Utrecht University

School of Economics

Networkd disruption and the common enemy effect

Britta Hoyer and Kris De Jaegher

U.S.E. Discussion Paper Series nr:12-06

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the first author.

Britta Hoyer
University of Paderborn
Warburger Str. 100
33098 Paderborn
Germany
E-mail: britta.hoyer@wiwi.upb.de

Kris De Jaegher
Utrecht University
Utrecht School of Economics
Kriekenpitplein 21-22
3584 TC Utrecht
The Netherlands.
E-mail: k.dejaegher@uu.nl

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 12-06

Cooperation and the common enemy effect: a network disruption model*

Britta Hoyer[a]
Kris De Jaegher[b]

[a]University of Paderborn

[b]Utrecht School of Economics
Utrecht University

This Version: November 2015

Abstract
The phenomenon that groups or people work together when they face an opponent, although they have little in common otherwise, has been termed the "common enemy effect". We study a model of network formation, where players can use links to build a network, knowing that they are facing a common enemy who can disrupt the links within the network, and whose goal it is to minimize the sum of the benefits of the network. We find that introducing a common enemy can lead to the formation of stable and efficient networks as well as fragmented networks and the empty network.
Keywords: strategic network disruption, strategic network design, non-cooperative network games

JEL classification: C72, D85

Acknowledgements
We would like to thank Hans Haller, Sanjeev Goyal, Fernando Vega Redondo, Arno Riedl, Jacques Siegers, Vincent Vannetelbosch, Stephanie Rosenkranz, as well as participants at the Nake Research Day 2011 in Utrecht, the IFORS 2011 in Melbourne, the Games 2012 in Istanbul, the IWGTS 2014 in Sao Paulo and the UECE Lisbon Meetings 2014 for comments and suggestions. This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Centre ''On-The-Fly Computing'' (SFB 901).

1. Introduction

A wide literature across several scientific disciplines hypothesizes what may broadly be termed as "the common enemy effect": individuals who face a common enemy are more likely to cooperate than if they did not face a common enemy. Examples include by-product mutualism in biology (Mesterton-Gibbons and Dugatkin, 1992), the in-group out-group hypothesis in sociology (Simmel, 1908; Coser, 1956), balancing theory in social psychology (Heider, 1982), and the backfiring effect of government repression in political science (Muller and Opp, 1986).[1] Typical rationales for the common enemy effect are that the presence of a common enemy changes either individual psychology, or changes group psychology.[2] Alternatively, the fact that they face a common enemy may inform rational individuals about the type of decision problem they face, and teach them that it is in their individual interests to cooperate.[3]

The focus of our paper is on a conceptually more parsimonious rationale for the common enemy effect, where rational players who face a common enemy cooperate without their (individual or group) psychology or their information being changed by the presence of the common enemy. For a broad intuition of the common enemy effect, consider the metaphor of a chain of cooperating players, where a non-cooperating player is a weak link in the chain. Once the chain has been formed, it is in the common enemy's interest to attack the weakest link. Yet, by doing this, he inadvertently punishes defection from joint cooperation, and thus makes joint cooperation more likely.

In this paper, we take this metaphor literally, and consider players as forming a network, and a disruptor as ex post attempting to disrupt this network. In a model of cooperation as information sharing, each player possesses a unit of information. He additionally obtains information from every player with whom he has a direct link, and moreover from each player to whom he is indirectly linked (connections of his connections, connections of his connections' connections, etc.). The players anticipate that after they have formed their network, the common enemy will attack their network in a manner that causes maximal damage. As we will show, this strategy may be counterproductive to the common enemy, as it may ensure that no player wants to delete a link in the network.

Yet, our model also allows for the possibility of an effect that is diametrically opposed to the common enemy effect, where the presence of a common enemy decreases the probability of players cooperating. Indeed, a smaller number of papers formulate such a competing hypothesis (e.g., Carroll et al. (2005); McLauchlin and Pearlman (2012)), and at an intuitive level, it is plausible that the presence of a common enemy creates inner division and stops individuals from cooperating, rather than encouraging them from doing so (Stein, 1976, p.144). A contribution of our paper is to identify circumstances in which the common enemy effect, and a competing effect should apply.

We show specifically that the common enemy effect applies for low linking costs, and the opposite effect for high linking costs. This may at first sight seem counterintuitive, as one may expect that high linking costs would make it more difficult for the players to act against a common enemy. To see the intuition for our results, we consider linking costs as low if players form a network in the absence of a common enemy, and we consider linking costs as high if they do not form a network in the absence of a common enemy (where forming a network is efficient both for low and for high linking costs). In the presence of a common enemy, the results are not affected by whether linking costs are high or low. Players may coordinate on forming an empty network, as any single link formed is automatically targeted. Alternatively, they may coordinate on a circle-shaped network. Such a network is fully protected against a common enemy who is able to remove only one link, as removal of any link leaves all players still connected. If a link in the circle is not formed, the network becomes a chain, which the common enemy can cut in two. Therefore, not forming a link has a large effect on individual payoffs, explaining why the circle is maintained even

[1] In economics, Bornstein et al. (2002) and Riechmann and Weimann (2008) show that in laboratory experiments, participants who face competing groups are more likely to coordinate on cooperation in Stag Hunts.

[2] E.g., political scientists invoke the relative deprivation theory from psychology: facing government repression makes citizens angry, and induces them to cooperate (e.g., Siegel (2011)). Others political scientists argue that common enemies create a collective identity and/or increase group solidarity (e.g., Koopmans (1997) or Chang (2008)). This is inherent in the in-group out-group hypothesis in sociology (see footnote 1), and in balancing theory in social psychology (Heider (1982)).

[3] For an example of such a mechanism, see , e.g., Pierskalla (2009) in the literature on government repression.

for high linking costs. Let us compare now the case with and without a common enemy, both for low an high linking costs. With high linking costs, in isolation players do not form a network, but when facing a common enemy it is possible that they form a network; therefore, a common enemy effect applies. With low linking costs, in isolation players always form a network, but with a common enemy they form an empty network (or a fragmented network) with positive probability; therefore the opposite effect applies.

A few game-theoretic models, without any features of network formation, provide alternative intuitions for the common enemy effect. The model by Münster and Staal (2011) features several competing groups. To the individual group, spending resources to compete with other groups has a beneficial effect, as fewer resources are then left to compete within the group. In Hugh-Jones and Zultan (2013), players within groups come to each other's help in order to maintain group reputation. In Kovenock and Roberson (2012), two players separately playing a Colonel Blotto game against the same enemy may benefit from making transfers to each other. De Jaegher and Hoyer (2014) finally consider two defenders who can only successfully ward off attacks when both defending. It is shown that the players are more likely to coordinate on joint defense if the attacks are performed by an intelligent attacker, rather than by nature.

In all of these game-theoretic models, the modeled type of cooperative action, namely defense, only has value to the players in the presence of a common enemy, but not when a common enemy is absent. Put otherwise, while the presence of a common enemy induces cooperation in these models, from the perspective of the players, it is not the case that they benefit from the presence of a common enemy, as they would be better of without a common enemy and without a reason to cooperate. The added value of our network model is to model an instance where the presence of a common enemy may make the players better off, because the cooperative activity is not a mere response to the presence of a common enemy, but is an activity that is efficient whether or not such an enemy is present.

The paper is structured as follows. Because of the links of our model to the literature on network formation, we separately relate our model to this literature in Section 2. Section 3 presents the model of network formation and disruption. Section 4 introduces the benchmark case and Sections 5 and 6 focus on the analysis of stable networks, where we first introduce a complete characterization for the case of a disruption budget of one link and then move on to more general results. Section 7 then concludes the paper.

2. Relation to the networks literature

While using one of the models of network formation introduced by Jackson and Wolinsky (1996) - the connections model - first and foremost, our paper adds to the part of networks research that deals with networks under attack. This problem has received more attention in physics and operational research (e.g., Albert et al. (2000), Bollobás and Riordan (2003), Lipsey (2006), and Taylor et al. (2006)) than in economics. However, whereas the focus in the physics research is on random attacks on graphs, here we are looking at intelligent network disruptors, thus shifting the focus of the research from designing a network that is proof against random attacks to one that is proof against targeted attacks. The literature in operational research has a similar focus, but mostly uses survival analysis. What we are investigating instead is what happens if the network is separated into different components, as in our model the distance between players does not matter.

Previous work on random and strategic disruption of networks (see, e.g., Dziubiński and Goyal (2013), Goyal and Vigier (2014), Hong (2009), Landwehr (2015) or Albert et al. (2000), as well as in Hoyer and De Jaegher (2016)) has mainly focused on robust designs against disruption. It does not take into account, however, the incentives of individual nodes within the network, as it only analyzes the value of the network as a whole and the incentive it creates for a network designer. Consequently, the effect of a network disruptor in this literature has always been strictly negative, as the introduction of a network disruptor by definition will lead to a lower value in the network if links and/or nodes are taken out of the network. Contrary to this type of analysis, in this paper we focus on the incentives of self-interested individual players to form a network when faced with a network disruptor.

3. Model

Modeling the structure of a network formation game with a disruptor can be achieved by using an undirected network model, consisting of links and nodes. The disruptor is modeled as having a disruption budget D consisting of a number of links that he can destroy within the network. In the game the nodes building the network move first.[4] They build a network by adding costly links bilaterally and deleting them unilaterally. The players know that an attack will ensue in the following stage of the game and they also know the size of the network disruptor's disruption budget. At stage two, the network disruptor who has observed the network chooses which links to disrupt in the network.

Formally, we study a two-stage game between a finite set of $N = 0, 1, ..., (n-1)$ self-interested, myopic players and a disruptor.[5] The relations between the players are represented by a graph g, the links of which capture the pairwise relations between players. Each player i has one unit of non-rival information w_i. For ease of exposition we assume that $\forall i \in N$: $w_i = 1$. Thus the information player i possesses has the same value to himself as it has to other players and the value of the information each node possesses is the same. To obtain the piece of information node j has, node i can either be directly or indirectly linked to j. To be directly linked, node i needs to form a costly link to j. However, node j needs to accept the link and both players incur the costs. Let g_{ij} denote a link between any two players i and j and let $g_{ij} = g_{ji}$, so that links are undirected. We say that a path in g connects i and j if a set of distinct nodes $i_1, ..., i_k \subset N$ exists, such that $g_{i_1,i_2}, g_{i_2,i_3}, ..., g_{i_{k-1},i_k} \subset g$. Only if there is a path between any two nodes do we refer to them as connected. In order to isolate the pure effect of the presence of a network disruptor on network formation, we assume that there is no decay.[6] A component C of the graph g is a connected subset. We use the notation C_i to denote a component including node i, and $|C_i|$ to denote its cardinality, or using the graph theoretic term, its order. A graph g plus link g_{ij}, is denoted as $g + g_{ij}$. A graph without this link is denoted as $g - g_{ij}$. A player's degree η is the number of links this player has. If a player is connected of degree $\eta = 1$, we refer to him as an end-player.

3.1. Payoffs and Value

To calculate the payoffs of the players, taking into account network disruption, we look at the graph g at two points in time: once before disruption and once after disruption. The pre-disruption graph is labeled g^1, where the set of all g^1_{ij} forms the pre-disruption network g^1. The post-disruption graph is labeled g^2, where the set of all g^2_{ij} forms the post-disruption network g^2. This is necessary, as the costs for linking depend on the pre-disruption network and will be borne even if a particular link will be disrupted by the network disruptor. The benefits of being linked to other players, however, depend on the post-disruption network. This can be justified by looking at the cost of linking as the costs of forming a connection, whereas the benefits will only be gained via actual contact between the players, which will not take place if the network disruptor takes out the link.

The overall payoff of node i can be derived from the value of its own information and the information it obtains from the members of the component it belongs to in g^2 minus the cost of its direct links in g^1. We define $g^2(g^1)$ as the disruptor's best response correspondence to g^1 and assume that if he is indifferent between targeting several links, he will choose randomly between them. Thus, there are possibly multiple g^2 corresponding to a g^1. Therefore, in the following, everything can be considered from the perspective of g^1 and the disruptor's best response correspondence, as the players anticipate the response of the disruptor. The costs player i incurs are then defined as $c_i(g^1) = \sum_{k:g^1_{ik} \in g^1} c_{ik}$, (where $c_{ik} = c$ for every g_{ik}, so that the cost of one link is constant and identical for each player) and include all direct links node i possesses in g^1. The expected benefits player i gains from the network after disruption are defined as $b_i(g^2(g^1)) =$

[4] Throughout the paper, nodes are also referred to as players in the network.

[5] The labeling of the nodes from 0 to $(n-1)$ instead of from 1 to n is not standard. However, we will need it when we introduce circulant networks.

[6] Decay refers to any information loss in the transmission between two players that is caused by the distance (length of the shortest path) between them.

$w_i + \sum_{j \in \mathcal{N}_i(g^2)} w_j$, where $\mathcal{N}_i(g^2)$ is the set of all nodes $j \neq i$ for which there is a path in the post-disruption network g^2 between i and j.[7] Since we assume that $\forall i \in N$: $w_i = 1$ we can rewrite this as $b_i(g^2(g^1)) = |C_i(g^2)|$, where $C_i(g^2)$ denotes the component in the post-disruption network g^2 to which i belongs. The overall expected payoff of node i, is then defined as $u_i(g^1) = b_i(g^2(g^1)) - c_i(g^1)$.

Following Myerson (1977) and Jackson and Wolinsky (1996), we assume that the value of a network is the sum of all the individual utilities of the players. Therefore we assume that the value of the network after disruption is given by $v(g^1) = \sum_{i \in N} u_i(g^1)$. Given the way we have defined the payoffs of node i above, it is clear that the anticipation of the network disruptor's strategy in the second stage of the game is already reflected in the expected payoff of each node and consequently, also in the overall value of the network. Since what we model is essentially a communications network and the information is non-rival, this assumption is reasonable. We call a network that maximizes $v(g^1)$ an efficient network.[8]

3.2. *Stability*

We next come to the equilibrium concept employed. To define which pre-disruption networks will be formed, we use the concept of pairwise stability, as introduced in Jackson and Wolinsky (1996). In our model the players forming the network move first, anticipating how adding or deleting a link will affect the response of the network disruptor, who moves second. Pairwise stability of the pre-disruption network consequently means that for any best response disruption strategy of the disruptor (i.e., for any (possibly mixed) strategy of the disruptor on which links to delete), the players form a pre-disruption network that is pairwise stable.[9] Accordingly, nodes have the freedom to add or delete links. However, while the decision on link deletion is unilateral, for link formation both players need to consent. Thus, a pre-disruption network is pairwise stable if and only if no player unilaterally wants to delete a link and no pair of players wants to add a link, always taking into account that the network will be disrupted and how the disruption will influence their respective expected payoffs. This can be stated as:
A pre-disruption network g^1 is pairwise stable iff

1. for all $g_{ij} \in g^1, u_i(g^1) \geq u_i(g^1 - g_{ij})$ and $u_j(g^1) \geq u_j(g^1 - g_{ij})$, and
2. for all $g_{ij} \notin g^1, if u_i(g^1 + g_{ij}) > u_i(g^1)$ then $u_j(g^1 + g_{ij}) < u_j(g^1)$.

Thus at least one player needs to strictly prefer forming a link over not forming a link, when the other player is indifferent. At the same time no player strictly prefers deleting a link. Pairwise stable networks are defined here with respect to the pre-disruption network, as networks are formed before disruption takes place. As we are interested in what networks players form to protect themselves against disruption, however, we are using the post-disruption expected payoffs as explained before.

3.3. *The Network Disruptor*

At Stage 2, the network disruptor observes the pre-disruption network g^1 and can then choose how to disrupt it. He has a disruption budget D, where $D \geq 0$, which refers to the number of links the disruptor will take out of the network. In the model, the goal of the network disruptor is to minimize the sum of the benefits of the nodes in the post-disruption network, $\sum_i^n b_i$. In our model, in order to approximate this goal in a tractable way, we assume that the network disruptor has lexicographic preferences in the following

[7]Note again that $b_i(g^2)$ calculates the expected benefits of a player, as those may be stochastic if the network disruptor follows a mixed strategy as is explained in the following. We take the expected benefits to calculate the payoffs to player i, as he has to decide on links before he knows which links will be disrupted.

[8]Note that this concept of efficiency does not take into account the disruptor's payoff. This is due to the fact that the network disruptor is seen as an external force and we analyze his influence on the network the nodes may form.

[9]We assume that the network disruptor will randomly decide on which link to take out, if he is indifferent between multiple links. In contrast to the other players, who are assumed to use pure strategies, the network disruptor is thus assumed to use mixed strategies. We chose this seemingly inconsistent approach, as, on the one hand, having the disruptor choose pure strategies would render the analysis more cumbersome without adding to the general conclusions. On the other hand, leaving the other players to use mixed strategies would make it difficult to define pairwise stability.

sense.[10] Comparing two post-disruption networks, the disruptor always prefers the network where the component with the largest order is smaller. If two networks have largest components of equal order, the disruptor always prefers the one whose components with the second largest order is smaller. Again, if those have the same order, he prefers the network where the order of the third largest component is lower, and so on. Formally this means that any post-disruption network g^2 can be characterized by the order of its components, i.e., the number of nodes in each component.[11] Consider a post-disruption network g^2 with a set of $N = 1, 2, 3, ..., n$ players. This can be characterized by ranking its components by their order. So the network is characterized as $C_1, C_2, ..., C_k, ..., C_m$ where it holds that $|C_1| \geq |C_2| \geq ... \geq |C_k| \geq ... \geq |C_m|$ and $|C_1| + |C_2| + ... + |C_k| + ... + |C_m| = n$. Given a pre-disruption network g^1 including n players, we can then compare two possible post-disruption networks g_a^2 and g_b^2. We assume that the disruptor has lexicographic preferences in the following way. Letting $\succ$ denote the preference relation, then for the network disruptor it holds that $g^{2_a} \succ g^{2_b}$ iff $|C_1^a| < |C_1^b|$ or $|C_k^a| = |C_k^b|$ and $|C_{k^*}^a| < |C_{k*}^b|$, where k^* is the largest rank for which a component in g_a^2 and a component in g_b^2 with identical rank have a different cardinality.[12] By assuming lexicographic preferences of the network disruptor, we can approach modeling him as minimizing the sum of the benefits of all nodes in the post-disruption network, given any pre-disruption network using l links, as a whole. The value of any post-disruption component C is $|C|^2$, because all players within the component get the information of themselves and all other players. Therefore, larger components tend to matter more for the value of the network. It is for this reason that we approximate a disruptor who minimizes the sum of the players' benefits, by one with lexicographic preferences.[13] Costs for linking need not be taken into account when comparing the values of the networks, as they will be the same across all post-disruption networks, given a disruptor with a disruption budget of $D = x$, since the disruptor will take out exactly x links from the pre-disruption network.

4. Benchmark Case

We begin our analysis by looking at the benchmark case in which there is no disruptor. This is the case analyzed by Jackson and Wolinsky (1996) in the symmetric connections model without decay. While Jackson and Wolinsky do not treat the case without decay (δ) explicitly, it is straightforward to adjust their results to such a case and in the following we adjust all results to $\delta = 1$ where there is no information lost along the way. They find for the case of low linking costs (i.e., linking costs for which it is worth to be linked to one node for the value of that one node alone) that all minimally connected networks are pairwise stable, but the only efficient network is the star network. For $\delta = 1$, the architecture of the minimally connected network does not matter anymore and therefore all minimally connected networks are pairwise stable and efficient.

For the case of high linking costs (i.e., linking costs for which it is not worth to be linked to one node for the value of that one node alone) Jackson and Wolinsky (1996) find that for $\delta < c$ networks that are pairwise stable are such that each player has at least two links, as it is not worth it to be linked to an end-player. For $\delta = 1$, however, the only pairwise stable network is the empty network. Here as well, no player is willing to link to an end-player, thus minimally connected networks are not pairwise stable. At the same time,

[10]While other papers (see, e.g., Dziubiński and Goyal (2013)) focus solely on the connectivity of the remaining network, we thus allow also for non-connected network structures after disruption.

[11]The order of the network is the size of the network in terms of the number of nodes it includes. This term is used rather than the size of the network so as to avoid confusion, as it is a commonly used term in graph theory.

[12]Rank refers here to the ordering of the components by their order. Thus Rank 1 is the largest component, Rank 2 the second largest and so on.

[13]This may not hold in extreme cases. Consider the case of a network consisting of 15 players which are either split up in a component with 10 players and 5 singleton players, or in a component with 9 players and a component with 6 players. According to his lexicographic preferences, he will prefer the second option. However, the value of the network in case 1 is 105 while it is 117 in the second case. Thus here the lexicographic preferences would not coincide with minimizing the sum of the benefits of all nodes in the network. In practice, the disruptor will not face such extreme choices, because ensuring that the post-disruption network has 5 singleton players will require a much higher disruption budget D than ensuring that the post-disruption network has two components of order 9 and 6.

because $\delta = 1$, the circle network is also not pairwise stable, as every player has an incentive to delete at least one link, as he will also get all the information in the network if it is a line instead of a circle, but his costs are lower. By deleting a link, a player becomes an end-player and the network will unravel. While the empty network is thus the only pairwise stable network, it is not necessarily efficient. It is easy to see that depending on the linking costs and the number of players, connected networks will have a higher value than the empty network, but they are never pairwise stable.

5. Full Characterization for a Disruption Budget of $D = 1$

Before turning to a general analysis of the case where there is a disruptor with a positive disruption budget, we will first provide a full characterization of the pairwise stable networks when the disruption budget is equal to one ($D = 1$). This will provide intuition for the analysis of the general case which will then follow. We do not explicitly split up the analysis between high and low linking costs, as the results and proofs are very similar. We do, however, show the differences in the effect a network disruptor has, depending on the linking costs. Since we have normalized the value of each node to $w_i = 1$, this means that we call $c < 1$ low linking costs and $c > 1$ high linking costs. To give a full characterization, we start by looking at all non-robust networks in Section 5.1 and then go on to look at all robust networks consisting of one component in Section 5.2, before finally turning to discuss all non-connected networks in Section 5.3.

5.1. Pairwise Stable Non-Robust Networks for $D = 1$

In any network in which at least one pair of nodes is connected via only one *link-independent path*,[14] a network disruptor with a disruption budget of $D = 1$ can disconnect at least one node from the network. Therefore, any such network is not robust against disruption with a disruption budget of $D = 1$ in the sense that after disruption the network will consist of at least two separate components. We will refer to these networks as *critical link networks*.

Definition 1. Critical link networks *are all networks in which at least one pair of nodes is connected via only one link-independent path. Thus, these are all networks in which there is at least one critical link upon deletion of which the network is split into two components.*

A subset of critical link networks is the set of minimally connected networks in which *all* pairs of nodes are connected via exactly one link-independent path. In terms of its architecture, the star network, g^*, is a special case within this set of networks, as there are exactly $(n-1)$ end-players and only one central node. Therefore, with *every* link *exactly one* node can be disconnected from the network. As the network disruptor is assumed to be indifferent between which link he will target, if there are multiple links that lead to the same post-disruption network architecture, the chance of any one end-player of being disconnected is thus lower than in any other minimally connected network. Consequently, each node has a lower incentive to add links in the star network than in any other minimally connected network. Additionally for $c < 1$ there are no incentives to delete a link. Thus, the star is the minimally connected network that is most likely to be pairwise stable.

Lemma 1. *For a disruption budget of $D = 1$, the star network g^* is pairwise stable only iff $1 - \frac{1}{n-1} < c < 1$ for $n > 3$.*

Proof The expected payoff to any end-player i in the star network is $u_i(g^*) = \frac{1}{n-1} * 1 + (1 - \frac{1}{n-1})(n-1) - c$, where the first part of the function denotes the payoff should node i be disconnected, and the second part denotes the payoff should node i not be disconnected. Adding a link to any other end-player j ensures node i that he will remain in the network.[15] His payoff is thus $u_i(g^* + g_{ij}) = n - 1 - 2c$. Comparing the

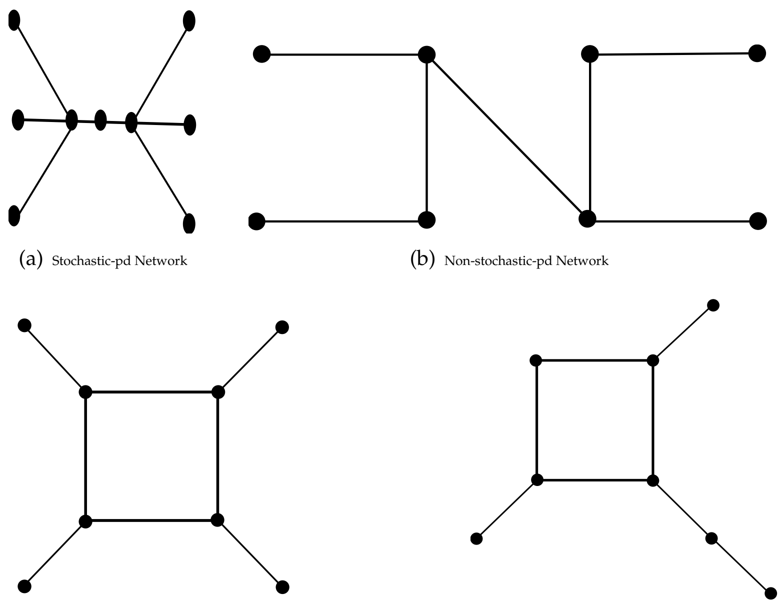
[14] The number of link independent paths between any two nodes in a network is the number of different paths between the two nodes that do not share a link.

[15] An exception here is the case of $n = 3$ as then no other node can be disconnected. However, then the star network is also equal to the line and can be analyzed as such.

two payoffs, we find that $u_i(g^*) > u_i(g^* + g_{ij})$ holds only if $c > 1 - \frac{1}{n-1}$. The payoff of the central player z is $u_z(g^*) = (n-1) - (n-1)c$. Should he delete a link to an end-player i, his payoff is $u_z(g^* - g_{zi}) = (n-2) - (n-2)c$. For any $c < 1$ it holds that $u_z(g^*) > u_z(g^* - g_{zi})$. However, for any $c > 1$, it holds that $u_z(g^* - g_{zi}) > u_z(g^*)$. Thus the star network is pairwise stable only for $1 - \frac{1}{n-1} < c < 1$.

Only for a very small cost range do we find the star network to be pairwise stable. Moreover, for an increasing number of players the cost range for pairwise stability gets smaller. Consequently, the overall range for which the star network is pairwise stable is very small. As the star network is the one with the highest likelihood of being pairwise stable, we do not expect other minimally connected structures, or critical link networks in general, to be pairwise stable. To analyze these types of networks, we will make a distinction between different types of pre-disruption networks, which we define as *stochastic-pd* networks and *non-stochastic-pd* networks, where 'pd' stands for pre-disruption, and where each pre-disruption network is either stochastic-pd or non-stochastic-pd. Informally speaking, stochastic-pd networks are all such networks as the star, where the disruptor is indifferent between deleting a number of links. Consequently, players in such a network do not know what their payoff after disruption will be, as the order of the component they belong to after disruption depends on which link the disruptor takes out. Non-stochastic-pd networks are then those networks where the network disruptor will certainly target one specific link. Thus, players know exactly what their payoff after disruption will be.

Definition 2. *A* stochastic-pd network *is any pre-disruption network in which the network disruptor is indifferent between disrupting multiple different links. A* non-stochastic-pd network *is any pre-disruption network where the network disruptor has a clear preference for targeting a specific (set of) link(s).*[16]



(a) Stochastic-pd Network

(b) Non-stochastic-pd Network

(c) Stochastic-pd Network

(d) Non-stochastic-pd Network

Figure 1: Stochastic - and Non-stochastic-pd Networks

Examples of such networks can be found in Figure 1, where Subfigures 1(a) and 1(b) are examples of minimally connected networks, and Subfigures 1(c) and 1(d) are examples of connected network structures

[16]This means that in non-stochastic-pd networks the network disruptor's best response is a pure strategy, whereas in stochastic-pd networks his best response is a mixed strategy.

which are not minimally connected but belong to the set of critical link networks. Subfigures 1(a) and 1(c), are examples of stochastic-pd networks. In these networks, the network disruptor is indifferent between targeting multiple links, as he cares only about the order of the remaining components, which is the same independent of which component he disconnects. On the other hand, in non-stochastic-pd networks, such as the ones in Subfigures 1(b) and 1(d), there is only one link that will definitely be disrupted. In the following, we show that neither of these types of networks is pairwise stable.

Lemma 2. *For a disruption budget of $D = 1$, no critical link network consisting of one connected component is pairwise stable unless it is the star network.*

Proof We first prove this for stochastic-pd and then for non-stochastic-pd critical link networks consisting of one connected component.

- In stochastic-pd networks there are $m \geq 2$ links that connect equally sized pd-components to the rest of the network.[17] Consequently, each one of these m links is an equally likely target for the network disruptor. Assume each of the pd-components is of order $|y|$. Then the expected payoff for any node k in one of the pd-components is $u_k(g) = \frac{1}{m}|y| + (1 - \frac{1}{m})(n - |y|) - vc$, where v denotes the number of links k has.
  1. Assume any of the pd-components is **not** minimally connected and that k and l are part of this pd-component. Then take any link kl upon deletion of which the pd-component still remains connected. Node k's payoff if this link is taken out is then given by $u_k(g - g_{kl}) = \frac{1}{m}|y| + (1 - \frac{1}{m})(n - |y|) - (v - 1)c$, which is always larger than $u_k(g)$. Thus the network is never pairwise stable.
  2. Assume pd-component C_k is minimally connected. Then let node k be linked to an end-player r. As the pd-component is minimally connected, it has at least 2 end-players.[18] If node k deletes the link to node r, the pd-component is only of order $(|y| - 1)$ and thus no longer a target of the network disruptor. Thus, the payoff of node k will be $u_k(g - g_{kr}) = n - |y| - 1 - (v - 1)c$. Comparing this payoff with $u_k(g)$ we find that there is always an incentive to delete the link unless $-\frac{n-2|y|}{m} + 1 > c$. Now consider node k adding a link to node i in one of the other pd-components of order $|y|$. To show that a stochastic critical link network is not pairwise stable, we need to show that when players do not have an incentive to delete a link in a minimally connected pd-component, there is an incentive for two players in different pd-components to form a link g_{ik} between them. To analyze if this is the case, we need to distinguish between 2 cases based on the number of pd-components in the network:
     - Assume $m \geq 3$. If there is a link g_{ki}, the network disruptor will then disrupt one of the pd-components of order $|y|$ that does not include k or i, as they have formed a circle by adding the link. The payoff to node k is thus $u_k(g + g_{ki}) = n - |y| - (v + 1)c$ and the network is pairwise stable only if $-\frac{n-2|y|}{m} + 1 > c > \frac{n-2|y|}{m}$. This may hold only if $\frac{m}{2} > n - 2|y|$ and at the same time we know that $n \geq m|y| + 1$. We can merge these two conditions as $\frac{m}{2} > m|y| + 1 - 2|y|$, which can be rewritten as $m - 2 > 2|y|(m - 2)$. This, however, may be fulfilled only if $2 > m$. But by assumption we have that $m \geq 3$ and consequently, such a network is not pairwise stable.
     - Assume $m = 2$. The link g_{ki} then ensures that neither of the two links that were previously possible target links can now be disconnected. Thus the payoff to node k is given by $u_k(g + g_{ki}) = n - z - (v + 1)c$, where z denotes the number of nodes that can now be disconnected and where it needs to hold that $z < |y|$, as otherwise the disruptor would have disconnected z nodes before. The network is then pairwise stable only if $-\frac{n-2|y|}{2} + 1 > c > \frac{n-2|y|}{2} + |y| - z$

[17] By pd-components, we denote those parts of the networks that might become disconnected components, but which are currently still part of the one connected component in the network.

[18] Thus, there is at least one end-player before the disruption takes place.

holds. This may hold only if $-(n-2|y|) > |y| - z - 1$. As $n \geq 2|y| + 1$, the left-hand side is smaller than 0 whereas the right-hand side is at least 0. Thus, the network is not pairwise stable.

- In non-stochastic-pd networks, by definition, there is a single link that the network disruptor strictly prefers to delete. Let this targeted link be g_{ij} and assume that without link g_{ij} the network will consist of two separate pd-components C_i and C_j, where C_i includes node i and C_j includes node j. We can then distinguish between two cases:
  1. Assume that without link g_{ij}, the network disruptor will disrupt in pd-component C_i. Then the payoff for node j if link g_{ij} is maintained is $u_j(g) = |C_j| - vc$, where v denotes the number of links node j has. If j cuts the link to player i on the other hand, his payoff is $u_j(g - g_{ij}) = |C_j| - (v-1)c$. Independent of linking costs, it thus always holds that $u_j(g - g_{ij}) > u_j(g)$ and the network is therefore not pairwise stable.
  2. Assume that without link g_{ij}, the network disruptor is indifferent between disrupting in pd-components C_i or C_j or not able to cause any damage in either pd-component. We can then distinguish between two cases.
     – Assume C_i and C_j are **not** minimally connected. Then take any link g_{kl} in C_i upon deletion of which the pd-component still remains connected. Node k's payoff if this link is taken out is then given by $u_k(g - g_{kl}) = |C_i| - (v-1)c$, whereas the payoff with the link is $u_k(g) = |C_i| - vc$ which is always smaller than $u_k(g - g_{kl})$. Thus the network is never pairwise stable.
     – Assume C_i and C_j are minimally connected. Then look at any node k connected to an end-player l in C_i. Node k's payoff in the original network is then given by $u_k(g) = |C_i| - vc$. If node k deletes the link to end-player l, his payoff depends on the possibly changed best response strategy of the network disruptor. Given that the network disruptor had a clear preference for disrupting link g_{ij} before and is indifferent between disrupting in C_i or C_j if g_{ij} is not kept, even if he changes his strategy now, node k's payoff is minimally $u_k(g - g_{kl}) = |C_i| - 1 - (v-1)c$. Node k then has an incentive to delete the link if $c > 1$. If node k adds a link to node t in C_j, they automatically form a circle, thereby ensuring that there is more than one link-independent path between them, thus making link g_{ij} no longer the targeted link. The payoff for node k after disruption will at least be $u_k(g + g_{kt}) = |C_i| + |C_j| - z - (v+1)c$, where z denotes the number of nodes that can be disconnected from from the new joint component if there is a link between k and t. There is an incentive to add the link for node k then only if $|C_j| - z > c$. For pairwise stability it would thus need to hold that $1 > c > |C_j| - z$. However, this never holds as $z < |C_j|$, because otherwise link g_{ij} would not have been the only target in the first place. Consequently, the network is not pairwise stable.

That the star network is pairwise stable for some cost ranges follows from Lemma 1.

We have hereby shown that unlike in the benchmark case, where minimally connected networks were stable and efficient for low linking costs, minimally connected networks consisting of one component are not stable when facing a network disruptor, with the exception of the star network (and there only for a very small range of linking costs, namely $1 - \frac{1}{n-1} < c < 1$). The same holds for any critical link network. For the case of high linking costs, in the benchmark case no minimally connected network is pairwise stable, and as we have shown above, the same holds for the case with a network disruptor.

Given that minimally connected networks except for the star are never pairwise stable, we need to look for different network structures that may be pairwise stable. In Hoyer and De Jaegher (2016), in a purely structural model of network disruption and defense between a network designer and disruptor, it was found that a class of networks, known as regular networks, which are completely symmetric in the sense that each player has the exact same number of links as any other player in the network, are completely disruption proof. As vulnerability of the network seems to be one of the main reasons for the lack of stability in the networks we have discussed so far, we will now analyze a generalization of regular networks, namely networks with a generalized circle structure (as defined in the following section), which are completely robust against disruption, to see if such networks will also be pairwise stable in a model where nodes face the network disruptor.

5.2. Pairwise Stable Robust Networks for $D = 1$

Having looked at networks that include at least one pair of nodes that is connected only via one link-independent path, we will now turn to the remaining possible networks consisting of one component, namely all those networks where each pair of players is linked via more than one link-independent path. In such networks there are no end-players. We term these networks, networks with a generalized circle structure, as defined below. As we do not make any claims on the structure of these networks, this set of networks includes all remaining connected pre-disruption networks, next to the ones we have already discussed in the previous section.

Definition 3. *A network has a* generalized circle structure *if each pair of nodes is connected via at least two link-independent paths.*

It is obvious that networks that have such a generalized circle structure are safe against disruption. However, they are not necessarily pairwise stable. Networks with a generalized circle structure may be pairwise stable only if there are no redundant links. Thus as soon as one link is taken out, the network disruptor can disconnect at least one node. Consequently, every link is critical to the robustness of the network. This means that there is at least one pair of nodes between which there are only exactly two link-independent paths. If this is not the case, there is at least one redundant link in the network.

Definition 4. Generalized circle networks without redundant links *are all those networks where each pair of players is connected via at least two link-independent paths and where each link is critical in the sense that if* ***any*** *of the links were to be deleted the network would be no longer safe against disruption.*

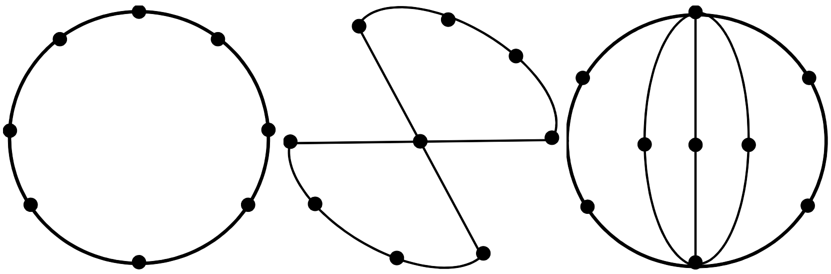
Lemma 3. *Any network that has a generalized circle structure and no redundant links will be pairwise stable against a network disruptor with a disruption budget of $D = 1$, for $|y| > c$, where $|y|$ denotes the order of any pd-component that can be disconnected if players decide to delete a link.*

Proof Take any node i in a network with a generalized circle structure without redundant links. Its payoff will be defined as $u_i(g) = n - vc$, where v denotes the number of links that node i possesses. Should player i delete a link, his payoff will be $u_i(g - g_{ij}) = (n - |y|) - (v - 1)c$, where $|y|$ denotes the order of the disconnected component.[19] If $c < |y|$, it always holds that $u_i(g) > u_i(g - g_{ij})$. Adding a link will only increase the costs, while keeping the size of the component the same. Consequently, $u_i(g) > u_i(g + g_{ij})$ and the network is pairwise stable for $|y| > c$.

Looking at the different networks in Figure 2, we can see that for the circle, the order of y will be $\frac{n}{2}$, no matter which node deletes a link. For the network in Figure 2(b) the order of y depends on which node decides to delete a link, but it will in any case be of at least order 2. For the network in Figure 2(c) the order of y is equal to 1. This means that the cost range for which such a network may be pairwise stable is maximal for the circle and it holds that $1 \leq y \leq \frac{n}{2}$. Put differently, this means that the circle network is pairwise stable for the highest cost range of any generalized circle networks. Additionally, the circle is also the most efficient of these networks, as it uses only nine links to achieve stability, whereas the spanning circles uses 10 links and the nested circle 12 links.

Any network with a generalized circle structure that has redundant links is not pairwise stable, as there is always an incentive to delete a link for players incident to the redundant link. As players are already

[19] $|y|$ could be stochastic. For example, if you take the circle with $n = 9$, a player will either remain in a component of order 4 or order 5, in which case for the sake of calculation we would assume $|y| = 4.5$. This does not change the results though.



(a) Circle (b) Spanning Circle (c) Nested Circle

Figure 2: Networks (n=9) without Redundant Links for $D_l = 1$

safe within the network in any form of a circle network, when facing a network disruptor with a disruption budget of $D = 1$, any additional link will only add costs but not benefits. Thus, players in such a network that have links additional to the links needed to form a generalized circle always have an incentive to delete a link and therefore the networks are never pairwise stable.

We have thus seen that networks in which players are completely safe against the disruption by a network disruptor are pairwise stable for low costs, and some also for higher costs. Here the cost range depends on the network structure and how many nodes may be disconnected if players decide to delete a link before facing the network disruptor. That such networks are pairwise stable in the presence of a network disruptor, but not otherwise (=common enemy effect), can be explained by what happens if players delete a link. Unlike in the benchmark case, where players also obtain all information if they delete a link in a generalized circle structure, the incentives to delete links are lower when facing a network disruptor, as players will obtain all information only if they keep all their links. Otherwise, at least one player can be disconnected from the network and the benefits of all players are lowered.

While for low and high linking costs the same network structures are pairwise stable, for an intermediate cost range (decreasing in n) the star network is also pairwise stable. This is caused by the fact that only for this intermediate cost range, there is no incentive for end-players to add a link to one another, or for central players to delete the link to end-players. However, as n increases, the cost range for which this is possible, is decreasing.

5.3. *Multiple Components*

After having looked at a complete characterization for the case of $D = 1$ for networks consisting of one component only, we will now look at networks consisting of multiple components. While in the benchmark case without a disruptor such networks are never pairwise stable, when introducing a network disruptor we find that the disruptor can indeed cause such segmentation to be pairwise stable. To characterize all networks consisting of multiple components, we again subdivide this set of networks into several categories.

We will first have a look at the empty network, g^e. Whereas in the benchmark case any single pair of players has an incentive to form a link for low costs but not for high costs, here this one link will become an automatic target for the network disruptor and will thus definitely be taken out. Therefore there is no incentive for players to add links for any range of linking costs and the empty network is pairwise stable. This result, while very straightforward, is also a marked distinction between this analysis and the benchmark case without a threat of disruption. In the benchmark case the empty network is pairwise stable only for high linking costs but not for low linking costs. Here we find that the empty network is also pairwise stable for low linking costs.

Lemma 4. *The empty network, g^e is pairwise stable when facing a network disruptor with a disruption budget of $D = 1$.*

Proof The payoff for any node i in the empty network is $u_i(g^e) = 1$. Should nodes i and j agree to add a link, this link will be the only link in the network and therefore the automatic target of the network disruptor. Thus the payoff for node i would be $u_i(g^e + g_{ij}) = 1 - c$. Since $u_i(g^e) > u_i(g^e + g_{ij})$, the empty network is pairwise stable.

The next set of networks that we analyze are *non-stochastic-pd networks* consisting of multiple components. This set of networks includes all those networks where the network disruptor has a clear preference about which component he will target. In general it includes all those networks that consist of components of different orders, or of components that have the same order but different architectures, such that the network disruptor can disconnect more nodes from one component than from another. We will show in Lemma 5 that such networks are only pairwise stable under very restrictive conditions, as players in the non-targeted components often have incentives to change their links if such a change does not influence the disruptor's best response. At the same time players in the targeted component often have incentives to change their links to either avoid being targeted or ensure that they remain in a post-disruption component of the largest possible order. Obviously not pairwise stable are those networks with sufficiently small components that are not targeted, as there is always an incentive for the players then to add a link between the components to increase its order, and therefore their payoffs, while ensuring that it is still small enough not to be targeted.

Lemma 5. *Non-stochastic-pd networks consisting of multiple separate components facing a network disruptor with a disruption budget of $D = 1$ are not pairwise stable unless the targeted component C_i is the star and*

1. $\forall j \neq i : |C_i| - 1 \leq |C_j|$ *and it holds that all other components are generalized circles. Then the network is pairwise stable if* $1 - \frac{1}{|C_i|-1} < c < 1$.
2. $\forall j \neq i : |C_i| - 1 > |C_j|$ *and it holds that all other components are minimally connected. Then the network is pairwise stable if* $\forall k, j \neq i : |C_i| < |C_k| + |C_j|$ *and linking costs fall in the range of* $1 - \frac{1}{|C_i|-1} < c < 1$.

Proof Assume the network consists of m different components and the network disruptor has a clear preference for disrupting in component C_i, which includes node i. Should any of the components include redundant links the network is clearly not pairwise stable, thus for the remainder we assume that none of the components includes redundant links. The payoff to node i is then given by $u_i(g) = |C_i| - |y| - vc$, where y denotes the number of nodes that will be disconnected from C_i and v denotes the number of links i has. We first prove statement (1) and then statement (2) of the lemma.

1. Assume that $\forall j \neq i : |C_i| - 1 \leq |C_j|$ holds. This means that all remaining components are of an order larger than the targeted component C_i. As we have already excluded components with redundant links, and the network disruptor has a clear preference about which component he targets, C_i needs to have a critical link structure as otherwise the disruptor would not target it, whereas all other components may have a critical link or a generalized circle structure.
   - Assume all other components, C_j, are generalized circles. Then as $|C_i| - 1 \leq |C_j|$, they will automatically be targeted should the players decide to remove one link. By Lemma 3 we know that the nodes will not have any incentive to do so if $y > c$, where y denotes the number of nodes that can be disconnected from C_j if one link has been removed. Whether the players in component C_i will have incentives to change their links, can be analyzed according to Lemma 2. There we find that only if C_i is a star and it holds that $1 - \frac{1}{|C_i|-1} < c < 1$, the network will be pairwise stable.
   - Assume at least one other non-targeted component, C_j, has a critical link structure. Then the network can be pairwise stable only if the targeted component, C_i, is a star, as seen above. However, as only one node can be disconnected from the star component, the other component C_j would automatically have been a target to begin with, as at least one node can be disconnected since it has a critical path structure and by assumption, $|C_i| - 1 \leq |C_j|$. Thus, such a case cannot exist.

2. Assume that $\forall j \neq i : |C_i| - 1 > |C_j|$ holds. This means that all non-targeted components are of a lower order than the targeted component C_i.
   - Assume that $|C_i| > |C_k| + |C_j|$. Then node k in C_k has an incentive to link to node j in C_j, as even with this link, the network disruptor will still prefer to disrupt in component C_i, as it is of a larger order than the new joint component. This incentive holds as long as $c < |C_j|$. Thus only for very high costs will k not have an incentive to add the link. C_k needs to be minimally connected, however, as the component is not targeted. Therefore, if it is not minimally connected, there is always an incentive for players to remove at least one link that is not needed to ensure connectivity. Then, node k's payoff if he deletes a link to an end-player t will be $u_k(g - g_{kt}) = |C_k| - 1 - (v-1)c$. Thus for pairwise stability it needs to hold that $1 > c > |C_j|$, which is never fulfilled. Consequently, the network is not pairwise stable.[20]
   - Assume that $|C_i| \leq |C_k| + |C_j|$. In this case, if the other components are connected and it holds that $|C_j| < |C_i| - 1$ for all $j \neq i$ they will not be targeted and do not have any incentives to add or delete links. Therefore, we can analyze C_i by Lemma 2, in which we have already seen that such a network will not be pairwise stable, unless it is the star and then only if $1 - \frac{1}{|C_i|-1} < c < 1$. As $|C_j| < |C_i| - 1$ for all $j \neq i$, this is not influenced by the presence of the other components, since even if the central player in C_i deletes a link, the disruptor will still target C_i.

Consequently, the only non-stochastic-pd network consisting of multiple separate components that is pairwise stable is a network where the targeted component is a star and all other components are minimally connected and it holds that $1 - \frac{1}{|C_i|-1} < c < 1$ and $|C_j| < |C_i| - 1$ for all $j \neq i$, while it does not hold that $|C_i| > |C_k| + |C_j|$ or where all other components are generalized circles and it holds that $1 - \frac{1}{|C_i|-1} < c < 1$ and $|C_j| \geq |C_i| - 1$ for all $j \neq i$.

We now turn to *stochastic-pd networks* consisting of multiple components. These are the networks where the network disruptor is indifferent about which component he targets. This is the case if he cannot cause any damage, independent of which component he targets: i.e., if all the components have a generalized circle structure or are empty components. The network disruptor is also indifferent about which component he targets if he can cause damage to the component (thus there is at least one pair of nodes that is connected only via one link-independent path within the component), but the network structure in terms of the order of the post-disruption components is the same independent of which link he targets. This, however, can be achieved only if those pre-disruption components, between which the disruptor is indifferent, are of the same order. It is of course possible that next to the m components between which the disruptor is indifferent, there are also components in the network that will not be targeted. However, we can then treat these components the same way as we treated the non-targeted components for non-stochastic-pd networks.

We have already seen in the proof on non-stochastic-pd networks that if there are components in the network that will not be targeted and these components are very small, the network is not pairwise stable. Consequently, the only cases we need to analyze are those where if there are any components in the network that will not be targeted, these are not so small that they would have incentives to link to one another. The network may then be pairwise stable only if there are at least 3 components. This is caused by the decrease in the possibility of being part of the component that will be targeted by the network disruptor. If there are only 2 components, the probability that one component will be targeted is 50% and consequently, incentives for adding a link to another component are rather high if linking costs are sufficiently low. So costs need to be sufficiently high. But if they are, then there are incentives to delete a link such that the component is a bit smaller and therefore will not be targeted anymore. The more components there are, the less likely it is that one specific component will be targeted and therefore the incentives for adding or deleting links decrease. To identify cost ranges for which such a network may be pairwise stable, we consequently identify the nodes that have the most to gain by adding a link to another component, and

[20] A special case of this is if any of the components is the empty component. However, the proof also holds then.

those that have the least to lose by deleting a link within the component. To show that the network may be pairwise stable for a certain cost range, it is then enough to show that the node(s) with the most to gain do not have an incentive to add a link, while at the same time the node(s) with the least to lose do not have any incentive to delete a link. As the proof consists of stating a number of different conditions, we will show this for some special cases and relegate the remainder of the proof to the appendix.

Lemma 6. *For certain cost ranges, stochastic-pd networks consisting of multiple different targeted components, C_i, are pairwise stable if for any non-targeted components C_k and C_j it holds that $|C_k| + |C_j| > |C_i|$ (where $|C_k| \wedge |C_j| > 1$), and*

- *each component has a generalized circle structure without redundant links or is an isolate node, or*
- *independent of the structure of the components, the network consists of at least 3 components with critical link structure and including end-players*

Proof As we have seen in Lemma 5, networks will not be pairwise stable if for any non-targeted components C_k and C_j it holds that $|C_k| + |C_j| < |C_i|$. We will thus omit this part of the proof and assume for the remainder of this proof that if there are non-targeted components it holds that $|C_k| + |C_j| > |C_i|$, in which case they do not influence the disruptor's strategy. We first look at those networks where the network disruptor cannot cause any damage in the components and then turn to those networks where he can cause damage within the components.

- Assume there are multiple components where each has a generalized circle structure or is an isolate node and at least one of the components is non-empty. Further assume there is a node i part of a non-empty component C_i. Its payoff is then given by $u_i(g) = |C_i| - vc$, where v denotes the number of links node i possesses.
  1. Assume there is at least one redundant link in C_i and it links node i to node k, which is also in C_i. Node i's payoff when deleting the link will be $u_i(g - g_{ik}) = |C_i| - (v-1)c$. Comparing this to $u_i(g)$ we find that the network is never pairwise stable.
  2. Assume there are no redundant links in any of the components. Adding a link within the component does not make sense, as the network disruptor cannot cause any damage within the components. However, for high costs, maybe deleting a link makes sense. The payoff if node i deletes its link to node t in C_i is $u_i(g - g_{it}) = |C_i| - |y| - (v-1)c$, where $|y|$ denotes the order of the component that can then be disconnected from C_i. Comparing the payoffs, we find that node i does not have an incentive to delete the link if $|y| > c$. There is no incentive for i to add a link to a player j in C_j as that link will directly be targeted. If not all components are of the same order, or $|y|$ differs between components, then the network is pairwise stable only for the lowest value of $|y|$ across all components.
  3. Assume that node h is in an isolated node. Any link that node h may add will be an automatic target. Therefore node h has no incentive to add a link and the analysis for pairwise stability of the previous step still holds.
- Assume that there are m different components for which the disruptor is indifferent between which one to target. To analyze this case we need to distinguish between components that are stochastic-pd and non-stochastic-pd and between components that are minimally connected and those that have a critical link structure but are not minimally connected. While the same type of logic is used throughout this proof, a number of subcases need to be distinguished. Here we will therefore show only the proof for the case of multiple star components and multiple non-stochastic components and move the remainder of the proof to the appendix.[21]

[21] The remainder of the proof encompasses the case of minimally connected stochastic-pd components, and of components with a critical link structure that are not minimally connected. This way we give a complete definition of stochastic-pd networks consisting of multiple components.

– Assume that the network is stochastic-pd and the components themselves are stars. Then the payoff to any node in the spoke can be calculated as $u_i(g) = \frac{1}{m}[\frac{1}{\frac{n}{m}-1} * 1 + (1 - \frac{1}{\frac{n}{m}-1})(\frac{n}{m} - 1)] + (1 - \frac{1}{m})\frac{n}{m} - c$. For the central node, the payoff is given by $u_c(g) = \frac{1}{m}(\frac{n}{m} - 1) + (1 - \frac{1}{m})\frac{n}{m} - (\frac{n}{m} - 1)c$. Comparing these two, we find that benefits (thus abstracting from any linking costs) for the central player are always bigger than those for the spokes. Consequently, the players in the spokes have more interest in adding a link to a node in another component, as that will ensure that they remain in a component of size $\frac{n}{m}$. As opposed to this, the central player has more interest in deleting a link, as this will ensure that the component is not targeted, as it is smaller than the other components. Stated differently, the marginal benefits of adding a link are highest for the spokes, whereas the marginal losses of deleting a link are lowest for the central player. Consequently, to find a general condition for pairwise stability, we need to analyze the case for which the spokes do not have an incentive to add a link to a node s in a different component, while at the same time the central player does not have an incentive to delete a link. We can state this as $u_i(g) > u_i(g + g_{is})$ has to hold at the same time as $u_c g > u_c(g - g_{ci})$, where $u_i(g + g_{is}) = \frac{n}{m} - 2c$ and $u_c(g - g_{ci}) = \frac{n}{m} - 1 - (\frac{n}{m} - 2)c$. Comparing these payoffs, we find that the network may be pairwise stable for $1 - \frac{1}{m} > c > \frac{\frac{2n}{m} - 3}{n-m}$. This can be fulfilled only if $m\frac{4-m}{3-m} < n$ for $m > 3$ and if $m\frac{4-m}{3-m} > n$ for $m \leq 2$. For $m > 3$ this is possible, as the left-hand side of the equation is lower than m and $m < n$ always holds. For $m \leq 2$ this can never be satisfied. Thus if we have more than 3 components we can always find cost ranges for which such networks are pairwise stable.
– Assume that while the network is stochastic-pd, the components themselves are non-stochastic and minimally connected. The payoff to node i in component C_i is then given by $u_i(g) = \frac{1}{m}(\frac{n}{m} - |y|) + (1 - \frac{1}{m})\frac{n}{m} - vc$, where v denotes the number of links i has and $|y|$ denotes the number of nodes that can be disconnected should the network disruptor target C_i. For non-stochastic components, the incentive to add a link are the same for all nodes, as the network will be split into equally sized parts. The incentives to delete a link are again highest for a node linked to an end-player.[22] Should node i add a link to node s in component C_s, its payoff will be $u_i(g + g_{is}) = \frac{n}{m} - (v + 1)c$. Assume i is connected to end-player r in C_i. If i deletes its link to node r, its payoff is given by $u_i(g - g_{ir}) = \frac{n}{m} - 1 - (v - 1)c$. Comparing these payoffs, we find that the network is pairwise stable if $1 - \frac{|y|}{m} > c > \frac{|y|}{m}$. For $m > 2$ we can find cost ranges that fulfill this condition. Thus, if there are at least 3 components there are cost ranges for which such a network is pairwise stable, but such networks consisting of only 2 components are not pairwise stable.

Thus for $m \geq 3$ we can find pairwise stable network structures.

We can thus conclude the analysis of the case of a disruption budget of $D = 1$, where we have found that minimally connected networks are in general not pairwise stable, with the exception of the star for a very small range of linking costs. Instead, networks that are regular and symmetric and therefore distribute benefits and costs of the network very equally are pairwise stable for larger cost ranges. At the same time, networks that are separated into multiple components may also be pairwise stable, which is a marked difference from the benchmark case. The complete characterization of pairwise stable networks when facing a network disruptor with a disruption budget of $D = 1$ is summarized in the following proposition.

Proposition 1. *For a disruption budget of $D = 1$ a network may only be pairwise stable, depending on costs and the requirements stated in the previous lemmata, if it has a generalized circle structure without redundant links, it is the star network, it is the empty network or if it consists of multiple separate components. Networks are **not** pairwise stable, if they include redundant links or they include at least one pair of nodes that is only connected via one link independent path (with the exception of the star network).*

[22] Note in the case of non-stochastic components that is not a central node.

Proof We have looked at all minimally connected networks or more generally all networks that include at least one pair of players that is only connected via one link independent paths and seen that those, with the exception of the star, which under very specific circumstances is pairwise stable, are not pairwise stable. Additionally we have analyzed any network that is based on a circle structure without redundant links and found that it is pairwise stable for certain cost ranges and that networks with redundant links are never pairwise stable. This provides a full characterization for networks consisting of one component. We have also looked at networks consisting of multiple components and found that they are generally not pairwise stable if they are non-stochastic-pd networks, unless the targeted component is the star network and certain conditions apply. If networks consisting of multiple components are stochastic-pd networks, there are cost ranges for which they are generally pairwise stable, if they consist of at least 3 separate components or if all components have a generalized circle structure without redundant links. Additionally we have found that the empty network is always pairwise stable. We have consequently characterized all possible network structures and shown under which circumstances they are pairwise stable for the case of $D = 1$. The results follow directly from Lemmata 1 to 6.

5.4. The Effect of a Network Disruptor - Value and Efficiency of Stable Networks

In the model described in Jackson and Wolinsky (1996), which we have used as a benchmark case in our analysis, there is a tension between stability and efficiency in the case of high linking costs, whereas for low linking costs in the model without decay, all minimally connected networks are pairwise stable and equally efficient. As they all use the same amount of links, the value of any minimally connected network can easily be calculated as $v(g) = n^2 - 2(n-1)c$. The only pairwise stable network in the benchmark case for high linking costs is the empty network, which has a value of $v(g^e) = n$. It is easy to see that other networks might have a higher value. Thus, for the case of high linking costs the only pairwise stable network is not generally the most efficient network.

When a network disruptor is introduced to the network formation setting, what we are interested in is the value of the network after disruption. Consequently, if we want to compare this with the value in the benchmark case, we need to be aware that we assume in the benchmark case a disruption budget of $D = 0$. We have seen in the previous section that connected pairwise stable networks, when facing a network disruptor with a disruption budget of $D = 1$, are the star, for a very small cost range, and generalized circles. The post-disruption value in the star network can easily be calculated as $v(g^*) = (n-1)^2 + 1 - 2(n-1)c$. In any network based on a generalized circle structure, no node can be disconnected. Thus, the post-disruption value for such networks is given by $v(g) = n^2 - 2vc$, where v denotes the number of links that are used in the network. It is straightforward to see that the circle network, which is the network based on a generalized circle structure that uses the least amount of links, will be the most efficient network and we can calculate the post-disruption value of the circle network as $v(g) = n^2 - 2nc$.

Comparing the values of the post-disruption networks for $D = 0$ and $D = 1$, we find that for low linking costs, the effect of a network disruptor on the value of the network is negative, independent of which equilibrium players coordinate on. Even when abstracting from the additional linking costs, which are necessary to keep the network connected for $D = 1$ (thus build a circle), overall the disruptor has a negative effect because there is a positive probability that the players coordinate on the empty equilibrium, the star or a separated equilibrium, in which case after disruption no full information sharing is possible. Thus, for low linking costs, the network disruptor has the expected negative effect on the value of the network. For the case of high linking costs we find that depending on which network players coordinate on, the network disruptor may actually have a positive effect on network formation. Comparing the post-disruption values of the circle and the empty network, we find that the circle has a higher value than the empty network as long as $\frac{n-1}{2} > c$. This means that if we look at the overall network value of the post-disruption network the value of the network is higher when facing a network disruptor than when no threat of a network disruptor is present. Abstracting from the additional links used, even if players do not coordinate on the circle but on other network structures based on a generalized circle structure or even on networks consisting of multiple separate components, the effect of the network disruptor can be seen as positive, as information sharing is possible in the presence of a network disruptor where it is not if no disruptor is present.

For high linking costs, we can thus see what we have described in the introduction as the common enemy effect. The network disruptor acts as a threat to the sum of the benefits of all nodes in the network, and players that are myopic and self-interested are able to coordinate on equilibria that increase the value of the network as compared to the only equilibrium they may coordinate on without the threat of such a common enemy. The result is in so far surprising, as we do not change the assumptions on the players or refer to any psychological effects. Instead we can show this effect using purely economic incentives. The assumptions on the players remain that they are self-interested and myopic. However, adding additional links that would be redundant without the presence of a network disruptor is in everyone's best interest and therefore leads to players protecting the network as a whole. We have documented in the introduction and literature review that the common enemy effect as such is a phenomenon that is studied in some way in a number of different disciplines. What we show here is that while psychological factors will definitely play a role in the emergence of a common enemy effect, it may also be in line with simple rational behavior that players manage to cooperate only when a common enemy is present.

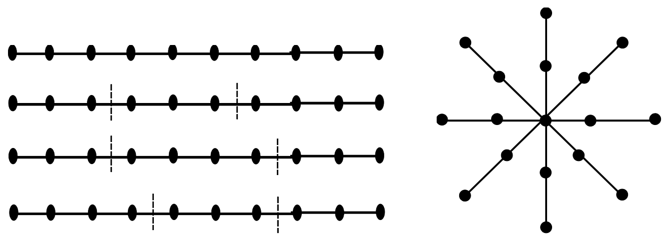
6. General Disruption Budget

After having given a complete characterization for the case of $D = 1$, we will now move to a general disruption budget of $D = x$. Here we cannot provide a full characterization of all networks that are pairwise stable for all cost ranges and numbers of nodes, due to the difficulty of describing the multitude of possible network structures. Having seen for the case of $D = 1$ which network structures are pairwise stable, we will now focus on showing that the results which we have found for networks that are pairwise stable, can be generalized to a general disruption budget.

6.1. The Star Network and Minimally Connected Networks

The payoff to any end-player i in the star network is given by $u_i(g) = \frac{x}{n-1} * 1 + (1 - \frac{x}{n-1})(n-x) - c$, if we assume that the network disruptor has a disruption budget of $D = x$, where $1 < x \leq n-3$. Should node i add a link to another end-player j he is always safe within the network and his payoff is given by $u_i(g + g_{ij}) = n - x - 2c$. Comparing these two payoffs we find that as long as $c > x - \frac{x^2}{n-1}$, the players do not have an incentive to add a link. At the same time, however, we know from Lemma 1 that only for $1 > c$ does the center player not have an incentive to delete a link, thus the network is pairwise stable if $1 > c > x - \frac{x^2}{n-1}$. This, however, can never be fulfilled for $1 < x \leq n-3$ and $n \geq 5$.[23] It can be easily verified that for the remaining cases of $x \geq n-2$, the network will not be pairwise stable.

We have seen that in the case of $D = 1$, other minimally connected networks apart from the star were not pairwise stable. Here we can show that this generalizes to a general disruption budget. When any minimally connected network faces a network disruptor with a disruption budget of $D = x$, it can be cut into $(x+1)$ separate components by a network disruptor. Along the line of the distinction between stochastic and non-stochastic-pd networks, which we introduced in the discussion of $D = 1$ (see Definition 2), we can now also divide the set of minimally connected networks into two categories. We will still call them stochastic- and non-stochastic-pd networks. However, this is only possible with some abuse of terminology, as is explained in the following. Take, for example, the line network depicted in Figure 3(a). The line network including 10 players can be split up in 3 different ways if a network disruptor has a disruption budget of $D = 2$. However, here we claim this network to be a non-stochastic-pd network, although the disruptor can take out several different sets of links. We do this, as here there is no chance of any node to remain in one large component or be disconnected. Instead, in such networks, the only issue is the divisibility, meaning that they would be non-stochastic if they had a divisible number of nodes. Otherwise, the set of links that will be disrupted would be fixed. Stochastic-pd networks are then such networks where a player could remain in one large connected component or be disconnected in a smaller one. An example of such a network is given in the generalized star in Figure 3(b). In such a graph, a



(a) Line Network (b) Generalized Star

Figure 3: Minimally Connected Non-stochastic and Stochastic-pd Networks

network disruptor with a disruption budget of 2 can disconnect any two of the spoke components, whereas the rest remains connected in one large component.

Having made this distinction, we can use it in the following lemma to prove that minimally connected networks are not pairwise stable for any disruption budget $D \geq 2$. The argumentation follows along similar lines as it did for the case of $D = 1$.

Lemma 7. *Minimally connected networks are not pairwise stable when facing a network disruptor with a disruption budget of $D = x$ for $x \geq 2$.*

Proof We will first look at non-stochastic-pd networks and then at stochastic-pd networks.

- Take any node i in a minimally connected **non-stochastic-pd** network, which, after disruption is in a pd-component C_i (the order of which may be stochastic as in the example above). The payoff to node i is then given by $u_i(g) = |C_i| - vc$, where v denotes the number of links i has. If node i adds a link to node j in pd-component C_j, the network disruptor can still cut the network into $x + 1$ pd-components. However, there would be a joint pre-disruption component $C_i + C_j$. As opposed to that, the network disruptor could also take out the link g_{ij} and then adapt his disruption strategy with the remaining $(x - 1)$ links such that he will cause maximal damage. But still in this case, node i's payoff is minimally given by $u_i(g + g_{ij}) = |C_i| + 1 - (v + 1)c$, as this action in effect leads to reducing the disruptor's budget from x to $(x - 1)$. Opposed to this, node i could also delete a link to end-player k in C_i.[24] His payoff would then be given by $u_i(g - g_{ik}) = |C_i| - 1 - (v - 1)c$. Comparing these payoffs to $u_i(g)$, we find that the network is pairwise stable only if $1 > c > 1$, which is never fulfilled.
- Take any node i in a minimally connected **stochastic-pd** network and assume that there are $m > x$ pd-components of order $|y|$ that could be disconnected by the network disruptor.[25] Then node i's payoff is given by $u_i(g) = \frac{x}{m} * |y| + (1 - \frac{x}{m})(n - x * |y|) - vc$.
  - Let $m \geq (x + 2)$. Should node i add a link to any node j in one of the other pd-components of order $|y|$, those two pd-components will not be disconnected, due to the network disruptor's lexicographic preferences. Node i's payoff is thus given by $u_i(g + g_{ij}) = n - x * |y| - (v + 1)c$. Should node i delete its link to end-player k in C_i, the pd-component is only of order $|y - 1|$, thus smaller than the others and will not be targeted.[26] His payoff is then given by $u_i(g - g_{ik}) = n - x * |y| - 1 - (v - 1)c$. Comparing these payoffs with $u_i(g)$, we find that the network is pairwise stable only if it holds that $\frac{x|y| - x(n - x|y|)}{m} + 1 > c > -\frac{x|y| - x(n - x|y|)}{m}$. This is possible only

[23] Any $n < 5$ is of no importance here as the condition $1 < x \leq n - 3$ cannot be fulfilled.
[24] For ease of exposition we use node i here again. However, it is not necessary that node i actually has a link to an end-player. It only needs to hold that any one player in C_i is linked to an end-player. As the network is minimally connected this is always given.
[25] For the case of $m \leq x$ we are in a non-stochastic-pd network.
[26] Again, here it is not necessary that this is the same node i as before, as long as it is a node i in component C_i.

if $\frac{m}{2} > x(n - x|y| - |y|)$, which can be rewritten as $n < \frac{m}{2x} + x|y| + |y|$. Combined with the condition that $n \geq m|y| + 1$, this means that it must be the case that $m|y| + 1 < \frac{m}{2x} + x|y| + |y|$, which is fulfilled only if $(x+1)|y| - 1 > m(y - \frac{1}{2x})$. This, however, can never be satisfied. Suppose that the right-hand side of the equation is positive. Then, if the inequality is not even valid for the minimal value of m, namely $m = x + 2$, this can never hold. For $m = x + 2$, the condition can be rewritten as $(x+1)|y| - 1 > (x+2)(y - \frac{1}{2x})$ iff $\frac{x+2}{2x} > y$. The left-hand side here is maximally equal to one, and $|y|$ is at least one. Thus, this is never fulfilled.

– Let $m = (x+1)$. Now let node i be linked to an end-player k in the same pd-component. Node k's payoff is then given by $u_k(g) = \frac{x}{x+1} * |y| + (1 - \frac{x}{x+1})(n - x * |y|) - c$. If node k adds a link to another end-player j in another pd-component, his payoff will at least be $u_k(g + g_{kj}) = |y| + 1 - 2c$, as by forming this link they form a circle. The disruptor will then disconnect according to his preferences, and this leads to possibly different links being disconnected than before. However, the pd-component that will include nodes k and j will be at least of order $(|y| + 1)$, as otherwise one of the other pd-components will become larger, as not all pd-components can be disconnected from the central component. Thus, node k will not add the link if it holds that $u_k(g) > u_k(g + g_{kj})$, which holds if $c > \frac{x|y|+|y|+x+1-n}{x+1}$. We have already discussed stars above, thus we know that $|y| \geq 2$. At the same time, there should not be an incentive for any other node to delete a link. If node i deletes its link to node k, the pd-component will be of order $(|y| - 1)$ and no longer be targeted. Node i's payoff is then given by $u_i(g - g_{ik}) = n - x|y| - 1 - (v-1)c$. Comparing this to $u_i(g)$, we find that node i will not delete the link if $\frac{x|y|-nx+x^2|y|+x+1}{x+1} > c$. For the network to be pairwise stable both conditions have to be met at the same time, which is the case only if $|y|(x+1) > n$. As $n \geq (x+1)|y| + 1$, this condition is never satisfied and therefore the network is not pairwise stable.

The result that the star network is pairwise stable holds only for a disruption budget of $D = 1$. For a larger disruption budget, minimally connected networks, without exceptions, are not pairwise stable, which is a marked distinction from the benchmark case without disruption, in which for low costs all minimally connected networks are pairwise stable.

6.2. *Generalized $(x+1)$ Paths Structures*

We will now turn to a generalization of the concept of generalized circles introduced in the analysis of the case of $D = 1$. In a network that has a generalized $(x+1)$ paths structure, a network disruptor with a disruption budget of $D = x$ cannot disconnect any players from the network, as they are all connected to one another via at least $(x+1)$ link-independent paths. However, not all such networks are pairwise stable. They are pairwise stable only if they do not include any redundant links as discussed for the case of $D = 1$.

Definition 5. *A network has a generalized $(x+1)$ paths structure if every pair of nodes is connected via at least $(x+1)$ link-independent paths.*

Definition 6. *Networks without redundant links are all those networks where each pair of players is connected via at least $(x+1)$ link-independent paths and that are therefore safe against disruption by a network disruptor with a disruption budget of $D = x$. In such networks each link is critical in the sense that if* ***any*** *of the links were to be deleted the network would be no longer safe against disruption.*

This set of networks includes many different network structures using different amounts of links. The set of networks that connects each pair of nodes via exactly $(x+1)$ link-independent paths achieves pairwise stability using the least amount of links possible and therefore achieves the highest value of the network. Such networks are $(x+1)$-regular networks - thus networks in which each player has exactly $(x+1)$ links. $(x+1)$ regular networks use exactly $n * \frac{x+1}{2}$ links to connect n nodes in one component. However,

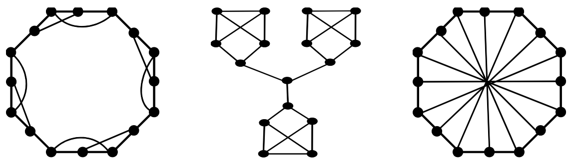


Figure 4: 3-regular Networks

not all $(x+1)$-regular networks also have $(x+1)$ link-independent paths between every pair of nodes, as can be seen in Figure 4, in which three 3-regular networks are depicted. It is straightforward to see that the only network in which there are 3 link-independent paths between any two nodes is the network to the right. By construction, in each one of these networks which have a generalized $(x+1)$ paths structure without redundant links, should any of the players delete a link, the network will no longer be robust against disruption.

Lemma 8. *Any network that has a generalized $(x+1)$ paths structure and no redundant links will be pairwise stable against a network disruptor with a disruption budget of $D = x$, for $|y| > c$, where $|y|$ denotes the order of any component that can be disconnected if players decide to delete a link.*

Proof Take any node i in a network with a generalized $(x+1)$ paths structure without redundant links. Its payoff will be defined as $u_i(g) = n - vc$, where v denotes the number of links that node i possesses. Should player i delete a link, his payoff will be $u_i(g - g_{ij}) = (n - |y|) - (v-1)c$, where $|y|$ denotes the order of the disconnected component.[27] If $c < |y|$, it always holds that $u_i(g) > u_i(g - g_{ij})$. Adding a link will only increase the costs, while keeping the size of the component the same. Consequently, $u_i(g) > u_i(g + g_{ij})$ and the network is pairwise stable.

One set of networks which has a generalized $(x+1)$ paths structure without redundant links and is $(x+1)$-regular, is the set of circulant graphs, which are defined below.[28]

Definition 7. *A circulant graph $C_n(a_1, a_2, ..., a_k)$, where $0 < a_1 < a_2 ... < a_k < \frac{n+1}{2}$, has node i adjacent to $i \pm a_1, i \pm a_2, ..., i \pm a_k$ (mod n). The sequence (a_i) is called the jump sequence and the a_i are called the jumps. The nodes of a graph are labeled $0, 1, 2, ..., n-1$.*

The circle is the circulant graph $C_n(1)$. In Hoyer and De Jaegher (2016), we formally show that circulant graphs that are regular of degree $(x+1)$ are disruption proof for a disruption budget of $D = x$, in the sense that no node can be disconnected from the main component and also no larger subset of nodes can be disconnected. This then also means that circulant graphs that are $(x+1)$-regular include only networks in which each pair of nodes is connected via exactly $(x+1)$ link-independent paths and therefore have a generalized $(x+1)$ paths structure without redundant links. As seen in Lemma 8, they are therefore pairwise stable. In the appendix, we use circulant networks to show that for a disruption budget of $D = x$, we can also find pairwise stable connected pre-disruption networks, for which the post-disruption network consists of multiple components.[29] This result is in line with Dziubiński and Goyal (2013), who look at a game between a network designer and attacker where single nodes can be defended by means of a defense mechanism, and the designer cares only about connectivity. They find that when defense costs are so high that no defense mechanism is used, but linking costs are low enough, circulant graphs are the best reply of the network designer.

[27] $|y|$ could be stochastic. This does not change the results though. It is also possible that i is in $|y|$. Then the cost range for pairwise stability is even larger.
[28] The definition and notation given below follows the one given by Boesch and Tindell (1984).
[29] For $D = 1$ an example of such a network is the star network.

6.3. Pairwise Stable Networks for $D = x$

Unlike for the case of $D = 1$, we unfortunately cannot define all networks that are pairwise stable for $D \geq 2$, but moving along the same classes of networks that we had definite results for when $D = 1$, we can analyze where these results generalize for a larger disruption budget and where they do not. Next to the networks we have already discussed above, another network for which we can give definite results is the empty network. As discussed in the analysis of the case of $D = 1$, the empty network is pairwise stable for any cost range. The logic for the case of $D = x$ is exactly the same as for the case of $D = 1$, as any link that might be added to the network will be an automatic target and, therefore, players never have an incentive to add such a link. We can thus directly state that the empty network is pairwise stable for any disruption budget x.

Corollary 1. *The empty network is pairwise stable for any positive cost range c and any positive disruption budget x.*

Finally, we will have a look at networks consisting of multiple components. As we have shown for the case of $D = 1$, there are cost ranges for which networks consisting of multiple components may be pairwise stable. Here we will focus on those type of network structures that we have shown for $D = 1$ to be pairwise stable and analyze if the results can be generalized for larger disruption budgets. While the results on networks consisting of multiple components that have a generalized circle structure can be extended to multiple components having a generalized $(x + 1)$-paths structure (potentially with isolate nodes) without redundant links, unfortunately the results on stochastic-pd networks consisting of at least 3 components that do not include any redundant links cannot be generalized.

Lemma 9. *Networks consisting of multiple components that each have a generalized $(x + 1)$-paths structure and do not include any redundant links or are isolate nodes are pairwise stable for low levels of linking costs.*

Proof Consider a network consisting of multiple components that each have a generalized $(x + 1)$-paths structure and do not include any redundant links. We know by Lemma 8 that for $|y| > c$, where $|y|$ denotes the order of any component that can be disconnected if players decide to delete a link, the network is pairwise stable.[30] The same holds here. However, if not all components are of the same order or if the order of $|y|$ differs between components - then the network is pairwise stable only for the lowest value of $|y|$ across all components. No player has any incentive to add a link to any of the other components, as such a link would be automatically targeted as the network disruptor cannot cause any damage within the components. That a network is also pairwise stable for the same cost range if there are isolated nodes as well is evident, as otherwise the link that might be added will automatically be disrupted as it is the only link the network disrupter can target and cause any actual damage to.

Unfortunately, we cannot claim any general results about networks consisting of multiple components if they do not fit the descriptions posed in the previous lemma. This is due to the change in the network disruptor's strategy, should an extra link be added or a link be deleted. For every specific structure, it is straightforward to tell if that structure is pairwise stable or not. However, we cannot draw any further general conclusions.

We can summarize our results on pairwise stable networks, when the players face a network disruptor with a general disruption budget of $x \geq 2$ in the following proposition.

Proposition 2. *When facing a network disruptor with a disruption budget of $x \geq 2$, minimally connected networks are not pairwise stable, independent of their structure. The set of pairwise stable networks includes:*

- *the empty network;*
- *networks that have a generalized $(x + 1)$-paths structure without redundant links;*

[30] Lemma 8 refers only to connected networks. However, as each component here has a generalized $(x + 1)$-paths structure and the disruptor thus cannot cause any damage to any of the components, we can apply the result here for each of the components.

- *networks consisting of multiple components that each have a generalized $(x+1)$-paths structure (or are isolate nodes) without redundant links for relatively low linking costs and,*
- *for an intermediate cost range x-regular circulant networks may also be pairwise stable.*

Proof We cannot provide a general characterization of all networks that are pairwise stable for a general disruption budget of $D = x$ and $x \geq 2$. However, that the mentioned networks fall into the set of pairwise stable networks follows directly from Lemma 8, Lemma A.1, Corollary 1, and Lemma 9. That minimally connected networks are not pairwise stable follows directly from Lemma 7.

Again, as for the case of $D = 1$, we have shown here that when a network disruptor is present, multiple equilibria can be reached. Also for the general case of $D = x$, the presence of a network disruptor can be interpreted differently for the case of low linking costs and the case of high linking costs. For low linking costs, the network disruptor can be seen as a possibly negative force again, should players coordinate on the empty network or a network split up into several separate components, since, as already discussed in Section 5.4, this leads to post-disruption networks in which information sharing between all players is no longer possible, whereas it is possible in the benchmark case. Yet for high linking costs, we again observe a common enemy effect. Whereas in the benchmark case without a disruptor, information sharing is not possible, as the only stable network is the empty network, when facing a network disruptor, there are also multiple stable networks in which post-disruption information sharing is possible. The most efficient of these structures is an $(x+1)$-regular circulant network, which has a generalized $(x+1)$ regular structure using the least amount of links, and which is completely robust against disruption so that in the post-disruption network complete information sharing is possible. Thus, also for a general disruption budget of $D = x$, we find that whereas for low linking costs the presence of a network disruptor can be seen as a rather negative occurrence, for high linking costs, we can observe a common enemy effect if players coordinate on a connected network.

7. Conclusion

What we have shown in this paper is that when facing an outside force that aims to destroy the value of the network as a whole, it is possible for a group of self-interested, myopic players to build a network that is stable and actually more efficient than they would be able to do without this outside threat. We have thus given a purely economic explanation for the psychological concept of a common enemy effect. However, we have also shown that the effect of a common enemy can also be negative, especially for low linking costs. This is because in the presence of a common enemy, players can lock each other into not forming any network, or into forming a network with multiple components.

This paper suggests that an effect rooted in psychology can also be explained by purely economic means. Future research should consequently aim at exploring the economic incentives for the common enemy effect more thoroughly in this setting by adding heterogeneous players, information asymmetry or looking at 'strong'-stability, where groups of more than two players can deviate at the same time (see Jackson and Van den Nouweland (2005)). Additionally, in the pairwise stability model, the focus lies purely on which networks are stable but not on how players can actually reach such networks. Future research should therefore focus on implementing actual mechanisms on how such a network can be formed (e.g., in an economic laboratory experiment) and whether myopic players can reach all equilibrium networks. Another avenue of research focusses on the farsighted behavior of players. Theoretical work on networks assuming farsightedness of players (see, e.g., the work by Morbitzer et al. (2011), Morbitzer et al. (2012) or Herings et al. (2009)) shows that a concept of (perfect) farsighted stability can be established. Network experiments (see, e.g., the work by Mantovani et al. (2011)) on the same topic show that players do tend to behave more farsightedly than myopic in such experiments. Future research may analyze how the assumption of farsightedness would change the results we have found here.

Bibliography

Albert, R., Jeong, H., and Barabasi, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382. M3: 10.1038/35019019; 10.1038/35019019.

Boesch, F. and Tindell, R. (1984). Circulants and their connectivities. *Journal of Graph Theory*, 8(4):487–499.

Bollobás, B. and Riordan, O. (2003). Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1(1):1–35.

Bornstein, G., Gneezy, U., and Nagel, R. (2002). The effect of intergroup competition on group coordination: An experimental study. *Games and Economic Behavior*, 41(1):1–25.

Carroll, M. S., Cohn, P. J., Seesholtz, D. N., and Higgins, L. L. (2005). Fire as a galvanizing and fragmenting influence on communities: the case of the rodeo–chediski fire. *Society and Natural Resources*, 18(4):301–320.

Chang, P. Y. (2008). Unintended consequences of repression: alliance formation in South Korea's democracy movement (1970–1979). *Social Forces*, 87(2):651–677.

Coser, L. (1956). *The Functions of Social Conflict*. Clencoe, IL: Free Press.

De Jaegher, K. and Hoyer, B. (published online, 2014). Collective action and the common enemy effect. *Defence and Peace Economics*.

Dziubiński, M. and Goyal, S. (2013). Network design and defence. *Games and Economic Behavior*, 79:30 – 43.

Goyal, S. and Vigier, A. (2014). Attack, defence, and contagion in networks. *The Review of Economic Studies*, 81(4):1518–1542.

Heider, F. (1982). *The psychology of interpersonal relations*. Lawrence Erlbaum.

Herings, P., Mauleon, A., and Vannetelbosch, V. (2009). Farsightedly stable networks. *Games and Economic Behavior*, 67(2):526–541.

Hong, S. (2009). Enhancing transportation security against terrorist attacks. *Working Paper, Vanderbilt University*.

Hoyer, B. and De Jaegher, K. (forthcoming, 2016). Strategic network disruption and defense. *Journal of Public Economic Theory*.

Hugh-Jones, D. and Zultan, R. (2013). Reputation and cooperation in defense. *Journal of Conflict Resolution*, 57(2):327–355.

Jackson, M. O. and Van den Nouweland, A. (2005). Strongly stable networks. *Games and Economic Behavior*, 51(2):420–444.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.

Koopmans, R. (1997). Dynamics of repression and mobilization: The german extreme right in the 1990s. *Mobilization: An International Quarterly*, 2(2):149–164.

Kovenock, D. and Roberson, B. (2012). Coalitional colonel blotto games with application to the economics of alliances. *Journal of Public Economic Theory*, 14(4):653–676.

Landwehr, J. (2015). Network design and imperfect defense. *Center for Mathematical Economics Working Paper*, (537).

Lipsey, R. A. (2006). Network warfare operations: Unleashing the potential. *Mimeo Center for Strategy and Technology, Air War College, Air University*.

Mantovani, M., Kirchsteiger, G., Mauleon, A., and Vannetelbosch, V. (2011). Myopic or farsighted? an experiment on network formation. *FEEM Working Papers 45*, (45).

McLauchlin, T. and Pearlman, W. (2012). Out-group conflict, in-group unity? exploring the effect of repression on intramovement cooperation. *Journal of Conflict Resolution*, 56(1):41–66.

Mesterton-Gibbons, M. and Dugatkin, L. A. (1992). Cooperation among unrelated individuals: Evolutionary factors. *The Quarterly Review of Biology*, 67(3):pp. 267–281.

Morbitzer, D., Buskens, V., and Rosenkranz, S. (2011). Network formation with limited forward-looking actors. *Working Paper*.

Morbitzer, D., Buskens, V., and Rosenkranz, S. (2012). Strategic formation of networks to obtain information when actors are limitedly farsighted. *Working Paper*.

Muller, E. N. and Opp, K.-D. (1986). Rational choice and rebellious collective action. *American Political Science Review*, 80(02):471–487.

Münster, J. and Staal, K. (2011). War with outsiders makes peace inside. *Conflict Management and Peace Science*, 28(2):91–110.

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of Operations Research*, 2(3):pp. 225–229.

Pierskalla, J. H. (2009). Protest, deterrence, and escalation: The strategic calculus of government repression. *Journal of Conflict Resolution*.

Riechmann, T. and Weimann, J. (2008). Competition as a coordination device: Experimental evidence from a minimum effort coordination game. *European Journal of Political Economy*, 24(2):437–454.

Siegel, D. A. (2011). When does repression work? collective action in social networks. *The Journal of Politics*, 73(04):993–1010.

Simmel, G. (1908). *Conflict*. Glencoe, IL: Free Press.

Stein, A. A. (1976). Conflict and cohesion a review of the literature. *Journal of Conflict Resolution*, 20(1):143–172.

Taylor, M., Sekhar, S., and D'Este, G. (2006). Application of accessibility based methods for vulnerability analysis of strategic road networks. *Networks and Spatial Economics*, 6(3):267–291. M3: 10.1007/s11067-006-9284-9.

Appendix

Additions to the proof of Lemma 6 on stochastic-pd networks consisting of multiple components. Here we first show the proof on networks consisting of multiple minimally connected stochastic-pd components. We will then show that networks consisting of components with a critical link structure are not pairwise stable unless they contain end-players and then finally show the conditions under which stochastic- and non-stochastic-pd components with a critical link structure are pairwise stable.

- Assume that the network is stochastic-pd, the components themselves are stochastic and minimally connected but not stars.[31] The payoff to node i in component C_i, if it is not a central node in the component, but part of one of the t (where $t > 1$) components that can be disconnected from the component if it is targeted, is given by $u_i(g) = \frac{1}{m}(\frac{1}{t} * |y| + (1 - \frac{1}{t})(\frac{n}{m} - |y|)) + (1 - \frac{1}{m})\frac{n}{m} - vc$, where t denotes the number of target links in the component and $|y|$ denotes the number of nodes that can be disconnected. Since the component is minimally connected, stochastic-pd, but not the star, we know that $|y| > 1$. Again determining the nodes that have the most incentives to add or delete links, we find that the incentive to add a link is highest for any node in a targeted component i, and the incentive to delete a link is highest for any node directly linked to an end-player k. This is caused by the fact that due to the disruptor's lexicographic preferences, he will not target this component any more if it contains one node less than the other components. If node i adds a link to node j in another component C_j the payoff to node i would be $u_i(g + g_{ij}) = \frac{n}{m} - (v + 1)c$. Deleting the link to an end-player k leads to $u_i(g - g_{ik}) = \frac{n}{m} - 1 - (v - 1)c$.[32] Comparing these payoffs with $u_i(g)$, we find that the network is pairwise stable only if $\frac{2|y| - |y|t - \frac{n}{m} + mt}{mt} > c > \frac{\frac{n}{m} - 2|y| + |y|t}{mt}$ holds. This may be fulfilled if $2|y|m - t|y|m + \frac{m^2 t}{2} > n$ holds. This can not be fulfilled for $m = 2$ as even for the lowest value of t, $t = 2$, the first two terms of the equation add up to 0 and for increasing t they get negative and $2t > n$ can never be fulfilled. However, for sufficiently large m this condition may hold.

- Assume that the network is stochastic-pd and that the components themselves are not minimally connected but still have a critical link structure. We can directly dismiss all such networks that do not have end-players, as here there is always an incentive for players to disconnect non-critical links that will not influence the disruptor's strategy. What we thus still need to analyze are networks that are not minimally connected but have a critical path structure with end-players. In such networks, it is straightforward to see that if the non-minimal part of the component is of a lower order than what could be disconnected, players in the non-minimal part always have an incentive to delete a non-critical link, as again it does not change the disruptor's strategy. The remaining cases are summarized below:

  - In a non-stochastic component with a critical path structure and end-players, where the component that may be disconnected is of order $|y|$ and it holds that $\frac{n}{m} - |y| > |y|$, the players with the most incentive to change links are any end-player k and any node i connected to end-player k.[33] The payoff to node k can be calculated as $u_k(g) = \frac{1}{m}|y| + (1 - \frac{1}{m})\frac{n}{m} - c$. If he adds a link to node j in C_j, the payoff would be given by $u_k(g + g_{kj}) = \frac{n}{m} - 2c$. The payoff to node i depends on the order of y.[34] If $|y| = 1$, the payoff to node i is given by $u_i(g) = \frac{1}{m}(\frac{n}{m} - 1) + (1 - \frac{1}{m})\frac{n}{m} - vc$. If $|y| > 1$, the payoff to node i is given by $u_i(g) = \frac{1}{m}|y| + (1 - \frac{1}{m})\frac{n}{m} - vc$. Deleting a link to

[31] The proof for the case where the components are stars is in the general part of the paper. The logic here is completely the same. The only difference is that the player who is linked to an end-player, and will therefore benefit most from deleting a link, is not the central player, as now more than one node can be disconnected.

[32] In any minimally connected stochastic-pd component that is not the star it needs to hold that $|y| \geq 2$ and, thus, the player directly linked to an end-player must be part of y.

[33] For $\frac{n}{m} - |y| \leq |y|$, there is always an incentive for players in the non-minimal part of the component to delete a non-critical link as it will not change the disruptor's strategy.

[34] If $|y| = 1$, node i is part of the central component. If $|y| > 1$, node i is part of the component that may be disconnected. Thus, the payoffs to i differ. This is parallel to the distinction between a star and other minimally connected networks in the previous analysis.

node k will in both cases lead to a payoff of $u_i(g - g_{ik}) = \frac{n}{m} - 1 - (v-1)c$. Comparing all these payoffs we find that the network is pairwise stable only if $1 - \frac{1}{m} > c > \frac{n}{m^2} - \frac{1}{m}$ for $|y| = 1$ and $1 + \frac{|y|}{m} - \frac{n}{m^2} > c > \frac{n}{m^2} - \frac{|y|}{m}$ for $|y| > 1$. This may be fulfilled if $m^2 > n$ and $\frac{m^2}{2} + |y|m > n$, respectively, which can both be fulfilled if $m \geq 3$.

– In a stochastic component with a critical path structure and end-players, where the component that may be disconnected is of order $|y|$ and it holds that $\frac{n}{m} - |y| > |y|$,[35] the players with the most incentive to change links are again any end-player k and any player connected to such end-players, i. The payoff to node k can then be given as $u_k(g) = \frac{1}{m}(\frac{1}{t}|y| + (1-\frac{1}{t})(\frac{n}{m} - 1)) + (1 - \frac{1}{m})\frac{n}{m} - c$, where t denotes the number of possible target links in C_i. The payoff to node i depends on the order of y. If $|y| = 1$, the payoff to node i is given by $u_i(g) = \frac{1}{m}(\frac{n}{m} - 1) + (1 - \frac{1}{m})\frac{n}{m} - vc$. If $|y| > 1$, the payoff to node i is given by $u_i(g) = \frac{1}{m}(\frac{1}{t}|y| + (1-\frac{1}{t})(\frac{n}{m} - |y|)) + (1-\frac{1}{m})\frac{n}{m} - vc$, where t denotes the number of possible target links in the component. Deleting a link to node k will in both cases lead to a payoff of $u_i(g - g_{ik}) = \frac{n}{m} - 1 - (v-1)c$. Comparing all these payoffs, we find that the network is pairwise stable only if $1 - \frac{1}{m} > c > \frac{\frac{n}{m} - 1 + t}{tm}$ for $y = 1$ and if $\frac{2|y| + tm - t|y| - \frac{n}{m}}{tm} > c > \frac{\frac{n}{m} - 2|y| + t|y|}{tm}$ for $y > 1$ holds. This is possible only if $m(mt - 2t + 1) > n$ and $2|y|m - t|y|m + \frac{m^2 t}{2} > n$ respectively holds. This can be fulfilled only if $m > 3$.

Circulants

Due to their very structured form, circulant graphs are also readily analyzed for other cases. Thus, we will use them here to show that just as for the case of $D = 1$, where the star network was pairwise stable for certain cost ranges, for a general disruption budget we can also find pairwise stable, connected pre-disruption networks that are not completely safe against disruption. In other words, we can find pairwise stable connected pre-disruption networks for which the post-disruption networks consist of multiple components. To show this, we analyze the case of x-regular circulant networks. We are able to do so due to the fact that for every circulant network we exactly know the structure of the post-disruption network. For the circle, which is the only 2-regular circulant (and in fact even the only connected 2-regular network), the post-disruption network for the case of $D = 2$ consists of two components of order $\frac{n}{2}$.[36] For any x-regular circulant network with $x \geq 3$, the network after disruption by a network disruptor with a disruption budget of $D = x$ consists of two components, where one is of order $(n-1)$ and the other one is of order 1. Comparing the incentives, players then have to add or delete links in the network. We find that such networks are indeed pairwise stable for certain cost ranges.

Lemma A.1. *For $D = x$, it holds that x-regular circulant networks, with $x \geq 2$, are pairwise stable if $\frac{n}{6} > c > 1$ for $x = 2$ and $n > 6$ and $\frac{(n-2)^2}{2n} > c > \frac{n-2}{n}$ for $x \geq 3$ and $n \geq 5$.*

Proof We first show under which conditions x-regular circulants are pairwise stable for $x = 2$ and then continue to discuss the conditions under which they are pairwise stable for $x \geq 3$.

- The only x-regular network for $x = 2$ is the circle network. Consider any node i in a circle network. When facing a network disruptor with a disruption budget of $D = 2$, its payoff is given by $u_i(g) = \frac{n}{2} - 2c$. Should node i delete a link to its neighbor node k, its payoff would be given by $u_i(g - g_{ik}) = \frac{n}{3} - c$. Should node i add a link to node j in such a way that it spans half of the network (thus a jump of $\frac{n}{2}$), its payoff would be given by $u_i(g + g_{ij}) = \frac{n}{2} + 1 - 3c$, as the disruptor can then disrupt only two of i or j's links other than g_{ij}. Comparing the payoffs we find that for pairwise stability it needs to hold that $\frac{n}{6} > c > 1$. For $n > 6$, we can thus find cost ranges for which the network is pairwise stable.

[35] Again if this does not hold, there is always an incentive for players in the non-minimal part to delete a non-critical link.
[36] This of course is abstracting from any issues with divisibility of n.

For $x \geq 3$ there are multiple different ways to build a circulant network for each x. However, the structures always have in common that only one node can be disconnected when facing a network disruptor with $D = x$. Thus the payoff to any node i in such a network is given by $u_i(g) = \frac{1}{n} * 1 + (1 - \frac{1}{n})(n-1) - xc$. If one player decides to delete one link, at least one node can be disconnected from the network and the chance is always $\frac{1}{2}$ to be in the larger pd-component and $\frac{1}{2}$ to be in the smaller pd-component. Consider any node i in a x-regular circulant network, facing a network disruptor with a disruption budget of $D = x$. Should node i add a link to node j it has previously not been linked to, it is automatically safe in the connected component and its payoff is given by $u_i(g + g_{ij}) = n - 1 - (x+1)c$. Should node i delete a link to its neighbor node k, the disruptor can either disconnect node i or node k plus possibly some neighbors, such that node i's payoff would be $u_i(g - g_{ik}) = \frac{1}{2} * y + \frac{1}{2}(n-y) - (x-1)c$. Comparing the payoffs to $u_i(g)$, we find that the network is pairwise stable if it holds that $\frac{(n-2)^2}{2n} > c > \frac{n-2}{n}$, which is fulfilled if $n > 4$. Thus, for every $n \geq 5$ this condition is fulfilled and the network is pairwise stable for the determined cost range.