UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge-INET Institute

Cambridge-INET Working Paper Series No: 2014/01

Cambridge Working Paper in Economics: 1450

# HOW TO DEFEND A NETWORK?

Marcin Dziubiński
(Warsaw University)

Sanjeev Goyal
(University of Cambridge)

## ABSTRACT

Modern economies rely heavily on their infrastructure networks. These networks face threats ranging from natural disasters to human attacks. As networks are pervasive, the investments needed to protect them are very large; this motivates the study of targeted defence. What are the 'key' nodes to defend to maximize functionality of the network? What are the incentives of individual nodes to protect themselves in a networked environment and how do these incentives correspond to collective welfare?

We provide a characterization of equilibrium attack and defence in terms of two classical concepts in graph theory – separators and transversals. We study the welfare costs of decentralized defence.

We apply our results to the defence of the US Airport Network and the London Underground.

# How to Defend a Network

Marcin Dziubiński[*]
Sanjeev Goyal[†]

July 26, 2015

**Abstract**

Modern economies rely heavily on their infrastructure networks. These networks face threats ranging from natural disasters to human attacks. As networks are pervasive, the investments needed to protect them are very large; this motivates the study of targeted defence. What are the 'key' nodes to defend to maximize functionality of the network? What are the incentives of individual nodes to protect themselves in a networked environment and how do these incentives correspond to collective welfare?

We first provide a characterization of optimal attack and defence in terms of two classical concepts in graph theory – *separators* and *transversals*. This characterization permits a systematic study of the *intensity of conflict* (the resources spent on attack and defence) and helps us identify a new class of networks, *windmill graphs*, that minimize conflict.

We then study security choices by individual nodes. Our analysis identifies the externalites and shows that the welfare costs of decentralized defence in networks can be very large.

[*] Institute of Informatics, Faculty of Mathematics, Informatics and Mechanics, Warsaw University. Email: `m.dziubinski@mimuw.edu.pl`

[†] Faculty of Economics & Christ's College, University of Cambridge. Email: `sg472@cam.ac.uk`
We thank Diego Cerdeiro, Syngjoo Choi, Amrita Dhillon, Matt Elliott, Julien Gagnon, Timo Hiller, Prem Chand Jain, Peter Kondor, Michael König, Botond Kőszegi, Francesco Nava, Anja Prummer, Pauline Rutsaert, Ádám Szeidl, and seminar participants at Bristol, Cambridge, CEU (Budapest), Microsoft Research (New England), UCLA, and WEHIA 2012 for comments. Both authors thank the European Research Area Complexity-Net for financial support, Marcin Dziubinski acknowledges support from the Foundation for Polish Science and Sanjeev Goyal acknowledges support from a Keynes Fellowship and the Cambridge-INET Institute.

# 1 Introduction

> *"Our nation's critical infrastructure is crucial to the functioning of the American economy... (It) is increasingly connected and interdependent and protecting it and enhancing its resilience is an economic and national security imperative"* Department of Homeland Security (2012).

Infrastructure networks – highways, aviation, shipping, pipelines, train systems, and posts – are a vital part of the modern economy. These networks face a variety of threats ranging from natural disasters to human attacks. The latter may take a violent form (guerrilla attacks, attacks by an enemy country, and terrorism) or a non-violent form (as in political protest that blocks transport services).[1] A network can be made robust to such threats through additional investments in equipment and in personnel. As networks are pervasive, the investments needed could be very large; this motivates the study of targeted defence. What are the 'key' parts of the network that should be protected to ensure maximal functionality? As defence is often a choice made by individual actors, we also wish to understand the relation between network structure and decentralized incentives. This paper develops a model to study these questions.

Consider a given infrastructure network consisting of nodes and links. The designer chooses to protect 'nodes' of the network against damage/attacks; protecting a node is costly. Protection include investments in security personnel, in training, in equipment and in cybersecurity. These protection measures typically take time to implement and so we focus on *ex-ante* investments in protection. We suppose that a defended node is immune to attack whereas an undefended node is eliminated by attack (along with all its links). The initial network, the defence and the attack together yield a set of surviving nodes and links – the residual network. The defender chooses a defence strategy that maximizes the value of the residual network net of the costs of defence.

Our model covers two scenarios. The first is that of an intelligent adversary who seeks to damage components and disrupt the flows in the network. The second is that of a natural threat: facing such a threat the defender focuses on the worst case scenario. In both cases the defender looks for the 'maximin' solution. For

[1] For an introduction to network based conflict, see Arquilla and Ronfeldt (2001) and Zhu and Levinson (2011); for news coverage of the effects of natural disasters and human attacks on infrastructure networks, see Eun (2010), Kliesen (1995), India Today (2011) and Luft (2005).

expositional simplicity, we use the language of an intelligent adversary throughout. We study a game between a defender and an adversary and analyze the sub-game perfect equilibrium of this game.

We consider network payoff functions in which the value to the designer of a network is component additive, and the payoff from each component is increasing and convex in the size of the component.[2] The convexity of value in component size is key to the appeal of connectivity in networks.

We begin with a study of optimal defence. Proposition 2, characterizes optimal defence and attack. Optimal attack targets two types of nodes: those that fragment the network into distinct components (the *separators*), and those that simply reduce the size of components (the *reducing attacks*). Anticipating this attack, optimal defence targets nodes that block the separators and reducing attacks. A set of nodes that block a collection of separators is referred to as a *transversal*. We prove that optimal defence either targets a minimal transversal or protects all nodes. Figure 3 illustrates these concepts.[3]

This characterization result allows us to study the relation between networks and conflict more closely. We find that the size of defence and attack are both non-monotonic in the cost of attack; even more surprisingly, the size of defence and the payoff of the defender may fall with the addition of links in the network (Proposition 3).

We then turn to the *intensity of conflict*: this is the sum of expenditures of defence and attack. For given conflict technology, we define minimal intensity of conflict and then describe networks that sustain it (Proposition 4). We then demonstrate that network architecture can create very large variations in the intensity of conflict. A feature of minimal conflict is that there is a single active player. We next discuss circumstances where both players devote resources to conflict in equilibrium.

An important insight of the analysis is the idea of *strategic exposure:* the defender may find it optimal to leave unprotected a 'key' node (the elimination of which dis-

[2] This specification is consistent with Metcalfe's Law (network value is proportional to the square of number of nodes) and Reed's Law (network value is exponentially increasing in number of nodes). It is also in line with the large theoretical literature on network externalities (Katz and Shapiro (1985), Farrell and Saloner (1986)) and network economics (Bala and Goyal (2000), Jackson and Wolinsky (1996)). One way of defining network value is the number of connected node pairs. This is a special case of our value function.

[3] Appendix C provides a detailed application of the concepts to well known families of networks (trees, core-periphery, interlinked stars).

connects the network) and instead protects a alternative larger set of nodes. We refer to this as the 'queen sacrifice'. This leads us to identify a class of networks – *windmill graphs* – that minimize conflict and are also attractive for the defender. Figure 6 below presents these networks.

In many situations, security decisions are made at the local level, e.g., individual airports choose their own security checks. This motivates the study of decentralized security.[4] Individual nodes care about surviving an attack and about being part of large connected components. Observe, that to block a 'separator' it is sufficient for one node in the separator to protect itself. So, in the game among the nodes, defence choices within a separator are strategic substitutes. But for the network to remain connected all separators must be blocked. So a node will protect itself only if other separators are being blocked: thus defence choices also exhibit strategic complementarity. We show that decentralized security choices is characterized by separators and transversals (Proposition 5). We establish that a combination of incentive and coordination issues can lead to very large costs of decentralization.

Our paper contributes to the economic study of networks. The research on networks has been concerned with the formation, structure and functioning of social and economic networks (Goyal (2007), Jackson (2008) and Vega-Redondo (2007)). The problem of 'key players' has traditionally been studied in terms of Bonacich centrality, betweenness, eigenvector, and degree centrality, see e.g., Bala and Goyal (2000), Ballester et al. (2006), Choi et al. (2013), DeMarzo et al. (2003), Elliot and Golub (2013), Galeotti et al. (2010), Golub and Jackson (2010). Our paper suggests that for the problem of attack and defence the 'key players' are nodes that lie in separators and transversals. These nodes are typically distinct from nodes that maximize familiar notions of centrality. Appendix B discusses this distinction in detail. Thus the principal contribution of our paper is to introduce two classical concepts from graph theory into economics and show how they address a problem of practical importance.

Individual defence is a public good, and so this conceptual contribution is also relevant for the study of games on network more generally. Bramoullé and Kranton (2007) draw attention to maximal independent sets. By contrast, our work brings out the role of minimal transversal of the separators. This set is generally different from maximal independent sets.[5]

[4] For an early contribution on inter-dependent security, see Kunreuther and Heal (2004).

[5] For example, in core-periphery network, core nodes are the minimal separators, while the maximal independent set can include at most one core node and must include peripheral nodes.

Our paper also contributes to the literature on network defence, see e.g., Bier et al. (2006), Baccara and Bar-Isaac (2008), Acemoglu et al. (2013),Dziubiński and Goyal (2013), Goyal and Vigier (2014) Clark and Konrad (2007) and Kovenock and Roberson (2012). To the best of our knowledge, our results on the role of separators and transversals in network conflict are novel, relative to the existing body of work. In particular, we note that the earlier work by Dziubiński and Goyal (2013) and Goyal and Vigier (2014) focuses on optimal design and defence. In these papers, the optimal network takes on a very simple form – it is a star – and so the optimal defence takes on a correspondingly simple structure – protect the central hub node. By contrast, in the present paper the network is exogenous and arbitrary: this is a much broader problem and requires new conceptual tools.

In passing it must be noted that the problem of network defence has traditionally been studied in operations research, electrical engineering and computer science; see e.g., Alpcan and Başar (2011), Aspnes et al. (2006), Smith (2008) and Grötschel et al. (1995). In an early paper, Cunningham (1985) looks at the problem of network design and defence with conflict on links. Relative to this literature, the novelty of our paper lies in the study of intensity of conflict and the externalities that arise in decentralized defence.

The rest of the paper is organized as follows. Section 2 presents the model of defence and attack. Section 3 introduces the main concepts and provides a characterization of equilibrium defence and attack. It also contains the study of comparative statics, active conflict and conflict intensity. Section 4 takes up the case of decentralized defence. Section 5 concludes. All proofs are presented in the Appendix A.

# 2 The model

We start with a given network. We consider a two-player sequential move game with a defender and an adversary. In the first stage, the defender chooses an allocation of defence resources. In the second stage, given a defended network, the adversary chooses the nodes to attack. Successfully attacked nodes (and their links) are removed from the network, yielding a residual network. The goal of the defender is to maximize the value of the residual network, while the goal of the adversary is to minimize this value.[6]

[6] The sequential move game formulation appears to be appropriate for the large scale and time consuming protection investments discussed in the introduction. This two stage model with observ-

Let $N = \{1, \ldots, n\}$, with $n \geq 3$, be a finite set of *nodes*. A *link* is a two element subset of $N$. The set of all possible links over $P \subseteq N$ is $g^P = \{ij : i, j \in P, i \neq j\}$ (where $ij$ is an abbreviation for $\{i, j\}$). A *network* is a set of links. Given the set of nodes $P \subseteq N$, $\mathcal{G}(P) = 2^{g^P}$ is the set of all networks over $P$. The set $\mathcal{G} = \bigcup_{P \subseteq N} \mathcal{G}(P)$ is the set of all networks that can be formed over any subset of nodes from $N$. Every network $g \in \mathcal{G}$ has a *value* $\varPhi(g)$, associated with it: $\varPhi : \mathcal{G} \to \mathbb{R}$ is called a *value function*.

The set of nodes $X \subseteq N$ chosen by the adversary is called an *attack*. The set $X = \varnothing$ is called the *empty attack*. A *defence* is a set of nodes $\varDelta \subseteq N$; node $i \in N$ is defended under $\varDelta$ if and only if $i \in \varDelta$. We assume that the defence is perfect: a protected node cannot be removed by an attack, while any attacked unprotected node is removed with certainty. Given a defence $\varDelta$ and an attack $X$, a set $Y = X \setminus \varDelta$ will be removed from the network. Removing a set of nodes $Y \subseteq N$ from a network creates a *residual network* $g - Y = \{ij \in g : i, j \in N \setminus Y\}$.

Defence resources are costly: the cost of defending a node is $c_{\mathrm{D}} > 0$. Given network $g$, defender's payoff from strategy $\varDelta \subseteq N$, when faced with adversary's strategy $X \subseteq N$, is

$$\Pi^{\mathrm{D}}(\varDelta, X; g, c_{\mathrm{D}}) = \varPhi(g - (X \setminus \varDelta)) - c_{\mathrm{D}}|\varDelta|. \tag{1}$$

Attack resources are costly: the cost of attacking a node is given by $c_{\mathrm{A}} > 0$. Given defended network $(g, \varDelta)$, payoff to the adversary from strategy $X \subseteq N$ is

$$\Pi^{\mathrm{A}}(\varDelta, X; g, c_{\mathrm{A}}) = -\varPhi(g - (X \setminus \varDelta)) - c_{\mathrm{A}}|X|. \tag{2}$$

We study the (sub-game perfect) equilibria of this game.

**Remarks on model:** We have assumed sequential moves; this is mainly for exposition. It is possible to show that our main results on characterization of conflict in terms of certain properties of the graph carries over with simultaneous moves. Perfect defence is a more substantial assumption. Smoother models of conflict such as the Tullock contest function would lead to modifications in parts of the main characterization results below. Appendix D discusses these points in greater detail.

A *component* is a minimal and non-empty set of nodes $C \subseteq N$ such that any two distinct nodes $i, j \in C$ are connected. Two nodes $i$ and $j$ are connected in network $g$ if

ability of first stage actions is consistent with the approach in the large literature on security and networks, see e.g., Tambe (2011) and Alpcan and Başar (2011).

there is a sequence of nodes $i_0, \ldots, i_m$ such that $i = i_0$, $j = i_m$ and for all $0 < k \leq m$, $i_{k-1} i_k \in g$. The set of components of $g$ is denoted by $\mathcal{C}(g)$.

We assume that $\Phi$ is component additive. Given a network $g$,

$$\Phi(g) = \sum_{C \in \mathcal{C}(g)} f(|C|), \tag{3}$$

where $f$ satisfies the following assumption:

**Assumption 1.** $f : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing, strictly convex, and $f(0) = 0$.

Since the game is finite and sequential, standard result guarantees existence of (subgame perfect) equilibria. These equilibria are usually not unique, but generically, equilibrium outcomes are equivalent with respect to player's payoffs, sizes of defence and attack, and the value of residual network. This is the content of the following result.

**Proposition 1.** *For any network $g$ and costs $c_{\mathrm{D}}$ and $c_{\mathrm{A}}$, there exists a sub-game perfect equilibrium. For generic values of $c_{\mathrm{A}}$ and $c_{\mathrm{D}}$ and generic $f$ the equilibrium attack and defence size and the payoffs of the players are unique.*

# 3 The analysis

This section develops our main results for the two person game between the defender and the adversary. Optimal attacks focus on nodes that fragment the network (the *separators*), while optimal defence targets a set of nodes that block these separators, (the *transversal*). The interest then moves on to the relation between network architecture and the intensity of conflict (the sum of resources allocated to attack and defence) and the prospects of active conflict (when the adversary eliminates some nodes while the defender protects others).

We begin with a study of a simple example that helps illustrate a number of interesting phenomena.

**Example 1.** *Defence and Attack on the Star*

Consider the star network with $n = 4$ and $\{a\}$ as central node (as in Figure 1). The value function is $f(x) = x^2$.

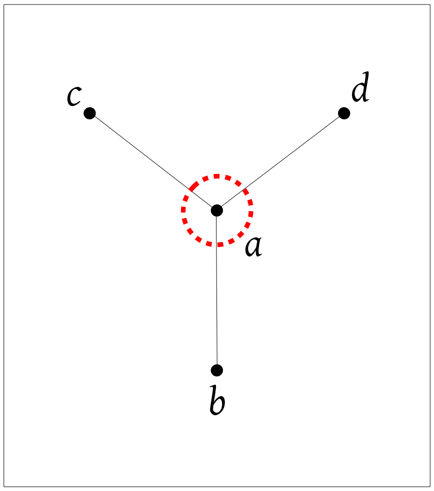

Figure 1: Star network ($n = 4$)

As is standard, we solve the game by working backward. For every defended network $(g, \mathit{\Delta})$ we characterize the optimal response of the adversary. We then compare the payoffs to the defender from different profiles, $(g, \mathit{\Delta})$, and compute the optimal defence strategy. Equilibrium outcomes are summarized in Figure 2.

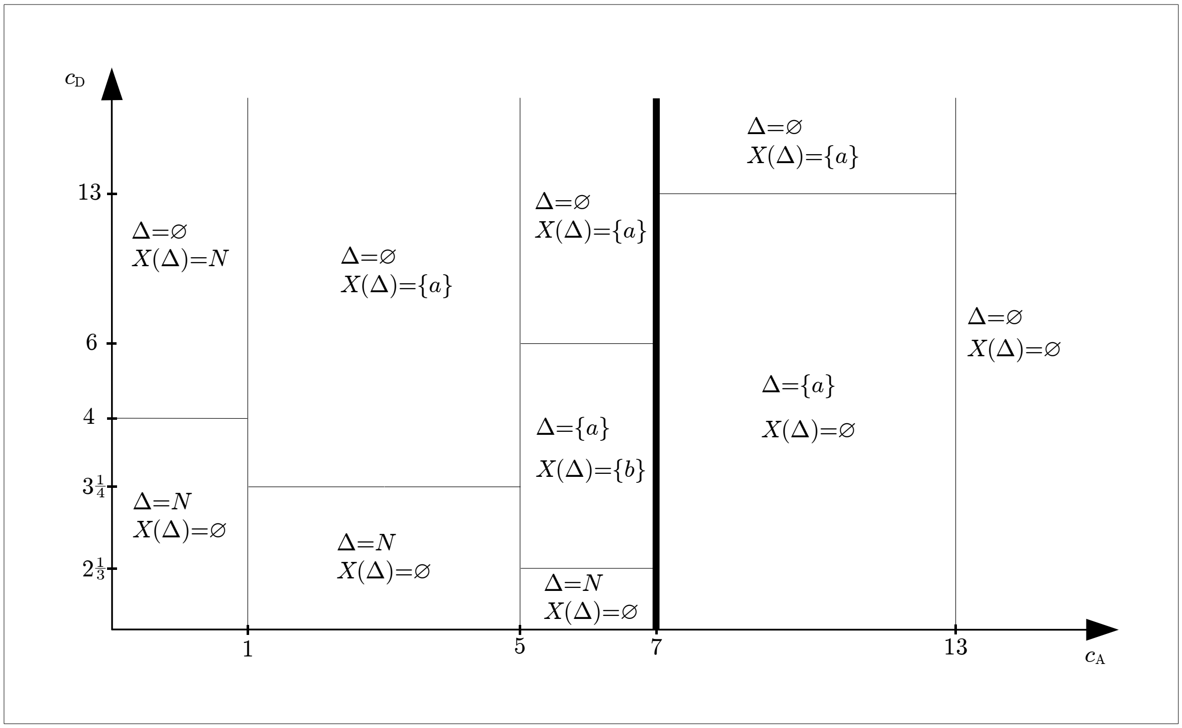

Figure 2: Equilibrium outcomes: star network ($n = 4$) and $f(x) = x^2$.

A number of points are worth noting.

1. Observe that removing node $a$ disconnects the network; this node is a *separator*. Moreover, there is a threshold level of cost of attack (7) such that the adversary either attacks $a$ or does not attack at all when $c_{\mathrm{A}} > 7$. Protecting this node is also central to network defence.

2. The intensity of conflict exhibits rich patterns: when cost of attack is very large there is no threat to the network and no need for defence. If the cost of attack is small, intensity of conflict hinges on the level of defence costs. When they are low all nodes are protected and there is no attack (the costs of conflict are $nc_{\mathrm{D}}$), if they are high then there is no defence but all nodes are eliminated (the costs of conflict are $nc_{\mathrm{A}}$). For intermediate cost of attack and defence, both defence and attack are seen in equilibrium.

3. The size of the defence may be non-monotonic in the cost of attack. Fix the cost of defence at $c_{\mathrm{D}} = 3.5$. At a low cost of attack ($c_{\mathrm{A}} < 1$) the defender protects all nodes, in the range $c_{\mathrm{A}} \in (1, 5)$ he protects 0 nodes, in the range $c_{\mathrm{A}} \in (5, 13)$ he protects $\{a\}$, and then in the range $c_{\mathrm{A}} > 13$, he stops all protection activity. Similarly, the size of the attack strategy may be non-monotonic in the cost of attack.

■

The starting point of the general analysis is the nature of optimal attack. Given the convexity in the value function of networks, disconnecting a network is especially damaging. A set $X \subseteq N$ is a *separator* if $|\mathcal{C}(g)| < |\mathcal{C}(g - X)|$. In other words, a separator is a set of nodes removing which strictly increases the number of components in the network. A network will normally possess multiple separators and the adversary should target the most effective ones. A separator $S \subseteq N$ is *essential* for network $g \in \mathcal{G}(N)$, if for every separator $S' \subsetneq S$, $|\mathcal{C}(g - S)| > |\mathcal{C}(g - S')|$. The set of all essential separators of a network $g$ is denoted by $\mathcal{E}(g)$. Figure 3 illustrates essential separators and their transversal in an example. We provide a detailed discussion of essential separators and their transversals in well known families of networks (trees, core-periphery, interlinked stars) in Appendix C.

The second element is the level of costs. As illustrated by Example 1, the network defence problem can be divided into two parts, depending on the cost of attack. Given $x \in \mathbb{N}$, $\Delta f(x) = f(x + 1) - f(x)$ is the marginal gain to a node in the value of a

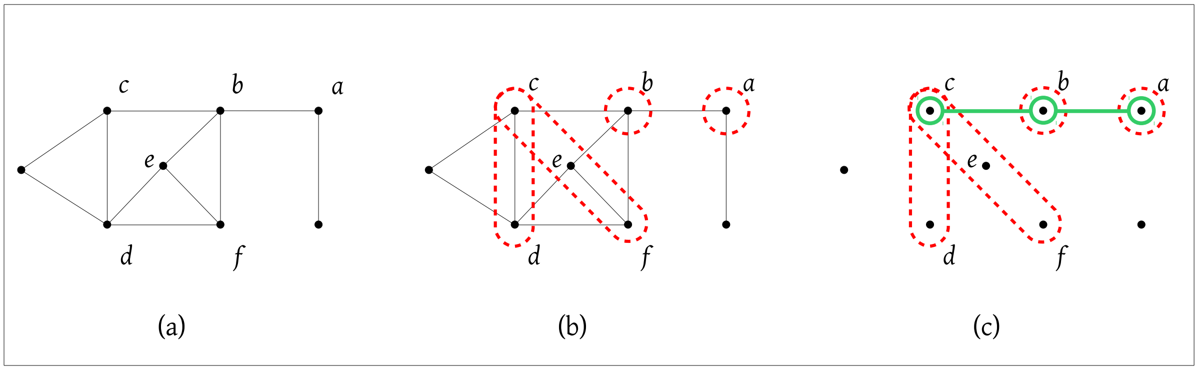

Figure 3: (a) network $g$, (b) essential separators, (c) minimum transversal of essential separators.

component of size $x$. Under Assumption 1, $\Delta f(x)$ is strictly increasing. It is useful to separate two levels of costs: one, high costs with $c_{\mathrm{A}} > \Delta f(n-1)$, and two, low costs with $c_{\mathrm{A}} < \Delta f(n-1)$.

We start with the case of high cost as it brings out some of the main general insights in a straightforward way. Facing a high cost, the adversary must disconnect the network, i.e., choose a separator or not attack the network at all. Clearly, the adversary would never use an essential separator that yields a lower payoff than the empty attack. Given cost of attack $c_{\mathrm{A}}$ and network $g$, the set of individually rational separators is $\mathcal{E}(g, c_{\mathrm{A}}) = \{X \in \mathcal{E}(g) : \Phi(g) - \Phi(g - X) \geq c_{\mathrm{A}}|X|\}$.

When cost of attack is low, it may be profitable for the adversary to use attacks that merely remove nodes from the network, without disconnecting it. A set $R \subseteq N$ is a *reducing attack* for a network $g$, if there is no $X \subseteq R$ such that $X$ is a separator for $g$. The set of all reducing attacks for a given network $g$ is denoted by $\mathcal{R}(g)$.

The following lemma characterizes all the possible attacks of the adversary in terms of essential separators and reducing attacks. In addition, it provides characterization of the attacks that are best responses in the adversary's sub-game.

**Lemma 1.** *Fix a connected network $g$. Let $\Delta \subseteq N$ be a defence selected by the defender in the first stage. Any attack $X \subseteq N$ can be decomposed into two disjoint sets: a set $E$ and a reminder set $R$ such that*

1. *$E$ is either empty or $E \in \mathcal{E}(g)$.*

2. *$R$ is a reducing attack for $g - E$.*

*Moreover, if $X$ is a best response to $\Delta$, then $E$ is either empty or $E \in \mathcal{E}(g, c_{\mathrm{A}})$.*

We now turn to equilibrium strategies of the designer. Again, it is instructive to start with the setting where cost of attack is high. An optimal strategy of the defender should block a subset of individually rational essential separators in the most economical way. Given a family of sets of nodes, $\mathcal{H}$, and a set of nodes $M$, $\mathcal{D}(M,\mathcal{H}) = \{X \in \mathcal{H} : X \cap M \neq \varnothing\}$ are the sets in $\mathcal{H}$ that are blocked (or *covered*) by $M$. Given a family of sets $\mathcal{F} \subseteq \mathcal{H}$, the set $M$ is called a *transversal* of $\mathcal{F}$, if $\mathcal{D}(M,\mathcal{H}) = \mathcal{F}$. The set of all transversals of $\mathcal{F}$ is denoted by $\mathcal{T}(\mathcal{F})$. Elements of $\mathcal{T}(\mathcal{F})$ with the smallest size are called *minimum* transversals of $\mathcal{F}$. Let $\tau(\mathcal{F})$ denote the *transversal number* of $\mathcal{F}$, i.e., the size of a minimum transversal of $\mathcal{F}$. Figure 3 illustrates essential separators and their transversal in a simple network. We provide more examples and discussion of essential separators and their transversals in some well known families of networks (trees, core-periphery, interlinked stars) in Appendix C.

We are now ready to state our first main result on optimal defence and attack.

**Proposition 2.** *Consider a connected network $g \in \mathcal{G}(N)$. Let $(\mathit{\Delta}^*, X^*)$ be an equilibrium.*

1. *If $c_{\mathrm{A}} < \Delta f(n-1)$ then*
   - *$\mathit{\Delta}^* = N$ or $\mathit{\Delta}^*$ is a minimal transversal of $\mathcal{D}(\mathit{\Delta}^*, \mathcal{E}(g, c_{\mathrm{A}}))$.*
   - *$X^*(\mathit{\Delta}) = E \cup R$, where $E \in \mathcal{E}(g, c_{\mathrm{A}})$ and $R \in \mathcal{R}(g - E)$, with $X^*(\mathit{\Delta}) \cap \mathit{\Delta} = \varnothing$.*
2. *If $c_{\mathrm{A}} > \Delta f(n-1)$ then*
   - *$|\mathit{\Delta}^*| \leq \tau(\mathcal{E}(g, c_{\mathrm{A}}))$ and $\mathit{\Delta}^*$ is a minimum transversal of $\mathcal{D}(\mathit{\Delta}^*, \mathcal{E}(g, c_{\mathrm{A}}))$.*
   - *$X^*(\mathit{\Delta}) = \varnothing$, if $\mathit{\Delta} \in \mathcal{T}(\mathcal{E}(g, c_{\mathrm{A}}))$; $X^*(\mathit{\Delta}) \in \mathcal{E}(g, c_{\mathrm{A}})$ with $X^*(\mathit{\Delta}) \cap \mathit{\Delta} = \varnothing$, otherwise.*

Optimal defence is characterized in terms of minimal transversal of the appropriate hypergraph of separators (or defence covers all nodes). If cost of attack is such that elimination of single nodes is not worthwhile, optimal attack is bounded above by the transversal number of the graph. Optimal attack is either empty or targets essential separators. If cost of attack justifies elimination of single nodes optimal defence can range from a minimal transversal to covering all nodes. Optimal attack is constituted of nodes that comprise reducing attacks and essential separators. A general feature of optimal defence is that it may be larger than the smallest possible transversal (even when it does not cover all nodes).

We now briefly describe the arguments underlying the proof. By Lemma 1, we know that any attack may be decomposed into two disjoint parts comprising an essential separator and a reducing attack.

In the range of costs covered by part 2, the adversary will not use reducing attacks. So, an optimal attack must be either empty or it must be an individually rational essential separator. Next consider the optimal defence strategy, $\mathit{\Delta}^*$. Clearly, $\mathit{\Delta}^*$ cannot be larger than the size of the minimum transversal of $\mathcal{E}(g, c_{\mathrm{A}})$, as that would be wasteful for the defender. If $|\mathit{\Delta}^*| = \tau(\mathcal{E}(g, c_{\mathrm{A}}))$, then $\mathit{\Delta}^*$ must be a minimum transversal of $\mathcal{E}(g, c_{\mathrm{A}})$; choosing a defence other than a minimum transversal would simply lower payoffs. If $|\mathit{\Delta}^*| < \tau(\mathcal{E}(g, c_{\mathrm{A}}))$, then $\mathit{\Delta}^*$ is a minimum transversal of $\mathcal{D}(\mathit{\Delta}^*, \mathcal{E}(g, c_{\mathrm{A}}))$.

We turn next to part 1 of Proposition 2. The proof proceeds by showing that a defence exceeding a minimal transversal (of covered essential separators) must include some node that is being protected purely to prevent it from removal. Hence the role of such a defence is to ensure the size of the component. This must mean that, in the absence of defence, the node would be eliminated in the subsequent optimal attack. We then exploit convexity of $f$ and linearity of costs of defence and attack to establish that the adversary must find it optimal to eliminate all other unprotected nodes in the surviving component. Extrapolating from this, we establish that this must apply to all essential separators and by convexity then to single nodes in those components as well. In other words, if the defender finds it optimal to go beyond a minimal transversal of blocked essential separators, then he must protect all nodes.

We now consider the general comparative statics with respect to the costs and the network. It is worth noting some patterns in Example 1 above. Figure 2 suggests that defence size is falling in defence costs and is non-monotonic in attack cost. The attack size is non-monotonic in both attack cost and defence cost. These patterns are true more generally. They have payoff implications. The following result summarizes our analysis.

**Proposition 3.** *The equilibrium comparative statics are as follows.*

1. *The size of defence and defender's payoff are both decreasing in the cost of defence. Defender's payoff increases in the cost of attack. However, depending on the costs and the network, the size of defence may increase or decrease when cost of attack increases.*

2. *Depending on the costs and the network, the size of attack and adversary's payoff may increase or decrease when cost of attack increases. Adversary's payoff increases in the cost of defence. However, depending on the costs and the network, the size of attack may increase or decrease when cost of defence increases.*

3. *Depending on the costs and the network, adding links may increase or decrease the size of optimal defence as well as defender's payoff.*

We note that the effect of defence cost on size of attack may be non-monotonic. This is because with higher cost of defence, the defender may uncover some essential separators which the adversary could switch to. Their size might be smaller or higher than the size of separators chosen by the adversary under lower cost of defence. As an example, consider the network $g$ in Figure 4 and suppose that $f(x) = x^2$, $c_{\mathrm{A}} \in (31, 54)$, and $c_{\mathrm{D}} \in (108, 121)$. Under these parameters in every equilibrium the defender defends node $a$ and the adversary responds with essential separator $\{b, c\}$. When cost of defence rises to 122, equilibrium defence of the defender is $\varnothing$ to which the adversary responds with essential separator $\{a\}$. On the other hand, Example 1 illustrates that size of attack might rise when cost of defence is rising (c.f. the case of $c_{\mathrm{A}} \in (7, 13)$ in Figure 2). Despite this non-monotonic behavior of equilibrium attack size, the payoff to the adversary increases when the cost of defence rises. A similar observation also holds on the effect of attack cost on defence size and on payoffs.

An increase in attack cost has non-monotonic effects on attack size and adversary's payoff. This is illustrated by Example 1, e.g. when cost of defence is in the range $(3.25, 4)$. The reason to these non-monotonicities is as follows. When cost of attack rises, some of the attacks stop being individually rational. This creates an opportunity for the defender to reduce defence, possibly on the expense of some value of the network. This, in turn, allows the adversary to execute attacks which were blocked when cost of attack was lower. In the example, when $c_{\mathrm{A}} \in (0, 1)$, it is individually rational for the adversary to remove any unprotected node. Therefore, with $c_{\mathrm{D}} \in (3.25, 4)$, the defender defends all the nodes. When $c_{\mathrm{A}} \in (1, 5)$, it is not individually rational for the adversary to remove single unprotected nodes. With costs of defence in $(3.25, 4)$, the defender prefers to leave the network undefended and loose the central node, saving on cost of defence and loosing some value of the network. Such an attack is better to the adversary than not removing any node. The size of attack rises from 0 to 1 and the payoff of the adversary rises from $-16$ to $-3 - c_{\mathrm{A}} \in (-9, -4)$. When

$c_{\mathrm{A}} > 7$ then size of attack falls back to 0 and payoff to the adversary falls back to $-16$.

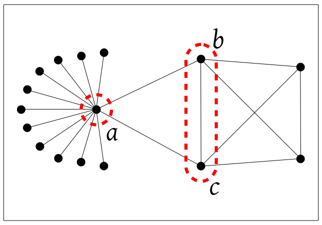

Figure 4: Network where rise in cost of defence reduces size of attack.

Finally, consider the effects of adding links. A first conjecture would be that adding links should always be good for the defender, as it creates more routes for connection and this should make the network easier to defend. The next example shows that this intuition is false: a denser network may induce a bigger optimal defence with lower defender payoffs!

**Example 2.** *Adding links may increase defence size and lower defender payoffs*

We consider network given in Figure 5. Suppose that payoff from component of size $x$ is $f(x) = x^2$.

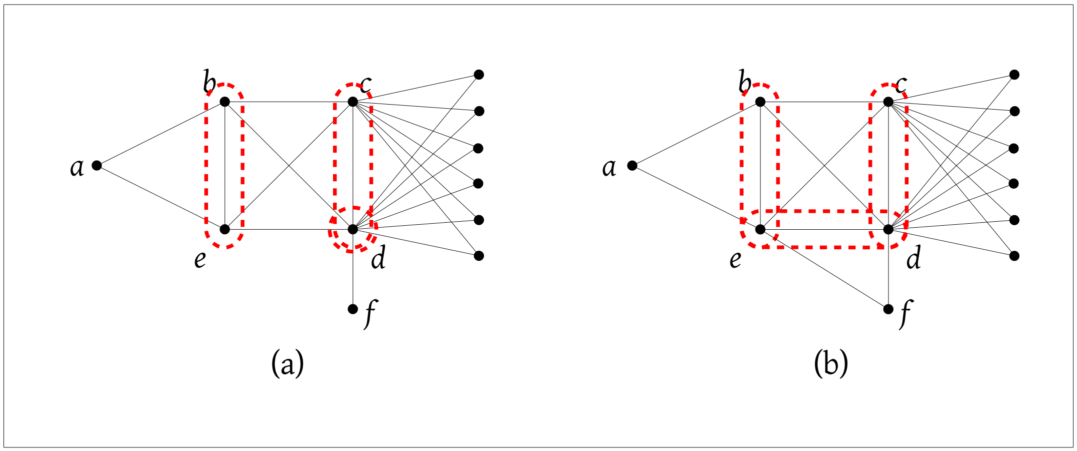

Figure 5: Example 2. (a) Original network. (b) Network with added link.

Assume that the cost of attack, $c_{\mathrm{A}} \in (23, 31)$, and the cost of defence, $c_{\mathrm{D}} \in (43, 85)$. The unique equilibrium outcome is $\Delta^* = \{c\}$, $X^* = \{d\}$. The equilibrium payoff to the defender is $101 - c_{\mathrm{D}}$.

Now consider a network $g' = g \cup \{ef\}$, with a link between the nodes $e$ and $f$ added. With this additional link, the separator $\{d\}$ is replaced by separator $\{d, e\}$.

Suppose that cost of defence is $c_\mathrm{D} \in (43, 62)$. Observe that with defence $\mathit{\Delta}^* = \{c\}$, there exists an attack $\{d, e\}$ that is optimal for the adversary and yields only $82 - c_\mathrm{D}$ to the defender. Thus the addition of a link, and retaining the same defence, may actually lower defender's payoffs.

In the new network, $g'$, the unique equilibrium outcome is $\mathit{\Delta}^* = \{d, e\}$ and $X^* = \varnothing$. The equilibrium payoff to the defender is $144 - 2c_\mathrm{D} < 101 - c_\mathrm{D}$. So, *the optimal defence size increases and the defender's payoff falls as the network becomes denser.*

On the other hand, it is clear that as we keep adding links and arrive at the complete network, the optimal attack is empty (as $c_\mathrm{A} > 23$) and so optimal defence is also the empty set. Defender's payoff is 144, which is the maximal attainable. Thus the effects of adding links are non-monotonic.

■

This non-monotonicity is not an artifact of the specifics of the network and the costs of attack and defence. It reflects a general feature of conflict in networks. To see this consider the case of the complete network. The first thought would be that a network that contains the most connections is the hardest to disrupt and always leads to the best outcomes for the defender. This is not true. The following example clarifies this point.

**Example 3.** *Complete network vs core-periphery network*

Suppose that n is large and that the cost of attack satisfies $f(n-2) - f(n-3) < c_A < f(n-1) - f(n-2)$. With this cost of attack, the adversary removes 2 nodes from complete network over $n$ nodes, 1 node from complete network containing $n-1$ nodes, and does not remove any nodes from complete network containing $n-2$ or less nodes. Finally, suppose that the cost of defense satisfies $(f(n) - f(n-2) - f(1))/n < c_D < (f(n) - f(n-2))/n$. With this cost of defence the defender will protect all the nodes in a complete network with $n$ nodes, because $f(n) - nc_D > f(n-2)$, (and we know that in a complete network the defender either protects all or no nodes, in equilibrium).

Now consider a network with $n-1$ nodes in a clique with one node linked to a single element of the core (let's call it $i$). This is a type of core-periphery network. If such a network is not protected, the adversary will remove node $i$ only, disconnecting the network into a clique of size $n-2$ and a single isolated node. Now, we know that the defender is either inactive, protects $i$, or protects all the nodes in equilibrium. With the above cost of defence the defender is inactive. First note that $f(n) - nc_D <$

$f(n-2)+f(1)$ (so protecting everything is worse than being inactive). It can be checked that protecting $i$ is worse, because in response the adversary would remove 2 nodes from the core of the network.

Thus in the core-periphery network the equilibrium payoff to the defender is $f(n-2)+f(1) > f(n) - nc_D$; so it is better than the complete network.

□

This example illustrates the attractiveness of *queen sacrifice* strategy: it is better to leave $i$ unprotected because there is greater loss in value if it is protected! The idea of 'queen sacrifice' and the sub-optimality of the complete network will resurface in other contexts below.

### 3.1 Networks and conflict

This section examines the relation between the network architecture and the nature of conflict more closely. We define the *intensity of conflict* as the sum of expenditures of defence and attack. Our analysis shows that for given costs of conflict, differences in network structure can lead to very large differences in conflict.

Proposition 1 tells us that the size of equilibrium attack and defence are generically unique. We start by defining the minimum intensity of conflict for given costs of attack and defence. Define minimal cost of conflict for given costs and $f$ as follows:

$$CC(c_{\mathrm{A}}, c_{\mathrm{D}}, f) = \min_{g \in \mathcal{G}(N)} c_{\mathrm{D}}|\Delta^*(g, c_{\mathrm{A}}, c_{\mathrm{D}}, f)| + c_{\mathrm{A}}|X^*(g, c_{\mathrm{A}}, c_{\mathrm{D}}, f)|. \tag{4}$$

Example 1 illustrates some of the forces at work. Observe that when cost of attack is very large, $c_{\mathrm{A}} > 13 = f(n) - (n-1)f(1)$, no attack is profitable, and anticipating this, the defender abstains from defence. The intensity of conflict is 0. This lack of conflict for large costs of attack is independent of the architecture of the network.

Turning to the lower cost of attack, an inspection of Figure 1 in Example 1 tells us that the intensity of conflict also depends on the cost of defence. It will be useful to define a special class of networks, *windmill graphs.* These graphs are denoted by $h_n^m$, where $n \geq 2$ and $m \in \{1, \ldots, n-1\}$. There is one critical node which, when removed, disconnects the network. The remaining nodes are partitioned into cliques of size $m$ and, possibly, one group of smaller size (this implies that there are $\lceil (n-1)/m \rceil$ such groups). Every member of a clique is connected to the critical node. We now define a key cost threshold for defence: this equates the payoff from full defence with the payoff from an unprotected $h_n^m$ network.

$$c(m,n) = \frac{f(n) - \lfloor\frac{n-1}{m}\rfloor f(m) - f((n-1) \bmod m)}{n} \tag{5}$$

Figure 6 illustrates windmill graphs.

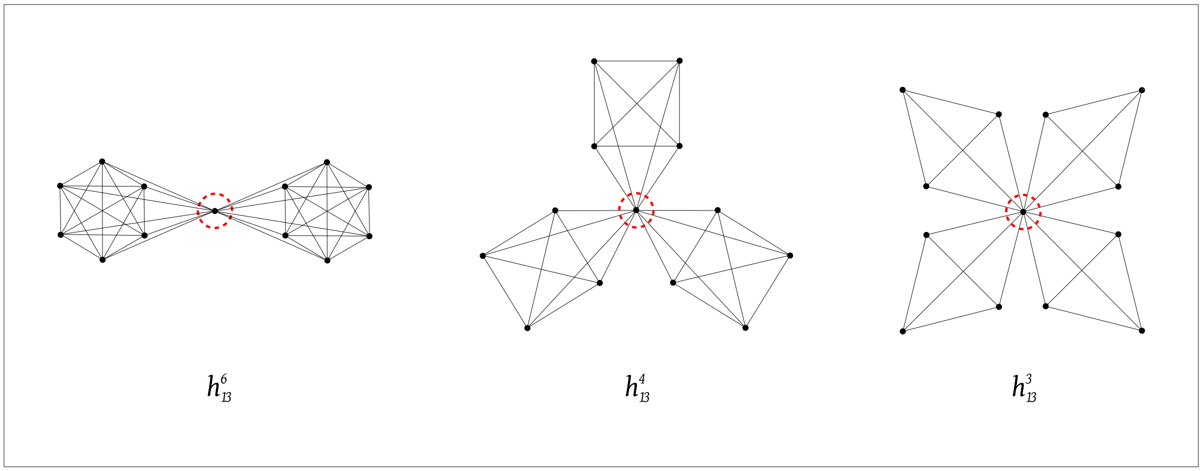

Figure 6: Windmill graphs ($h_n^m$): $n = 13$, $m = 6, 4, 3$.

We are now ready to provide a general characterization of minimal conflict levels.

**Proposition 4.**

1. *If* $c_{\mathrm{A}} > f(n) - (n-1)f(1)$*, then* $CC(c_{\mathrm{A}}, c_{\mathrm{D}}, f) = 0$*. It is attained on any connected network.*

2. *If* $c_{\mathrm{A}} \in (\Delta f(n-1), f(n) - (n-1)f(1))$*, then* $CC(c_{\mathrm{A}}, c_{\mathrm{D}}, f) = 0$*. It is attained on any connected network* $g$ *with* $\mathcal{E}(g, c_{\mathrm{A}}) = 0$*.*

3. *If* $c_{\mathrm{A}} \in (\Delta f(m-1), \Delta f(m))$ *with* $m \in \{1, \ldots, n-1\}$ *and:*

   - $c_{\mathrm{D}} > c(m,n)$*, then* $CC(c_{\mathrm{A}}, c_{\mathrm{D}}, f) = c_{\mathrm{A}}$*. It is attained on a windmill network,* $h_n^m$*.*

   - $c_{\mathrm{D}} < c(m,n)$ *with* $m \in \{1, \ldots, n-1\}$*, then* $CC(c_{\mathrm{A}}, c_{\mathrm{D}}, f) = nc_{\mathrm{D}}$*. It is attained on any connected network.*

In case 2, when cost of attack is high, $c_{\mathrm{A}} > \Delta f(n-1)$, the minimal costs of conflict are 0, as it is not profitable for the adversary to attack any network with $\mathcal{E}(g, c_{\mathrm{A}}) = \varnothing$. Such networks include the complete network, as well as networks which are robust to nodes removal in the sense that they require large number of nodes to be removed to get disconnected. More generally, for any integer $t \geq 1$, a network is *t-connected* if it can be disconnected by removing $t$ nodes and cannot be disconnected by removing less than $t$ nodes. Any $t$-connected network with $t \geq (f(n) - nf(1)/(c_{\mathrm{A}} - f(1))$ has empty $\mathcal{E}(g, c_{\mathrm{A}})$. Menger (1927) provides a characterization of such networks: a network is a least $t$-connected if and only if any two nodes which are not neighbours are connected with at least $t$ node independent paths.[7] Thus such networks have many redundant connections between nodes.

The last case with lower attack costs $c_{\mathrm{A}} < \Delta f(n-1)$, is much richer. Suppose that $c_{\mathrm{A}} \in (\Delta f(m-1), \Delta f(m))$, where $m \in \{1, \ldots, n-1\}$. Now it is profitable to the adversary to attack any undefended node in a component of size greater than $m$. *Hence, the lower bound on costs of conflict is* $\min(c_{\mathrm{A}}, nc_{\mathrm{D}})$. If the cost of defence is sufficiently low, $c_{\mathrm{D}} < c(m, n)$, then complete defence is better than any other defence and the minimal cost of conflict is $nc_{\mathrm{D}}$. If $c_{\mathrm{D}} > c(m, n)$, then complete defence has higher cost as compared to the outcome with no defence and one attacked node. This leads to total cost of conflict of $c_{\mathrm{A}}$. To sustain such an equilibrium, we need a network that has a separator of size 1 and that all components in the residual network have size at most $m$. The windmill graph possesses exactly this characteristic. This motivates the windmill network: for $m \in \{1, \ldots, n-2\}$, the windmill network $h_n^m$ has such an equilibrium and yields the minimal costs of conflict, $c_{\mathrm{A}}$.

We now turn to the role of networks in shaping the intensity of conflict. Proposition 4 tells us that network architecture matters only if costs are in cases 2 or 3.

Consider case 2. Proposition 4 tells us that $CC(c_{\mathrm{A}}, c_{\mathrm{D}}, f) = 0$ in this range. To see the impact of network architecture, consider a star network. If $c_{\mathrm{D}} < f(n)-(n-1)f(1)$, then in equilibrium the defender protects the center of the star and costs of conflict are $c_{\mathrm{D}}$. On the other hand, if $c_{\mathrm{D}} > f(n)-(n-1)f(1)$, then in equilibrium the defender chooses the empty defence, the adversary attacks the center of the star and costs of conflict are $c_{\mathrm{A}}$. So, when the costs of attack and defence reach their upper bound, the difference in costs of conflict between the star network and minimal attainable is $f(n) - (n-1)f(1)$. It is easy to see that this can grow without bound as $n$ gets large.

[7] Two paths are *node independent* if the only nodes they have in common are the starting and the ending node.

Next consider case 3, with $m \in \{1, \ldots, n-2\}$. Proposition 4 tells us that the minimum conflict, attained on network $h_n^m$ (for example) is $c_\mathrm{A}$. Suppose $c_\mathrm{D} \in (c(m,n), (f(n) - f(m))/n)$ and consider a complete network. The unique equilibrium outcome is full protection and so costs of conflict are $nc_\mathrm{D}$. When cost of defence reaches its upper bound and cost of attack reaches its lower bound, the difference in costs between this minimum and the complete network reaches $f(n) + f(m-1) - 2f(m)$, which is maximal, $f(n) - 2f(1)$, for $m = 1$. Again, the network architecture can have very large effects on the intensity of conflict.

*Active conflict:* In Proposition 4, minimal conflict is associated with a single active player. An inspection of Figure 2, in Example 1 above, shows us that both players can be active in equilibrium. This motivates the study of circumstances under which we should expect to see active conflict. Example 1 draws attention the role of costs: neither the attack nor the defence costs can be too high. Here we briefly discuss the role of the network architecture and the network value function.

We start with an observation that draws upon Proposition 2: for active conflict to arise there must exist an individually rational essential separator. If such a separator does not exist, then convexity of function $f$ together with linearity of costs implies that either none or all nodes are defended. In particular, if $g$ is a complete network, then for all costs and all functions $f$ (satisfying our assumptions), there is no equilibrium with active conflict.

Are there any other (connected) networks with the same property as complete networks? If marginal value of $f$ is growing sufficiently fast, then no active conflict is possible. Let $f$ satisfy the following property, for $x \geq 0$:

$$\Delta f(x) > x f(x), \tag{6}$$

where $\Delta f(x) = f(x+1) - f(x)$.

The property is satisfied by functions $f(x) = (x+1)! - 1$ and $(x+1)^x - 1$, for example. Marginal value in these functions grows so rapidly, that adding a single node to a component of size $m$ increases its value more than $m$ times. In effect, the returns from protecting $m < n$ nodes are smaller than average returns from protecting additional $m - n$ nodes. Thus if the defender prefers protecting the first $m$ nodes to no protection, he is even more willing to protect the whole network. Formally, let

$$\varPhi^*(m; g, c_\mathrm{A}) = \max_{\varDelta \subseteq N, |\varDelta| \leq m} \min_{X \in BR(\varDelta; g, c_\mathrm{A})} \varPhi(g - X(\varDelta) \setminus \varDelta), \tag{7}$$

be a function giving maximum value of the residual network that can be attained from network $g$ when up to $m$ units of defence are used and cost of attack is $c_{\mathrm{A}}$ ($BR(\mathit{\Delta}; g, c_{\mathrm{A}})$ denotes the set of best responses of the adversary to $\mathit{\Delta}$, given $g$ and $c_{\mathrm{A}}$). Suppose that there is an equilibrium, $(\mathit{\Delta}^*, X^*)$, featuring active conflict. Let $|\mathit{\Delta}^*| = m$, Since there is active conflict, so $1 \leq m \leq n-1$ and $|X^*(\mathit{\Delta}^*)| \geq 1$. Since $\mathit{\Delta}^*$ is better than $\varnothing$, so $c_{\mathrm{D}} \leq (\mathit{\Phi}^*(m; g, c_{\mathrm{A}}) - \mathit{\Phi}^*(0; g, c_{\mathrm{A}}))/m \leq f(n-1)$. On the other hand, since $\mathit{\Delta}^*$ is better than $N$, so $c_{\mathrm{D}} \geq (f(n) - \mathit{\Phi}^*(m; g, c_{\mathrm{A}}))/(n-m) \geq (f(n) - f(n-1))/(n-1)$. Combining both the inequalities we get $f(n) \leq nf(n-1)$, which contradicts Equation (6).

# 4 Decentralized defence

In many applications, security decisions are made at the individual node level. This section studies decentralized security choices in a network that is under attack. We begin by showing that the equilibrium choices of nodes and the adversary can be characterized in terms of transversals and separators of the underlying network. We then show that the welfare gap between decentralized equilibrium and first best outcomes is unbounded: interestingly, individual choice may lead to too little and to too much protection, relative to the choice of a single (centralized) defender.

We consider a two-stage game. In the first stage, each of the nodes in the network decides whether to protect itself or to stay unprotected. These choices are observed by the adversary who then chooses the nodes to attack.

Let $N = \{1, 2, \ldots, n\}$, with $n \geq 3$ be the set of players and let $S_i = \{0, 1\}$ denote the strategy set of node $i \in N$. Here $s_i = 1$ means that the node chooses to defend itself and $s_i = 0$ refers to the case of no-defence. These choices are made simultaneously. There is a one-to-one correspondence between a strategy profile of the nodes, $\mathbf{s} \in \{0, 1\}^N$, and the resulting set of defended nodes $\mathit{\Delta} \subseteq N$. So we will use $\mathit{\Delta}$ to refer to the strategy profile of the nodes in the first stage.

In the second stage the adversary observes the defended network $(g, \mathit{\Delta})$ and chooses an attack $X \subseteq N$, which leads to a residual network $g - (X \setminus \mathit{\Delta})$. The payoff to the adversary remains as in the case of the centralized defence and is defined in Equation (2). The payoff to a node depends on whether the node is removed by the attack or not. A removed node receives payoff 0. Each of the surviving nodes receives an equal share of the value if its component in the residual network.

$$\Pi^i(\mathit{\Delta}, X; g) = \begin{cases} 0 & \text{if } i \in X \setminus \mathit{\Delta}, \\ \frac{f(|C(i)|)}{|C(i)|} - s_i c_{\mathrm{D}} & \text{otherwise,} \end{cases} \tag{8}$$

where $C(i)$ is the component in the residual network $g - (X \setminus \mathit{\Delta})$ containing $i$.

This completes the description of the *decentralized defence game*. We study the sub-game perfect equilibria of this game, restricting attention to those without active conflict.

Let us solve the game starting from the second stage. As in the two player game, the adversary chooses either the empty attack or an attack being a combination of an essential separator and a reducing attack. If cost of attack is low and there is no active conflict, then either the adversary removes all the nodes, or all nodes are protected. In any other outcome the adversary must remove at least one node. If cost of attack is high and there is no active conflict then either none of the nodes protects or, anticipating the strategy of the adversary, the nodes choose a defence configuration that blocks all the individually rational essential separators. Therefore, in equilibrium, they must choose a minimal transversal of $\mathcal{E}(g, c_{\mathrm{A}})$. We build on these observations to provide the following characterization of equilibria with no active conflict in the decentralized defence game.[8]

**Proposition 5.** *Consider a connected network $g \in \mathcal{G}(N)$. Let $\mathit{\Delta}^*$ be the equilibrium defence.*

1. *If $c_{\mathrm{D}} > \frac{f(n)}{n}$, then $\mathit{\Delta}^* = \varnothing$ is the unique equilibrium defence.*
2. *If $c_{\mathrm{D}} \leq \frac{f(n)}{n}$, and*
   (a) *$c_{\mathrm{A}} < f(n) - f(n-1)$, then $\mathit{\Delta}^* = N$ is an equilibrium defence.*
   (b) *$c_{\mathrm{A}} > f(n) - f(n-1)$, then any minimal transversal of $\mathcal{E}(g, c_{\mathrm{A}})$, is an equilibrium defence.*

*The equilibrium strategy of the adversary is as in Proposition 2.*

[8] We concentrate on equilibria with no active conflict, because, on one hand, it allows for providing a clean characterization and, on the other hand, it provides a sufficiently rich platform for discussing the sources if inefficiencies when defence decisions are decentralized. All other equilibria in decentralized defence game could be characterized in the same spirit as the characterization provided in Proposition 2 for the centralized defence game.

We now turn to discussing inefficiencies that may arise due to decentralized protection, as well as their sources. We compare the aggregate welfare of the nodes in the equilibrium of the two-player game with the aggregate welfare in the decentralized defence game. Aggregate welfare in the 2-player game, starting from network $g$, and costs $c_{\mathrm{A}}$ and $c_{\mathrm{D}}$, is defined as:

$$W^{F}\left(g, c_{\mathrm{A}}, c_{\mathrm{D}}\right)=\Pi^{\mathrm{D}}\left(\mathit{\Delta}^{*}, X^{*} ; g, c_{\mathrm{D}}\right), \tag{9}$$

where $(\mathit{\Delta}^{*}, X^{*})$ is an equilibrium of the two person game. Aggregate welfare in an equilibrium, $s=(\mathit{\Delta}, X)$, of the $n+1$ player game starting from network $g$, and given costs $c_{\mathrm{A}}$ and $c_{\mathrm{D}}$, is defined as:

$$W^{D}\left(s ; g, c_{\mathrm{D}}\right)=\sum_{i \in N} \Pi^{i}\left(\mathit{\Delta}, X ; g, c_{\mathrm{D}}\right). \tag{10}$$

Proposition 2 and Proposition 5 allow us to assess the costs of decentralization. We compare the aggregate welfare of the nodes in the equilibrium of the two-player game with the aggregate welfare in the decentralized defence game. Aggregate welfare in the 2-player game, starting from network $g$, and costs $c_{\mathrm{A}}$ and $c_{\mathrm{D}}$, is defined as:

$$W^{F}\left(g, c_{\mathrm{A}}, c_{\mathrm{D}}\right)=\Pi^{\mathrm{D}}\left(\Delta^{*}, X^{*} ; g, c_{\mathrm{D}}\right), \tag{11}$$

where $(\Delta^{*}, X^{*})$ is an equilibrium of the two person game.

Aggregate welfare in an equilibrium, $s=(\Delta, X)$, of the $n+1$ player game starting from network $g$, and given costs $c_{\mathrm{A}}$ and $c_{\mathrm{D}}$, is defined as:

$$W^{D}\left(s ; g, c_{\mathrm{D}}\right)=\sum_{i \in N} \Pi^{i}\left(\Delta, X ; g, c_{\mathrm{D}}\right). \tag{12}$$

Following Koutsoupias and Papadimitriou (1999), we study the costs in terms of the *price of anarchy*: the ratio of welfare in the two player game to the welfare in the worst equilibrium of the decentralized defence game.

$$\mathrm{PoA}=\max_{g, c_{\mathrm{A}}, c_{\mathrm{D}}} \frac{W^{F}\left(g, c_{\mathrm{A}}, c_{\mathrm{D}}\right)}{\min_{\Delta, X} W^{D}\left(s ; g, c_{\mathrm{A}}, c_{\mathrm{D}}\right)} \tag{13}$$

Our analysis highlights externalities and points to sources of inefficiency in decentralized defence. The first source is the familiar one of positive externalities: an individual's protection decision creates benefits for other nodes, which she does not take into account. Consider a star network and suppose that cost of attack is high, $c_{\mathrm{A}}>f(n)-f(n-1)$, and $c_{\mathrm{D}} \in(f(n)/n, f(n))$. In the equilibrium of the two player game, the aggregate welfare $W^F(g, c_{\mathrm{A}}, c_{\mathrm{D}}) = f(n) - c_{\mathrm{D}}$. However, in the equilibrium of the decentralized game, the central player does not find it profitable to defend itself, as $c_{\mathrm{D}} > f(n)/n$. So aggregate welfare $W^D(s; g, c_{\mathrm{A}}, c_{\mathrm{D}}) = 0$. The ratio of the two is unbounded, for $c_{\mathrm{D}} \in (f(n)/n, f(n))$.

Protection choices exhibit a threshold property: for a node to find it profitable to protect it is necessary that other nodes belonging to the same minimal transversal protect. Thus protection decisions are strategic complements. This can generate coordination failures, resulting in large welfare losses. To see this consider a tree with two hubs each of whom are linked to $(n-2)/2$ distinct nodes. Suppose that $f(n) - f(n-1) < c_{\mathrm{A}} < f(n/2) - (n-2)f(1)/2$, so the adversary will only attack hub nodes. If $2f(n/2)/n < c_D < f(n)/n$ then the first best outcome is to defend the two hubs. One hub protecting itself gives incentives to the other hub to protect: two protected hubs is an equilibrium. However, a hub node does not have unilateral incentives to protect: zero protection it is also an equilibrium. In this equilibrium the aggregate payoffs, $(n-2)f(1)$ as compared to first best outcome of $f(n) - 2c_D$. The cost of decentralization can be unbounded.

Thirdly, at the local level, the game is clearly one of strategic substitutes. A node in a separator has incentives to protect only if no other node in the separator protects itself. Like public good games on networks (c.f. Bramoullé and Kranton (2007)), the network protection game therefore displays multiple equilibria that. This can generate very large efficiency losses. As an example consider network $g$ depicted in Figure 7.

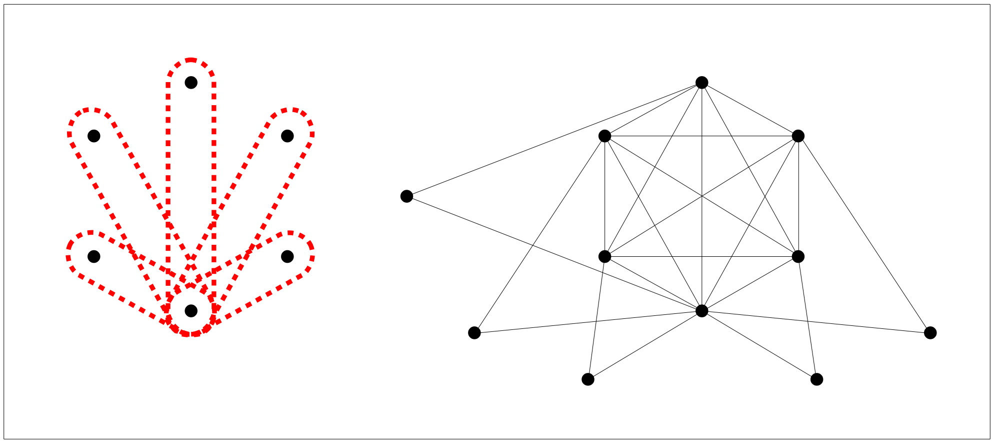

Figure 7: Essential separators with minimal transversals of sizes 1 and 5 ($n = 11$).

Suppose that $f(x) = x^2$, $c_{\mathrm{A}} \in (21, 28)$ and $c_{\mathrm{D}} < 11$. Since cost of attack is high, the adversary will not remove a node without disconnecting the network. The set of individually rational essential separators is a combination of sets depicted in Figure 7. Notice that the minimum transversal of $\mathcal{E}(g, c_{\mathrm{A}})$ is the node belonging to each of the separators, while the largest minimal transversal consists of one distinct node from each of the two element separators. Hence the modified PoA in this case is $|\mathcal{E}(g, c_{\mathrm{A}})|$ and as the example in Figure 7 suggests, it is possible to have a graph $g$ such that $|\mathcal{E}(g, c_{\mathrm{A}})| \geq (n-1)/2$. Again, the cost of decentralization is unbounded.

The idea that personal security exhibits positive externalities is well known in the economic epidemiology literature (and has been noted in the recent research in this area, see e.g., Acemoglu et al. (2013); Cerdeiro et al. (2014); Zawadowski (2013). Moreover, in the standard disease setting security choices are strategic substitutes. Our model departs from this standard setting in two important ways: one, we have an intelligent adversary and two, agents in our model care about the size of the component (and not just about survival). This means that security choices exhibit features both of complements and substitutes. In addition due to the role of size effects, security choices can exhibit large coordination failures. These features of the model distinguish it from the existing literature and call for new methods of analysis and yield fresh insights.

# 5 Concluding Remarks

Infrastructure networks are a key feature of an economy. These networks face a variety of threats ranging from natural disasters to intelligent attacks. This paper develops a strategic model of defence and attack in networks.

We provide a characterization of equilibrium attack and defence in terms of two classical concepts in graph theory – *separators* and *transversals*. We show that the intensity of conflict (the resources spent on attack and defence) and the possibility of active conflict (when both adversary and defender target nodes for action) are both intimately related to the architecture of the network. Finally, we show that the welfare costs of decentralized defence can be very large.

We have assumed that the defender moves first and is followed by the defender and that the defence is perfect: it would be more natural to allow for outcomes of conflict to vary with resources of attack and defence allocated to a node. Appendix D presents a preliminary analysis of models where we relax these assumptions. A general analysis remains an important problem for future research.

Finally, we have assumed that payoff depends only on the size of the networks (or their components). In future work, it would be important to study a model where payoffs depend on the details of the architecture of the components.

# References

D. Acemoglu, A. Malekian, and A. Ozdaglar. Network security and contagion. *MIT Mimeo*, 2013.

T. Alpcan and T. Başar. *Network Security: A Decision and Game Theoretic Approach.* Cambridge University Press, Cambridge, England, 2011.

J. Arquilla and D. Ronfeldt. *Networks and Netwars: The Future of Terror, Crime, and Militancy.* Rand, Santa Monica, CA, 2001.

J. Aspnes, K. Chang, and A. Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *Journal of Computer and System Sciences*, 72(6):1077–1093, 2006.

M. Baccara and H. Bar-Isaac. How to organize crime? *Review of Economic Studies*, 75(4):1039–1067, 2008.

V. Bala and S. Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1230, 2000.

C. Ballester, A. Calvó-Armengol, and Y. Zenou. Who's who in networks. Wanted: The key player. *Econometrica*, 74(5):1403–1417, 09 2006.

V. M. Bier, S. Oliveros, and L. Samuelson. Choosing what to protect: Strategic defensive allocation against an unknown attacker. *Journal of Publc Economic Theory*, 9:1–25, 2006.

Y. Bramoullé and R. Kranton. Public goods in networks. *Journal Economic Theory*, 135(1):478–494, 2007.

D. Cerdeiro, M. Dziubiński, and S. Goyal. Individual security and network design. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pages 205–206, New York, NY, USA, 2014. ACM.

S. Choi, A. Galeotti, and S. Goyal. Trading in networks: Theory and experiment, 2013. Working Paper.

D. Clark and K. A. Konrad. Asymmetric conflict: weakest link against bestshot. *Journal of Conflict Resolution*, 51:457–469, 2007.

W. Cunningham. Optimal attack and reinforcement of a network. *Journal of the ACM*, 32(3):549–61, 1985.

P. DeMarzo, D. Vayanos, and J. Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118(3):909–968, 2003.

Department of Homeland Security. *Office of Infrastructure ProtectionStrategic Plan: 2012-2016*. Washington, DC, 2012.

M. Dziubiński and S. Goyal. Network design and defence. *Games and Economic Behavior*, 79(1):30–43, 2013.

M. Elliot and B. Golub. A network approach to public goods. *Caltech Mimeo*, 2013.

H. Eun. *PhD. Dissertation:Impact analysis of natural disasters on critical infrastructure, associated industries, and communities*. Purdue University, West Lafayette, 2010.

J. Farrell and G. Saloner. Installed base and compatibility: Innovation, product preannouncements, and predation. *American Economic Review*, 76:940–955, 1986.

A. Galeotti, S. Goyal, M. Jackson, F. Vega-Redondo, and L. Yariv. Network games. *Review of Economic Studies*, 77(1):218–244, 2010.

B. Golub and M. O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, 2010.

S. Goyal. *Connections: an introduction to the economics of networks*. Princeton University Press, 2007.

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

M. Grötschel, C. Monma, and M. Stoer. Design of survivable networks. In M. Ball, T. Magnanti, C. Monma, and G. Nemhauser, editors, *Hanbooks of Operations Research and Management Science*, Handbooks in Operations Research and Management Science. North Holland, Amsterdam, 1995.

India Today. Political agitations affect railway service. (March 26), 2011.

M. Jackson. *Social and economic networks.* Princeton University Press, Princeton, New Jersey, 2008.

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

M. Katz and C. Shapiro. Network externalities, competition and compatibility. *American Economic Review*, 75(3):424–440, 1985.

K. Kliesen. The economics of natural disasters. *The Regional Economist*, April, 1995.

E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413, 1999.

D. Kovenock and B. Roberson. Conflicts with multiple battlefields. In M. Garfinkel and S. Skaperdas, editors, *Oxford Handbook of the Economics of Peace and Conflict.* Oxford University Press, Oxford, 2012.

H. Kunreuther and G. Heal. Interdependent security. *The Journal of Risk and Uncertainty*, 26(3):231–249, 2004.

G. Luft. Pipeline sabotage is terrorists weapon of choice. *Energy Security*, March 28, 2005.

K. Menger. Zur allgemainen kurventheorie. *Fundamenta Mathematicae*, 10:96–115, 1927.

C. Smith. Preface to special issue on *Networks: Games, Interdiction, and human interaction problems on networks. Networks*, 52(3):109–110, 2008.

M. Tambe. *Security and Game Theory.* Cambridge University Press, 2011.

G. Tullock. *Efficient Rent Seeking*, pages 97–112. Texas A&M University Press, College Station, TX, 1980.

F. Vega-Redondo. *Complex Social Networks.* Cambridge University Press, Cambridge, England, 2007.

A. Zawadowski. Entangled financial systems. *Review of Financial Studies*, 26(5): 1291–1323, 2013.

S. Zhu and D. Levinson. Disruptions to transportation networks: A review. *Working Paper, University of Minnesota*, 2011.

# Appendix A: Proofs

We start with proving Proposition 1 that states generic equivalence of equilibrium outcomes of the defender-adversary game in terms of payoffs, size of defence and size of attack. We start with the following auxiliary lemmata.

**Lemma 2.** *Let $g$ be a network over set of nodes $N$ and $\mathit{\Delta} \subseteq N$ be a set of defended nodes. Generically, for any best responses $X^*$ and $X^{**}$ to defence $\mathit{\Delta}$, $\mathit{\Phi}(g - X^*) = \mathit{\Phi}(g - X^{**})$ and $|X^*| = |X^{**}|$.*

*Proof.* Let $g$ be a network and $\mathit{\Delta}$ be a defence, as stated in the lemma. Let $X^*$ and $X^{**}$ be best responses to $(g, \mathit{\Delta})$. Then it holds that

$$-\mathit{\Phi}(g - X^*) - |X^*|c_{\mathrm{D}} = -\mathit{\Phi}(g - X^{**}) - |X^{**}|c_{\mathrm{D}}. \tag{14}$$

If $|X^*| = |X^{**}|$, then it follows that $\mathit{\Phi}(g - X^*) = \mathit{\Phi}(g - X^{**})$ and we are done. Otherwise, the equality is equivalent to

$$c_{\mathrm{D}} = \frac{\mathit{\Phi}(g - X^*) - \mathit{\Phi}(g - X^{**})}{|X^{**}| - |X^*|} \tag{15}$$

The set of values on the right hand side of the equality is finite (there are at most $2^{n+1} - 1$ values there). Hence the equality can be satisfied for a finite number of values of $c_{\mathrm{D}} \in \mathbb{R}_{++}$ and it is not satisfied for almost any value of $c_{\mathrm{D}} \in \mathbb{R}_{++}$. This completes the proof. □

**Lemma 3.** *Let $g$ be a network over set of nodes $N$. Generically, for any two equilibria $(\mathit{\Delta}^*, X^*)$ and $(\mathit{\Delta}^{**}, X^{**})$, $\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) = \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**}))$ and $|\mathit{\Delta}^*| = |\mathit{\Delta}^{**}|$.*

*Proof.* Let $g$, $\mathit{\Delta}^*$, $\mathit{\Delta}^{**}$, $X^*$ and $X^{**}$ be as stated in the lemma. Since $\mathit{\Delta}^*$ is a best response to $X^*$ so

$$\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - |\mathit{\Delta}^*|c_{\mathrm{D}} \geq \mathit{\Phi}(g - X^*(\mathit{\Delta}^{**})) - |\mathit{\Delta}^{**}|c_{\mathrm{D}} \tag{16}$$

and since $\mathit{\Delta}^{**}$ is a best response to $X^{**}$ so

$$\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**})) - |\mathit{\Delta}^{**}|c_{\mathrm{D}} \geq \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) - |\mathit{\Delta}^*|c_{\mathrm{D}}. \tag{17}$$

By Lemma 2, generically, $\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) = \mathit{\Phi}(g - X^*(\mathit{\Delta}^*))$ (as both $X^{**}(\mathit{\Delta}^*)$ and $X^*(\mathit{\Delta}^*)$ are best responses to $\mathit{\Delta}^*$. This, together withwith (16) and (17) implies

$$\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**})) - |\mathit{\Delta}^{**}|c_{\mathrm{D}} \geq \mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - |\mathit{\Delta}^*|c_{\mathrm{D}}. \tag{18}$$

Similarly, by Lemma 2, generically, $\Phi(g - X^*(\Delta^*)) = \Phi(g - X^{**}(\Delta^*))$. This together with (16) and (17) implies

$$\Phi(g - X^{**}(\Delta^{**})) - |\Delta^{**}|c_{\mathrm{D}} = \Phi(g - X^*(\Delta^*)) - |\Delta^*|c_{\mathrm{D}}. \tag{19}$$

If $|\Delta^*| = |\Delta^{**}|$, then $\Phi(g - X^*(\Delta^*)) = \Phi(g - X^{**}(\Delta^{**}))$ and we are done. Otherwise, (19) can be rewritten as

$$c_{\mathrm{D}} = \frac{\Phi(g - X^*(\Delta^*)) - \Phi(g - X^{**}(\Delta^{**}))}{|\Delta^*| - |\Delta^{**}|}. \tag{20}$$

Since the number of values on the right hand side is finite, for almost every value of $c_{\mathrm{D}} \in \mathbb{R}_{++}$ this equality is not satisfied. Hence, generically, $|\Delta^*| = |\Delta^{**}|$ and $\Phi(g - X^*(\Delta^*)) = \Phi(g - X^{**}(\Delta^{**}))$. □

**Lemma 4.** *Let $g$ be a network over set of nodes $N$ and $X, Y \subseteq N$ be two attacks such that $|X| \neq |Y|$. Generically, $\Phi(g - X) \neq \Phi(g - Y)$.*

*Proof.* Let $g$, $X$ and $Y$ be as stated in the lemma. Suppose that $\Phi(g-X) = \Phi(g-Y)$. This equality can be rewritten as

$$\sum_{C \in \mathcal{C}(g-X)} f(|C|) = \sum_{C \in \mathcal{C}(g-Y)} f(|C|). \tag{21}$$

Since $X \neq Y$ so there exists $s > 0$ such that $g - X$ has component of size $s$ and $g - Y$ has not or $g - Y$ has a component of such a size and $g - X$ has not. Suppose that $\Phi(g - X) = \Phi(g - Y)$. Hence the equality above reduces to

$$f(s_1) + \ldots + f(s_p) = f(z_1) + \cdots + f(z_q) \tag{22}$$

where $s_1, \ldots, s_p$ and $z_1, \ldots, z_q$ are sizes of components such that $\{s_1, \ldots, s_p\} \cap \{z_1, \ldots, z_q\} = \varnothing$. Equation (22) puts very strict constraints on function $f$ and perturbing $f$ it slightly (within the set of functions satisfying Assumption 1) destroys the equality. Thus $\Phi(g - X) = \Phi(g - Y)$ for $|X| \neq |Y|$ is a non-generic property of $f$. □

With Lemmas 2, 3 and 4 in hand, we are ready to prove Proposition 1.

***Proof of Proposition 1:*** Generic equivalence of defence size and of payoff to the defender follow directly from Lemma 3. Consider equivalence of attack size and of payoff to the adversary. By Lemmata 3 and 4, generically $\Phi(g - X^*(\Delta^*)) = \Phi(g - X^{**}(\Delta^{**}))$ and $|X^*(\Delta^*)| = |X^{**}(\Delta^{**})|$. Thus the points follow as well. □

Proofs of Lemma 1 and Proposition 2 exploit some properties of graphs. The first step is to establish these properties. Lemma 5 characterizes the essential separators as those separators which are 'thin': every node of such separators neighbours at least two components of the residual network. Given a set of nodes $X \subseteq N$ and a network $g$ over $N$, $\partial_g(X) = \{k \in N \setminus X : \text{ there is } j \in X \text{ such that } jk \in g\}$, is the *neighbourhood of $X$ in $g$*. If $X$ is a singleton, that is $X = \{j\}$, then we will write $\partial_g(j)$ instead of $\partial(\{j\})$ ($\partial_g(j)$ is the set of neighbours of $j$ in $g$). We will drop the subscript $g$ in the notation if network $g$ is clear from the context.

**Lemma 5.** *Let $g \in \mathcal{G}(N)$ be a network over a set of nodes $N$. A set $X \subseteq N$ is an essential separator if and only if $X \neq \varnothing$ and for every $i \in X$ there exist two distinct components $C_1, C_2 \in \mathcal{C}(g - X)$, $C_1 \neq C_2$, such that $\partial_{g-X}(i) \cap C_1 \neq \varnothing$ and $\partial_{g-X}(i) \cap C_2 \neq \varnothing$.*

*Proof.* Let $g \in \mathcal{G}(N)$ be a network over a set of nodes $N$, and $X \subseteq N$.

*The necessary part:* Assume that $X$ is an essential separator. Since $X$ is a separator, so $X \neq \varnothing$. Assume, to the contrary, that there exists $i \in X$ such that there is at most one component $C \in \mathcal{C}(g - X)$ such that $\partial_{g-X}(i) \cap C \neq \varnothing$. Suppose first there is no such component. Then the attack $X' = X \setminus \{i\}$ results in the set of components $\mathcal{C}(g - X') = \mathcal{C}(g - X) \cup \{\{i\}\}$, larger than $\mathcal{C}(g - X)$, which contradicts the assumption that $X$ is essential. Secondly, suppose that there is exactly one component $C \in \mathcal{C}(g - X)$ such that $\partial_{g-X}(i) \cap C \neq \varnothing$. Taking attack $X'$, as before, leads to residual network with set of components $\mathcal{C}(g - X') = (\mathcal{C}(g - X) \setminus \{C\}) \cup \{C \cup \{i\}\}$, which has the same cardinality as $\mathcal{C}(g - X)$. Therefore $X$ is not essential, a contradiction.

*Sufficiency part:* Assume that $X \neq \varnothing$, and for every $i \in X$ there exist two distinct components $C_1, C_2 \in \mathcal{C}(g - X)$ such that $C_1 \cap \partial_{g-X}(i) \neq \varnothing$ and $C_2 \cap \partial_{g-X}(i) \neq \varnothing$. Then there exist two nodes, $j_1 \in C_1 \cap \partial_{g-X}(i)$ and $j_2 \in C_2 \cap \partial_{g-X}(i)$, which are connected in $g$ and not connected in $g - X$. Hence $X$ is a separator and we have to show that it is essential. Suppose $X' \subsetneq X$, so there is some $i$ such that $i \in X$ but $i \notin X'$. Given the definition of $i \in X$ it follows that $|C(X')| \leq |C(X)| - 1$. Since $X'$ was arbitrary, the claim is established. □

We now develop a characterization of optimal attack strategies in terms of essential (affordable) separators and reducing attacks.

***Proof of Lemma 1:*** The proof of the first part is by induction on the number of nodes in $X$ that violate the condition from Lemma 5. For the induction basis consider the set of all $X \subseteq N$ for which there are no nodes that violate the condition. Then, by Lemma 5, $X$ is essential and so the reminder is $\varnothing$ and $E = X$ (in particular it may be that $E = X = \varnothing$). The claim holds.

For the induction step, take any $X \subseteq N$ for which there are exactly $m$ nodes that violate the condition from Lemma 5. Suppose that the claim holds for any $Y \subseteq N$ for which there are $l < m$ nodes that violate the condition. Let $i \in X$ be a node that violates the condition and let $Y = X \setminus \{i\}$. Since the condition is violated for $i \in X$, so $g - Y$ contains either one component more than $g - X$ (namely component $\{i\}$), or it has the same number of components with one component $C$ in $g - X$ replaced with $C \cup \{i\}$ in $g - Y$. Hence the condition is violated for $l < m$ nodes from $Y$ in $g - Y$. Thus, by the induction hypothesis, $Y$ can be decomposed into two disjoint sets $E$ and $R$ as claimed. Since, as we argued above, adding $i$ to $Y$ does not increase the number of components in the residual network, so $R \cup \{i\}$ does not contain a separator of $g - E$ and so the decomposition of $X$ into $E$ and $R \cup \{i\}$ satisfies the conditions from the claim. Thus points 1 and 2 are shown.

Now we show that if $g$ is connected and $X$ is a best response to some defence $\Delta \subseteq N$, then either $E = \varnothing$ or $E \in \mathcal{E}(g, c_{\mathrm{A}})$.

We show first, for any attack $X$ and any decomposition of $X$ into two disjoint sets $E$ and $R$ satisfying points 1 and 2, that

$$\Phi(g - E) - \Phi(g - X) \leq \Phi(g) - \Phi(g - R) \tag{23}$$

We use induction on $R$. For the induction basis let $R = \varnothing$. Then Equation (23) trivially holds. For the induction step, suppose that Equation (23) holds for any $T \subsetneq R$. Take any $i \in R$, let $T = R \setminus \{i\}$ and $Y = X \setminus \{i\}$. Let $C \in \mathcal{C}(g - Y)$ be the component with $i \in C$. Since $R$ does not contain an essential separator of $g - X$ so $\mathcal{C}(g - X)$ and $\mathcal{C}(g - Y)$ differ at component $C$ only: either $C \setminus \{i\} \in \mathcal{C}(g - X)$ or $C \setminus \{i\} = \varnothing$. Hence

$$\Phi(g - X) = \Phi(g - Y) - (f(|C|) - f(|C| - 1)). \tag{24}$$

Now let $C' \in \mathcal{C}(g - T)$ be the component with $i \in C'$. Applying attack $\{i\}$ to $g - T$ replaces $C'$ with components $C'_1, \ldots, C'_m$ such that $\bigcup_{i=1}^{m} C'_i = C' \setminus \{i\}$. Hence

$$\Phi(g - R) = \Phi(g - T) - \left( f(|C'|) - \sum_{i=1}^{m} f\left(|C'_i|\right) \right) \leq \Phi(g - T) - (f(|C'|) - f(|C'| - 1)) \tag{25}$$

(by the fact that $f$ is strictly convex). By the induction hypothesis

$$\Phi(g - E) - \Phi(g - Y) + (f(|C|) - f(|C| - 1)) \leq \Phi(g) - \Phi(g - T) + (f(|C|) - f(|C| - 1)) \tag{26}$$

and, by the fact that $C \subseteq C'$ and by convexity of $f$,

$$\varPhi(g-E)-\varPhi(g-Y)-(f(|C|)-f(|C|-1)) \leq \varPhi(g)-\varPhi(g-T)+(f(|C'|)-f(|C'|-1)). \tag{27}$$

Thus, by (24) and (25),

$$\varPhi(g-E)-\varPhi(g-X) \leq \varPhi(g)-\varPhi(g-R). \tag{28}$$

This shows the induction step. Hence we have shown Equation (23).

Now, let $\varDelta \subseteq N$ be a defence chosen in the first stage and suppose that $X$ is a best response to $\varDelta$. $X$ is a better response to $\varDelta$ than $R$, so

$$-\varPhi(g-X)-c_{\mathrm{A}}|X| \geq -\varPhi(g-R)-c_{\mathrm{A}}|R| \tag{29}$$

and, consequently,

$$\varPhi(g-R) \geq \varPhi(g-X)+c_{\mathrm{A}}(|X|-|R|)=\varPhi(g-X)+c_{\mathrm{A}}|E| \tag{30}$$

From 23, we have

$$\varPhi(g-X) \geq \varPhi(g-E)+\varPhi(g-R)-\varPhi(g). \tag{31}$$

Putting the last two inequalities together, we arrive at

$$\varPhi(g-R) \geq \varPhi(g-E)+\varPhi(g-R)-\varPhi(g)+c_{\mathrm{A}}|E|. \tag{32}$$

Simplifying this yields:

$$-\varPhi(g-E)-c_{\mathrm{A}}|E| \geq -\varPhi(g). \tag{33}$$

In other words, $E \in \mathcal{E}(g, c_{\mathrm{A}})$. □

The proof of Part 2 of Proposition 2 now follows from the lemmata above and the arguments in the main text. We turn next to proving Part 1 of Proposition 2.

To simplify some parts of the argument, we will make a tie-breaking assumption on the behavior of the adversary. It says that if two strategies yield equal payoffs to the adversary then he will choose the strategy that yields a lower payoff to the defender.

**Assumption 2.** Given a network $g$ and defence $\mathit{\Delta}$, if two strategies $X \subseteq N$ and $X' \subseteq N$ yield the same payoff to the adversary then he chooses the the strategy that results in residual network of lower value.

The first step here is to state and prove the following lemma.

**Lemma 6.** *Let $g \in \mathcal{G}(N)$ be a connected network over $N$, $c_{\mathrm{D}}$ and $c_{\mathrm{A}}$ be costs of defence and attack, respectively. Suppose that $\mathit{\Delta} \subseteq N$ is an equilibrium defence and $X \subseteq N$ is a best response to it. Suppose that there exists $i \in \mathit{\Delta}$ such that $\mathcal{D}(\mathit{\Delta}, \mathcal{E}(g, c_{\mathrm{A}})) = \mathcal{D}(\mathit{\Delta} \setminus \{i\}, \mathcal{E}(g, c_{\mathrm{A}}))$. Let $X' \subseteq N$ be a best response to $\mathit{\Delta}' = \mathit{\Delta} \setminus \{i\}$.*

*Then there exists a component $C \in \mathcal{C}(g - X)$ such that $C \subseteq \mathit{\Delta}$ and either $C = \{i\}$ or $C \setminus \{i\} \in \mathcal{C}(g - X')$. Moreover*

$$\Pi^{\mathrm{D}}(\mathit{\Delta}, X; g) = \Pi^{\mathrm{D}}(\mathit{\Delta}', X'; g) + f(|C|) - f(|C| - 1) - c_{\mathrm{D}} \tag{34}$$

*and*

$$c_{\mathrm{A}} \leq f(|C|) - f(|C| - 1). \tag{35}$$

*Proof.* Let $\mathit{\Delta} \subseteq N$ be a defence, $i \in \mathit{\Delta}$ and $\mathit{\Delta}' = \mathit{\Delta} \setminus \{i\}$. Let $X$ be a best response to $\mathit{\Delta}$ and $X'$ be a best response to $\mathit{\Delta}'$.

Since $X$ is a best response to $\mathit{\Delta}$ so $X \cap \mathit{\Delta} = \varnothing$ and $\mathit{\Phi}(g - (X \setminus \mathit{\Delta})) = \mathit{\Phi}(g - X)$. Analogously with $X'$ and $\mathit{\Delta}'$. We prove the lemma in seven steps below.

**(i). $\mathit{\Phi}(g - X) > \mathit{\Phi}(g - X')$.**

Since $\mathit{\Delta}$ is an equilibrium strategy of the defender, so $\Pi^{\mathrm{D}}(\mathit{\Delta}, X; g) \geq \Pi^{\mathrm{D}}(\mathit{\Delta}', X'; g)$, that is $\mathit{\Phi}(g - X) - c_{\mathrm{D}}|\mathit{\Delta}| \geq \mathit{\Phi}(g - X') - c_{\mathrm{D}}(|\mathit{\Delta}| - 1)$. Hence $\mathit{\Phi}(g - X) > \mathit{\Phi}(g - X')$.

**(ii). $i \in X'$.**

Assume, to the contrary, that $i \notin X'$. Then $X' \cap \mathit{\Delta} = X' \cap \mathit{\Delta}' = \varnothing$. Similarly, since $X \cap \mathit{\Delta} = \varnothing$, so $X \cap \mathit{\Delta}' = \varnothing$. Hence $\Pi^{\mathrm{A}}(\mathit{\Delta}', X'; g) = \Pi^{\mathrm{A}}(\mathit{\Delta}, X'; g)$ and $\Pi^{\mathrm{A}}(\mathit{\Delta}', X; g) = \Pi^{\mathrm{A}}(\mathit{\Delta}, X; g)$. By the fact that $\Pi^{\mathrm{A}}(\mathit{\Delta}', X'; g) \geq \Pi^{\mathrm{A}}(\mathit{\Delta}', X; g)$, as $X'$ is a best response to $\mathit{\Delta}'$, this yields $\Pi^{\mathrm{A}}(\mathit{\Delta}, X'; g) \geq \Pi^{\mathrm{A}}(\mathit{\Delta}, X; g)$. Additionally, by point (i), $\mathit{\Phi}(g - X) > \mathit{\Phi}(g - X')$, so $X'$ results in residual network of lower value than in the case of $X$. Hence, by the tie-breaking Assumption 2, $X'$ is an equilibrium response to $\mathit{\Delta}$, a contradiction. Thus it must be that $i \in X'$.

**(iii). Let $Y \subseteq N$ be an attack with $i \in Y$ and such that $Y \cap \mathit{\Delta}' = \varnothing$. Then either $\{i\} \cup \partial(i) \subseteq Y$ or there exists exactly one $C \in \mathcal{C}(g - Y)$ such that $i \in \partial(C)$.**

Take any decomposition $E \cup R$ of $Y$, as described in Lemma 1. It cannot be that $i \in E$, as otherwise we would have $E \in \mathcal{D}(\Delta, \mathcal{E}(g, c_{\mathrm{A}}))$, while $E \notin \mathcal{D}(\Delta', \mathcal{E}(g, c_{\mathrm{A}}))$, as $Y \cap \Delta' = \varnothing$, and we would have a contradiction with the assumption that $\mathcal{D}(\Delta, \mathcal{E}(g, c_{\mathrm{A}})) = \mathcal{D}(\Delta', \mathcal{E}(g, c_{\mathrm{A}}))$. Hence $i \in R$ and there exists a component $\widetilde{C} \in \mathcal{C}(g - E)$ such that $i \in \widetilde{C}$. Let $C = \widetilde{C} \setminus R$ be what remains of $\widetilde{C}$ after the remainder $R$ of $Y$ is applied to $g - E$. By the definition of the remainder, $R$ does not contain a separator for $g - E$. Therefore either $C = \varnothing$ (i.e. it is completely removed by $R$) or $C \in \mathcal{C}(g - Y)$, (i.e. it is a component in $g - Y$). Suppose that $C = \varnothing$, that is $C \subseteq R$. Then $\partial_{g-E}(i) \subseteq R$ and $\partial_g(i) \subseteq E \cup R = Y$. Since $i \in Y$, so $\{i\} \cup \partial_g(i) \subseteq Y$. Suppose now that $C$ is a component in $\mathcal{C}(g - Y)$. We will show that $i \in \partial_{g-E}(C)$. For assume the opposite. Then $\partial_{g-E}(C)$ must be a separator in $g - E$, as it separates $C$ from a component containing $i$. But then $\partial_{g-E}(C)$ contains an essential separator for $g - E$. Since $\partial_{g-E}(C) \subseteq R$, so this contradicts the assumption that $R$ is a remainder and does not contain any essential separators of $g - E$. Hence it must be that $i \in \partial_{g-E}(C)$ and, consequently, $i \in \partial_g(C)$.

**(iv). For all $C' \in \mathcal{C}(g - X')$ with $i \in \partial(C')$, $C' \subseteq \Delta$.**

For assume the opposite. Then there exists $C' \in \mathcal{C}(g - X')$ with $i \in \partial(C')$ (and consequently $i \notin C'$) such that $i' \in C' \setminus \Delta$. Consider a strategy $X'' = (X' \setminus \{i\}) \cup \{i'\}$. Since $X' \cap \Delta' = \varnothing$ and $i' \notin \Delta$ so $X'' \cap \Delta' = \varnothing$. Notice that $\Phi(g - X'') \leq \Phi(g - X')$, as both the residual networks agree at all the components apart from what remains of $C' \cup \{i\}$ after $i'$ is removed (at the least it is one component of the same size as $C'$). Since $|X'| = |X''|$ so $\Pi^{\mathrm{A}}(\Delta', X''; g) \geq \Pi^{\mathrm{A}}(\Delta', X'; g)$ and so $X''$ is a best response to $\Delta'$. But then we get a contradiction with point (ii), as $i \notin X''$. Hence it must be that $C' \subseteq \Delta$.

**(v). There exists $C' \in \mathcal{C}(g - X') \cup \{\varnothing\}$ such that $C = C' \cup \{i\} \in \mathcal{C}(g - X)$ and $C \subseteq \Delta$.**

Let $C' = \varnothing$, if $\{i\} \cup \partial(i) \subseteq X'$, or let $C'$ be the unique $C' \in \mathcal{C}(g - X')$ with $i \in \partial(C')$, otherwise. By point (iii) such $C'$ exists. By point (iv) and by the fact that $i \in \Delta$, $C \subseteq \Delta$. Thus there exists a component $C'' \in \mathcal{C}(g - X)$ such that $C \subseteq C''$. Suppose that $C \subsetneq C''$. We will show that in this case $X \cup \{i\}$ is a better response to $\Delta'$ than $X'$, a contradiction.

Notice that since $X \cap \Delta = \varnothing$ and $\Delta' = \Delta \setminus \{i\}$ so $(X \cup \{i\}) \cap \Delta' = \varnothing$. By point (iii) either $\{i\} \cup \partial(i) \subseteq X \cup \{i\}$ or there exists exactly one component $C''' \in \mathcal{C}(g - (X \cup \{i\}))$ such that $i \in \partial(C''')$. Hence $C'' = C''' \cup \{i\}$ and $C''$ must be unique in $\mathcal{C}(g - X)$ with $i \in \partial(C'')$. The residual network $g - (X \cup \{i\})$ differs from $g - X$ at one component only: instead of $C''$ it has $C'' \setminus \{i\}$. Thus the value of residual network $g - (X \cup \{i\})$ is

$$\mathit{\Phi}(g - (X \cup \{i\})) = \mathit{\Phi}(g - X) - (f(|C''|) - f(|C''| - 1)) \tag{36}$$

Similarly, since either $C' = \varnothing$ or $i \in \partial(C')$, so residual network when using $X' \setminus \{i\}$ against $\mathit{\Delta}'$, $g - (X' \setminus \{i\})$, differs from $g - X'$ by one component: it has $C$ instead of $C'$. Additionally, since $\mathit{\Delta} = \mathit{\Delta}' \cup \{i\}$ and $X' \cap \mathit{\Delta}' = \varnothing$ so $X' \setminus \{i\} = X' \setminus \mathit{\Delta}$. Thus the value of residual network $g - (X' \setminus \{i\})$ can be written as

$$\mathit{\Phi}(g - (X' \setminus \{i\})) = \mathit{\Phi}(g - X') + f(|C|) - f(|C| - 1) \tag{37}$$

Since $X$ is a best response to $\mathit{\Delta}$, so it is not worse than $X' \setminus \{i\}$. Hence

$$-\mathit{\Phi}(g - X) - c_{\mathrm{A}}|X| \geq -\mathit{\Phi}(g - (X' \setminus \{i\})) - c_{\mathrm{A}}(|X'| - 1). \tag{38}$$

This, together with Equation (37), implies

$$\mathit{\Phi}(g - X) \leq \mathit{\Phi}(g - X') + f(|C|) - f(|C| - 1) - c_{\mathrm{A}}(|X| - |X'| + 1). \tag{39}$$

Similarly, since $X'$ is a best response to $\mathit{\Delta}'$, so it is not worse than $X \cup \{i\}$. Hence

$$-\mathit{\Phi}(g - X') - c_{\mathrm{A}}|X'| \geq -\mathit{\Phi}(g - (X \cup \{i\})) - c_{\mathrm{A}}(|X| + 1) \tag{40}$$

This, together with Equation (36), implies

$$\mathit{\Phi}(g - X) \geq \mathit{\Phi}(g - X') + (f(|C''|) - f(|C''| - 1)) - c_{\mathrm{A}}(|X| - |X'| + 1) \tag{41}$$

From (39) and (41) we get

$$f(|C''|) - f(|C''| - 1) - (f(|C|) - f(|C| - 1) \leq c_{\mathrm{A}}(|X| + 1) - c_{\mathrm{A}}|X| - (c_{\mathrm{A}}|X'| - c_{\mathrm{A}}(|X'| - 1)) = 0 \tag{42}$$

If $C \subsetneq C''$, then $|C| < |C''|$, and, by strict convexity of $f$, LHS $> 0$, a contradiction. Thus it must be that $C'' = C$.

**(vi).** $\Pi^{\mathrm{D}}(\mathit{\Delta}, X; g) = \Pi^{\mathrm{D}}(\mathit{\Delta}', X'; g) + f(|C|) - f(|C| - 1) - c_{\mathrm{D}}$**.**

Since $X$ is a best response to $\mathit{\Delta}$ so it is not worse than $X' \setminus \{i\}$. Hence

$$-\mathit{\Phi}(g - X) - c_{\mathrm{A}}|X| \geq -\mathit{\Phi}(g - (X' \setminus \{i\})) - c_{\mathrm{A}}(|X'| - 1). \tag{43}$$

Adding $f(|C|) - f(|C|-1)$ to both sides we get

$$\begin{multline} -\left(\mathit{\Phi}(g - X) - (f(|C|) - f(|C|-1))\right) - c_{\mathrm{A}}|X| \geq \\ -\left(\mathit{\Phi}(g - (X' \setminus \{i\})) - (f(|C|) - f(|C|-1))\right) - c_{\mathrm{A}}(|X'|-1). \end{multline} \tag{44}$$

As we observed in proof of point (v) (Equations (36) and (37) and the fact that $C'' = C$)

$$\mathit{\Phi}(g - (X \cup \{i\})) = \mathit{\Phi}(g - X) - (f(|C|) - f(|C|-1)) \tag{45}$$

$$\mathit{\Phi}(g - X') = \mathit{\Phi}(g - (X' \setminus \{i\})) - (f(|C|) - f(|C|-1)) \tag{46}$$

Hence, from (44), we get

$$-\mathit{\Phi}(g - (X \cup \{i\})) - c_{\mathrm{A}}(|X|+1) \geq -\mathit{\Phi}(g - X') - c_{\mathrm{A}}|X'|. \tag{47}$$

On the other hand, since $X'$ is a best response to $\mathit{\Delta}'$ so

$$-\mathit{\Phi}(g - (X \cup \{i\})) - c_{\mathrm{A}}(|X|+1) \leq -\mathit{\Phi}(g - X') - c_{\mathrm{A}}|X'|. \tag{48}$$

Combining these two inequalities we get

$$-\mathit{\Phi}(g - (X \cup \{i\})) - c_{\mathrm{A}}(|X|+1) = -\mathit{\Phi}(g - X') - c_{\mathrm{A}}|X'|. \tag{49}$$

Since $X'$ is the equilibrium response to $\mathit{\Delta}'$ and by tie-breaking Assumption 2,

$$\mathit{\Phi}(g - X') \leq \mathit{\Phi}(g - (X \cup \{i\})). \tag{50}$$

Additionally this, together with Equations (45,46), implies

$$\mathit{\Phi}(g - (X' \setminus \{i\})) \leq \mathit{\Phi}(g - X). \tag{51}$$

From Equation (49) and Equations (45,46) we get

$$-\mathit{\Phi}(g - X) - c_{\mathrm{A}}|X| = -\mathit{\Phi}(g - (X' \setminus \{i\})) - c_{\mathrm{A}}(|X'|-1). \tag{52}$$

Again, since $X$ is the equilibrium response to $\mathit{\Delta}$ and by tie-breaking Assumption 2,

$$\mathit{\Phi}(g - X) \leq \mathit{\Phi}(g - (X' \setminus \{i\})), \tag{53}$$

and, by Equations (45,46) and (51),

$$\mathit{\Phi}(g - X) = \mathit{\Phi}(g - (X' \setminus \{i\})), \tag{54}$$

$$\mathit{\Phi}(g - (X \cup \{i\})) = \mathit{\Phi}(g - X'). \tag{55}$$

Thus both both $X$ and $X' \setminus \{i\}$ are best responses to $\varDelta$ and both $X'$ and $X \cup \{i\}$ are best responses to $\varDelta'$. This, together with Equation (46), implies

$$\Pi^{\mathrm{D}}(\varDelta, X; g) = \Pi^{\mathrm{D}}(\varDelta', X'; g) + f(|C|) - f(|C| - 1) - c_{\mathrm{D}}. \tag{56}$$

**(vii).** $c_{\mathrm{A}} \leq f(|C|) - f(|C| - 1)$**.**

Since $X'$ is a better response to $\varDelta'$ than $X' \setminus \{i\}$ so

$$-\varPhi(g - X') - c_{\mathrm{A}}|X'| \geq -\varPhi(g - (X' \setminus \{i\})) - c_{\mathrm{A}}(|X'| - 1) \tag{57}$$

and, consequently,

$$c_{\mathrm{A}} \leq \varPhi(g - (X' \setminus \{i\})) - \varPhi(g - X'). \tag{58}$$

By Equation (45),

$$c_{\mathrm{A}} \leq f(|C|) - f(|C| - 1). \tag{59}$$

□

***Proof of Part 1 of Proposition 2:*** Characterization of optimal strategies of the adversary follows directly from Lemma 1. Thus in what follows we concentrate on equilibrium defence.

Let $\varDelta$ be an equilibrium defence. We will show first that if $\varDelta \subsetneq N$, then $\varDelta$ must be a minimal transversal of $\mathcal{D}(\varDelta, \mathcal{E}(g, c_{\mathrm{A}}))$.

Assume the opposite. Then there exists $i \in \varDelta$ such that $\mathcal{D}(\varDelta \setminus \{i\}, \mathcal{E}(g, c_{\mathrm{A}})) = \mathcal{D}(\varDelta, \mathcal{E}(g, c_{\mathrm{A}}))$. Let $X$ be the equilibrium response to $\varDelta$ and $X'$ be the equilibrium response to $\varDelta' = \varDelta \setminus \{i\}$. Clearly $X \cap \varDelta = \varnothing$ and $X' \cap \varDelta' = \varnothing$.

Recall that $\mathcal{C}(g - X')$ is the set of components in the residual network when the strategies $\varDelta'$ and $X'$ are used by the players, and $\mathcal{C}(g - X)$ is the set of components in the residual network when $\varDelta$ and $X$ are used. By the assumption that $\varDelta \subsetneq N$, both these sets are non empty. We will show that either $\varDelta'$ or $\varDelta''$ (described below) is a better strategy for the defender than $\varDelta$, which will contradict the assumption that $\varDelta$ is an equilibrium strategy.

Let $C \in \mathcal{C}(g - X)$ be a component such that $C \subseteq \varDelta$ and either $C = \{i\}$ or $C \setminus \{i\} \in \mathcal{C}(g - X')$. By Lemma 6 such $C$ exists.

Since for all $j \in \partial_g(C)$, $\mathcal{D}(\varDelta, \mathcal{E}(g, c_{\mathrm{A}})) \subsetneq \mathcal{D}(\varDelta \cup \{j\}, \mathcal{E}(g, c_{\mathrm{A}}))$, so any such $j$ belongs to an essential separator not covered by $\varDelta$. Take any $j \in \partial_g(C)$ and let $\{C_1, \ldots, C_m\} \subseteq \mathcal{C}(g - X)$ be all the components in $g - X$ such that $j \in \partial_g(C_l)$ for all $l \in \{1, \ldots, m\}$ (assume, without loss of generality, that $C_1 = C$; notice that in particular it may be that $m = 1$ and the argument below works for that case as well). Consider defence $\varDelta' = \varDelta \setminus \{i\}$ and $\varDelta'' = \varDelta \cup \{j\} \cup \bigcup_{l=2}^{m} C_l$. We will show that either $\varDelta'$ or $\varDelta''$ is a better strategy for the defender than $\varDelta$.

Let $X''$ be the equilibrium response of the adversary to $\varDelta''$ and let $C'' = \{j\} \cup \bigcup_{l=1}^{m} C_l$. We show first that $C'' \in \mathcal{C}(g - X'')$. Since $\varDelta''$ protects $C''$, so there is component $C''' \in \mathcal{C}(g - X'')$ such that $C'' \subseteq C'''$. Suppose that $C'' \subsetneq C'''$. Then there exists $v \in C'''$ such that $v \notin C''$. We will show that $v \notin \varDelta''$. If $v \in \partial_g(C_l)$, for some $l \in \{1, \ldots, m\}$, then it cannot be that $v \in \varDelta$ (because these components are separated by $X$ used as an equilibrium response to $\varDelta$). Thus the only possibility is that $v \in \partial_g(\{j\})$. But then $v$ would be one of the components $C_l$ created by applying $X$ to $g$ and, consequently, it would be $v \in C''$, a contradiction with the assumption that $v \notin C''$. Since $v \notin \varDelta$ and $v \notin C''$ so $v \notin \varDelta''$. Now, consider a response $X'' \cup \{v\}$ to $\varDelta''$. At the very least it removes a node from component $C'''$ (it may additionally disconnect the component). Hence $\varPhi(g - (X'' \cup \{v\})) \leq \varPhi(g - X'') - f(|C'''|) + f(|C'''| - 1)$. On the other hand, by Lemma 6, Equation (35), $c_{\mathrm{A}} \leq f(|C|) - f(|C| - 1) < f(|C'''| - f(|C'''| - 1))$ (by convexity of $f$ and $|C'''| \geq |C| + 1$). Thus it follows that

$$-\varPhi(g - (X'' \cup \{v\})) - c_{\mathrm{A}}(|X''| + 1) > -\varPhi(g - X'') - c_{\mathrm{A}}|X''|, \tag{60}$$

which contradicts the assumption that $X''$ is a best response to $\varDelta''$. Therefore it must be $C''' = C''$.

As we have shown above, $C'' = \{j\} \cup \bigcup_{l=1}^{m} C_l \in \mathcal{C}(g - X'')$. After attack $X'' \cup \{j\}$ is applied to $g$, component $C''$ is replaced with components $C_1, \ldots, C_m$. Hence

$$\varPhi(g - X'') = \varPhi(g - (X'' \cup \{j\})) + f\left(1 + \sum_{l=1}^{m} |C_l|\right) - \left(\sum_{l=1}^{m} f\left(|C_l|\right)\right). \tag{61}$$

On the other hand, since $C$ is a component in $g - X$, so every node in $\partial_g(C)$ is removed by $X$. Thus, when nodes in $\varDelta \cup \{j\} \cup \bigcup_{l=2}^{m} C_l$ are defended, the residual network $g - (X \setminus \{j\})$ differs from $g - X$ by having component $C''$ instead of components $C_1, \ldots, C_m$. Hence

$$\varPhi(g - (X \setminus \{j\})) = \varPhi(g - X) + f\left(1 + \sum_{l=1}^{m} |C_l|\right) - \left(\sum_{l=1}^{m} f\left(|C_l|\right)\right). \tag{62}$$

Since $X''$ is a better response to $\varDelta''$ than $X \setminus \{j\}$ so

$$-\varPhi(g - X'') - c_{\mathrm{A}}|X''| \geq -\varPhi(g - (X \setminus \{j\})) - c_{\mathrm{A}}(|X| - 1) \tag{63}$$

and $\varPhi(g-X'') \leq \varPhi(g-(X\setminus\{j\}))$ in the case of equality (notice that $(X\setminus\{j\})\cap\varDelta'' = \varnothing$ as $X \cap C_l = \varnothing$, for all $l \in \{1,\dots,m\}$, and $X \cap \varDelta = \varnothing$).

Equations (61,62) and (63) imply

$$-\varPhi(g-(X''\cup\{j\})) - c_{\mathrm{A}}|X''| \geq -\varPhi(g-X) - c_{\mathrm{A}}(|X|-1). \tag{64}$$

Subtracting $c_{\mathrm{A}}$ from both sides we get

$$-\varPhi(g-(X''\cup\{j\})) - c_{\mathrm{A}}(|X''|+1) \geq -\varPhi(g-X) - c_{\mathrm{A}}|X|. \tag{65}$$

On the other hand, since $X$ is a best response to $\varDelta$ than $X''\cup\{j\}$, we have

$$-\varPhi(g-X) - c_{\mathrm{A}}|X| \geq -\varPhi(g-(X''\cup\{j\})) - c_{\mathrm{A}}(|X''|+1) \tag{66}$$

and $\varPhi(g-X) \leq \varPhi(g-(X''\cup\{j\}))$, in the case of equality.

By (65) and (66), $X''\cup\{j\}$ is a best response to $\varDelta$ as well, and since $X$ is an equilibrium response to $\varDelta$, it must be that $\varPhi(g-X) \leq \varPhi(g-(X''\cup\{j\}))$. Combining this with Equation (61) we get

$$\varPhi(g-X'') \geq \varPhi(g-X) + f\left(1+\sum_{l=1}^{m}|C_l|\right) - \left(\sum_{l=1}^{m} f\left(|C_l|\right)\right). \tag{67}$$

From (34) and (67) it follows that

$$\begin{aligned} \Pi^{\mathrm{D}}(\varDelta,X;g) &= \Pi^{\mathrm{D}}(\varDelta',X';g) + f(|C|) - f(|C|-1) - c_{\mathrm{D}} \qquad (68)\\ \Pi^{\mathrm{D}}(\varDelta'',X'';g) &\geq \Pi^{\mathrm{D}}(\varDelta,X;g) + f\left(1+\sum_{l=1}^{m}|C_l|\right) - \left(\sum_{l=1}^{m} f\left(|C_l|\right)\right) - c_{\mathrm{D}}. \end{aligned}$$

Since $\varDelta$ is a better strategy than $\varDelta'$, so $f(|C|) - f(|C|-1) \geq c_{\mathrm{D}}$. On the other hand, since $\varDelta$ is a better strategy than $\varDelta''$ so $c_{\mathrm{D}} \geq f(1+\sum_{l=1}^{m}|C_l|) - (\sum_{l=1}^{m} f(|C_l|))$. Hence $f(|C|)-f(|C|-1) \geq f(1+\sum_{l=1}^{m}|C_l|) - (\sum_{l=1}^{m} f(|C_l|))$, which contradicts strict convexity of $f$.

Thus we have shown that $\varDelta \subsetneq N$, then $\varDelta$ must be a minimal transversal of $\mathcal{D}(\varDelta,\mathcal{E}(g,c_{\mathrm{A}}))$. □

***Proof of Proposition 3:*** The non-monotonicities have been established in the text. Here we establish monotonicity of defender's payoff in cost of attack and monotonicity of adversary's payoff in cost of defence.

We start with monotonicity of payoff to the defender in cost of attack. The argument here is straightforward in the generic case, where equilibrium payoffs are unique: suppose $(\mathit{\Delta}^*, X^*)$ is an equilibrium with network $g$ and costs $(c_{\mathrm{A}}, c_{\mathrm{D}})$. Let $c'_{\mathrm{A}} > c_{\mathrm{A}}$. If defender retains defence strategy $\mathit{\Delta}^*$, it must be the case that attack strategy will be weakly smaller under high cost $c'_{\mathrm{A}}$. This in turn implies that defender's payoff must be weakly larger if he maintains original strategy $\mathit{\Delta}^*$. So, in equilibrium under $(c'_{\mathrm{A}}.c_{\mathrm{D}})$, he must also do better. However, the monotonicity holds for any values of the parameters. The problem here is that non-uniqueness of equilibrium payoffs. However, this is not a concern, because if this was the case, the more costly attacks would cease being equally good for the adversary as the less costly ones. The precise argument is as follows. Let $c_{\mathrm{A}}$ and $c'_{\mathrm{A}}$ be costs of attack such that $c'_{\mathrm{A}} > c_{\mathrm{A}}$. Let $(\mathit{\Delta}^*, X^*)$ be an equilibrium under $c_{\mathrm{A}}$ and let $(\mathit{\Delta}^{**}, X^{**})$ be an equilibrium under $c'_{\mathrm{A}}$. Since $X^*(\mathit{\Delta}^*)$ is a best response to $\mathit{\Delta}^*$ under $c_{\mathrm{A}}$, so it is not worse than $X^{**}(\mathit{\Delta}^*)$, hence

$$-\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - c_{\mathrm{A}}|X^*(\mathit{\Delta}^*)| \geq -\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) - c_{\mathrm{A}}|X^{**}(\mathit{\Delta}^*)| \tag{69}$$

which yields

$$\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) \leq c_{\mathrm{A}}\left(|X^{**}(\mathit{\Delta}^*)| - |X^*(\mathit{\Delta}^*)|\right). \tag{70}$$

Similarly, since $X^{**}(\mathit{\Delta}^*)$ is a best response to $\mathit{\Delta}^*$ under $c'_{\mathrm{A}}$, so it is not worse than $X^*(\mathit{\Delta}^*)$. This yields

$$\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) \geq c'_{\mathrm{A}}\left(|X^{**}(\mathit{\Delta}^*)| - |X^*(\mathit{\Delta}^*)|\right). \tag{71}$$

Equations (70) and (71) imply $c_{\mathrm{A}}(|X^{**}(\mathit{\Delta}^*)| - |X^*(\mathit{\Delta}^*)|) \geq c'_{\mathrm{A}}(|X^{**}(\mathit{\Delta}^*)| - |X^*(\mathit{\Delta}^*)|)$. By $c'_{\mathrm{A}} > c_{\mathrm{A}}$ it follows that

$$|X^{**}(\mathit{\Delta}^*)| \leq |X^*(\mathit{\Delta}^*)|. \tag{72}$$

Now, assume to the contrary, that

$$\Pi^{\mathrm{D}}\left(\mathit{\Delta}^*, X^*(\mathit{\Delta}^*); g, c_{\mathrm{D}}\right) > \Pi^{\mathrm{D}}\left(\mathit{\Delta}^{**}, X^{**}(\mathit{\Delta}^{**}); g, c_{\mathrm{D}}\right). \tag{73}$$

Since $\mathit{\Delta}^{**}$ is an equilibrium defence under $c'_{\mathrm{A}}$ so

$$\Pi^{\mathrm{D}}\left(\mathit{\Delta}^{**}, X^{**}(\mathit{\Delta}^{**}); g, c_{\mathrm{D}}\right) \geq \Pi^{\mathrm{D}}\left(\mathit{\Delta}^*, X^{**}(\mathit{\Delta}^*); g, c_{\mathrm{D}}\right). \tag{74}$$

The two equations above imply

$$\Pi^{\mathrm{D}}\left(\mathit{\Delta}^*, X^*(\mathit{\Delta}^*); g, c_{\mathrm{D}}\right) > \Pi^{\mathrm{D}}\left(\mathit{\Delta}^*, X^{**}(\mathit{\Delta}^*); g, c_{\mathrm{D}}\right), \tag{75}$$

that is

$$\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - c_{\mathrm{D}}|\mathit{\Delta}^*| > \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) - c_{\mathrm{D}}|\mathit{\Delta}^*| \tag{76}$$

and, consequently,

$$\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) > 0. \tag{77}$$

Equations (70) and (77) imply $c_{\mathrm{A}}(|X^{**}(\mathit{\Delta}^*)| - |X^*(\mathit{\Delta}^*)|) > 0$. By $c_{\mathrm{A}} > 0$, it follows that $|X^{**}(\mathit{\Delta}^*)| > |X^*(\mathit{\Delta}^*)|$, a contradiction with Equation (72). Thus it must be that $\Pi^{\mathrm{D}}\left(\mathit{\Delta}^*, X^*(\mathit{\Delta}^*); g, c_{\mathrm{D}}\right) \leq \Pi^{\mathrm{D}}\left(\mathit{\Delta}^{**}, X^{**}(\mathit{\Delta}^{**}); g, c_{\mathrm{D}}\right)$ Notice that this argument holds for any parameters of the model, not only in the generic case.

We now turn to the monotonicity of payoff to the adversary in cost of defence. Let $c_{\mathrm{D}}$ and $c'_{\mathrm{D}}$ be costs of defence such that $c'_{\mathrm{D}} > c_{\mathrm{D}}$. Let $(\mathit{\Delta}^*, X^*)$ be an equilibrium under $c_{\mathrm{D}}$ and let $(\mathit{\Delta}^{**}, X^{**})$ be an equilibrium under $c'_{\mathrm{D}}$. Since $X^*(\mathit{\Delta}^*)$ is a best response to $\mathit{\Delta}^*$ and $X^{**}(\mathit{\Delta}^*)$ is a best response to $\mathit{\Delta}^*$ in adversary's subgame, so

$$-\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - |X^*(\mathit{\Delta}^*)|c_{\mathrm{A}} = -\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) - |X^{**}(\mathit{\Delta}^*)|c_{\mathrm{A}}. \tag{78}$$

Thus another equilibrium under $c_{\mathrm{D}}$ is $(\mathit{\Delta}^*, X')$ where $X'$ equals to $X^*$ at all defence profiles but $\mathit{\Delta}^*$, where it is equal to $\mathit{\Delta}^{**}$. By Lemma 3, generically, $\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) = \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*))$. By analogous arguments, $\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**})) = \mathit{\Phi}(g - X^*(\mathit{\Delta}^{**}))$. Since $\mathit{\Delta}^*$ is an equilibrium defence under $c_{\mathrm{D}}$ and $\mathit{\Delta}^{**}$ is an equilibrium defence under $c'_{\mathrm{D}}$, so

$$\begin{align} \mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) - |\mathit{\Delta}^*|c_{\mathrm{D}} &\geq \mathit{\Phi}(g - X^*(\mathit{\Delta}^{**})) - |\mathit{\Delta}^{**}|c_{\mathrm{D}} \tag{79}\\ \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**})) - |\mathit{\Delta}^{**}|c'_{\mathrm{D}} &\geq \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) - |\mathit{\Delta}^*|c'_{\mathrm{D}} \tag{80}\end{align}$$

which can be rewritten as

$$\begin{align} \mathit{\Phi}(g - X^*(\mathit{\Delta}^{**})) - \mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) &\leq (|\mathit{\Delta}^{**}| - |\mathit{\Delta}^*|)c_{\mathrm{D}} \tag{81}\\ \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**})) - \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) &\geq (|\mathit{\Delta}^{**}| - |\mathit{\Delta}^*|)c'_{\mathrm{D}} \tag{82}\end{align}$$

Since $c'_{\mathrm{D}} > c_{\mathrm{D}}$, these inequalities imply

$$\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**})) - \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*)) > \mathit{\Phi}(g - X^*(\mathit{\Delta}^{**})) - \mathit{\Phi}(g - X^*(\mathit{\Delta}^*)). \tag{83}$$

This, combined with $\mathit{\Phi}(g - X^*(\mathit{\Delta}^*)) = \mathit{\Phi}(g - X^{**}(\mathit{\Delta}^*))$ and $\mathit{\Phi}(g - X^{**}(\mathit{\Delta}^{**})) = \mathit{\Phi}(g - X^*(\mathit{\Delta}^{**}))$ leads to contradiction. Hence it must be that payoff to the adversary increases when cost of defence increases. Notice that this argument holds for generic values of parameters of the model. There are non-generic examples where payoff to the adversary decreases when cost of defence increases. □

Before proving Proposition **??**, we need the following auxiliary lemma, stating a useful property of of a convex function.

**Lemma 7.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly convex and differentiable function. Then function*

$$g(x, y) = \frac{yf(x) - xf(y)}{x - y} \tag{84}$$

*is strictly increasing in both arguments, as long as $x > y$.*

*Proof.* To show the result we compute partial derivatives of $h$:

$$\begin{aligned} g_x(x, y) &= \left(\frac{y}{x - y}\right)\left(f'(x) - \frac{f(x) - f(y)}{x - y}\right) \\ g_y(x, y) &= \left(\frac{x}{x - y}\right)\left(\frac{f(x) - f(y)}{x - y} - f'(y)\right). \end{aligned}$$

By strict convexity of $f$, $f'(y) < (f(x) - f(y))/(x - y) < f'(x)$, as long as $x > y$, hence $g_x, g_y > 0$ and so $g$ is strictly increasing in $x$ and in $y$. This completes the proof. □

Now we are ready to prove Proposition **??**.

***Proof of Proposition* ??**. Part 1 follows directly; we omit a proof. For Part 2, observe from Proposition 2 that with $c_{\mathrm{A}} \in (\Delta f(n-1), f(n) - (n-1)f(1))$, and $g \in \mathcal{E}(g, c_{\mathrm{A}}) = 0$, the optimal attack targets no nodes. So the optimal defence also consists of defending no node. Thus costs of conflict are 0.

For points **??** and **??**, assume that $c_{\mathrm{A}} \in (\Delta f(m-1), \Delta f(m))$ with $m \in \{1, \ldots, n-1\}$. With such cost of attack, on any connected network, the adversary best responds to any incomplete defence by removing at least one node. Therefore the lower bound for costs of conflict are $\min(c_{\mathrm{A}}, nc_{\mathrm{D}})$ in this case.

**Part 3:** Suppose that $c_{\mathrm{D}} > c(n, m)$ (point **??**). We show first that in every equilibrium on $h_n^m$ the defender chooses the empty defence and the adversary responds to it with attack $\{1\}$ (the separator of $h_n^m$). By Proposition 2, an equilibrium defence must be either empty, or complete, or equal to $\{1\}$. Moreover, best response of the adversary to empty defence either contains $\{1\}$, in which case the reducing attack part of it must be empty (because components of $h_n^m - \{1\}$ have sizes at most $m$), or does not contain $\{1\}$, in which case it must be a reducing attack leaving a residual network consisting of a single component of size $m$. It is easy to check that the former is the best response to the empty defence and the latter is the best response to defence $\{1\}$. Hence empty defence is better than $\{1\}$. Payoff to the defender from using empty defence is

$$\Pi^{\mathrm{D}}(\varnothing, \{1\}; h_n^m, c_{\mathrm{D}}) = \Phi(h_n^m - \{1\}) = \left\lfloor \frac{n-1}{m} \right\rfloor f(m) + f((n-1) \bmod m). \quad (85)$$

With cost of defence $c_{\mathrm{D}} > c(m, n)$, payoff to the defender from complete defence,

$$\Pi^{\mathrm{D}}(N, \varnothing; h_n^m, c_{\mathrm{D}}) = f(n) - nc_{\mathrm{D}}, \quad (86)$$

is lower than the payoff from empty defence. Hence on equilibrium path the defender chooses $\varnothing$ and the adversary responds with $\{1\}$.

Second, we show that for the ranges of costs in question, $nc_{\mathrm{D}} > c_{\mathrm{A}}$. Since $c_{\mathrm{D}} > c(n, m)$ so

$$nc_{\mathrm{D}} > f(n) - \left\lfloor \frac{n-1}{m} \right\rfloor f(m) - f((n-1) \bmod m). \quad (87)$$

Right hand side of this inequality can be rewritten as

$$f(n) - \frac{n-1-(n-1) \bmod m}{m} f(m) - f((n-1) \bmod m). \quad (88)$$

Since $f$ is strictly convex and $(n-1) \bmod m < m$ so $((n-1) \bmod m)f(m) > mf((n-1) \bmod n)$. Therefore

$$c_{\mathrm{D}} > f(n) - \left(\frac{n-1}{m}\right) f(m). \quad (89)$$

Right hand side can be rewritten as

$$f(n) - \left(\frac{n-1}{m}\right) f(m) = \sum_{j=m}^{n-1} \Delta f(j) - (n-m-1)\left(\frac{f(m)}{m}\right) = \Delta f(m) + \sum_{j=m+1}^{n-1} \left(\Delta f(j) - \frac{f(m)}{m}\right). \quad (90)$$

By convexity of $f$, for all $j > m$, $\Delta f(j) > \frac{f(m)}{m}$. Thus $nc_{\mathrm{D}} > \Delta f(m)$ and, since $c_{\mathrm{A}} \in (\Delta f(m-1), \Delta f(m))$, $nc_{\mathrm{D}} > c_{\mathrm{A}}$. Hence the minimal costs of conflict are $c_{\mathrm{A}}$.

**Part 4** suppose that $c_{\mathrm{D}} < c(n, m)$ (point **??**). We will show that with such cost of defence, in any equilibrium on a connected network the defender chooses the complete defence. Notice that with $c_{\mathrm{D}} < c(n, m)$, on any connected network $g$, any defence

$\mathit{\Delta}$ of size $|\mathit{\Delta}| \leq m$ is worse to the defender than the complete defence. This is because the residual network after the adversary best responding to $\mathit{\Delta}$ consist of components of sizes at most $m$ and the upper bound on value of such residual networks is $\lfloor (n-1)/m \rfloor f(m) + f((n-1) \bmod m)$ (this upper bound is attained by $h^m_n$). With $c_{\mathrm{D}} < c(n,m)$ the defender prefers complete defence to $\mathit{\Delta}$.

Consider defence $\mathit{\Delta}$ of size $d = |\mathit{\Delta}|$ such that $m < d < n$. Let $X$ be a best response to $\mathit{\Delta}$. Payoff to the defender from $\mathit{\Delta}$ and $X$ is

$$\Pi^{\mathrm{D}}(\mathit{\Delta}, X; g, c_{\mathrm{D}}) = \mathit{\Phi}(g-X) - dc_{\mathrm{D}} \leq f(d) + \left\lfloor \frac{n-d-1}{m} \right\rfloor + f((n-d-1) \bmod m) - dc_{\mathrm{D}}. \tag{91}$$

The upper bound on the value of the residual network above comes from the following observation. With $c_{\mathrm{A}} \in (\Delta f(m-1), \Delta f(m))$, in any best response the adversary removes unprotected nodes from any component of size greater than $m$. Therefore in the best case the adversary removes one node and the only component of size greater than $m$ in the residual network is a fully protected component of size $d$ (by convexity of $f$ it is better to have one fully protected component of size $d$ than several fully protected and smaller ones summing up to $d$). Thus if

$$c_{\mathrm{D}} < \frac{f(n) - f(d) - \lfloor \frac{n-d-1}{m} \rfloor - f((n-d-1) \bmod m)}{n-d} \tag{92}$$

then complete defence is better to $\mathit{\Delta}$ for the defender. We will show that $c(n,m)$ is lower than the RHS of the inequality above, which will imply that for costs of defence under consideration the complete defence is better for the defender. The inequality

$$c(n,m) = \frac{f(n) - \lfloor \frac{n-1}{m} \rfloor - f((n-1) \bmod m)}{n} < \frac{f(n) - f(d) - \lfloor \frac{n-d-1}{m} \rfloor - f((n-d-1) \bmod m)}{n-d} \tag{93}$$

can be rewritten as

$$df(n) - nf(d) - \left(\frac{n-d}{m}\right)(r_1 f(m) - mf(r_1)) + \left(\frac{n}{m}\right)(r_2 f(m) - mf(r_2)) > 0, \tag{94}$$

where $r_1 = (n-1) \bmod m$ and $r_2 = (n-d-1) \bmod m$.[9] Since $r_2 < m$ and $f$ is convex, so $r_2 f(m) - mf(r_2) > 0$ and to show that the inequality above holds it suffices to show that

$$df(n) - nf(d) - \left(\frac{n-d}{m}\right)(r_1 f(m) - mf(r_1)) > 0. \tag{95}$$

[9] Recall that for integer $x$ and $y$, $\left\lfloor \frac{x}{y} \right\rfloor = \frac{x - x \bmod y}{y}$.

Since $d < n$ and $f$ is convex, so $df(n) - nf(d) > 0$. Moreover, by Lemma 7,

$$\frac{df(n) - nf(d)}{n - d} > \frac{r_1 f(m) - mf(r_1)}{m - r_1} \tag{96}$$

(as $n > d > m > r_1$). Hence

$$\frac{df(n) - nf(d)}{n - d} - \left(\frac{m - r_1}{m}\right)\left(\frac{r_1 f(m) - mf(r_1)}{m - r_1}\right) > 0, \tag{97}$$

which implies (95), by multiplying both sides by $(n - d)$. Hence any equilibrium defence is complete and costs of conflict are $nc_{\mathrm{D}}$. □

**Example of no conflict networks:** Even if marginal of $f$ do not grow very rapidly, there may exist networks (other than complete network) that do not feature active conflict. Take $f(x) = x^2$, for example. Consider a family of core-periphery networks, $\{cp_k\}_{k \in \mathbb{N}}$. Given $k \in \mathbb{N}$, network $cp_k$ has $2k$ nodes: a fully connected core of $k$ nodes, and a periphery of $k$ nodes. Each core node is connected to exactly one, unique, periphery node (c.f. Figure 8).

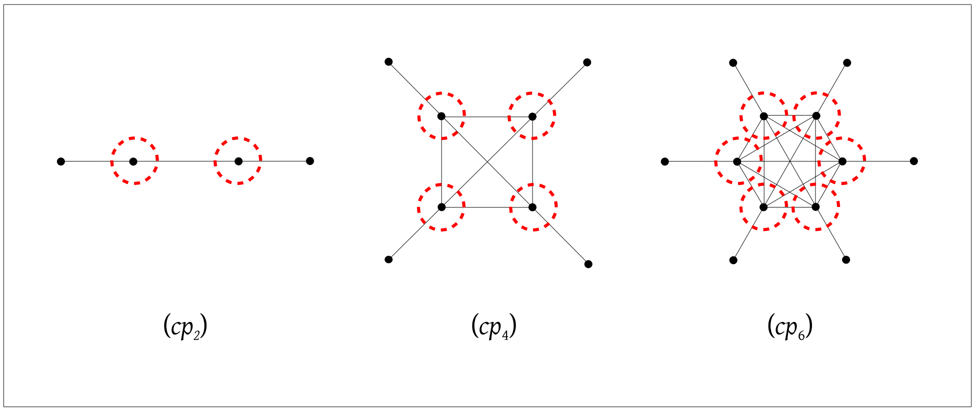

Figure 8: Core-periphery networks $cp_2$, $cp_4$ and $cp_6$.

When cost of attack is high, $c_{\mathrm{A}} > 4m - 1$, then it is easy to verify that in equilibrium the defender will either defend all the core nodes or use an empty defence. When cost of attack is low, $c_{\mathrm{A}} < 4m-1$, then, again, there are two types of equilibrium defence: either no node is defended or all nodes are defended. It is easy to verify that three types of defence would be candidates for equilibrium defence here: empty defence, complete defence and defence with all core nodes protected. To rule out the last one, suppose that $2(2m - k) - 1 \leq c_{\mathrm{A}} < 2(2m - k) + 1$, where $1 \leq k \leq m - 1$. With such cost of attack, the adversary would remove a single node from a component of size greater than $2m-k$ and would not remove single nodes from a component of size $2m-k$ or less. Thus, when all the core nodes are protected, the adversary removes $k$ spokes. When no node is protected, the adversary removes $k$ core nodes. Since defending only the core nodes is better than defending all the nodes, so $mc_{\mathrm{D}} \geq (2m)^2 - (2m-k)^2$. On the other hand, protecting all the core nodes is better than no protection, so $mc_{\mathrm{D}} \leq (2m-k)^2 - (2m-2k)^2 - k$. The two inequalities imply $2k^2 - k \leq 0$, a contradiction for $k \geq 1$.

Notice, in the example above, that if each core node was connected to a higher number of periphery nodes, active conflict would be possible (as illustrated by Example 1). With more periphery nodes per core node (and with suitable costs of defence and attack), protecting the separators may create enough value for such a defence to be attractive. Increasing the value of residual network requires defending all the nodes, which is too high an investment and to low a gain to be profitable. This illustrates one reason for possibility of active conflict in the model: blocking all the individually rational essential separators may secure high value of the residual network at a relatively low cost, while increasing the value further may require a much higher cost.

To get more insight into why active conflict is possible, despite the convexity of $f$ and linearity of costs, consider the network in Figure 9. Figure 10 illustrates function

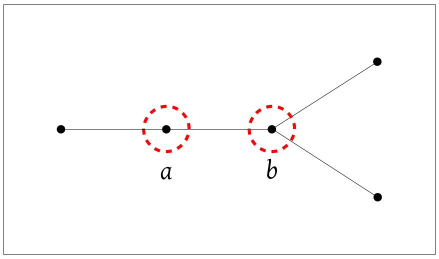

Figure 9: Network that allows for active conflict (under $f(x) = x^2$).

$\Phi^*(m; g, c_{\mathrm{A}})$ under different ranges of costs of attack. The dotted line is an upper convex hull of that function. Optimal size of defence is at a point of that hull adjacent to a line with slope $c_{\mathrm{D}}$. In the case of low cost of attack, if the convex hull contains any points of $\Phi^*(m; g, c_{\mathrm{A}})$, for $0 < m < n$, then active conflict is possible for some suitable range of costs of defence. In the case of high cost of attack, active conflict is possible, if the convex hull contains any points of $\Phi^*(m; g, c_{\mathrm{A}})$, for $0 < m < \tau(\mathcal{E}(g, c_{\mathrm{A}}))$.

In Figure 10, low cost of attack is $c_{\mathrm{A}} < 9$ and $c_{\mathrm{A}} > 9$ is high cost of attack. Active conflict is possible for $c_{\mathrm{A}} \in (5, 9)$. When $c_{\mathrm{A}} \in (5, 7)$ and $c_{\mathrm{D}} \in (3.75, 4)$,

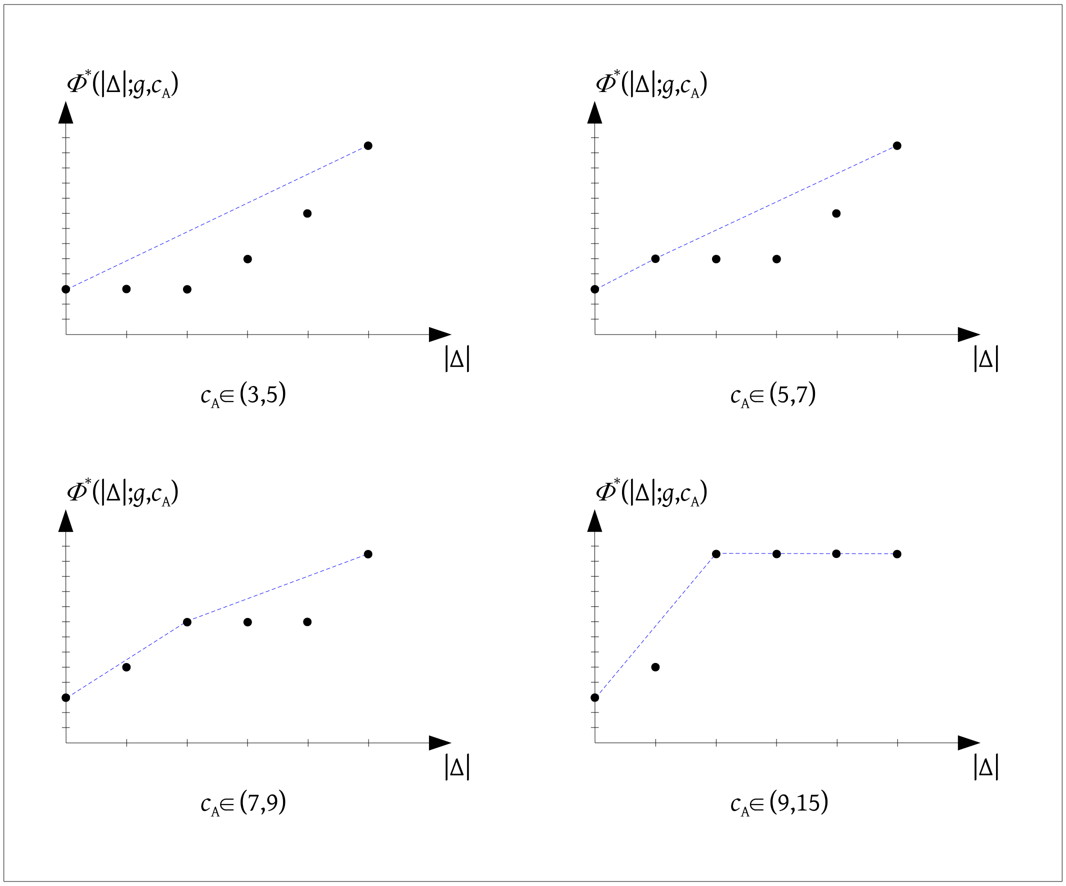

Figure 10: Optimal defences of different sizes for network in Figure 9.

then the unique equilibrium defence is $\mathit{\Delta}^* = \{b\}$, and the best response to it in the adversary's subgame is $X^*(\mathit{\Delta}^*) = \{a\}$. When $c_{\mathrm{A}} \in (7,9)$ and $c_{\mathrm{D}} \in (5,9)$, then the unique equilibrium defence is $\mathit{\Delta}^* = \{a,b\}$ and removing any unprotected node is a best response to it in the adversary's subgame. When $c_{\mathrm{A}} \in (9,15)$, $\tau(\mathcal{E}(g,c_{\mathrm{A}})) = 2$ and there is no equilibrium outcome with active conflict.

***Proof of Proposition 5:*** For point 1, suppose that $c_{\mathrm{D}} > f(n)/n$. We will show that in this case equilibrium defence $\mathit{\Delta} = \varnothing$. Assume, to the contrary, that $\mathit{\Delta} \neq \varnothing$ and let $X$ be the equilibrium response to $\mathit{\Delta}$. Pick any $i \in \mathit{\Delta}$ and let $C(i)$ be the component of $i$ in the residual network $g - X$. Payoff to $i$ is

$$\Pi_i(\mathit{\Delta}, X; g) = \frac{f(|C(i)|)}{|C(i)|} - c_{\mathrm{D}}. \tag{98}$$

By the fact that $f$ is strictly increasing and strictly convex, $f(x)/x$ is increasing. Hence $\Pi_i(\mathit{\Delta}, X; g) \leq f(n)/n - c_{\mathrm{D}} < 0$. Thus $i$ is better off by not protecting, a contradiction with the assumption that $\mathit{\Delta}$ is an equilibrium defence. Hence it must be that $\mathit{\Delta} = \varnothing$.

For point 2, suppose that $c_{\mathrm{D}} < f(n)/n$.

Assume that $c_{\mathrm{A}} < f(n) - f(n-1)$. We will show that $\mathit{\Delta} = N$ is an equilibrium defence. For assume otherwise. Then there exists $i \in \mathit{\Delta}$ which is better off by deviating and choosing no protection. Since $c_{\mathrm{A}} < f(n) - f(n-1)$, so the best response to $\mathit{\Delta} \setminus \{i\}$ is $X = \{i\}$, and so the deviating node gets removed, obtaining payoff 0 instead of $f(n)/n - c_{\mathrm{D}} \geq 0$. Hence $i$ is not better off by deviating and so $\mathit{\Delta} = N$ is an equilibrium defence. This proves point 2a.

Assume that $c_{\mathrm{A}} > f(n) - f(n-1)$. Let $\mathit{\Delta}$ be minimal transversal of $\mathcal{E}(g, c_{\mathrm{A}})$. We will show that $\mathit{\Delta}$ is an equilibrium defence. By Lemma 1, the best response to $\mathit{\Delta}$ is the empty attack $X = \varnothing$. Assume, to the contrary, that $\mathit{\Delta}$ is not an equilibrium defence. Then there exists $i \in \mathit{\Delta}$ that is better off by choosing no protection instead of protection. Since $\mathit{\Delta}$ is a minimal transversal, it must be that there exists an essential separator $E \in \mathcal{E}(g, c_{\mathrm{A}})$ such that $\mathit{\Delta} \setminus \{i\} \cap E = \varnothing$. Moreover, any such separator contains $i$. Since any such separator is better than the empty attack, the adversary responds to $\mathit{\Delta} \setminus \{i\}$ with one of these separators, removing $i$. But then $i$ gets payoff 0 instead of $f(n)/n - c_{\mathrm{D}} \geq 0$. Hence it is not better of by deviating, a contradiction. Therefore $\mathit{\Delta}$ must be an equilibrium defence. This proves point 2b.

Since adversary's sub-game remains like in the centralized defence game, so an equilibrium response $X^*$ is as described in Proposition 2. □

# Appendix B: Key players and Centrality

Essential separators and their transversals determine the key nodes in our study of attack and defence. These key groups of nodes give rise to new notions of centrality distinct from other notions such as closeness, betweenness or eigenvector centralities. To see how these notions are different, consider the network in Figure 11 (for simplicity the example is based on individual, rather than group, notions of centrality). Assume that network value is based on function $f(x) = x^2$ and suppose that cost of attack, $c_{\mathrm{A}} \in (25, 89)$, so that the adversary attacks only the nodes that separate the network and so that removing node 2 is better than not attacking at all. Suppose also that $c_{\mathrm{D}} \in (0, 89)$, so that defending node 2 constitutes an optimal defence as well. However, this node is less central than node 1 in the sense of degree, closeness, betweenness, eigenvector, Bonacich and inter-centrality measures.[10] The numerical values for these centralities are summarized in Table 1. For Bonacich centrality, we consider three values of the parameters: high ($\alpha = 0.237$), intermediate ($\alpha = 0.1$) and low ($\alpha = 0.01$).

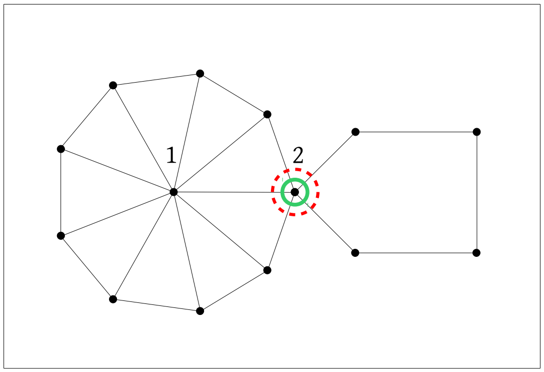

Figure 11: Separators and other centrality measures

| **Centrality** | **Node** 1 | **Node** 2 |
|---|---|---|
| Degree | 9 | 5 |
| Closeness | 0.684 | 0.619 |
| Betweenness | 42.5 | 37.5 |
| Eigenvector | 0.5765 | 0.3036 |
| Bonacich, high | 532.2 | 281.18 |
| Bonacich, medium | 2.4311 | 1.8208 |
| Boncacich, low | 1.093 | 1.0519 |
| Inter-centrality, high | 2940.9 | 2863.2 |
| Inter-centrality, medium | 5.2438 | 3.1263 |
| Inter-centrality, low | 1.1936 | 1.1061 |

Table 1: Centralities of nodes 1 and 2 in network from Figure 11.

[10] Following Ballester et al. (2006), we define for a parameter $\alpha \in \mathbb{R}$: $\mathbf{b}(g,\alpha) = \mathbf{M}(g,\alpha)\mathbf{1}$, where $\mathbf{M}(g,\alpha) = (\mathbf{I} - \alpha\mathbf{G})^{-1}$, $\mathbf{I}$ is the identity matrix and $\mathbf{G}$ is the adjacency matrix of the network. We require $\alpha$ to be relatively small so that $\mathbf{M}(g,\alpha)$ is well defined and non-negative. The inter-centrality measure we consider, also defined in that paper, is $c_i(g,\alpha) = \frac{b_i(g,\alpha)^2}{M_{ii}(g,\alpha)}$. We define closeness as $cl_i(g) = \frac{n-1}{\sum_{j \neq i} d(i,j;g)}$ where $d(i,j;g)$ is the length of the shortest path between $i$ and $j$ in $g$.

# Appendix C: Separators and transversals in families of networks

**Interlinked stars:** Interlinked stars are networks with two disjoint non-empty sets of nodes: the set of *centers* $C$ and the set of *periphery nodes* $P$. The centers are fully connected forming a clique. Each of the periphery nodes is connected to all the centers. Interlinked stars have one essential separator: the set of all the centers, $\mathcal{E}(g) = \{C\}$. All minimal transversals of $\mathcal{E}(g)$ are singleton sets consisting of one central node. The essential separator and a minimal transversal for interlinked star are illustrated in Figure 12.

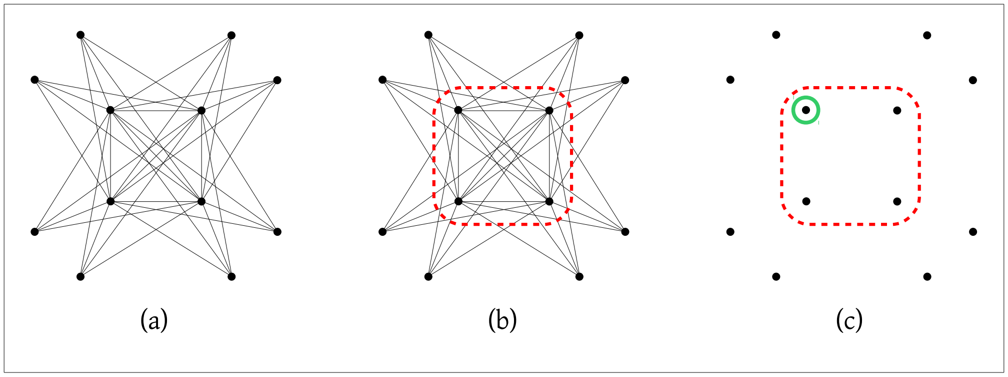

Figure 12: Separators and transversals in interlinked stars ($n = 12$)

**Complete bipartite networks:** In a complete bipartite network the set of nodes, $N$, can be partitioned into two disjoint sets, $N_1$ and $N_2$, $N_1 \cap N_2 = \varnothing$, such that the set of links is the set of all possible links connecting nodes from $N_1$ and nodes from $N_2$. There are two essential separators in these networks, $\mathcal{E}(g) = \{N_1, N_2\}$. Every transversal, consists of one node from $N_1$ and one node from $N_2$. Minimal essential separators and transversal for complete bipartite networks are illustrated in Figure 13.

**Trees:** In any tree network, every non-empty set of internal nodes (nodes which are not leaves) constitutes a separator. Essential separators are sets of internal nodes such that no two of them are neighbours. Transversals of essential separators are subsets of internal nodes. In particular, there is a unique transversal of the set of all essential separators: the set of all internal nodes. Minimal essential separators and transversal for tree networks are illustrated in Figure 14.

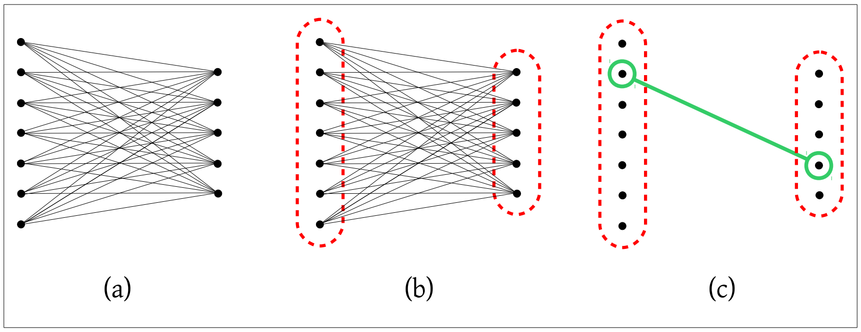

Figure 13: Separators and transversals in complete bipartite networks ($n = 12$)

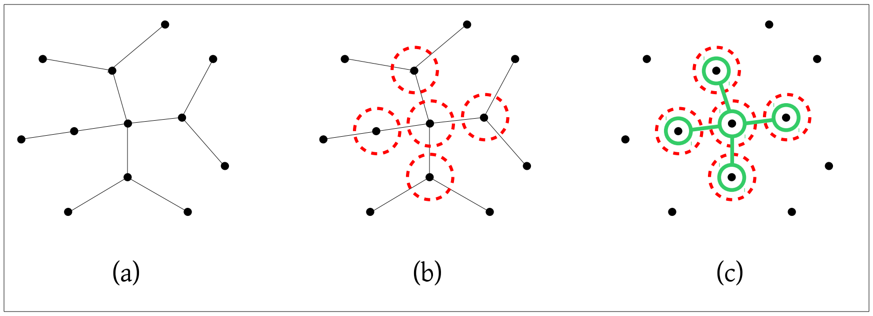

Figure 14: Separators and transversals in trees ($n = 12$)

**Core-periphery networks:** Nodes are divided in two disjoint sets: the *core* and the *periphery*. Each node of the periphery is connected to exactly one node of the core, while the nodes of the core are connected with periphery nodes and the core constitutes a clique. Essential separators are subset of the core. There is a unique transversal: the set of all core nodes. Minimal essential separators and transversal for core-periphery networks are illustrated in Figure 15.

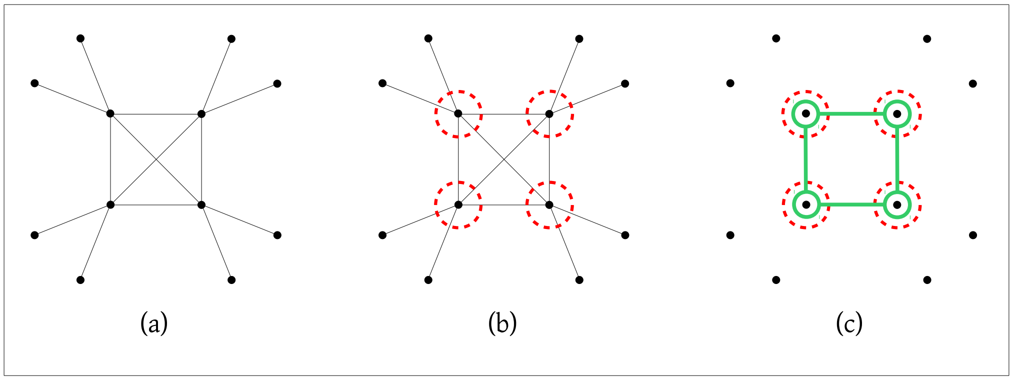

Figure 15: Separators and transversals in core periphery networks ($n = 12$)

# Appendix D: Order of moves and nature of conflict

This section explores the role of sequential choice and perfect defence.

**Simultaneous moves:** Consider a variant of the model studied in the paper where the players make their choice simultaneously. In this case the set of strategies of the defender remains unchanged. A pure strategy of the adversary is now a set of nodes, $X \subseteq N$, chosen to attack. It is important to note that the timing of moves does not affect Lemma 1 which remains unchanged. Suppose that cost of attack is high. Any strategy, $X$, in the support of equilibrium strategy of the adversary must be an individually rational essential separator, i.e. $X \in \mathcal{E}(g, c_{\mathrm{A}})$. Similarly, any strategy, $\varDelta$, in the support of equilibrium strategy of the defender must be a minimum transversal of the set of essential separators it blocks, $\mathcal{D}(\varDelta, \mathcal{E}(g, c_{\mathrm{A}}))$.

The second observation is that depending on the network, the players may use pure or mixed strategies in equilibrium. This is a departure from our existing results, where equilibrium always exist in pure strategies. But note that the use of mixed strategies is sensitive to the network. In particular, if the network is such that one unit of defence is sufficient to block all the individually rational essential separators of the adversary, then in equilibrium both players use pure strategies and equilibrium outcomes are the same as in the sequential model studied in the paper. When $\tau(\mathcal{E}(g, c_{\mathrm{A}})) > 1$, the defender may choose to block more individually rational essential separators, by mixing across several transversals.

**The model of conflict:** We have assumed perfect defence. A more natural way to proceed would be to suppose that the number of resources assigned by each player to a node determines the probability of winning/loosing the node. Following Tullock (1980), suppose that the probability of successfully attacking the node is given by a *contest success function* (CSF)

$$\pi(a, d) = \begin{cases} 0, & \text{if } a = 0 \\ \frac{d^\gamma}{a^\gamma + d^\gamma}, & \text{otherwise,} \end{cases} \tag{99}$$

where $\gamma \in \mathbb{R}_+$ and $a$ and $d$ are resources assigned by the adversary and defender, respectively. The probability of successfully defending the node is $\pi(d, a) = 1 - \pi(a, d)$.[11]

[11] The perfect defence model studied in the paper can be seen as a limiting case of the general contest model: the probability of successful attack is given by $\frac{\alpha a^\gamma}{\delta d^\gamma + \alpha a^\gamma}$ with $\alpha = 1$ and $\delta \to +\infty$.

A strategy of the defender is a vector $\mathbf{d} \in \mathbb{N}^N$ such that $d_i$ is the number of defence resources assigned to node $i$. A strategy of the adversary is a function $X : \mathbb{N}^N \to \mathbb{N}^N$ such that, given vector of defence allocation $\mathbf{d}$, maps it to a vector of attack allocation $\mathbf{a} = X(\mathbf{d})$ such that $a_i$ is the number of attack resources assigned to node $i$. We will call the set of nodes receiving positive number of defence resources the *defended nodes* and the set of nodes receiving positive number of attack resources the *attacked nodes*. Given defence and attack allocations, $(\mathbf{d}, \mathbf{a})$, the probability that set $M \subseteq N$ of nodes is won by the adversary and removed from $g$ is

$$w(M|\mathbf{a}, \mathbf{d}) = \prod_{j \in M} \pi(a_j, d_j). \tag{100}$$

The expected payoffs to the defender and the adversary from defence and attack allocations, $(\mathbf{a}, \mathbf{d})$, are

$$\begin{aligned} \Pi^{\mathrm{A}}(\mathbf{a}, \mathbf{d}|g, c_{\mathrm{A}}) &= -\textstyle\sum_{M \subseteq N} w(M|\mathbf{a}, \mathbf{d})(1 - w(N \setminus M|\mathbf{a}, \mathbf{d}))\varPhi(g - M) - c_{\mathrm{A}} \sum_{j \in N} a_i \\ \Pi^{\mathrm{D}}(\mathbf{a}, \mathbf{d}|g, c_{\mathrm{D}}) &= \textstyle\sum_{M \subseteq N} w(M|\mathbf{a}, \mathbf{d})(1 - w(N \setminus M|\mathbf{a}, \mathbf{d}))\varPhi(g - M) - c_{\mathrm{D}} \sum_{j \in N} d_i. \end{aligned}$$

Lemma 1 still obtains. The set of attacked nodes can be decomposed into an essential separator and a reducing cut. In what follows we restrict attention to high costs of attack and we focus on the benchmark model of linear contests: $\gamma = 1$. The main point we wish to make is that with Tullock contests optimal defence will extend beyond minimal transversals and may cover multiple nodes in the same separator.

Consider an interlinked star with two core nodes: 1, 2, and $n - 2$ periphery nodes ($n \geq 4$). Suppose that cost of attack is high, $c_{\mathrm{A}} > \Delta f(n - 1)$. The unique essential separator of $g$ is the set of core nodes, $\{1, 2\}$. Let $a_1$, $a_2$ be the amount of resources assigned by the adversary to the two core nodes and $d_1$, $d_2$ be the defence resources assigned by the defender to the two core nodes. Expected payoff to the adversary from assignment $(a_1, a_2, d_1, d_2)$ is:

$$\begin{aligned} \Pi^{\mathrm{A}}(\mathbf{d}, \mathbf{a}|g, c_{\mathrm{A}}) = &- \pi(a_1, d_2)\pi(a_2, d_2)(n - 2)f(1) \\ &- (\pi(a_1, d_1) + \pi(d_2, a_2) - 2\pi(d_1, a_1)\pi(a_2, d_2))f(n - 1) \\ &- (1 - \pi(a_1, d_1) - \pi(d_2, a_2) + \pi(d_1, a_1)\pi(a_2, d_2))f(n) \\ &- c_{\mathrm{A}}(a_1 + a_2) \\ = &- f(n) + \pi(a_1, d_1)\pi(a_2, d_2)V_1(n) \\ &+ (\pi(a_1, d_1) + \pi(a_2, d_2) - \pi(a_1, d_1)\pi(a_2, d_2))V_2(n) \\ &- c_{\mathrm{A}}(a_1 + a_2) \end{aligned} \tag{101}$$

where $V_1(n) = f(n-1) - (n-2)f(1)$ and $V_2(n) = f(n) - f(n-1)$. Notice that $V_2(n)$ is the gain from removing the first node of the core, and $V_1(n)$ is the gain from removing the second node of the core. Since cost of attack is high, $V_2(n) < c_{\mathrm{A}}$. Hence if $V_1(n) \leq V_2(n)$, then it is not profitable for the adversary to attack, and both players assign no resources to the nodes in equilibrium. Consider now the more interesting case, where $V_1(n) > V_2(n)$.

The expected payoff to the defender is:

$$\begin{aligned}\Pi^{\mathrm{D}}(\mathbf{d}, \mathbf{a}|g, c_{\mathrm{A}}) = & f(n) - \pi(a_1, d_1)\pi(a_2, d_2)V_1(n) \\ & - (\pi(a_1, d_1) + \pi(a_2, d_2) - \pi(a_1, d_1)\pi(a_2, d_2))V_2(n) \\ & - c_{\mathrm{D}}(d_1 + d_2) \end{aligned} \tag{102}$$

The defender chooses $(d_1, d_2)$ to maximize his expected payoff subject to the constraints that $d_1, d_2 \geq 0$ and that the adversary chooses $(a_1, a_2)$ to maximise his expected payoff, subject to $a_1, a_2 \geq 0$.

It is simpler to begin with the case where the defender is given $2d \geq 0$ defence resources and the adversary is given $2a \geq 0$ attack resources. This turns the optimization problem above into a zero-sum bi-level optimization problem, where the defender chooses an allocation of $2d$ to maximize:

$$\pi(a_1, d_1)\pi(a_2, d_2)V_1(n) + (\pi(a_1, d_1) + \pi(a_2, d_2) - \pi(a_1, d_1)\pi(a_2, d_2))V_2(n). \tag{103}$$

It is possible to show the partition $(d, d)$ is a maximizer of both $\pi(a_1, d_1)\pi(a_2, d_2)V_1(n)$ and $(\pi(a_1, d_1) + \pi(a_2, d_2) - \pi(a_1, d_1)\pi(a_2, d_2))V_2(n)$, and hence of the whole expression above. In response the adversary chooses the partition $(a, a)$. Thus $(d, d)$ and $(a, a)$ are the equilibrium defence and attack strategies as well.

When both players distribute their resources evenly, the payoff to the adversary is

$$\begin{aligned}\Pi^{\mathrm{A}}(\mathbf{d}, \mathbf{a}|g, c_{\mathrm{A}}) = & - f(n) + \pi(a, d)^2 V_1(n) \\ & + (2\pi(a, d) - \pi(a, d)^2)V_2(n) \\ & - 2c_{\mathrm{A}}a \end{aligned} \tag{104}$$

If $d \geq V_2(n)/c_{\mathrm{A}}$, it is not profitable for the adversary to attack. Thus with sufficiently low ratio $c_{\mathrm{D}}/c_{\mathrm{A}}$, the adversary distributes his resources evenly and the adversary does not attack. Otherwise both players compete, choosing optimal levels of attack and defence resources and distributing them evenly.