all players $j \neq i$ defect but player $i$ can either cooperate or defect. The two equilibrium profiles are payoff-equivalent and the empty network is established in all periods of any subgame-perfect equilibrium of the repeated dilemma game.
3. If $d = 0$ then any profile $(D, p) \in J(k)$ is a Nash equilibrium of $\Gamma(k)$.

Proof. Let $(a^*, p^*)$ be a Nash equilibrium of $\Gamma(k)$ and let $N_C = \{i | a^*_i = C\}$ be the set of cooperative players. Each player $i \in N_C$ is isolated as otherwise defection would strictly increase her payoff. Let $d > 0$ and let $k_i > 0$. If $i \in N_C$ then this player could strictly increase the payoff by defecting and proposing the link to any other player. In equilibrium this should not be possible, hence $i \notin N_C$. Now, $i$ can strictly increase her payoff by proposing the link to any player in $N_C$, hence $N_C$ must be empty. Consider now $d < 0$ and let $i \in N_C$ and $j \neq i$ such that $k_j > 0$. Player $j$ could strictly increase her payoff by defecting, removing all her proposed links and proposing the link to $i$. Again, this implies that $N_C$ is empty.

(1) Let $d < 0$. Consider $(D, p)$ with redundant $p$. If all links are proposed by both players then no player can change her payoff by removing some of her links, while players decrease their payoff by adding links. Hence, $(D, p)$ is a Nash equilibrium. On the other hand, if some link is proposed only by one of the two players, this player could increase her payoff by removing the link. If $(D, p)$ is a Nash equilibrium, $p$ must be redundant.

(2) Let $d > 0$. Let $(D, p)$ be a Nash equilibrium and let $i$ be such that $l_i(g(p)) < n - 1$. It must be that $m_i = k_i$ and $p_{ij}p_{ji} = 0$ for all $j$, as otherwise $i$ could increase her payoff by proposing more links or moving some of the links already proposed by another player. Hence, $p$ is maximal. Consider now $(D, p)$ and let $p$ be maximal. If all players defect then a player can increase her payoff only by adding links. If $p$ is maximal then no player can unilaterally increase the number of her links. Hence, $(D, p)$ is Nash equilibrium.

(3) If $d = 0$ then players earn the same payoff in each $(D, p) \in J(k)$. ■

If $(D, p^*) \in J(k)$ is a Nash equilibrium of $\Gamma(k)$ we say that $g(p^*)$ is an equilibrium network for $d$. There are many redundant linking profiles and many maximal ones. Hence, for any value of $d$ several equilibrium networks can exist.\(^\text{15}\)

Assume $d < 0$. Under the mutual link formation model only the empty network is established in a Nash equilibrium. The reason for existence of nonempty equilibrium networks under unilateral link formation is that unilateral exclusion is not possible. If two players propose the link to each other each of them is indifferent between proposing the link or not. Two players separate only if both mutually exclude each other. However, the equilibrium in which all players defect and no links are proposed is the unique strict Nash equilibrium.

Example 4.10 Consider dilemma game $\Gamma(k)$ with $n = 4$ and $k = (1, 1, 1, 1)$. Selected equilibrium networks for the game $\Gamma(k)$ are illustrated in Figure 4.6. The star network $g^s$, the wheel network $g^w$, and the flower network $g^{f_{ij}}$ are equilibrium networks when $d > 0$. The empty network $g^e$ and the pairs networks $g^{p_{ij}}$ and $g^p$ are equilibrium networks when $d < 0$. Note that each link in $g^{p_{ij}}$ and $g^p$ is proposed by both linked players. Other equilibrium networks exist.

\(^{15}\)In fact, for $d < 0$ the set of equilibrium networks coincides with $G(k)$, defined by (4.4).
4.4.2 Cooperation in subgame perfect equilibria of the repeated game

Just as in section 4.3 above we can construct cooperative subgame perfect equilibria of the repeated dilemma game with $d > 0$, using threats with exclusion of early defectors. Such threats are effective if there exists an equilibrium network $g^*$ in which each player has more links than she has in her specific punishment equilibrium network $g^i$. In proposition 4.12 below we give sufficient conditions on the vector of linking constraints $k$ for the existence of networks $g^*$ and $\{g^i\}_{i \in N}$ that satisfy this requirement.

Furthermore, cooperative subgame perfect equilibria can be constructed also for dilemma games with $d < 0$. This is possible because non-empty equilibrium networks exist.

Recall Definition 2.5 of the trigger strategy profile $\sigma(r, q^*, \{q^i\}_{i \in N}, T, t^*)$ and define

$$\phi(p, p^*, \{p^i\}_{i \in N}, T, t^*) = \sigma((C, p), (D, p^*), \{(D, p^i)\}_{i \in N}, T, t^*).$$ (4.10)

Along the equilibrium path of $\phi(p, p^*, \{p^i\}_{i \in N}, T, t^*)$: all players cooperate and establish network $g(p)$ during the early periods $1, \ldots, T - t^*$; and establish network $g(p^*)$ and defect during the remaining periods. Any deviation during the early periods triggers a change of the network: all players defect and establish $g(p^i)$ forever, where $i$ is one of the players which deviated. Profile $\phi(p, p^*, \{p^i\}_{i \in N}, T, t^*)$ is cooperative by definition, and can be subgame perfect only if $g(p^*)$ and all $g(p^i)$ are equilibrium networks.
Proposition 4.12 Consider dilemma game $\Gamma(k)$ with $d > 0$ and let $1 \leq k_i \leq \frac{n}{2} - 1$ for each player $i$. There exist maximal profiles $p^*, \{p^i\}_{i \in N} \in P(k)$ such that (i) $l_i(g(p^*)) > k_i$, and (ii) $p^i_{ji} = 0$ for all $i$ and $j$. Furthermore, for any maximal profile $p \in P(k)$ there exists $\gamma$ such that the cooperative strategy profile $\phi(p, p^*, \{p^i\}_{i \in N}, T, t^*)$ is subgame perfect for $\Gamma^T(k)$ for all $T, t^*$ such that $T > t^* \geq \gamma$.

Proof. Given in the appendix to this chapter. 

Let $d > 0$. Proposition 4.12 gives sufficient conditions for the existence of equilibrium networks $g(p^*), \{g(p^{f:i})\}_{i \in N}$ such that (i) in $g(p^*)$ each player $i$ establishes more than $k_i$ links, and (ii) in $g(p^{f:i})$ player $i$ establishes only the $k_i$ links proposed by herself. If any player $j$ deviates from the equilibrium path of $\phi(p, p^*, \{p^{f:i}\}_{i \in N}, T, t^*)$ she is punished when other players remove their links with her. The following example demonstrates a construction of such strategy profile. Let $i' = i$ for $i \in \{1, \ldots, n\}$ and let $n+1 = 1, n+2 = 2, \text{ and } \overrightarrow{0} = n$.

Example 4.11 Consider dilemma game $\Gamma(k)$ of example 4.10, with $n = 4$ and $k = (1, 1, 1, 1)$, and let outside option be low, $d > 0$. Let the game be repeated twice and let $f < c+d/2$. Consider the following strategy for player $i$: "Cooperate in the first and defect in the second period. In both periods propose a link to player $i+1$, unless $i+1$ was the only defector in the first period, in which case propose the link to player $i+2$ in the second period." If all players follow this strategy, or if more than one player defects in the first period, the wheel network $g^w$ is established in both periods. If in the first period player $i$ is the only to defect, player $i-1$ relocates her link and the flower network $g^{f:i,i+1}$ is established. See Figure 4.7 for an illustration.

Conditions of proposition 4.12 are not necessary for existence of a cooperative and subgame perfect equilibrium for a dilemma game with $d > 0$. Just as in section 4.3, recursive trigger strategies may be used to construct cooperative subgame perfect equilibria even if this is not possible with the trigger strategies (4.10). The following proposition shows that such construction is possible for almost all dilemma games.

Theorem 4.2 Consider dilemma game $\Gamma(k)$ with $d > 0$. If (i) $1 \leq k_j$ for all $j$ and (ii) $k_i, k_{i'} < n - 1$ for some distinct $i$ and $i'$, then there exists a positive number $T^*$ such that for any integer $T > T^*$ a cooperative subgame perfect profile for $\Gamma^T(k)$ can be constructed.

Proof. It is straightforward to define sets $N^0, N^1, \ldots, N^m$ and rewrite definition 4.3 of recursive trigger strategy profiles for the set of equilibrium profiles of linking choices $F(k, d)$. The proof of Theorem 4.1 can then be applied to show that a subgame perfect recursive trigger profile for $\Gamma^T(k)$ with sufficiently large

---

16 This characterizes a version of the indexing function $(\cdot)$, defined by (4.11) in the appendix to this chapter.
Figure 4.7: An illustration of a cooperative and subgame perfect strategy profile for the repeated game of example 4.10, with \( n = 4, k = (1, 1, 1, 1), d > 0 \) and \( f < c + d/2 \).

\( T \) exists if \( l_i(g(p^{i})) < l_i(g(p^*)) \) for some \( i \) and some equilibrium profiles \( p^* \) and \( p^i \) such that \( g(p^*) \) is connected. We show that there exist such \( p^* \) and \( p^i \) in the continuation of this proof, given in the appendix to this chapter.

Finally, trigger strategies (4.10) can be used to construct cooperative subgame perfect equilibria for dilemma games with \( d < 0 \). Let \( p^e \) be the profile in which no links are proposed and an empty network is established, that is, \( p^e_{ij} = 0 \) for all \( i,j \). Let \( p^{p^i} \) be the redundant profile in which the link between players \( i \) and \( i + 1 \) is mutually proposed and no other links are proposed, that is, \( p^{p^i}_{i,i+1} = 1 \) and \( p^{p^i}_{ij} = 0 \) otherwise.

**Proposition 4.13** Consider dilemma game \( \Gamma(k) \) with \( d < 0 \). If \( k_i \geq 1 \) for each player \( i \), then there exists \( p \in P(k) \) and \( \gamma \) such that cooperative trigger profile \( \phi(p,p^*,\{p^{p^i}\}_{i \in N,T,t^*}) \) is subgame perfect for \( \Gamma_T(k) \) for all \( T,t^* \) such that \( T > t^* \geq \gamma \).

**Proof.** If \( k_i \geq 1 \) for each player \( i \) there exists a profile \( p \in P(k) \) such that the network \( g(p) \) is connected. Consider, for example, any profile establishing a star network \( g^s \). Clearly \( \pi_i(C,p) > 0 = \pi_i(D,p^e) > \pi_i(D,p^{p^i}) \) for each \( i \). The proposition then follows from Theorem 2.1.

If linking is unilateral, then exclusion has to be mutual. By consequence non-empty equilibrium networks exist even if \( d < 0 \). Hence, cooperative subgame perfect equilibria may be constructed where any early defection is punished by
inclusion into one of the non-empty equilibrium networks.\footnote{This, however, crucially relies on the active participation of a punished player in her own punishment.}

**Proposition 4.14** Consider dilemma game $\Gamma(k)$ with $d > 0$ and no linking constraints, i.e. $k_i = n - 1$ for each $i$. All subgame perfect equilibria of $\Gamma^T(k)$, for any $T$, are defective.

**Proof.** If $k = n - 1$ and $d > 0$ then a profile is maximal only when $l_i(g(p)) = n - 1$ for each $i$. Hence, following Proposition 4.11, the complete network is the unique equilibrium network. All players defect in a static equilibrium. Hence, all static equilibria are payoff-equivalent. The assertion of the proposition is now implied by Lemma 2.1. 

### 4.5 Network dilemma game with linking costs

In the sections above we assumed that linking is exogenously constrained. We argue in the introduction that such an assumption is not too artificial. Social relations, for example, have to be nurtured frequently, which requires time. The amount of time which can be spent on relations is limited and people cannot support any number of relations. In a sense, we have modeled a world in which each player has a limited amount of available time to be spent on time-consuming relations.

In this section we explicitly model the relation between the number of links and the associated costs. Throughout the section one can think of these costs as the opportunity costs of devoting a part of available time to exchange relations, rather than to some other potentially profitable activity. In particular, we assume that players are not constrained in the number of links, but that each link is costly.

A cost of a link can be incurred by both linked players or only by one of them. We say that a player sponsors a link if she bears its costs. In particular, we assume that the player sponsors a link if she proposes and establishes the link. The cost of a new link increases with the number of links already sponsored but is otherwise independent of the identities of the players it links. The benefit of having a new link, however, depends on the behavior of the new neighbors, captured here by their actions in the prisoner’s dilemma game.

In terms of opportunity costs, imagine that by sponsoring a new link the player forgoes the value of some alternative activity. When deciding whether to add a link the player compares its expected benefit with the value of the least profitable activity among those she would undertake if she does not establish any new links. The larger the number of established links, the smaller the set of activities pursued in the remaining time, and the more valuable is the cheapest of them. In this sense, the unit of available time becomes more valuable when available time is scarce. The opportunity cost of sponsoring a new link is therefore independent of who the link is established with, but increases with the number of links already sponsored.
Figure 4.8: Illustration of a cost function $b_i$ with an increasing marginal cost $\Delta b_i$. (a) If all players defect, player $i$ earns $d$ from each sponsored link and is willing to sponsor up to $k_i$ links. If, however, all players cooperate, player $i$ earns $c$ from each sponsored link and is willing to sponsor up to $k_i \geq k_i$ links. (b) The cost function $b_i$ is convex because the marginal cost $\Delta b_i$ is increasing.

In general, sponsoring is represented by a function $x_i : P \rightarrow \{1, ..., n - 1\}$, which counts the number of links sponsored by player $i$ given the profile of proposed links. Assuming that a player sponsors only links which she both proposed and established, sponsoring for the two models of link formation is defined as follows:

**MUTUAL LINK FORMATION:** Player must have proposed each link that she established. Hence, she sponsors each of her established links: $x_i(p) = l_i(g(p))$.

**UNILATERAL LINK FORMATION:** Player established each link she proposed. Hence, she sponsors each of her proposed links: $x_i(p) = m_i(p)$.

Let $b_i(x_i)$ be the cost that player $i$ incurs if she sponsors $x_i$ links. We assume that for each player $i$ her cost function $b_i : \{1, ..., n - 1\} \rightarrow \mathbb{R}$ is non-decreasing and that $b_i(0) = 0$. We also assume that for each $i$ the marginal cost of establishing the $x_i$-th link, $\Delta b_i(x_i) = b_i(x_i) - b_i(x_i - 1)$, is not decreasing with $x_i$, which implies that $b_i$ is a convex function. For convenience we assume that there is no $i$ and $x_i$ such that $\Delta b_i(x_i) \in \{c, d, e, f\}$.

Given the link formation model $g(\cdot)$ and the sponsoring model $x(\cdot)$ the payoff function for player $i$ is defined by

$$\varphi_i(a, p) = \sum_{j \in L_i(p)} v(a_i, a_j) - b_i(x_i(p)).$$

We refer to the stage game $\Omega = \langle N, J, \varphi \rangle$ as the dilemma game with linking costs.
4.5.1 Stage game equilibria

The theory developed in the previous sections for dilemma games with linking constraints is very useful and instrumental in the analysis of the dilemma game with linking costs. We first characterize the close relation between the Nash equilibria of dilemma game with linking costs and the Nash equilibria of an associated dilemma game with linking constraints. Then we show that this relation extends also to subgame-perfect equilibria of the repeated dilemma games.

Let
\[ k_i = \max \{ x_i \mid \Delta b_i(x_i) < d \} \]
be the maximal number of links player \( i \) is willing to support if all players defect. We refer to the vector \( k = (k_1, \ldots, k_n) \) as the minimal linking support. In a Nash equilibrium of the dilemma game with linking costs each non-isolated player defects. Hence, no player \( i \) is willing to support more than \( k_i \) links. The following Proposition shows that, in a Nash equilibrium, players behave as if their linking was constrained by \( k \).

Proposition 4.15 Fix the prisoner's dilemma game payoffs \( v \). Consider a dilemma game with linking costs \( \Omega = \langle N, J, \varphi \rangle \), with the minimal linking support \( k \). Let \( \Gamma(k) = \langle N, J(k), \pi \rangle \) be the dilemma game with linking constraints \( k \).

1. If \( d < 0 \) then, for both mutual and unilateral link formation models, only the empty network is established in Nash equilibria of \( \Omega \).

2. If \( d > 0 \) then, for both mutual and unilateral link formation models, the set of networks established in Nash equilibria of \( \Omega \) coincides with the set of networks established in Nash equilibria of \( \Gamma(k) \). In particular,

a. for the mutual link formation model: the profile \( (a, p) \) is a Nash equilibrium of \( \Omega \) if and only if \( (a, p') \) is a Nash equilibrium of \( \Gamma(k) \) for some \( p' \in J(k) \) such that \( g(p) = g(p') \),

b. for the unilateral link formation model: the profile \( (a, p) \) is a Nash equilibrium of \( \Omega \) if and only if it is a Nash equilibrium of \( \Gamma(k) \) and \( p_{ij}p_{ji} = 0 \) for each \( i, j \).

Proof. In a Nash equilibrium each player with at least one established link defects. Furthermore, if link formation is unilateral then proposing a link is always costly, and thus no player is willing to propose a link to another player if the other player already proposed that link. In the following we use results of Propositions 4.2 and 4.11.

(1) Let \( d < 0 \). In a Nash equilibrium a player may have proposed a link only if it is not established. Hence, no links are established.

(2.a) Consider mutual link formation and let \( d > 0 \). Let \( (a^*, p^*) \) be a Nash equilibrium of \( \Omega \) and let \( g^* = g(p^*) \) be the corresponding network. All linked players defect, hence no player \( i \) has more than \( k_i \) links. Therefore, \( g^* \in G(k) \).
Then, \((a^*, p(g^*))\) is a Nash equilibrium of \(\Gamma(k)\): \(g^*\) is feasible and if \(a^*_i = C\) for some \(i\) then \(i\) is isolated in \(g^*\). This proves the “only if” part of the claim.

To prove the “if” part of the claim, let \((a^*, p^*)\) be a Nash equilibrium of \(\Gamma(k)\) and let \(g^* = g(p^*)\). Because this is a Nash equilibrium no player wants to remove any links. If \(a^*_i = C\) then \(p^*_{ij} = p^*_{ji} = 0\) for all \(j\). For each player \(i\) such that \(l_i(g^*) < k_i\), it must be that \(p^*_{ji} = 0\) or for all \(j\) with whom \(i\) has no link, as otherwise \(j\) could add a link to \(i\). If \(l_i(g^*) = k_i\) then \(i\) does not want to establish more links with defective players. Hence, no player can or wants to establish more links. The \((a^*, p^*)\) is then also a Nash equilibrium of \(\Omega\).

It follows from the above that the sets of networks established in Nash equilibria of \(\Omega\) and, respectively, of \(\Gamma(k)\), coincide.

(2.b) Consider unilateral link formation and let \(d > 0\). Let \((a^*, p^*)\) be a Nash equilibrium of \(\Omega\) and let \(g^* = g(p^*)\). If \(a^*_i = C\) for some \(i\) then she must be isolated, but then other players can profitably deviate by proposing a link to \(i\). Hence, \(a^* = D\). Consequently, no player proposes more than \(k_i\) links. No link is proposed by two players as otherwise each could profitably deviate by not proposing and still keeping that link. Hence, \(p_{ij}p_{ji} = 0\) for each \(i, j\). If \(l_i(g^*) < n - 1\) for some \(i\) then it must be that \(i\) proposed exactly \(k_i\) links: if she proposed less she could profitably add another. This implies that, in relation to \(\Gamma(k)\), \(p\) is maximal. The profile \((a^*, p^*) = (D, p^*)\) is therefore a Nash equilibrium of \(\Gamma(k)\). By implication, any \(g^*\) established in \(T(k)\) is established in a Nash equilibrium of \(\Omega\) is established in a Nash equilibrium of \(\Gamma(k)\). This proves the “only if” part of the claim.

To prove the “if” part of the claim, let \((a^*, p^*)\) be a Nash equilibrium of \(\Gamma(k)\) such that \(p_{ij}p_{ji} = 0\) for each \(i, j\) and let \(g^* = g(p^*)\). All players defect. No player \(i\) proposes more than \(k_i\) links. Player \(i\) proposes less than \(k_i\) links only if \(l_i(g^*) = n - 1\). Hence, in relation to \(\Omega\), no player can profitably deviate by adding, moving or removing links, thus \((a^*, p^*)\) is a Nash equilibrium of \(\Omega\).

Finally, we prove now that for any \(g^*\) established in a Nash equilibrium \((D, p^*)\) of \(\Gamma(k)\) we can construct a Nash equilibrium \((D, p^{**})\) of \(\Omega\) supporting \(g^*\). Take an arbitrary Nash equilibrium \((D, p^*)\) of \(\Gamma(k)\) and let \(g^* = g(p^*)\). If \(p^*_{ij}p^*_{ji} = 1\) for some \(i, j\) then it must be that \(l_i(g^*) = l_j(g^*) = n - 1\). Let \(p^{**}\) be obtained from \(p^*\) by setting \(p^{**}_{ij} = 0\) for each \(i < j\) such that \(p^*_{ij}p^*_{ji} = 1\), and \(p^{**}_{ij} = p^*_{ij}\) otherwise. The linking profile \(p^{**}\) satisfies (i) \(l_i(g^*) = l_i(g(p^{**}))\) for each player \(i\), (ii) \(p^{**}_{ij} = 0\) for all \(i, j\), and (iii) \(g(p^{**}) = g^*\).

If \(m_i(p^{**}) < m_i(p^*)\) for some \(i\) then it must be that she is one of players who double-proposed a link in \(p^*\), and hence \(l_i(g^{**}) = l_i(g^*) = n - 1\). If \(m_i(p^{**}) = m_i(p^*) < k_i\) it must also be that \(l_i(g^*) = n - 1\), because \((D, p^*)\) is a Nash equilibrium of \(\Gamma(k)\). Hence, \(l_i(g^{**}) = l_i(g^*) = n - 1\) for all \(i\) such that \(m_i(p^{**}) < k_i\). This proves that \((D, p^{**})\) is a Nash equilibrium of \(\Gamma(k)\), and therefore also of \(\Omega\), such that \(p_{ij}p_{ji} = 0\) for each \(i, j\). ■

For the mutual link formation model there is a similar relation, between the dilemma games with linking costs and with linking constraints, with regard to the linking-proof networks. We define linking-proofness for dilemma games with linking costs in line with Definition 4.1.
Definition 4.4 Consider a mutual link formation model. A profile of moves \((a^*, p^*) \in J\) is a **linking-proof equilibrium** (LP equilibrium) of the game \(\Omega = (N, J, \varphi)\) if (i) for each player \(i\) the move \((a_i^*, p_i^*)\) is a best response to \((a^*, p^*)\), and (ii) for each pair of players \(i, j\) either \((a_i^*, p_i^* \oplus ji)\) or \((a_j^*, p_j^* \oplus ij)\) is a best response to \((a^*, p^* \oplus ij)\). Network \(g^*\) is **linking-proof** if it is established in a linking-proof equilibrium.

Proposition 4.16 Consider a mutual link formation model. Fix the prisoner’s dilemma game payoffs \(v\). The set of linking-proof networks of a dilemma game with linking costs \(\Omega = (N, J, \varphi)\), with the minimal linking support \(k\), coincides with the set of linking-proof networks of the dilemma game with linking constraints \(\Gamma(k) = (N, J(k), \pi)\).

**Proof.** In the following we use results of Proposition 4.3. Let \(d < 0\). The Nash equilibrium \((D, 0)\) of \(\Omega\) is linking-proof as no player wants to reciprocate a link proposed by a defector. Hence the empty network is the unique LPN of \(\Omega\). It is also the unique LPN of \(\Gamma(k)\).

Consider now \(d > 0\). Let \(g^*\) be a LPN for \(\Omega\). As it is established in a Nash equilibrium of \(\Omega\), it is also established in a Nash equilibrium of \(\Gamma(k)\), and thus \(g^* \in G(k)\). Take a pair of players \(i \neq j\) such that \(l_i(g^*) < k_i\) and \(l_j(g^*) < k_j\). They must have an established link: if not, each would strictly prefer to add the link given that it is reciprocated, and \(g^*\) would not be LPN for \(\Omega\). Hence, there is no pair of such separated players, which implies that \(g^*\) is LPN for \(\Gamma(k)\).

Let now \(g^*\) be a LPN for \(\Gamma(k)\). Consider the profile \((D, p(g^*))\). This profile is a Nash equilibrium of \(\Gamma(k)\) and, therefore, of \(\Omega\). No player can profit from relocating a link because all players defect and only player \(i\) such that \(l_i(g^*) < k_i\) wants to add a link. However, there are no separated players \(i\) and \(j\) such that \(l_i(g^*) < k_i\) and \(l_j(g^*) < k_j\) (\(g^*\) is LPN of \(\Gamma(k)\)), hence \((D, p(g^*))\) is a linking-proof equilibrium of \(\Omega\).

4.5.2 Cooperation in subgame perfect equilibria of the repeated game

In sections 4.3 and 4.4 we have shown that, given the sufficient variety of equilibrium networks, it is possible to construct cooperative and subgame perfect equilibria for finitely repeated dilemma games with linking constraints. Propositions 4.15 and 4.16 assert that the variety of equilibrium networks in dilemma games with linking costs is related to that in dilemma games with linking constraints. Hence, there should also be a relation between the sets of subgame perfect equilibria of different dilemma games. We characterize this relation next.

Proposition 4.17 Fix the prisoner’s dilemma game payoffs \(v\). Consider a dilemma game with linking costs \(\Omega = (N, J, \varphi)\), with the minimal linking support \(k\). Let \(\Gamma(k) = (N, J(k), \pi)\) be the dilemma game with linking constraints \(k\).
1. If $d \leq 0$ then, for both mutual and unilateral link formation models and for any $T$, the empty network is established in each period along the equilibrium path of any subgame perfect equilibrium of $\Omega^T$.

2. Let $d > 0$.

(i) Consider the mutual link formation model. If the cooperative trigger strategy profile $\rho(g^*, g^*, \{g^i\}_{i \in N}, T', t')$ is subgame perfect for $\Gamma^T(k)$ for some $T' > t'$, then there exists a finite $\gamma$ such that a cooperative trigger strategy profile $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ for $\Omega^T$ is subgame perfect for all $T > t^* \geq \gamma$.

(ii) Consider the mutual link formation model. If the cooperative trigger strategy profile $\rho(g^*, g^*, \{g^i\}_{i \in N}, T', t')$ is subgame perfect and linking-proof for $\Gamma^T(k)$ for some $T' > t'$, then there exists a finite $\gamma$ such that a cooperative, linking-proof, trigger strategy profile $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ for $\Omega^T$ is subgame perfect for all $T > t^* \geq \gamma$.

(iii) Consider the unilateral link formation model. If for some $T' > t'$ there exists a subgame perfect and cooperative trigger strategy profile $\phi(p^*, p^*, \{p^i\}_{i \in N}, T', t')$ for $\Gamma^T(k)$, then there exists a finite $\gamma$ such that a cooperative trigger strategy profile $\phi(p^*, p^*, \{p^i\}_{i \in N}, T, t^*)$ for $\Omega^T$ is subgame perfect for all $T > t^* \geq \gamma$.

Proof. (1) Let $d \leq 0$. Following Lemma 2.1, since all Nash equilibria involve the empty network and are payoff-equivalent, one must be selected in each period along the equilibrium path of any subgame perfect equilibrium.

(2.i) Let $d > 0$. If $\rho(g^*, g^*, \{g^i\}_{i \in N}, T', t')$ is subgame perfect for $\Gamma^T(k)$, then $(D, p(g^*))$ and all $(D, p(g^i))$ are Nash equilibria of $\Gamma^T(k)$ and, consequently, of $\Omega$. Furthermore, if threats are effective, then for each $i$, $\pi_i(D, p(g^*)) > \pi_i(D, p(g^i))$, implying $l_i(g^*) > l_i(g^i)$. As $g^* \in G(k)$ it must be that $l_i(g^i) < k_i$ for each $i$. Consider now the payoff function with linking costs, $\varphi_i$. By definition of $k_i$, given that all players defect, player $i$ strictly prefers having $x_i$ links from having $x_i - 1$ as long as $x_i \leq k_i$. Hence, $\varphi_i(D, p(g^*)) > \varphi_i(D, p(g^i))$ for each $i$.

Following Theorem 2.1, any profile $(a, p)$ such that $\varphi_i(a, p) \geq \varphi_i(D, p(g^*))$ can be sustained in the early periods of the game $\Omega^T$ if $T$ is sufficiently large. Profile $(C, p(g^*))$ is one such profile. Hence, there exists a $\gamma$ such that $\rho(g^*, g^*, \{g^i\}_{i \in N}, T, t^*)$ is subgame perfect for $\Omega^T$ for all $T > t^* \geq \gamma$.

(2.ii) and (2.iii) are proven following the same steps as above.

Proposition 4.17 can be summarized as follows: a cooperative and subgame perfect trigger strategy for a dilemma game with linking costs, with the minimal linking support $k$, repeated sufficiently many times, exists if it exists for a finitely repeated dilemma game with linking constraints $k$. We can now use results of Propositions 4.8 and 4.12 to describe several classes of cost functions which permit construction of cooperative subgame perfect equilibria for finitely repeated dilemma games with linking costs.

72
Proposition 4.18 Let $\Omega = (N, J, \varphi)$ be an $n$-player dilemma game with linking costs and $d > 0$.

1. Consider the mutual link formation model. A (i) cooperative, (ii) linking-proof, and (iii) subgame perfect equilibrium for $\Omega^T$ with sufficiently large $T$ exists if $\Delta b_i(2) < d$ for each $i$ and $d < \Delta b_j(n - 2)$ for some $j$. If, in addition, any of the following two conditions holds then such an equilibrium can be constructed using trigger strategies:

   a. $b_i = b$ for each $i$ and $d < \Delta b(n - 1)$, or
   b. $d < \Delta b_i(\sqrt{2n - 9})$ for each $i$.

2. Consider the unilateral link formation model. A cooperative and subgame perfect equilibrium for $\Omega^T$ with sufficiently large $T$ exists if $\Delta b_i(1) < d$ for each $i$ and $d < \min\{\Delta b_j(n - 2), \Delta b_j(n - 2)\}$ for some distinct $j$ and $j'$. If, furthermore, $d < \Delta b_i(\frac{n}{2} - 1)$ for each $i$ then such equilibrium can be constructed using trigger strategies.

Proof. Conditions in Propositions 4.8, 4.10 and 4.12 and in Theorem 4.2 are given in terms of $k$. Condition $k_t < y$ can be rewritten as $\Delta b_i(y) > d$ and condition $k_t > y$ as $\Delta b_i(y) < d$. Now apply assertions of propositions 4.8, 4.12 and 4.17 to prove the statements about the trigger strategies. Next we prove the statements about general (no-trigger) strategy profiles.

Combining Proposition 4.17 with either Proposition 4.10 or Theorem 4.2 we can conclude that for some player $i$ there exists (linking-proof) equilibrium profiles $(D, p^*)$ and $(D, p^I)$ such that $h_i(g(p^I)) < h_i(g(p^*))$. Now we can define sets $N^0, \ldots, N^m$ given $g(p^*)$, in the same way as in section 4.3.3, and construct recursive trigger strategy profiles in line with constructions in the proofs of Theorems 4.1 and 4.2. □

Cooperation in a finitely repeated dilemma game with linking costs can therefore be sustained whenever the linking costs are such that players are willing to establish a first few links even with defectors, but want to add more links only if other players begin to cooperate. We conclude this section with a couple of examples.

Example 4.12 Let $b^k$ be the linking cost function such that $b^k(x) = 0$ for $x = 0, 1, \ldots, k$ and $b^k(x) = \infty$ for $x > k$. It is an extreme example of a cost function, as the first $k$ links are costless, but all subsequent links are costlier than any potential benefit. The $n$-player dilemma game with linking constraints $k$ is similar to the $n$-player dilemma game with linking costs $b_i = b^k$: in the latter no rational player $i$, optimizing her total payoff, would ever establish more than $k_i$ links. In fact, both the one-shot and the repeated game equilibrium payoffs coincide between the two games.

Consider an arbitrary dilemma game with linking costs. If all players cooperate they may be willing to establish more than $k$ links. Namely, by the
definition of the prisoner’s dilemma game, the benefit of mutual cooperation $c$ is higher than the benefit of mutual defection $d$. Hence, given that all players cooperate, some player $i$ may find that the benefit of cooperation exceeds the cost of the $(k_i + 1)$-st link. Consequently, networks among cooperative players may have more links than the networks among defective players. Figure 4.8, discussed earlier, illustrates an instance of a cost function of player $i$ such that when all players cooperate she is willing to support more than $k_i$ links.

**Example 4.13** In this example we demonstrate that there exist subgame perfect equilibria of the repeated dilemma game with linking costs such that in the early periods all players cooperate and establish a network with more links than in any equilibrium network. We consider $n = 4$ and the prisoner’s dilemma game of Figure 2.1.

(a) Let $\Omega$ be the dilemma game with mutual link formation and the following cost function for each player $i$: $b_i(0) = 0; b_i(1) = 1; b_i(2) = 3; b_i(3) = 7$. The minimal linking support is $k = (2, 2, 2, 2)$. The marginal cost of supporting the third link is 4 and a player is willing to support it if all other players cooperate. The linking-proof networks for $\Omega$ are the same as given in Figure 4.1 for the game $\Gamma(2, 2, 2, 2)$. However, if all players cooperate a complete network could be established. Consider the twice repeated game $\Omega$ and take the following strategy: "Cooperate in the first and defect in the second period. Support the complete network in the first period. If in the first period there was no defection, support network $g^1$ in the second period; otherwise support network $g^i$, where $i$ is the defector with the lowest index." See Figure 4.9 for an illustration. The profile of these strategies is subgame perfect for $\Omega^2$. In each period, no pair of separated
In the first period all players cooperate and establish the complete network.  

(b) Consider now the dilemma game \( \Omega \) with unilateral link formation and the following cost function for each player \( i \): \( b_i(0) = 0; b_i(1) = 2; b_i(2) = 6; b_i(3) = 10 \). The minimal linking support is \( k = (1,1,1,1) \), thus the equilibrium networks for \( \Omega \) coincide with those for the game \( \Gamma(1,1,1,1) \). Some equilibrium networks are shown in the upper row of Figure 4.6. If all players cooperate the complete network \( g^c \), shown in Figure 4.10, can also be supported. Let the game \( \Omega \) be repeated three times and take the following strategy: "Cooperate in the first and defect in the remaining periods. Support the network \( g^c \) in the first period. If in the first period there was no defection, support network \( g^w \) in the remaining periods; otherwise support network \( g^{i-1} \), where \( i \) is the defector with the lowest index." See Figure 4.10 for an illustration. The profile of these strategies is cooperative and subgame perfect.

### 4.6 Discussion

#### 4.6.1 Summary of the results

In this chapter we consider finitely repeated \( n \)-player games in which each player in each period selects her interaction partners and chooses whether to cooperate
or to defect. By means of individual partner selection an interaction network is established in each period. A player interacts with her immediate neighbors: she plays a prisoner's dilemma game with each of them. We call this game a network dilemma game.

Players may be constrained in the number of partners they can select, in which case we refer to a game with linking constraints. Interaction can be costly, in which case we refer to a game with linking costs. If no reference to linking constraints or costs is made we refer to general network dilemma games, with or without linking costs or constraints.

We consider two link formation models. Under unilateral link formation a pair of players interact whenever at least one selects the other. Under mutual link formation a pair of players interact only if both select each other.

We say that cooperation can be sustained if there exists a subgame perfect equilibrium such that in one or more periods: (i) no player is isolated and (ii) all players cooperate. We say that a network dilemma game $\Delta$ permits cooperation if cooperation can be sustained in $\Delta^T$ for some finite $T$.

Result 4.1 Cooperation and Endogenous Network Formation

Cooperation can be sustained in many finitely repeated prisoner's dilemma games with endogenous formation of the interaction network. In contrast, there is no cooperation in finitely repeated prisoner's dilemma games played on an exogenously fixed network.

An $n$-player prisoner's dilemma game played on a fixed network has a unique Nash equilibrium (see section 2.2.4): each player defects. Following Lemma 2.1-2, if the game is repeated finitely many times this implies a unique subgame perfect equilibrium: all players always defect. On the other hand, if players have the possibility to modify the network the stage game may have different Nash equilibria. This is a necessary condition to sustain non-equilibrium outcomes, such as cooperation, in the repeated game. To demonstrate that cooperation can be sustained if partner selection and, consequently, the interaction network are endogenous, we have given in this chapter a number of examples of cooperative subgame perfect equilibria for the finitely repeated network dilemma games.

The possibility for endogenous partner selection is, however, not sufficient to sustain cooperation. To begin with, a network dilemma game must be repeated sufficiently many times. For example, there are no cooperative relations in any one-shot dilemma game, or in the last period of any repeated dilemma game. Next, the payoff for mutual defection should be sufficiently high: players should be willing to establish some links even if they all defect. Finally, linking constraints or costs may be crucial. For example, a network dilemma game with unconstrained and costless linking may not permit cooperation. Alternatively, if linking constraints are too strict, or if linking costs are too high, cooperation may be limited only to subgroups of players.

Result 4.2 Outside Option

It is difficult, and in most games impossible, to sustain cooperation when the
outside option is high. If linking is mutual and \( d < 0 \), then in any Nash equi-
librium the empty network is established in each period of the repeated network
dilemma game.

We have shown that the set of subgame perfect equilibria crucially depends
on whether the outside option is low or high. Namely, all relations in a one-
shot network dilemma game are defective. If \( d < 0 \) then each player prefers
to opt out from all defective relations. The one-shot game may thus have only
one equilibrium network: no links are established. By backwards induction,
the empty network is established in each period of any finitely repeated game.
This is the case for all network dilemma games considered, except for the games
with unilateral linking and no linking costs, for which non-empty equilibrium
networks exist. These non-empty equilibrium networks exist because exclusion
has to be mutual when linking is unilateral (see the bottom row of Figure 4.6).
By consequence, cooperative subgame perfect equilibria can be constructed for
finitely repeated network dilemma games with unilateral linking, no linking
costs, and high outside option. However, in these cooperative equilibria the
punishments crucially rely on active participation of punished players.

In the results 4.3-4.5 we assume that the outside option is low, \( d > 0 \). Let
\( k = (k_1, \ldots, k_n) \) be the vector of linking constraints.

**Result 4.3 Linking Constraints**

Let \( d > 0 \) and let linking be costless. A network dilemma game with linking
constraints permits cooperation if constraints are neither extremely severe nor
extremely lax. In particular, given a sufficient number of periods,

1. for mutual link formation: a linking-proof, cooperative, subgame perfect equi-
librium exists if \( 2 \leq k_i \) for all \( i \) and \( k_j < n-1 \) for some \( j \); (matching model)
   if \( k_i = 1 \) for each \( i \) then such equilibrium exists if and only if \( n \) is odd;
   (unconstrained linking) no such equilibrium exists if \( k_i = n - 1 \) for each \( i \);

2. for unilateral link formation: a cooperative subgame perfect equilibrium exists
   if \( 1 \leq k_i \) for all \( i \) and \( k_j, k_j' < n-1 \) for some \( j \neq j' \); (unconstrained linking)
   no such equilibrium exists if \( k_i = n - 1 \) for each \( i \).

Support for this result is given by Propositions 4.10, 4.9, and 4.14, and by
Theorem 4.2. To assure, for network dilemma games with mutual linking, that
punishments never rely on active participation of the punished player we require
that strategy profiles are linking-proof (see section 4.3.3). In absence of linking-
proofness condition cooperation is permitted also in games with mutual linking
and no linking constraints (see section 4.3.3).

**Result 4.4 Linking Costs**

Let \( d > 0 \) and let linking be unconstrained. Assume that marginal linking costs,
given by functions \( \Delta b_i \), are non-decreasing. A network dilemma game with
linking costs permits cooperation unless linking is extremely cheap or extremely
expensive. In particular, given a sufficient number of periods,
1. for mutual link formation: a linking-proof, cooperative, subgame perfect equilibrium exists if $\Delta b_i(2) < d$ for all $i$ and $d < \Delta b_j(n - 2)$ for some $j$.

2. for unilateral link formation: a cooperative subgame perfect equilibrium exists if $\Delta b_i(1) < d$ for all $i$ and $d < \min\{\Delta b_j(n - 2), \Delta b_{j'}(n - 2)\}$ for some $j \neq j'$.

To permit cooperation, the cost of adding a new link must increase with the number of already established links. The first few links should be relatively cheap, so that players are willing to establish the first links even with defectors. However, once a player sponsors many links the new links should be relatively costly, so that the player is willing to add them only if her neighbors cooperate. By consequence, networks of cooperative players may be denser than networks of defective players.

A network dilemma game with linking costs can be directly related to a game with linking constraints. In fact, result 4.3 about games with linking constraints is instrumental in proving result 4.4 about games with linking costs. See Propositions 4.15 and 4.17 for details of the relation between linking costs and linking constraints.

The final result explains when cooperation can be sustained solely by threats of exclusion. For this we can use trigger strategies: any deviation by defection triggers partial or complete exclusion of defectors from the network. For the mutual link formation model the trigger strategy profiles are defined by (4.5). For the unilateral link formation model they are defined by (4.10). The following is a summary of Propositions 4.8, 4.12 and 4.18.

**Result 4.5 Punishment with Exclusion**

Let $d > 0$. If the number of players is relatively large compared to the number of links that players can establish, or if linking costs quickly exceed the mutual defection payoff $d$, then each player can be punished by exclusion using trigger strategies. Trigger strategy profiles can sustain cooperation:

1. for mutual link formation, and satisfying linking-proofness,
   - for games with linking constraints but no costs: if $2 \leq k_i \leq \sqrt{2n-9}$ for each $i$;
   - for games with linking costs but no constraints: if $\Delta b_i(2) < d < \Delta b_i(\sqrt{2n-9})$ for each $i$;

2. for unilateral link formation,
   - for games with linking constraints but no costs: if $1 \leq k_i \leq n/2 - 1$ for each $i$;
   - for games with linking costs but no constraints: if $\Delta b_i(1) < d < \Delta b_i(n/2 - 1)$ for each $i$. 

78
It is important to note that the conditions in results 4.3-4.5 are sufficient, but most likely not necessary. We have thus characterized only a part of network dilemma games that permit cooperation. The complete characterization seems to be a challenging combinatorial task and the ideas in this chapter may find use in future search for an improved characterization. Our results nevertheless serve to advance the following main conclusions:

- cooperation can be part of a subgame perfect equilibrium of a finitely repeated prisoner’s dilemma game if it is played on an endogenous network,
- to achieve cooperation it may be necessary that there is some competition for partners: either players are strictly constrained in the number of links, or linking is costly and the cost function is convex,
- cooperation can be sustained solely via exclusion of defectors if the number of all players is substantially larger than the number of links that players can or are willing to support.

4.6.2 Beyond the network dilemma game

When real-world situations are modeled by simple games this typically leads to crude approximations of the richness and complexity of analyzed situations. A typical game is a stylized model of interaction emphasizing some particular real phenomenon. Obviously, when abstracting from a real situation one should seek to avoid over-simplifications, especially if there is the danger that they drive the results. Unavoidably, more realism often leads to increased complexity and may compromise analytical tractability. To address this concern we discuss the assumptions made in the course of the design and analysis of the network dilemma games in this chapter.

Through the design of the game we assume that discrimination in the action choice across different neighbors is not permitted, and that information about past actions of all players, and of the whole network structure, are common knowledge. In our analysis we assume that players are rational, that the payoffs represent their preferences, and that this is also common knowledge.

In addition, we consider in this chapter only the equilibria in pure strategies. The reason is that deviations from mixed strategies cannot always be detected and so equilibria in mixed strategies are difficult to enforce. In this chapter, in fact, we suggest to restrict, rather than relax, the set of strategies in repeated network dilemma games. In particular, we consider restriction to linking-proof strategy profiles and to trigger strategy profiles. Our results suggest that in most of the games either there are no cooperative subgame perfect equilibria or there exists a linking-proof one. In this sense the restriction of linking-proofness is mild. The restriction to trigger strategies may, however, substantially reduce possibilities for cooperation.

Smith (1995) proposes a workaround by assuming a public randomization device: in every period players can condition their moves on the outcome of some publicly observed exogenous random variable.
Trigger strategy profiles were considered because they seem more intuitive than recursive trigger strategies: if a player defects too early, she is excluded from the network. However, even trigger strategies may be unrealistic if they require global transformations of the network, that is, if players need to transform the whole existing network in order to exclude one player. Future research may identify network dilemma games in which cooperation is sustained only via local transformation of the network: to exclude a player only her neighbors change their links.

In chapter 3 we give some examples of real-world situations that may be modeled by the network dilemma game. In each example the agents must either act cooperatively in all their interactions or in none. However, there are also many situations where agents are free to simultaneously cooperate with one interaction partner, defect with another, and refuse to interact with a third agent. This seems especially true for social interaction between people. It would thus be interesting to analyze an alternative model, in which players may discriminate in the action choice across their interaction partners. A few conclusions about such an alternative model can be made using our results. Suppose that players in a network dilemma game are free to discriminate in their action choice across their neighbors, and that information about past moves, including the actions of each player in interactions with each of her neighbors, is complete. Defection of any player on any of her neighbors is then observed by everyone and can trigger exclusion, just as in games without discrimination in actions. The only addition to strategies described in this chapter is that defection in any relation counts as deviation. This implies that, given complete information, the set of network dilemma games that support cooperation is the same whether or not discrimination across neighbors in respect to actions is permitted.

Relaxations of the complete information assumption are more difficult to address. In order to exclude a player that defected, her defection should be known among players that are supposed to adjust their links. The information condition thus limits the possibilities for exclusion by permitting only the transformations of the network that involve players which are aware of the defection. For example, if a player knows only the actions of her neighbors, exclusion can take place only via local transformation of the network.

Special cases of information conditions may be particularly interesting for future research. One such condition is that the actions of a player are known only by her neighbors. Another condition, in which a network structure may play an important role, is that actions of a player are known also by neighbors of her neighbors, that is, there is a limited flow of information through the network, e.g. through gossip.

Information about the network structure could also be limited, which certainly is the case for social networks. Previous research has shown that, in spite of understanding only their local neighborhood, people are surprisingly successful in guessing network characteristics, such as the direction of the closest path between themselves and a random other person (see Milgram, 1976, Kleinberg, 2000, Newman et al., 2003, and Watts, 2003). In relation to the net-
work dilemma games, a local knowledge of a network may be sufficient to induce exclusion of defectors. For example, if a player is aware of the links of her neighbor she may be able to make conjectures about her cooperation: if the neighbor suddenly looses links it may be that she defected with another player, is hence unreliable, and should be excluded. Again, further research should look into the effects for cooperation of various relaxations of the complete information assumption.

Finally, our assumptions about common knowledge of rationality and preferences are in tradition with standard non-cooperative game theory. However, previous research has shown that, especially in social dilemma games, theoretical results may change if any of these assumptions is relaxed even slightly. See e.g. Kreps et al. (1982) for relaxation of the common knowledge assumption, Axelrod (1987) and Anderson et al. (1998) for models with boundedly rational players, and Fehr and Schmidt (1999) for a model with other-regarding players.

In particular, it is unlikely that these traditional assumptions are satisfied in experiments. For example, in experiments the game is imposed upon the subjects using monetary payoffs, which may not correctly represent player’s preferences. Therefore, in order to discuss our experimental results, we analyze in chapter 5 the solutions of the repeated network dilemma game played by other-regarding players. Furthermore, in chapter 6 we discuss an alternative model of behavior in repeated network dilemma game, assuming boundedly rational players.

4.7 Conclusions

It is well known that a finitely repeated prisoner’s dilemma game has a unique subgame perfect equilibrium: players defect in each period. The same holds for $n$-player prisoner’s dilemma games played in fixed groups or on a fixed network. In this chapter, however, we prove that a different result holds if the network is endogenously generated by the players. In particular, we show that cooperation can be sustained in a subgame perfect equilibrium of a finitely repeated prisoner’s dilemma game played on an endogenous network.

We construct such equilibria under the traditional assumptions of common knowledge of rationality, complete information and selfishness. The main rationale behind our result is that players can be discouraged from defection if they fear to be punished by exclusion from the interaction network. In particular, threats with exclusion can be credible if players are constrained in the number of links, or if linking is costly. In such cases players can organize themselves into different equilibrium networks. For example, the equilibrium network may be such that some player is isolated, thus foregoing payoffs from playing the game.

We identify sufficient conditions for the existence of cooperative subgame perfect equilibria for finitely repeated prisoner’s dilemma games on endogenous networks. These depend on the game length, the outside option value, the linking constraints and the linking costs. It is interesting that introducing endogenous network formation itself is not sufficient for cooperation, but that
assuming, in addition, very weak constraints on the number of links or linking costs may be sufficient. This assumption is satisfied in many real situations, for example in social networks where people are constrained in the number of their acquaintances.

4.8 Appendix: Proofs

For convenience we introduce the indexing function \( \overleftarrow{i} : \{-n + 1, \ldots, 2n\} \rightarrow \{1, 2, \ldots, n\} \), defined by

\[
\overleftarrow{i} = \begin{cases} 
  i - n & \text{if } i \in \{n + 1, \ldots, 2n\} \\
  i & \text{if } i \in \{1, \ldots, n\} \\
  i + n & \text{if } i \in \{-n + 1, \ldots, 0\}
\end{cases}.
\]

The application of this function is as follows. Place players on the circle such that player \( i \) follows player \( i - 1 \) and that player 1 follows player \( n \), clockwise. For any \( i \) the indices of her \( m \) immediate neighbors, going in one direction along the circle, are given by \( \overleftarrow{i - 1}, \ldots, \overleftarrow{i - m} \), and by \( \overleftarrow{i + 1}, \ldots, \overleftarrow{i + m} \), going in the other direction. For even \( n \) the players \( i \) and \( i + n/2 \) are directly opposite on the circle.

The proof of Proposition 4.8 relies on the following lemmas.

Lemma 4.1 Take \( k \) such that \( 2 \leq k < n \). Consider \( \Gamma(k) \) with \( d > 0 \) and \( k_i = k \) for each \( i \).

a. If \( k \) and \( n \) are odd, there exists a linking-proof network \( g \) such that \( l_j(g) = k - 1 \) for some \( j \) and \( l_i(g) = k \) for all \( i \neq j \).

b. Otherwise, there exists a linking-proof network \( g \) such that \( l_i(g) = k \) for all players \( i \).

Proof. (b.1) If \( k \) is even then \( g \) is constructed by linking each player \( i \) to players \( \overleftarrow{i - 1}, \ldots, \overleftarrow{i - k/2} \) and \( \overleftarrow{i + 1}, \ldots, \overleftarrow{i + k/2} \). All players thus establish \( k \) links, hence network is linking-proof.

(b.2) If \( k \) is odd but \( n \) is even then \( g \) is constructed by linking each player \( i \) to players \( \overleftarrow{i - 1}, \ldots, \overleftarrow{i - (k - 1)/2} \), to players \( \overleftarrow{i + 1}, \ldots, \overleftarrow{i + (k - 1)/2} \), and to the player \( \overleftarrow{i + n/2} \). All players thus establish \( k \) links, hence network is linking-proof.

(a) If \( k \) and \( n \) are both odd then a network \( g' \) is constructed by linking each player \( i \) to players \( \overleftarrow{i - 1}, \ldots, \overleftarrow{i - (k - 1)/2} \) and \( \overleftarrow{i + 1}, \ldots, \overleftarrow{i + (k - 1)/2} \). All players thus establish \( k - 1 \) links. Network \( g \) is constructed from \( g' \) by linking each of the \((n - 1)/2 \) disjoint pairs of separated players \( \{i, (n - 1)/2 + i\}\}_{i=1}^{(n-1)/2} \). Player \( n \) was not assigned to any pair and thus establishes \( k - 1 \) links in \( g \), while all other players establish \( k \) links, hence network \( g \) is linking-proof.

Lemma 4.2 Consider \( \Gamma(k) \) with \( d > 0 \) and \( 2 \leq k_i < \sqrt{2n - 7} \) for each \( i \). Take any stable-equilibrium network \( g \in S(\Gamma(k)) \). For any pair of players \( i \) and \( j \)
there exists another pair of players \( i' \) and \( j' \) such that \( g_{ii'} = 0, g_{jj'} = 0 \), and \( g_{i'j'} = 1 \).

**Proof.** The proof proceeds as follows. Let \( k = \max_{i \in N} k_i \). Fix \( i \) and \( j \).

We characterize the set of players \( M_{ij} \) that cannot be in the role of \( i' \) or \( j' \). We show that the size of this set depends on \( k \). We characterize this dependence and prove that the size of \( M_{ij} \) is less or equal to \( n - 2 \) if \( k < \sqrt{2n - 7} \). If the size of \( M_{ij} \) is less or equal to \( n - 2 \) then there must be two players \( i' \) and \( j' \) satisfying the conditions of the Lemma.

We say that a pair of linked players \( i' \) and \( j' \) complement \( ij \) if \( g_{ix} = g_{iy} = 1 \) and \( g_{jx} = g_{jy} = 1 \). If \( xy \) is not separated from \( ij \) it must be that either (a) \( g_{ix} = g_{iy} = 1 \), or (b) \( g_{jx} = g_{jy} = 1 \), or (c) \( g_{ix} = g_{jx} = 1 \), or (d) \( g_{iy} = g_{jy} = 1 \), that is, either one of \( i, j \) is linked to both \( x, y \), or both \( i, j \) are linked to one of \( x, y \). Define sets \( O_{ij}, N_{ij}, N_i, \) and \( N_j \) as follows: \( N_{ij} = L_i(g) \cap L_j(g) \), \( N_i = L_i(g) \setminus N_{ij} \), \( N_j = L_j(g) \setminus N_{ij} \), \( O_{ij} = N \setminus (L_i(g) \cup L_j(g)) \). \( N_{ij} \) is the set of players linked to both \( i \) and \( j \), \( N_i \) and \( N_j \) are sets of players linked to one of \( i, j \), and \( O_{ij} \) is the set of remaining players, separated from both \( i \) and \( j \). Let \( M_{ij} = \{i, j\} \cup L_i(g) \cup L_j(g) \).

If a pair of linked players \( x, y \) does not complement \( ij \), then either (i) both of them belong to \( N_i \) or both to \( N_j \), or (ii) at least one of them belongs to \( N_j \). Let \( N_i = \{x \mid y \in N_j \text{ for each } y \text{ s.t. } g_{xy} = 1 \} \setminus M_{ij} \) be the set of players, separated from \( i \) and \( j \), but whose neighbors all belong to \( N_{ij} \). Let \( M_{ij} = M_{ij} \cup N_{ij} \) be the set of players whose neighbors are either \( i \) or \( j \) or belong all to \( N_{ij} \), including players \( i \) and \( j \). If a pair of neighbors \( x, y \), \( g_{xy} = 1 \) does not complement \( ij \), both \( x \) and \( y \) belong to \( M_{ij} \).

All but at most one player in \( N_{ij} \) establish their maximal number of links in \( g \). To see this assume \( x, y \in N_{ij} \) such that \( l_x(g) < k_x \) and \( l_y(g) < k_y \). Because \( g \) is a linking-proof network, Proposition 4.3 implies that \( g_{xy} = 1 \). Then, by definition of \( N_{ij} \), both \( x, y \) should belong to \( N_{ij} \). This, however, is not possible, since they both belong to \( N_{ij} \).

Since \( k_x \geq 2 \) for all \( x \in N_{ij} \) and since at most one of them does not establish all her links, all but at most one of them is linked to at least two players in \( N_{ij} \). This restricts the maximal possible number of players in \( N_{ij} \). Let \( n_{ij} = |N_{ij}| \), \( n_i = |N_i| \), and \( n_j = |N_j| \). Each player in \( N_{ij} \) is already linked to two players, \( i \) and \( j \), and can have at most \( k - 2 \) more links. Number \( n_{ij} \) is maximized if each player in \( N_{ij} \) makes \( k - 2 \) links with players in \( N_{ij} \), each of whom makes two links, apart from one who makes one link. The joint number of links made by players in \( N_{ij} \) is thus \( n_{ij}(k - 2) \) and each player in \( N_{ij} \) shares two of these, aside from at most one who shares only one link. The number of players in \( N_{ij} \) is therefore bounded above by

\[
|N_{ij}| \leq \left( n_{ij}(k - 2) - 1 \right)/2 + 1.
\]

Note that \( M_{ij} = N_{ij} \cup N_i \cup N_j \cup \{i, j\} \) with no overlap between pairs of sets. This implies

\[
|M_{ij}| \leq \left( n_{ij}(k - 2) - 1 \right)/2 + 1 + n_{ij} + n_i + n_j + 2 \leq (n_{ij}(k - 2) - 1)/2 + 3 + k,
\]

83
since \( n_{ij} + n_i + n_j \leq k \). This bound is maximized when \( n_{ij} = k \), in which case it becomes

\[
|\overline{M}_{ij}| \leq \frac{(k^2 + 5)}{2}
\]

By assumption, \( k < \sqrt{2n - 7} \), which implies that \( |\overline{M}_{ij}| \leq n - 2 \). At least two players do not belong to \( \overline{M}_{ij} \). Following Proposition 4.3 at least one of them makes at least one link. There must then be a pair of neighbors which complement \( ij \).

**Lemma 4.3** Consider \( \Gamma(k) \) with \( d > 0 \) and \( 2 \leq k_i < \sqrt{2n - 7} \) for each \( i \). There exists a linking-proof network \( g \) such that \( 2 \leq l_j(g) \leq k_j \) for some player \( j \) and \( l_i(g) = k_i \) for all other players \( i \neq j \).

**Proof.** The proof proceeds constructively as follows. We create, by an iterative procedure, an initial linking-proof network (LPN) \( g^0 \) in which each player establishes at least two links. In this network there may be several players that miss one or more links. We then use Lemma 4.2 iteratively to create a finite sequence of LPN \( g^1, ..., g^\Phi \) such that the total number of links strictly increases. At most one player is missing links in \( g^\Phi \).

Let \( g^0 \) be the network generated by the following procedure:

1. establish links between each pair \( \{i, i + 1\} \) and between the pair \( \{n, 1\} \).
2. repeat: if there exists a pair of separated players \( i \) and \( j \) such that \( l_i < k_i \), \( l_j < k_j \), establish the link \( ij \); until no such pair exists.

Because the number of established links in the above procedure strictly increases, but the maximal possible number of links is finite, the procedure stops in a finite number of steps. By Proposition 4.3, the network \( g^0 \) is LPN as there exists no pair of separated players \( i \) and \( j \) such that \( l_i(g^0) < k_i \), \( l_j(g^0) < k_j \).

Let networks \( g^1, g^2, ... \) be generated by the following procedure for \( \phi = 0, 1, ... \):

As long as in \( g^\phi \) there exist two players \( i \) and \( j \) such that \( l_i(g^\phi) < k_i \), \( l_j(g^\phi) < k_j \), then

1. take a pair of neighbors \( i'j' \), \( g^\phi_{ij} = 0 \) which complement \( ij \), that is, \( g^\phi_{i'j'} = 1 \) and \( g^\phi_{ij} = 0 \) (we show in Lemma 4.2 that such pair exists),
2. let \( g^{\phi+1} \) coincide in all links with \( g^\phi \), aside from \( g^{\phi+1}_{i'j'} = 1 \), \( g^{\phi+1}_{ij} = 1 \), and \( g^{\phi+1}_{ij} = 0 \).

The procedure ends when at most one player is still missing links. Again, the number of established links strictly increases with steps and the procedure stops in finite number \( \Phi \) of steps. In \( g^\Phi \) all players establish their maximal number of links, aside from at most one player who has at least two links.

To see that each \( g^{\phi+1} \) is LPN, proceed again iteratively. Let \( g^\phi \) be LPN and satisfy the linking constraints. Let \( ij \) and \( i'j' \) be the two pairs of players chosen by the above procedure in the corresponding step. Then, \( g^{\phi+1} \) also satisfies linking constraints. Furthermore, by Proposition 4.3, \( g^{\phi+1}_{ij} = 1 \), and
thus $g_{ij}^{\Phi+1} = 1$. Also by Proposition 4.3, because $g^\Phi$ is LPN and $l_i(g^\Phi) < k_i$, and $g_{ij}^{\Phi} = 0$, then $l_{ij}(g^\Phi) = k_{ij}$. Similarly, $l_{ij}(g^\Phi) = k_{ij}$. The only link that is removed is $i'j'$, but as players $i'$ and $j'$ establish the same number of links in both $g^\Phi$ and $g^\Phi+1$, the $g^\Phi+1$ is a LPN. Hence, network $g^\Phi$ is also LPN. ■

Proof of Proposition 4.8. (1) Assume that $k_i = k$ for each $i$. Let $2 \leq k \leq n - 2$. In the following we apply Lemma 4.1.

(a) Let $k$ be even. For every $i$ the remaining players $N \setminus \{i\}$ can form a linking-proof network (LPN) $g^i$ such that $l_j(g^i) = k$ for each $j \neq i$ and $l_i(g^i) = 0$. Furthermore, a LPN $g$ can be established such that $l_i(g) = k$ for each $i$, see Lemma 4.1 (b). (b) Let $k$ be odd and $n$ even. For every $i$ the remaining $n - 1$ players $N \setminus \{i\}$ can form a LPN $g^i$ such that $l_j(g^i) = k$ for each $j \neq i$ and $l_i(g^i) = 1$, see Lemma 4.1 (a). A LPN $g$ can also be established such that $l_i(g) = k$ for each $i$. (c) Let $k$ and $n$ both be odd. For every $i$ the remaining $n - 1$ players $N \setminus \{i\}$ can form a LPN $g^i$ such that $l_j(g^i) = k$ for each $j \neq i$ and $l_i(g^i) = 0$. A LPN $g$ can also be established such that $l_i(g) \geq k - 1$ for each $i$.

In all three cases (a), (b), and (c), $l_j(g^i) < l_i(g)$ for each $i$. This fulfills the conditions of Proposition 4.7.

(2) Let $2 \leq k_i < \sqrt{2n - 9} = \sqrt{2n - 1} - 1$ for each $i$. According to Lemma 4.3, for any subgroup of players $N \setminus \{i\}$ of size $(n - 1)$ there exists a network $g^{N \setminus \{i\}}$ such that $l_j(g^{N \setminus \{i\}}) \leq k_j$ for some $j \in N \setminus \{i\}$ and $l_x(g^{N \setminus \{i\}}) = k_x$ for all other $x \in N \setminus \{i, j\}$. Let $g^i$ be the network among players $N$ which is obtained from $g^{N \setminus \{i\}}$ by adding player $i$ and (i) adding the link $ij$ if $l_j(g^{N \setminus \{i\}}) < k_j$ or (ii) adding no links if $l_j(g^{N \setminus \{i\}}) = k_j$. Aside from player $i$ who establishes at most one link, all other players establish their maximal number of links in $g^i$, hence, following Proposition 4.3, $g^i$ is a linking-proof network (LPN).

We have shown that for each player $i$ there exists a LPN $g^i$ such that $l_i(g^i) \leq 1$. Furthermore, according to Lemma 4.3, there also exists a network $g$ such that $l_i(g) \geq 2$ for each $i$. The conditions of Proposition 4.7 are thus satisfied. ■

Proof of Theorem 4.1. All threats of the recursive trigger profile are credible, because they consist of a repetition of a static equilibrium. We thus need to verify that there exist positive integers $t^0, ..., t^m$ such that all threats are effective.

A threat against an early defection of player $i \in N^0$ is effective if the one-period profit from the defection is offset by the loss incurred through subsequent exclusion. A threat against an early deviation of player $i \notin N^0$ in period $t \in \{1, ..., T - t^{\eta(i)}\}$ is effective if the profit from the deviation collected during $\{t, ..., T - t^{\eta(i)}\}$ is offset by the loss of reward during $\{T - t^{\eta(i)} + 1, ..., T - t^{\eta(i)-1}\}$. By an early defection a player may avoid that other players free-ride on her, and may thus gain for more than one period. Namely, if some player defects early then all players defect in all subsequent periods. Player $i \in N^{\eta(i)}$ may thus avoid being free-rided upon by her neighbors from $N^{\eta(i)+1}$ during $\{T - t^{\eta(i)+1} + 1, ..., T - t^{\eta(i)}\}$. We need to take this potential profit into account to determine the effectiveness of threats.

85
Let \( \gamma_i \) be the per-period difference in payoff to player \( i \) for avoiding being free-ridden upon during \( \{ T - t^{\eta(i)} + 1, \ldots, T - t^{\eta(i)} \} \). This value depends on the number of neighbors of \( i \) in sets \( L_i = L_i \cap (N^\eta(i) \cup N^{\eta(i)} - 1) \) and \( \overline{L}_i = L_i \cap N^{\eta(i) + 1} \). If there were no deviations then all players in \( L_i \) cooperate and all players in \( \overline{L}_i \) defect during the above mentioned periods. However, if \( i \) deviated early then all players defect during these periods. Player \( i \) thus loses because her neighbors in \( \overline{L}_i \) do not cooperate, and gains because her neighbors in \( L_i \) do not free-ride on her. Hence, \( \gamma_i \) can be positive, negative or 0. Let \( \gamma_i = \max\{0, \gamma_i\} \).

Let \( \alpha_i \) be the maximal potential one-period profit to player \( i \) from an early defection. Let \( \beta_i \) be the per-period loss to player \( i \) from an exercised threat in any of the periods \( \{2, \ldots, T - t^{\eta(i)}\} \) (the first deviation can take place in period 1 and the first threat is then exercised in period 2). Finally, let \( \delta_i \) be the per-period loss to player \( i \) when foregoing the reward of free-riding during \( \{ T - t^{\eta(i)} + 1, \ldots, T - t^{\eta(i) - 1}\} \), or when \( i \in N^\eta \) is punished by exclusion during \( \{ T - t^\eta + 1, \ldots, T\} \). The following holds for each \( i \): \( \alpha_i, \beta_i, \delta_i > 0 \), while \( \gamma_i \geq 0 \).

Define \( \Delta_0 = t^\delta \) and \( \Delta_\eta = t^\eta - t^{\eta-1} \) for \( \eta = 1, \ldots, m \). Let \( \Lambda(i, t) \) be the difference in the total payoff to player \( i \) for an early defection in period \( t \in \{1, \ldots, T - t^{\eta(i)}\} \).

\[
\Lambda(i, t) \leq \alpha_i - (T - t^{\eta(i)} + 1 - 1)\beta_i + \Delta_{\eta+1} - \Delta_\eta \delta_i \\
\leq \alpha_i + \Delta_{\eta+1}  - \Delta_\eta \delta_i.
\]

All threats are surely effective if \( \alpha_i + \Delta_{\eta+1}  - \Delta_\eta \delta_i < 0 \) for each \( i \).

Players in \( N^m \) do not award any player by letting her free ride. Hence, \( \gamma_i = 0 \) for \( i \in N^m \). The threats for these players are thus effective if \( \alpha_i - \Delta_m \delta_i < 0 \), which rewrites as

\[
\Delta_m > \max_{i \in N^m} \frac{\alpha_i}{\delta_i}.
\]

There exists a finite positive integer \( \Delta_m \) which satisfies (4.12). Similarly, for \( \eta \in \{0, \ldots, m - 1\} \), the threats for players in \( N^\eta \) are effective if

\[
\Delta_\eta > \max_{i \in N^\eta} \frac{\alpha_i + \Delta_{\eta+1}}{\delta_i}.
\]

Given a finite \( \Delta_{\eta+1} \), there exists a finite positive integer \( \Delta_\eta \) which satisfies (4.13). Starting with \( \Delta_m \) it is now possible to recursively determine the minimal \( \Delta_\eta \) such that (4.12) and (4.13) are satisfied for all \( \eta \). Setting \( t^\eta = \sum_{\xi=0}^\eta \Delta_\xi \) we get positive integers \( t^0, \ldots, t^m \) such that all threats are effective. Setting \( T^* = t^m \) we conclude that \( p(g^*, g^*, \{g^i\}_{i \in N^0, T, (t^\eta)_{\eta=0}} \) is subgame perfect for any \( T > T^* + 1 \).

**Proof of Proposition 4.10.** Consider the following construction of a connected LPN. Begin with a wheel network connecting all players. Now recursively add links between pairs of separated players who can still add links. By
Proposition 4.3 the network is LPN when there is no such pair of players. This construction stops in finite time as there is a finite number of possible links.

The complete network is not feasible because \( k_j \leq n-2 \). Let \( g^* \) be a connected LPN. There exist separated players \( i' \) and \( i'' \). There also exists a shortest path \((i' = i^0, i^1, \ldots, i^\kappa = i'')\) between each such pair. Define a partition of \( N \) into players \( M^1 \) who establish all their links and players \( M^0 \) who do not,

\[
M^0 = \{i \in N \mid l_i(g^*) < k_i \} \quad \text{and} \quad M^1 = \{i \in N \mid l_i(g^*) = k_i \}.
\]

By Proposition 4.3 all players in \( M^0 \) are linked. Set \( M^1 \) cannot be empty as otherwise all players would be linked and the network complete. In the rest of the proof we consider all possible cases of how pairs of separated players are distributed between \( M^0 \) and \( M^1 \). For each case we construct another LPN in which one player has less neighbors than in \( g^* \).

a) Assume that \( M^0 \) is empty. Take separated players \( i' \) and \( i'' \) and a shortest path \((i' = i^0, i^1, \ldots, i^\kappa = i'')\). A network \( g^i_1 \) can be constructed from \( g^* \) by removing links \( i^0i^1 \) and \( i^1i^2 \) and adding the link \( i^0i^2 \). By Proposition 4.3 \( g^i_1 \) is LPN because \( i^1 \) is the only player who does not establish all her links.

b) Assume that both \( M^0 \) and \( M^1 \) are non-empty. Assume also that \( g^{i^i} = 1 \) for all \( i \in M^0 \) and \( i' \in M^1 \). Then there exist separated players \( i^0, i^2 \in M^1 \) and a shortest path \((i^0, i^1, i^2)\) with \( i^1 \in M^0 \). A network \( g^i_1 \) can be constructed from \( g^* \) by removing links \( i^0i^1 \) and \( i^1i^2 \) and adding the link \( i^0i^2 \). The sets \( M^0 \) and \( M^1 \) thus remain the same and no links between players in \( M^0 \) change. In \( g^i_1 \) all players in \( M^0 \) are still linked. By Proposition 4.3 \( g^i_1 \) is LPN.

c) Assume now that both \( M^0 \) and \( M^1 \) are non-empty and take separated players \( i \in M^0 \) and \( i' \in M^1 \) with a shortest path \((i = i^0, i^1, \ldots, i^\kappa = i')\). Because players \( i \) and \( i^2 \) are not linked, but all players in \( M^0 \) must be linked, it must be that \( i^2 \in M^1 \). W.l.o.g. let \( i^2 = i' \).

c.1) Assume that \( i^1 \in M^0 \). Player \( i^0 \) has no link to \( i^2 \) and can add links to those she has in \( g^* \). A network \( g^i_1 \) can be constructed from \( g^* \) by removing link \( i^1i^2 \) and adding the link \( i^0i^2 \). The new set \( M^0(g^i_1) \) is a subset of \( M^0 \) and no links between players in \( M^0 \) change. All players in \( M^0(g^i_1) \) are still linked. By Proposition 4.3 \( g^i_1 \) is LPN.

c.2) Assume that \( i^1 \in M^1 \) and that \( g^{i^i} = 1 \) for all \( i'' \in M^0 \). Hence, all players in set \( M^0 \cup \{i^1\} \) are linked. Again, a network \( g^i_1 \) can be constructed from \( g^* \) by removing the link \( i^1i^2 \) and adding the link \( i^0i^2 \). The new set \( M^0(g^i_1) \) is a subset of \( M^0 \cup \{i^1\} \) and no links between players in \( M^0 \cup \{i^1\} \) change. All players in \( M^0(g^i_1) \) are linked. By Proposition 4.3 \( g^i_1 \) is LPN.

Finally, assume that \( i^1 \in M^1 \) and let \( g^{i^i} = 0 \) for some \( i'' \in M^0 \). Set \( j^0 = i'', j^1 = i \) and \( j^2 = i^1 \) and construct a LPN \( g^{j^1} \) as in (c.1).

In cases a) - c.2) above we constructed a LPN \( g^i_1 \) such that \( l_{i^3}(g^i_1) < l_{i^3}(g^*) \). In (c.2) we constructed a LPN \( g^{j^1} \) such that \( l_{j^3}(g^{j^1}) < l_{j^3}(g^*) \). In all cases above the conditions of Theorem 4.1 are thus satisfied. 

87
Proof of Proposition 4.12. The construction of equilibrium profiles \( p^*, \{p^{f,i}\}_{i \in N} \) satisfying (i) and (ii) is as follows. Recall the indexing function \((\cdot)\) defined in (4.11). Place players on a circle such that each player \( i \) follows player \( i - 1 \) and player 1 follows player \( n \), clockwise. Let \( p^* \) be the linking profile obtained when each player \( i \) proposes links with her immediate \( k_i \) neighbors clockwise. That is, \( p^*_{ij} = 1 \) if and only if \( j \in \{i + 1, \ldots, i + k_i\} \). Let \( p^{f,i} \) be the profile obtained from \( p^* \) when all players that proposed a link to \( i \) remove it and propose a link to another player as follows: for each \( j \) such that \( p^*_{ij} = 1 \) let \( p^{f,i}_{j + kj + 1} = 1 \). Let \( p^{F,i}_{ij'} = p^{*}_{ij'} \) otherwise.

To see that \( p^* \) is an equilibrium profile note that \( p^* \) is maximal. Namely, \( m_i(p^*) = k_i \) for each \( i \) and no link is proposed mutually: if \( p^*_{ij} = p^*_{ij'} = 1 \) then \( i = j + x \) for \( x \leq k_j \) and \( j = i + y \) for \( y \leq k_i \), and since \( i = i + x + y \) it must be that \( n = x + y \leq k_j + k_i \leq n - 2 \), which is impossible. Now consider \( p^{f,i} \) for some fixed \( i \) and note that \( m_j(p^{f,i}) = k_j \) for each \( j \). In \( p^{f,i} \) no link is proposed mutually: if \( p^{f,i}_{j} = p^{f,i}_{j'} = 1 \) then each \( i' \neq i \) has a link proposed either by \( j \) or \( j' \), which implies that \( j \) and \( j' \) together proposed at least \( n - 1 \) links, which is impossible. Hence, \( p^{f,i} \) is an equilibrium profile.

Each player \( i \) has at least \( k_i + 1 \) neighbors in \( g(p^*) \): she proposes links to \( \{i + 1, \ldots, i + k_i\} \) and player \( i - 1 \) proposes a link to \( i \). On the other hand, under \( p^f \) no links to player \( i \) are proposed, hence players \( \{i + 1, \ldots, i + k_i\} \) are her only neighbors. Properties (i) and (ii) are therefore satisfied.

The payoff to each player \( i \) in a Nash equilibrium is proportional to the number of her neighbors. In particular, \( \pi_i(D,p^*) = d(k_i + 1) > dk_i = \pi_i(D,p^{f,i}) \). Furthermore, \( \pi_i(C,p) = ck_i > dk_i = \pi_i(D,p^{f,i}) \). The conditions of Theorem 2.1 are therefore satisfied and a cooperative trigger strategy profile \( \phi(p,p^*,\{p^{f,i}\}_{i \in N},T,t^*) \) is subgame perfect for sufficiently large \( T \) and \( t^* \). ■

Proof of Theorem 4.2, continued. For a profile \( p^0 \in P(k) \) with no mutually proposed links, i.e. \( p^0_{jj'}p^0_{j'j} = 0 \) for all \( j,j' \), let \( p^0* \) be the profile obtained from \( p^0 \) via the following iterative procedure. Let \( p^{00} = p^0 \). For \( x \geq 1 \) let \( p^{0,x} \) be obtained from \( p^{0,x-1} \) by an addition of one link: set \( p^{0,x}_{ij} = p^{0,x-1}_{ij} \); take the player with the smallest index \( j \) such that \( m(p^{0,x}) < k_j \) and \( l_j(g(p^{0,x})) < n - 1 \) and then find a player with the smallest index \( j' \neq j \) such that \( g(p^{0,x})_{jj'} = 0 \) and set \( p^{0,x}_{j'j'} = 1 \). Stop the procedure when no such \( j \) exists and let \( p^0* \) be the resulting profile. \( (D,p^0*) \) is an equilibrium profile: clearly \( p^0* \in P(k) \), no link is proposed mutually, and for each \( j \) either \( m(p^{0,x}) = k_j \) or \( l_j(g(p^{0,x})) = n - 1 \).

Now construct \( p^* \) and \( p^i \) as follows. Define \( p^0 \) by setting \( p^0_{jj} = 1, p^0_{ij} = 0 \) and \( p^0_{j'j'} = 0 \) for each \( j,j' \in N \setminus \{i\} \) and \( p^0_{ii} = 0 \). No link in \( p^0 \) is proposed mutually. Let \( p^* \) be the profile obtained from \( p^0 \) via the above procedure. Define \( p^{i0} \) by setting \( p^{i0}_{ij} = p^{i0}_{j'j} = 1 \) for \( k_i \) players with smallest index \( j \in N \setminus \{i,i'\} \) and \( p^{i0}_{ij'} = 0 \) otherwise. Let \( p^i \) be the profile obtained from \( p^{i0} \) via the above procedure.
Both \((D,p^*)\) and \((D,p^i)\) are equilibrium profiles. Players \(i\) and \(i'\) are separated in \(g(p^i)\). In \(g(p^*)\), however, \(i\) is linked to each other player and \(g(p^*)\) is connected. Hence, \(l_i(g(p^i)) < n - 1 = l_i(g(p^*))\). \(\blacksquare\)