# Alliance Formation in a Multipolar World*

Peter Devine†
George Washington University

Sumit Joshi‡
George Washington University

Ahmed Saber Mahmud§
Virginia Polytechnic Institute and State University

May 2024

**Abstract**

We propose a multilayer network approach to alliance formation. In a signed affinity layer, agents are partitioned into clusters, with friendly relations within and hostile connections across clusters. Agents then form defensive collaborations in an alliance layer as follows: Agents in the same cluster form a nested split graph with degree inversely correlated to the level of hostility, and agents from disparate clusters with high-degree and low-hostility form cliques. Within cliques, agents from a cluster that is "intermediate" in terms of discord serve as a bridge to interconnect agents from more "extreme" clusters.

**Key words and phrases:** Alliance formation, signed graphs, nested split graphs, pairwise stability, cliques.
**JEL Classification**: C72, D74, D85

## 1 Introduction

This paper is a contribution to the literature on alliance formation under conflict. It explores the incentives of agents (individuals, groups or nations) to form defense alliances when they are embedded in a pre-existing network of bilateral affinities that are friendly, hostile or neutral, and for these affinities in turn to be revised following the formation of defense collaborations. This primitive non-empty network (equivalently *layer* or *graph*) of *affinities* is assumed to be a possible consequence of political, religious, ideological, cultural or historical factors.[1] It can be formally represented as a *signed* network in which a positive link between two

*We would like to express our sincere thanks to the editor, associate editor, and two anonymous referees whose comments and suggestions have significantly improved the paper. We remain responsible for any remaining errors.

†Fellow, National Economic Council, The White House, Washington DC, USA. Email: peter.a.devine@who.eop.gov.

‡*Corresponding author*: Department of Economics, George Washington University, 271 Monroe Hall, 2115 G Street NW, Washington DC 20052, USA. Phone: 202-994-6154; Fax: 202-994-6147; Email: sumjos@gwu.edu.

§Department of Economics, Virginia Polytechnic Institute and State University, Pamplin Hall, Blacksburg, VA 24060. USA. Email: asaber1973@gmail.com

[1]Each of these factors could in itself constitute a network or layer that identifies agent-pairs as friendly, hostile or neutral along the given dimension. There is no reason why these various layers would be congruent. However, in many instances, one of these layers would be dominant in terms of dictating agents' incentives for alliance formation. For example, it can be argued that ideological differences motivated the formation of NATO and the Warsaw Pact. Consequently, for ease of analysis, we restrict attention to a single layer of *affinities*.

agents denotes friends, a negative link denotes enemies, and lack of a link denotes a neutral relationship.[2] While there is a large literature on network formation under conflict[3], our point of departure is an explicit two-way interaction between the signed network of affinities and the network of defense *alliances* among friends to thwart potential conflicts with enemies. This interaction is examined through the lens of a *multilayer* network. Figure 1 depicts a multilayer network in which the base layer is the affinity network (denoted by $\mathbf{H}$) and the accompanying layer is the network of defense alliances (denoted by $\mathbf{G}$). The affinity layer is a signed network in which a solid line connecting two agents denotes friendship (a positive relationship), a dashed line denotes hostility (a negative relationship), and lack of a connection denotes a neutral (zero) relationship. The alliance layer is an unsigned network in which a (solid) line connecting two agents denotes a defense collaboration and the absence of a line implies no such collaboration.

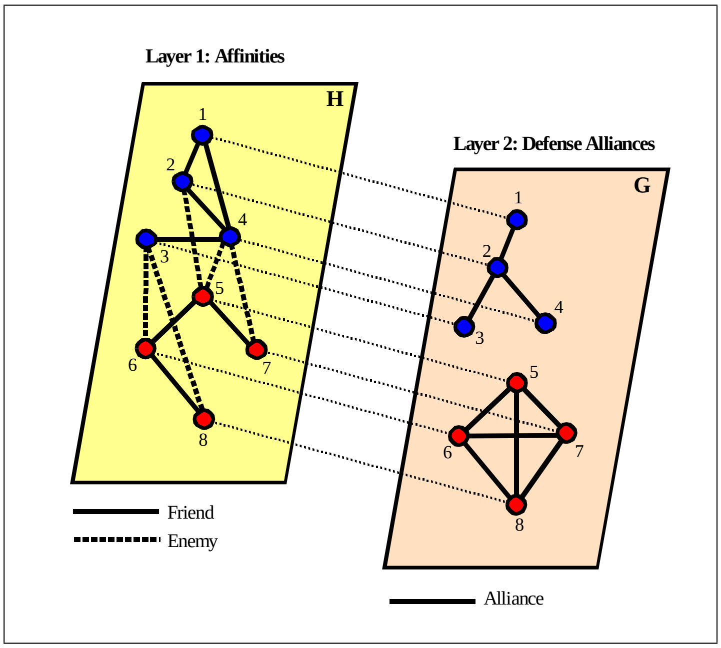


Figure 1: A Multilayer Network

Our paper is motivated by the fact that the complex web of interlocking defense alliances that characterize the world today can best be understood as a multilayer network building up from base affinities. The period of the Cold War was characterized by an affinity network in which countries were broadly divided into an Eastern and a Western bloc based on opposing political ideologies. The corresponding alliance network was *bipolar*: the Eastern bloc formed the Warsaw Pact while the Western bloc formed NATO, with no overlap between the two security pacts. The fall of the Berlin Wall altered the affinity network with former Eastern bloc countries recalibrating their relationships with the Western bloc. The resulting alliance network was *unipolar* with former Warsaw Pact members such as Poland, Hungary, Bulgaria,

[2] Signed networks are discussed in Cartwright and Harary (1956), Davis (1967), and Easley and Kleinberg (2010, Chapter 5).

[3] Please see Bloch (2012) and Goyal et al. (2016) for an excellent description of the main lines of research on alliance formation under conflict.

Romania and the Czech Republic joining NATO. The current alliance network is sometimes described as *multipolar*, which is inaccurate since nations cannot be divided into mutually exclusive coalitions that jointly coordinate their actions as a set. Instead, we see nations forming alliances across affinities. For example, the Economist[4] has noted that the United States has established bilateral alliances with Australia, Japan, Philippines, South Korea and Thailand in a hub-and-spoke network and quoted the prime minister of Japan, Kishida Fumio, as saying that promoting alliances among the spokes "will lead to the establishment of a *multilayered network* (our emphasis), and by expanding that we can improve deterrence."

Nations are close to each other in the affinity network due to commonality in their "norms" and condition their alliances on these norms. The United States provides a ready example. Its closest ally during the American Revolution was France. Immediately after the Revolution, John Adams was sent to London to highlight to the two countries commonalities, shared heritage, and similarities to mollify the contentious relationship. In response, King George replied, "I was the last to consent to the Separation, but ... let the Circumstances of Language; Religion and Blood have their natural and full Effect."[5] Separated by force, the two countries were intertwined first by culture, then similar legal and political institutions, related economies and trade, and by the time of the World War II, a formal military alliance. In the last century, that high-order defense network has evolved into NATO, improved upon by the intelligence sharing FIVE EYES group, the nuclear-powered submarine fleet-sharing AUKUS, and the ultimate subset, "special relationship." Meanwhile France (still a United States ally) has repeatedly threatened or actually withdrawn from NATO's command structures.

The norms themselves are not immutable but subject to change, which in turn precipitates changes in the alliance network. In Western Europe during the Cold War, a common norm centered on security threat from the Eastern bloc resulted in a mutually connected alliance network in the form of NATO. Contemporaneously, the more distributed views of a norm in the Western Pacific and the lack of a pressing threat prevented the emergence of a NATO-like Pacific Pact and instead created a hub-and-spoke architecture with the United States as the center and no security alliances among spokes (Australia, Japan, Philippines, South Korea). However, the increasing military threat from China has led to greater consonance of norms among the spokes and moved them towards exploring bilateral military alliances. The Philippines provides an excellent case study.[6] In 2022, China's coast guard fired a military-grade laser at a Filipino coast guard ship. This converged the Philippine's norm closer to the United States and other regional states and prompted it to sign defense agreements with the European Union, India and Britain, and discuss visiting-forces agreement with Japan, Canada and France.

These issues lead us to address the following questions in the context of a multilayer network. *First*, how does the architecture of the affinity network influence the structure of defense collaborations in the alliance network? In particular, based on their positions in the affinity network, who are the agents assuming a more central position in the alliance network and who are the agents that are more peripheral?

[4] "America's Asian Allies are Trying to Trump-Proof their Policies", *Economist*, April 9, 2024.
[5] Letter from John Adams, Minister to Britain, to John Jay, U.S. Secretary of State, 1785, The National Archives.
[6] "As tensions rise with China at sea, Philippines strikes deals", *The Washington Post*, Monday, March 11, 2024.

*Second*, under what circumstances will agents have an incentive to revise their relationships in the affinity network? Specifically, what are the incentives of agents to mend fences with enemies and transform hostile relationships into friendships? *Third*, how will any change in the affinity network impact defensive collaborations in the alliance network? *Fourth*, and finally, who are the agents that serve as "bridges" in the affinity network to connect agents who would otherwise remain disconnected due to their mutual hostility?

We begin with a description of the architecture of the affinity network **H** in the initial position. We assume that the distribution of positive and negative links is such that we can partition agents into non-empty *clusters* such that relationships within a cluster are friendly or neutral while relations across clusters are hostile or neutral. In the terminology of signed graphs following Davis (1967), a network with this particular distribution of positive and negative links is called *weakly balanced* (henceforth, simply *balanced*). In balanced affinity networks, each cluster is composed of agents who are friends, friends of friends, friends of friends of friends... etcetera. Any links connecting agents across distinct clusters are always negative indicating that the agents are enemies. We assume that the partition of agents into clusters is a consequence of their disagreement over some norm. Agents within the same cluster subscribe to a common norm or core belief (for example, ideology, religion or politics) and thus any links that exist within the cluster are always friendly. Agents in distinct clusters differ in their perception of the norm and this dissonance implies that any links in **H** connecting an agent-pair from two separate clusters is always hostile. The norm or belief is captured by a scalar and thus permits classifying clusters as "close" or "distant" depending on the difference between their adoptive norms or beliefs.

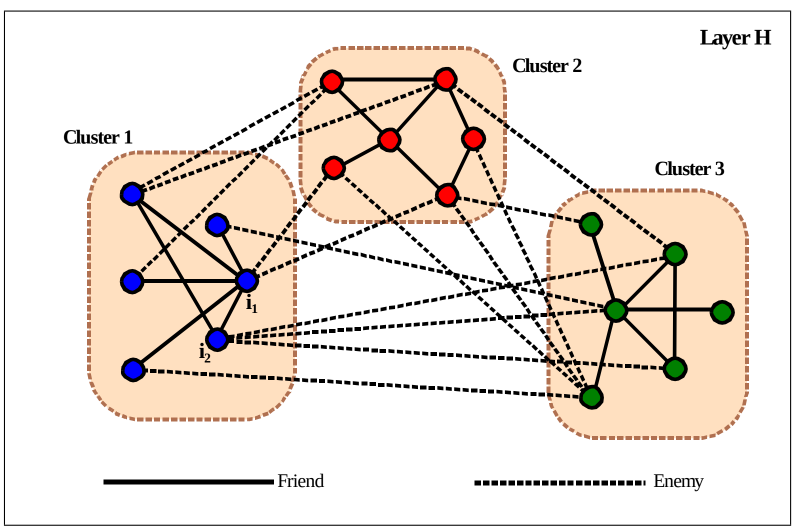


Figure 2: A Balanced Affinity Network

Figure 2 illustrates a balanced affinity network with three clusters. Each agent $i$ in **H** is indexed by a *friendship* measure, $\delta_i$, which is the number of $i$'s friends minus the number of $i$'s enemies. The *higher* the value of $\delta_i$, the more friends agent $i$ has relative to enemies, and thus the *lower* is the level of hostility faced by agent $i$ in the affinity network. For example, within cluster 1 in Figure 2, agent $i_1$ is characterized by $\delta_{i_1} = 3$ and faces the least hostility while agent $i_2$ satisfies $\delta_{i_2} = -1$ and faces the most hostility. This indexation by friendship is important in characterizing the identity of agents occupying the different vertices of the alliance network **G**. We will need two constructs to enable this characterization. We define an agent $i$'s *degree* as the number of its alliances in **G**, and the set of allies with whom agent $i$ has a defense agreement as $i$'s *neighborhood* in **G**. We call a subnetwork in **G** a *nested split graph* (NSG) if the set of agents in the subnetwork together with their links satisfy the property that the neighborhood of a lower degree agent is contained within the neighborhood of a higher degree agent. Figure 3 provides examples from the class of NSG networks. It includes at one end the star network in figure 3(a) with agent $j_1$ as the hub and the other agents ($i_1$ through $i_6$) as spokes, and at the other the complete network in figure 3(d) in which all pairs of agents are mutually linked. All examples display the property that the neighborhood of lower degree agents are nested within the neighborhood of higher degree agents. In particular, the neighborhoods of $i$-indexed agents are contained within the neighborhood of $k$-indexed agents which in turn are subsumed within the neighborhood of $j$- indexed agents.

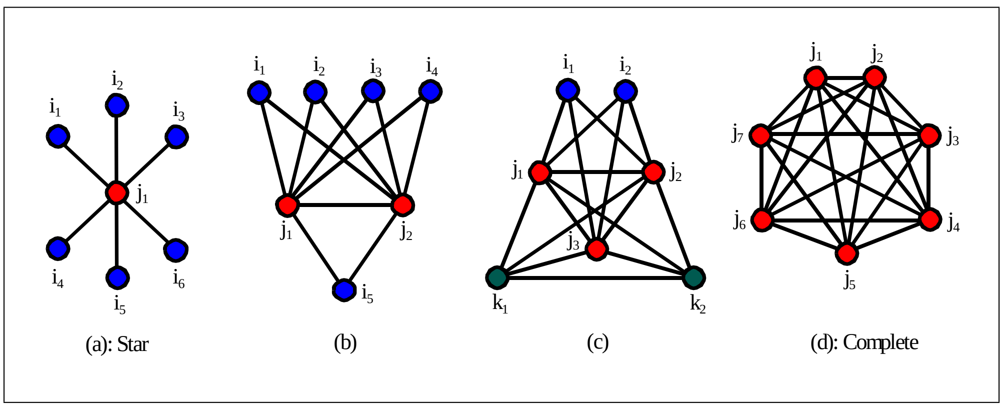


Figure 3: Nested Split Graphs

Given an initial non-empty balanced affinity network **H**, we show that agents within the same cluster form an independent, or disconnected, subnetwork in **G** with the following two characteristics: (i) agents facing less hostility in the affinity network (i.e., with a higher friendship measure) have a greater degree, and (ii) all the allies of an agent with a lower friendship measure are also the allies of an agent with a higher friendship measure. Therefore, we are able to relate the architecture of the affinity network to the structure of defense links in the alliance network: *Given a balanced affinity network, agents within a cluster establish an NSG subnetwork in the alliance network* **G** *with degrees positively correlated with their friendship levels, and the neighborhood of lower friendship agents nested within the neighborhood of higher friendship agents.* Figure 4 illustrates this result. It shows a cluster in the affinity network **H** with the friendship measures $\delta_1 = 5 > \delta_2 = 2 > \delta_3 = 1 > \delta_4 = -1$. The corresponding distribution of degrees – the degree partition – in the subnetwork of the alliance network **G** follows the same order and has an NSG architecture. For instance, the $\delta_1$-agent facing the least hostility in **H** is the most connected in **G** and has a neighborhood that contains the neighborhoods of all other agents in the cluster. At the other extreme, the $\delta_4$-agent facing the most hostility in **H** is the least connected in **G** with a neighborhood contained in the neighborhoods of all other agents.

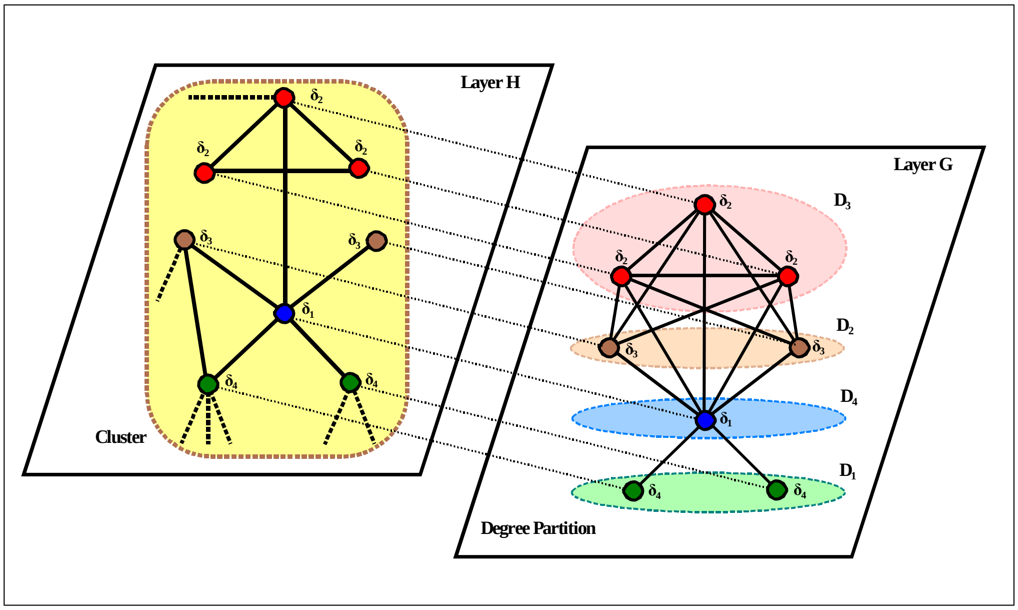


Figure 4: Architecture of the Alliance Network

Once the alliance network is formed, we allow agents to revisit their relationships in the affinity network. Therefore, we permit a two-way interaction between the affinity and alliance networks. Once agents have formed sufficient alliances within their own cluster, then the ensuing gains from these links can provide an incentive for well-connected agents in two separate clusters to change an existing hostile relationship or a neutral one in **H** into a friendly one. This will be particularly true if the difference between their perceived norms is sufficiently small. Of course, agents could also transform a neutral relationship into a hostile one but in our model there is no incentive to do so. Once these changes in the affinity network are implemented, then this revised (and potentially unbalanced) affinity network will spur a new round of alliances in the network **G**. Since new pathways of positive (friendly) links have been created across disparate pairs of clusters, two sufficiently well-connected agents from disparate clusters have an incentive to ally with each other. In particular, we show that *if two clusters are sufficiently close in their perceived norms, then sufficiently well-connected agents in the two clusters form a clique in the alliance network.* A *clique* in **G** is a set of agents such that every pair of agents in the sets are mutually linked. Thus, despite their dissonance over the norm, erstwhile hostile agents will have an incentive to ally if their disagreement over the norm is small.

Since the affinity network in the initial position is composed of separate clusters connected only by hostile links, it leads in the first iteration to disconnected subnetworks in the alliance network since defense alliances are formed within clusters and do not straddle distinct clusters. Thus, both the affinity and alliance layers are characterized by *structural holes* in the following sense: the affinity network is partitioned into clusters separated by hostile links and the alliance network is partitioned into subnetworks of alliances. Thus, both networks are fragmented. Despite the scale economies generated by alliances within a subnetwork, agents in two clusters may continue to remain hostile in the affinity network if the difference in their norms is large and consequently unlinked in the alliance network as well. It is in this context that the role of moderating agents who serve as *structural bridges* in the affinity network comes to the fore. These are agents who form a friendly path between two disparate clusters. We show that *it is agents who hold an intermediate position in their perceived norm between two clusters with widely divergent norms that serve as bridges to connect them in the form of either interlocking or all-inclusive cliques.* The case of an all-inclusive clique is particularly interesting. It is possible that two clusters would never form an alliance since the difference in their norms makes it unprofitable to incur the cost of transforming their existing relationship to a friendly one. However, through the auspices of an intermediate cluster with whom both extremes have developed a positive affinity link, the well-connected agents in the two extremes may end up participating in a three-way inter-cluster clique with agents in the intermediate cluster forming the bridge.

We now place our paper in the context of the existing literature. The literature examining *strategic* behavior in signed networks is relatively sparse. The definitive papers are Franke and Öztürk (2015), Huremovic (2015), Hiller (2017) and Kundu and Pandey (2023). Franke and Öztürk (2015) do not explicitly consider a signed graph but they interpret a link between two agents in an exogenously fixed conflict network as one of enmity. Thus their conflict network has the flavor of a signed graph in which all existing links are negative. Two agents connected by a link are involved in a bilateral contest over resources with success determined by their respective investment in conflict-specific technologies. Their objective is to relate the architecture of enmity links to the intensity of conflict which is measured as the aggregate investment by agents in conflict-specific technologies. In a similar vein, König et al. (2017) consider conflict on a fixed explicit signed graph in which success in a bilateral contest between agents is increasing in their fighting effort and that of friends and decreasing in the fighting effort of enemies. They examine how the architecture of the signed graph impacts individual and aggregate fighting efforts. In contrast to these papers, our focus is not on the investment (effort) involved in conflict but rather on the architecture of alliances that are engendered in the course of conflict. Specifically we focus on how the topology of signed relationships influences the endogenous formation of alliances; moreover, we allow changes in the signed network.

Huremovic (2015) and Hiller (2017) consider the endogenous formation of signed graphs and characterize the architecture of stable (Nash) networks. Both papers explicitly allow the formation of negative or enmity links and demonstrate, in substantiation of the thesis of Cartwright and Harary (1956), that stable (Nash) networks display the balanced property. In contrast, our paper takes a balanced signed network as given and allows limited changes in its structure while putting the focus on the endogenous formation of defense alliances. Further, our paper emphasizes a two-way interaction between the signed network of affinity relationships and the network of alliances: it examines the impact of signed relationships on the architecture of defense collaborations, and conversely the role of defensive alliances in shaping affinities. Kundu and Pandey (2023) also examine a two-layer network. Similar to our paper, one layer is a signed graph. In contrast to our paper, the second layer is also signed and formed endogenously in the course of pairwise cooperation games between agents. Two agents end up as friends (respectively, enemies) in the second layer if mutual cooperation is the equilibrium (respectively, not the equilibrium) of the cooperation game. Their focus is on whether balance in the first layer leads to balance in the second layer. Our emphasis is not on balance. In fact, an important preoccupation of our paper is how the density of intra-cluster defensive alliances and the congruence in perceived norms could render the affinity network unbalanced as agents transform affinities in pursuit of allies across clusters.

Our paper also contributes to the limited literature on multigraph formation. In contrast to Joshi et al. (2020) and Joshi et al. (2023), our paper explicitly incorporates signed graphs into the analysis. Our paper also contributes to the larger literature on alliance formation under conflict, a small sample of which includes Bloch (2012), Jackson and Nei (2015) and Goyal et al. (2016).While this literature focuses exclusively on a single network, our paper emphasizes the multi-network dimension of alliance formation under conflict and the role of preexisting affinities.

The paper is organized as follows. The model is presented in Section 2. The characterization of the stable alliance network for a fixed affinity network is the objective of Section 3. The two-way interaction between the affinity and alliance networks is considered in Section 4. Section 5 presents a selected set of examples to motivate our main results. Section 6 discusses in greater depth specific aspects of the model. Our conclusions are contained in Section 7. All proofs are collected in an appendix.

## 2 The Model

Let $\mathcal{N} = \{1, 2, ..., N\}$, $N \geq 3$, denote the set of agents. A *network* (equivalently, *layer*) comprises of the set $\mathcal{N}$ and the collection of bilateral links connecting members of $\mathcal{N}$. When $\mathcal{N}$ is unambiguous, the network is defined by the set of bilateral links. A *multilayer network* in our context is a set of two layers, with the same set $\mathcal{N}$ of agents inhabiting each layer, but with different bilateral connections on each layer. There are two types of layers that will preoccupy us. The first layer – the base or primitive layer – denoted by $\mathbf{H}$ is a network of affinities (historic, cultural or political). The second layer, denoted by $\mathbf{G}$, is a network of defensive alliances. When $\mathcal{N}$ is unambiguous, the multilayer network is given by the tuple $(\mathbf{G}, \mathbf{H})$. In the initial position it is assumed that $\mathbf{H}$ is non-empty and $\mathbf{G}$ is empty. Subsequently, in a manner to be specified, agents can affect changes in both networks.

### 2.1 The Signed Affinity Network

The agents are assumed to be connected to each other in an initial *signed network* $\mathbf{H} = (h_{ij})$, where $h_{ij} \neq 0$ denotes that agents $i$ and $j$ are linked while $h_{ij} = 0$ indicates that they do not have a link ($i$ and $j$ are *neutral* towards each other). If agents $i$ and $j$ are linked, then $h_{ij} = +1$ denotes that they are *friends*, while $h_{ij} = -1$ denotes that they are *enemies*. In other words, a positive link in $\mathbf{H}$ indicates that the involved agents have an amicable relationship while a negative link reflects their hostility. There are no self loops and thus $h_{ii} = 0$. Let:

$$h_{ij}^{+} = \max\{h_{ij}, 0\}, \quad h_{ij}^{-} = -\min\{h_{ij}, 0\}$$

The number of friends of $i$ in $\mathbf{H}$ is $d_{ij}^{+}(\mathbf{H}) = \sum_{j=1}^{N} h_{ij}^{+}$, and the number of enemies is $d_{ij}^{-}(\mathbf{H}) = \sum_{j=1}^{N} h_{ij}^{-}$. The difference, $\delta_i(\mathbf{H}) \equiv d_{ij}^{+}(\mathbf{H}) - d_{ij}^{-}(\mathbf{H})$, is a measure of *friendship* faced by agent $i$ in $\mathbf{H}$. The *higher* the value of $\delta_i(\mathbf{H})$, i.e., the more net friends agent $i$ has in $\mathbf{H}$, the *lower* the level of hostility faced by agent $i$ in $\mathbf{H}$.

A *complete* network, $\mathbf{H}^c$, is one in which each pair of agents is linked as either friend or enemy; otherwise, the network is *incomplete*. An *empty* network, $\mathbf{H}^e$, is one in which all agents are unlinked and thus neutral. A *path* in $\mathbf{H}$ connecting $i$ and $j$, denoted by $\rho_{ij}(\mathbf{H})$, is a sequence of distinct non-zero links $h_{ii_1}, h_{i_1i_2}, ..., h_{i_{n-1}i_n}, h_{i_nj}$. A network is *connected* if there exists a path between any pair of agents; otherwise, the network is *unconnected*. We will assume that $\mathbf{H}$ is connected. We will let $\rho_{ij}^{+}(\mathbf{H})$ denote a path connecting agents $i$ and $j$ in which all links are between friends, i..e., $h_{ii_1} = h_{i_1i_2} = \cdots = h_{i_{n-1}i_n} = h_{i_nj} = +1$. Any two agents $i$ and $j$ who are not directly linked are *distant friends* if there exists at least one path $\rho_{ij}^{+}(\mathbf{H})$ connecting them; otherwise, they are *distant enemies*, i.e., all paths connecting them involve at least one link among enemies. We will let $\mathcal{F}_{ij}(\mathbf{H}) = +1$ if agents $i$ and $j$ are either friends or distant friends, and $\mathcal{F}_{ij}(\mathbf{H}) = -1$ if they are enemies or distant enemies. Since $\mathbf{H}$ is connected, even if two agents are neutral, they will either be distant friends or enemies. Thus, $\mathcal{F}_{ij}(\mathbf{H}) \in \{-1, +1\}$ for each pair of agents $i, j \in \mathcal{N}$.

A *cluster* in $\mathbf{H}$ is a subnetwork $\mathcal{C} = \left(\mathcal{N}', (h_{ij})_{i,j\in\mathcal{N}'}\right)$ such that $h_{ij} \in \{0, 1\}$ for all $i, j \in \mathcal{N}'$. Thus agents belong to a cluster if they are either mutual friends or neutral. With some abuse of notation, we will let $i \in \mathcal{C}$ denote that agent $i$ belongs to cluster $\mathcal{C}$. Inspired by Davis (1967), $\mathbf{H}$ is *balanced* if it can be partitioned into a set of clusters $(\mathcal{C}_\alpha : \alpha \in \{1, 2, .., M\})$ such that if $i \in \mathcal{C}_\alpha$ and $j \in \mathcal{C}_{\alpha'}$, where $\alpha \neq \alpha'$, then either (i) $h_{ij} = -1$, or (ii) $h_{ij} = 0$ and $\mathcal{F}_{ij}(\mathbf{H}) = -1$. Therefore, $\mathbf{H}$ is balanced if agents can be separated into clusters such that any link that exists between agents of two distinct clusters is of an enemy. Figure 2 illustrates a balanced signed network. We will assume that the base network $\mathbf{H}$ in the initial position is balanced and denote it as $\mathbf{H}_0$.[7] Clusters are defined with respect to this primitive $\mathbf{H}_0$ and are thereafter fixed. We will draw attention to this fact by writing clusters as $\mathcal{C}_\alpha(\mathbf{H}_0)$. Even when the affinity network $\mathbf{H}$ is subsequently transformed with some negative links transformed into positive, the *identity* of agents will continue to be determined by the cluster they belonged to in $\mathbf{H}_0$. Our assumptions with respect to $\mathbf{H}_0$ are:

**Assumption (A.1):** *The base network* $\mathbf{H}_0$ *in the initial position is non-empty, connected, composed of at least two clusters, and balanced.*

[7] This is a realistic description of many real-world situations where the universe of agents belong to distinct factions and there is animosity among factions.

Agents are assumed to belong to different clusters because of discord over what they believe should be the norm. The norm is captured by a scalar taking values over an interval $[\underline{\theta}, \overline{\theta}]$, where $0 \leq \underline{\theta} < \overline{\theta} < \infty$. Agents *within* a cluster $\mathcal{C}_\alpha(\mathbf{H}_0)$, irrespective of whether they are neutral or friends in $\mathbf{H}_0$, subscribe to a common norm $\theta_\alpha(\mathbf{H}_0) \in [\underline{\theta}, \overline{\theta}]$, and $\theta_\alpha(\mathbf{H}_0) \neq \theta_{\alpha'}(\mathbf{H}_0)$ if $\alpha \neq \alpha'$. This norm is assumed immutable and does not change. The greater the difference, $|\theta_\alpha(\mathbf{H}_0) - \theta_{\alpha'}(\mathbf{H}_0)|$, the more agents in clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ differ in terms of core beliefs. We will define for $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ and $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$:

$$\Theta_{ij}(\mathbf{H}_0) = \frac{1}{1 + |\theta_\alpha(\mathbf{H}_0) - \theta_{\alpha'}(\mathbf{H}_0)|} \tag{1}$$

We will use $\Theta_{ij}(\mathbf{H}_0)$ as a *measure of discord* between clusters and suppress reference to $\mathbf{H}_0$ for brevity. If agents $i$ and $j$ belong to the same cluster, then $\Theta_{ij} = 1$ and there is no discord; if they belong to different clusters, then $\Theta_{ij} < 1$. Thus, the greater the dispersion in subscribed norms, the lower the value of $\Theta_{ij}$. Note that the measure of discord is a property of two clusters and not specifically of agents; in other words, for distinct agents $\{i, j, k, l\}$ where $i, k \in \mathcal{C}_\alpha(\mathbf{H}_0)$ and $j, l \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$, we have $\Theta_{ij} = \Theta_{kl}$. Also note that $\Theta_{ij} = \Theta_{ji}$. It is important to note once again that the discord between agents is fixed with respect to their position in $\mathbf{H}_0$. Even if subsequently two agents $i$ and $j$ from different clusters establish a friendly relationship, their mutual discord $\Theta_{ij}$ is not equal to 1, i.e., they are still not in consonance with respect to their respective subscribed norms.

Agents will be permitted to make limited changes to the primitive $\mathbf{H}_0$. A pair of agents $i$ and $j$ can change the relationship from neutral or enemy to friend, by each side incurring a cost that captures the effort required to build the necessary trust. Thus, the formation of a friendly link requires *bilateral* consent of the pair of agents involved. The individual cost to agents $i$ and $j$ of converting $h_{ij} \in \{-1, 0\}$ to $h_{ij} = +1$ is $\tau > 0$.[8] We will let $\mathbf{H} \oplus h_{ij}$ denote the network in which $h_{ij} \in \{-1, 0\}$ is transformed to $h_{ij} = +1$ in $\mathbf{H}$. An agent $i$ can also *unilaterally* dissolve a neutral or friendly relationship with agent $j$ and convert it into a hostile link. We will let $\mathbf{H} \ominus h_{ij}$ denote the network in which $h_{ij} \in \{0, +1\}$ is transformed to $h_{ij} = -1$ in $\mathbf{H}$. Note that any changes to the primitive link structure of $\mathbf{H}_0$ can render it unbalanced.

### 2.2 The Network of Alliances

Agents can form a network of defensive *alliances* which is denoted by $\mathbf{G} = (g_{ij})$, where $g_{ij} = 1$ denotes that agents $i$ and $j$ are allies while $g_{ij} = 0$ denotes they are unallied. An alliance can be any formal agreement between two agents relating to issues such as trade or defense. The formation of an alliance between two agents requires the acquiescence of both. An existing alliance can also be dissolved. We are assuming that the dissolution of an existing alliance can be carried out unilaterally. The *neighborhood* of agent $i$, $\mathbf{N}_i(\mathbf{G}) = \{j \in \mathcal{N}\backslash\{i\} : g_{ij} = 1\}$, is the set of allies of $i$ in $\mathbf{G}$, and $\eta_i(\mathbf{G})$ denotes the cardinality of this set and called $i$'s *degree*.[9] Suppose the distinct positive degrees in $\mathbf{G}$ are $\eta_{(1)} < \eta_{(2)} < \cdots < \eta_{(m)}$

[8] The cost can be expected to be lower if the pair was neutral rather than enemies. However, this complication will not change the qualitative results.

[9] Note by definition that $i \notin \mathbf{N}_i(\mathbf{G})$, i.e., $i$ is not included in its own neighborhood. This is in keeping with the standard formulation in network theory that there are no self-loops and thus $g_{ii} = 0$. Including $i$ in its own neighborhood would not

and let $\eta_{(0)} = 0$ even if there are no isolated agents in $\mathbf{G}$. The *degree partition* of $\mathbf{G}$ is denoted by $\mathcal{D}(\mathbf{G}) = \{D_0(\mathbf{G}), D_1(\mathbf{G}), ..., D_m(\mathbf{G})\}$, where all agents in the element $D_k(\mathbf{G})$ of the partition have the same degree $\eta_{(k)}$, $k \in \{0, 1, ..., m\}$. The definition of path and connectedness are defined analogous to the case of signed networks. A maximally connected subnetwork $\mathbf{G}'$ in $\mathbf{G}$ is called a *component* of $\mathbf{G}$. Given networks $\mathbf{G}$ and $\mathbf{G}'$, we will say that $\mathbf{G}$ is *denser* than $\mathbf{G}'$ if $\mathbf{G}' \subseteq \mathbf{G}$. We will let $\mathbf{G} - ij$ (respectively, $\mathbf{G} + ij$) denote the network obtained from $\mathbf{G}$ by deleting (respectively, adding) the link $ij$.

An important network architecture that we will consider is a *nested split graph*. This network has the property that if $\eta_i(\mathbf{G}) \leq \eta_j(\mathbf{G})$, then $\mathbf{N}_i(\mathbf{G}) \subseteq \mathbf{N}_j(\mathbf{G}) \cup \{j\}$. In other words, the neighborhood of a lower degree agent is contained within the neighborhood of a higher degree agent. Therefore, all allies of a less connected agent are also the allies of a more connected agent. Figure 3 illustrates this class of networks.

### 2.3 Gross Benefits from Alliances

Let $\mathbb{Z}$ denote the set of integers and consider the functions $\nu : \mathbb{Z}_+ \to \mathbb{R}_+$ and $w : \mathbb{Z}_+ \to \mathbb{R}_+$. The function $v$ captures return from own degree while $w$ captures the return from the partner's degree. Suppose agent $i$ with degree $\eta_i$ forms a link with agent $j$ with degree $\eta_j$. The incremental gross benefit to agent $i$ from this link with agent $j$ depends on the degree of both agents involved and is assumed to be given by:

$$[v(\eta_i + 1) - v(\eta_i)] + w(\eta_j + 1) \tag{2}$$

noting that the degree of each agent involved in the link increases by 1. The functions $v$ and $w$ are assumed to satisfiy the following conditions:

**Assumption (A.2):** $v(0) = w(0) = 0$. *Further, for all* $1 \leq \eta < N - 1$,
(**a**) $\nu(\eta + 1) > \nu(\eta)$, $w(\eta + 1) > w(\eta)$
(**b**) $\nu(\eta + 1) - \nu(\eta) > \nu(\eta) - \nu(\eta - 1)$.

The rationale behind these restrictions is as follows. An agent benefits from having more alliances (degree) due to "strength in numbers". Moreover, incremental gross benefits are increasing in own degree because defensive investment in fixed resources do not have to be replicated when additional alliances are formed thus yielding economies of scale. An agent also prefers to link with a partner who has more alliances because it reinforces strength in numbers and also permits indirect access to a larger number of allies. We now substantiate this assumption with some real-world examples. Within the AUKUS (Australia, UK and US) military and technology alliance, each nation is able to share the costs of producing an out-the-door forward deployable nuclear submarine fleet. Australia bears a high marginal share of production costs and base sharing, but receives unparalleled benefits in technology and knowledge transfers that would otherwise be unobtainable. Britain receives economies of scale in its defense industry. The United States gains priority basing rights, and an auxiliary fleet to complement its own, already permanently deployed in

alter any result.

the region dramatically reducing operating and maintenance costs for its own fleet. Furthermore, there are indirect benefits from aligning with a higher degree node in **G**. The AUKUS military alliance came about as Australia was set to join a looser and more transactional military industrial agreement with France. Australia's post-cold war defense strategy concluded that "[Australia is] one of the most secure countries in the world... distant from the main centres of global military confrontation"[10]. Therefore a relatively inexpensive and limited agreement with France suited both. However, when Australia perceived China as a more present threat to Australia's homeland, the country reneged on the agreement with France and opted for AUKUS with the two most central nodes in NATO and the western alliance at approximately five times the monetary cost and incurring significant obligations on its military autonomy and sovereignty. The benefits of joining the tripartite AUKUS with stronger ties between and emanating from each node were, ceteris paribus, significantly greater than a bilateral agreement with France.[11]

**Remark:** (*Separable versus non-separable gross benefits*) We have postulated an additively separable in degrees specification for gross benefits in (2). Such a separable specification allows a transparent exposition of the main results. A non-separable formulation, in which gross benefit to agent $i$ from a link with agent $j$ is more generally specified as $\psi\left(\eta_i, \eta_j\right)$, would also yield the same set of results under appropriate conditions on $\psi$. We demonstrate this in Section 6.3.

### 2.4 Cost of Hostility

Recall that agents can only form an alliance in **G** with those who are friends or distant friends in **H** and that an alliance requires mutual consent. By forming an alliance, an agent incurs a cost of linking which is a function of the hostility faced by the potential partner in **H**. Letting $c : \mathbb{Z} \to \mathbb{R}_+$ denote this linking cost function, we will impose the following assumptions on $c$.

**Assumption (A.3):** *For all* $\delta \in \mathbb{Z}$:
(**a**) $c\left(\delta+1\right) < c\left(\delta\right)$.
(**b**) $c\left(\delta\right) - c\left(\delta+1\right) > c\left(\delta-1\right) - c\left(\delta\right)$.

Therefore, the cost to an agent is lower when the potential partner faces less hostility. Further, the cost reduction realized with a higher friendship partner is greater than with a lower friendship partner. With a link, each agent assumes some of the risks posed by the hostile relationships of the potential partner. These risks are consequently lower if each partner has more friends and less enemies. This also explains

[10] Protection by Projection, *The Economist*, April 25, 2023.

[11] Another example substantiating assumption (A.2) is the Nordic fighter fleet agreements signed in spring 2023. The Nordic nations (Norway, Sweden, Denmark, and Finland), agreed to pool resources creating an integrated air defense. The Scandinavian peninsula is a great example of this effect. Because the shortest distance from Russian air bases to the allied coast is from the north, each nation has a relatively small geographic slice of detection zones, but their entire geography is collectively exposed around the clock. Therefore, the increased number of participants in resource-sharing greatly expands the benefit to each individual member through burden sharing in time, distance and capacity by eliminating duplications in detection, early warning, rotating alert and surge units, and command and control infrastructure. A similar dynamic is also observed in the NATO VJTF, a rotating multinational task force kept on ready alert to spearhead an immediate counter attack to an invasion. All NATO members rotate units through the VJTF to defend the collective's eastern perimeter.

why NATO does not permit admitting nations that are engaged in territorial disputes. A dispute indicates a high level of hostility towards that potential member and, therefore, a risk of conflict that commits the entire alliance. A more detailed look at this is NATO's admission of Finland, before the (presumptive) admission of Sweden. While both Finland and Sweden distrust Russia, only Sweden has some measure of hostile links with Turkey (also a NATO member). Finland, with its lower level of hostility, was therefore prioritized for admission to the military alliance ($\mathbf{G}$) while the hostile link between Sweden and Turkey is being flipped in $\mathbf{H}$.

Each agent will also face some cost due to its own hostile relations. We will let $c_0 : \mathbb{Z} \rightarrow \mathbb{R}_+$ denote the cost to an agent from these hostile relations and assume that $c_0$ also satisfies conditions (a) and (b) of assumption A.3.

**Remark:** (*Costs that are functions of both hostility and degree*) The cost to agent $i$ of linking with agent $j$ is assumed to be $c(\delta_j)$. It can be argued that this cost to agent $i$ could be lower if $j$ had greater degree, i.e., the cost function should be $c(\delta_j, \eta_j)$, with $c(\delta_j, \eta_j + 1) < c(\delta_j, \eta_j)$. Our results would continue to obtain under this specification as well as demonstrated in Section 6.3. We note, however, that this effect is already captured in our basic model if we construe the net cost to agent $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ from linking with agent $j$ as:

$$\widetilde{c}_i(\delta_j, \eta_j) = \begin{cases} c_0(\delta_j) - v(\eta_j), & j = i \\ c(\delta_j) - w(\eta_j), & j \in \mathcal{C}_\alpha(\mathbf{H}_0) \\ c(\delta_j) - \Theta_{ij} w(\eta_j) - \tau h_{ij}^+, & j \notin \mathcal{C}_\alpha(\mathbf{H}_0) \end{cases} \tag{3}$$

Recalling assumption A.2(a), $\widetilde{c}_i(\delta_j, \eta_j)$ is decreasing in $\eta_j$.

Finally, we impose a joint restriction on the cost of forming an alliance in $\mathbf{G}$ and the cost of transforming a relationship in $\mathbf{H}$. This assumption bounds the reduction in alliance costs that can be achieved with agents changing their affinity relationship *within* a cluster from neutral to friendly. The logic is that agents are already friends or distant friends within a cluster. Thus, any reduction in alliance costs attained within a cluster by forming a more direct friendly relation is less than the cost $\tau$ of transforming an affinity relationship. This shifts the impetus of agents to revise links *outside* rather than inside the cluster in an affinity network.

**Assumption (A.4):** *For all* $\delta \in \mathbb{Z}$:

$$c_0(\delta) - c_0(\delta + 1) < \tau$$

Henceforth we will use the following notation to denote a unit increase in degree and friendship:

$$\begin{aligned} \Delta v(\eta) &\equiv v(\eta + 1) - v(\eta) \\ \Delta c(\delta) &\equiv c(\delta) - c(\delta + 1) \\ \Delta c_0(\delta) &\equiv c_0(\delta) - c_0(\delta + 1) \end{aligned}$$

Note that we define $\Delta c(\delta)$ and $\Delta c_0(\delta)$ such that they are positive due to A.3(a).

### 2.5 Payoffs

In contrast to the contest function approach of the traditional literature, we adopt a reduced form additive specification of payoffs that reflect the tradeoffs present in the model. There are essentially four factors at play: (i) the "economies of scale" from allying with those who have high degree in $\mathbf{G}$; (ii) the cost of allying with a partner facing a given hostility in $\mathbf{H}$; (iii) the level of discord from the potential partner; and (iv) whether the potential partner is a (distant) friend or (distant) enemy. Recall that $\mathcal{F}_{ij} = 1$ and $\Theta_{ij} = 1$ for agents $i$ and $j$ in the same cluster. Further, $\mathcal{F}_{ij} = 1$ if $i$ and $j$ belong to different clusters if and only if there is at least one friendly link between the two distinct clusters. If this friendly link was formed by two other distinct agents, then $i$ and $j$ can "free ride" on the path of friendly links connecting them and not incur costs other than those of linking in $\mathbf{G}$; if $i$ and $j$ were involved in changing a neutral or hostile relationship to one of friends in $\mathbf{H}$, then each would respectively incur the cost $\tau > 0$. The payoff of agent $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ is given by:

$$\begin{aligned}\pi_i(\mathbf{G},\mathbf{H}) &= \nu(\eta_i(\mathbf{G})) - c_0(\delta_i(\mathbf{H})) + \sum_{j \in \mathbf{N}_i(\mathbf{G}) \cap \mathcal{C}_\alpha(\mathbf{H}_0)} \left[w\left(\eta_j(\mathbf{G})\right) - c(\delta_j(\mathbf{H}))\right] \\ &\quad + \sum_{j \in \mathbf{N}_i(\mathbf{G}) \cap \mathcal{C}_{\alpha'}(\mathbf{H}_0); \mathcal{F}_{ij}(\mathbf{H})=1} \left[\Theta_{ij} w\left(\eta_j(\mathbf{G})\right) - c(\delta_j(\mathbf{H})) - \tau h_{ij}^{+}\right] \end{aligned} \tag{4}$$

The first term captures the net payoff to agent $i$ from its own degree and own hostility level while the second term captures the net gain from its links in $\mathbf{G}$ with members of the same cluster. The third term captures the net payoff from linking in $\mathbf{G}$ to those in other clusters with whom the relationship in $\mathbf{H}$ has been transformed into one of friends or distant friends. This formulation presumes that a link can only be formed between two agents in the alliance network if they are friends or distant friends. In particular, two agents $i$ and $j$ cannot form an alliance in $\mathbf{G}$ if they continue to be enemies or distant enemies. Further, the gross benefit from an alliance across clusters is scaled down by the level of discord between the clusters.

**Remarks:** We record two facts regarding the payoff specification.
**a.** (*Comparison to payoffs from a standard conflict model*) Given our focus on the structure of alliance networks, we have directly postulated a reduced form payoff function that captures the essential features of conflict. We demonstrate in Section 6.1 that these payoffs can be derived from a conflict model such as a hawk-dove game between agents with a negative link. Thus, there is no loss of generality in working with the specification given by (4).

**b.** (*The free riding assumption*) We have assumed that once a friendly link has been formed to connect two distinct clusters, then other agents in the two clusters can free ride on this friendly link to form alliances spanning the two clusters. Our qualitative results would continue to hold if we rule out free riding and require instead that each pair of agents from the two clusters who form an alliance have to first incur the cost $\tau$ of transforming their relationship of neutrality or enmity. The only difference between the two cases will be in the *density* of alliances that are formed. Under free riding, a friendly relationship that is established between two clusters confers a positive externality on all other agents in the two clusters and, therefore, leads to more alliances being formed than would be the case if all agents had to bear a private cost of reaching out beyond their cluster. Specifically, if $\widetilde{\mathbf{G}}$ (respectively, $\widehat{\mathbf{G}}$) is the stable alliance network under free riding (respectively, without free riding), then $\widehat{\mathbf{G}} \subseteq \widetilde{\mathbf{G}}$.

## 3 Fixed Affinity Network

We begin our analysis with the case of a fixed affinity network $\mathbf{H}_0$. We then examine the implications of this fixed affinity network on the topology of alliances in $\mathbf{G}$. Therefore, we consider a one-way interaction between the affinity and alliance networks. This section can be construed as a *short run* analysis when the horizon is sufficiently small for agents to effect a change in relationships in the affinity network. We are assuming here that relationships (whether friend or enemy) embodied in the network $\mathbf{H}_0$ have taken time to coalesce. Within the time frame of the short run, new relations cannot be established in the affinity network. The incremental utility to agent $i$ from forging an alliance in $\mathbf{G}$ with a member $j$ in its own cluster is given by:

$$\pi_i(\mathbf{G} + g_{ij}, \mathbf{H}_0) - \pi_i(\mathbf{G}, \mathbf{H}_0) = [v(\eta_i(\mathbf{G}) + 1) - v(\eta_i(\mathbf{G}))] + \left[w\left(\eta_j(\mathbf{G}) + 1\right) - c(\delta_j(\mathbf{H}_0))\right]$$

We will use a definition of stability inspired by Jackson and Wolinsky (1996).

**Definition (Pairwise-stability for monolayer networks):** Given $\mathbf{H}_0$, a network $\mathbf{G}^*$ is *pairwise-stable* if:

- No agent $i \in \mathcal{N}$ has an incentive to unilaterally delete an existing link with agent $j$ in $\mathbf{G}^*$, i.e., $\pi_i(\mathbf{G}^*, \mathbf{H}_0) - \pi_i(\mathbf{G}^* - g_{ij}, \mathbf{H}_0) \geq 0$.
- No pair of agents $i, j \in \mathcal{N}$ who are unlinked in $\mathbf{G}^*$ should have an incentive to bilaterally form a link, i.e., $\pi_i(\mathbf{G}^* + g_{ij}, \mathbf{H}_0) - \pi_i(\mathbf{G}^*, \mathbf{H}_0) > 0$ implies $\pi_j(\mathbf{G}^* + g_{ij}, \mathbf{H}_0) - \pi_j(\mathbf{G}^*, \mathbf{H}_0) < 0$.

Note that agents can dissolve alliances unilaterally, but the formation of an alliance requires mutual consent. Further, according to the given definition of pairwise-stability, if one agent has a strict incentive to form a link and the other agent has a weak incentive to reciprocate, then the link will be established.

**Remark:** (*Static versus dynamic link formation games*) As noted in Jackson and Wolinsky (1996), the pairwise stability notion is a relatively weak restriction and not tied to any link formation game. For our main analysis we formulate an explicit *dynamic* alliance formation game which, starting from an empty network $\mathbf{G}^e$, shows the emergence of a pairwise-stable network and its relation to the affinity network $\mathbf{H}_0$. The advantage of such a dynamic formulation is that it outlines in an intuitive step-by-step manner the evolution of a pairwise-stable network and its underlying connection to the affinity network via preferential attachment. An alternative *static* formulation could be provided via a link announcement game (for example, Dutta et al. 1998) in which agents simultaneously announce alliances with other agents and only those alliances that are reciprocated are formed. For the sake of completeness, we analyze such a static game in Section 6.2.

### 3.1 The Basic Link Formation Game

We will examine link formation in $\mathbf{G}$ through a dynamic game inspired by Aumann and Myerson (1988). The advantage of this approach is that it selects one among potentially multiple pairwise stable networks.[12] The agents move sequentially in the order of their index (from lowest to highest) to add new links or delete existing links in $\mathbf{G}$. The game terminates when profitable opportunities to change links are exhausted. We will assume that information is perfect and agents observe all links that are formed or deleted. Thus, the information available to an agent with the move – the *state* of the dynamic game – is described by the 3-tuple of the current architecture of $\mathbf{G}$ on which links are being changed, the (fixed) architecture of $\mathbf{H}_0$, and the identity of the agent with the move (the *active* agent). Agents are myopic and make decisions to change links on a network depending on its current architecture.[13] We characterize this link formation game as *basic* since it only considers the one-way impact of $\mathbf{H}$ on $\mathbf{G}$. The details of this basic game on $\mathbf{G}$ is as follows:

- The game starts from an empty network $\mathbf{G}_0(0) = \mathbf{G}^e$ with agent 1 as the active agent, i.e., the state is $(\mathbf{G}_0(0), \mathbf{H}_0, 1)$. The *action set* of the active agent on network $\mathbf{G}$ comprises of two activities in the following order: deletion of any existing links followed by the possible proposal of a link to another agent $i$, the *passive* agent, who is a friend or distant friend.[14] Since $\mathbf{G}_0(0)$ is empty, agent 1 has no links to delete and can propose a link to an agent $i$. The active agent is assumed to choose an ally generating the highest payoff; if there is more than one yielding the same highest payoff, then the active agent chooses the one with the highest index. The passive agent $i$, given the state $(\mathbf{G}_0(0), \mathbf{H}_0, 1)$, has an action set comprising of two possible actions – accept or decline the proferred link. If $i$ declines, then the network remains empty, $\mathbf{G}_0(1) = \mathbf{G}_0(0) = \mathbf{G}^e$, while if $i$ accepts, then the network becomes $\mathbf{G}_0(1) = \mathbf{G}^e + g_{1i}$. We will say that $\mathbf{G}_0(1)$ is *reachable* from $\mathbf{G}_0(0)$.

- The game moves to the state $(\mathbf{G}_0(1), \mathbf{H}_0, 2)$ with agent 2 as the active agent and then to agents with successive higher indices. After all agents have moved once, the process restarts from state $(\mathbf{G}_0(N), \mathbf{H}_0, 1)$. The sequence of networks, $\{\mathbf{G}_0(\kappa); \kappa \in \mathbb{Z}_+\}$, generated in this manner is called an *improving path*. Along this path, each network is reachable from the preceding one and is the consequence of an active agent deleting unprofitable links and/or forming a mutually profitable link with another agent.

- Following Jackson and Watts (2002, Lemma 1), an improving path can lead to one of two possible outcomes: (i) a *limit* network $\mathbf{G}^*(\mathbf{H}_0)$ such that, for each state $\{(\mathbf{G}^*(\mathbf{H}_0), \mathbf{H}_0, i); i \in \mathcal{N}\}$, no active agent $i$ has an incentive to delete links or form a mutually profitable link with another agent, or (ii) a closed cycle. A *closed cycle* is a set of networks such that there is an improving path between any

[12] Similar link formation games have been employed by Jackson and Watts (2002), Joshi et al (2020) and Joshi et al (2023), to characterize pairwise stable networks.

[13] A myopic formulation is fairly standard in examining dynamic games of network formation due to reasons of tractability (for example, Jackson and Watts 2002). It presumes that agents highly discount the future.

[14] Recall that a friend or distant friend is an agent who belongs to the same cluster in $\mathbf{H}_0$. Since $\mathbf{H}_0$ is assumed fixed in this section, alliances can only be formed among agents within the same cluster.

pair of networks in the set and no improving path leading to a network outside the set. We will show below (Theorem 1) that a closed cycle is not possible in our link formation game. Thus, the only outcome is convergence to a limit network $\mathbf{G}^*(\mathbf{H}_0)$.

**Theorem 1** *The basic link formation game converges to a limit network* $\mathbf{G}^*(\mathbf{H}_0)$ *which is a pairwise-stable network.*

Therefore, Theorem 1 also shows the *existence* of a pairwise-stable network. The proof is based on the fact that no agent has an incentive to delete a link that it formed along an improving path in $\mathbf{G}$. With deletions of links ruled out, cycles cannot emerge along an improving path. Therefore, since the number of network architectures are finite, the link formation game will converge to a pairwise-stable network.

**Remark:** (*Salient features of the basic game*) We note two facts about the dynamic game. *First*, we have the active agent deleting any unprofitable links and then proposing a new link to a passive agent. However, it is immaterial in our framework whether agents first delete links and then form a link, or first form a link and then delete links. This is because the incremental payoff from links that are formed will only increase by virtue of assumption A.2 as the degree of agents increase. Thus, as noted earlier, formed links are never subsequently deleted. *Second*, we allow the active agent to propose at most *one* link to a potential ally. We address in the next subsection the proposal of multiple links by an active agent.

## 3.2 The Pairwise-Stable Architecture of $\mathbf{G}^*(\mathbf{H}_0)$

Since friends and distant friends are contained within a cluster from assumption A.1, all alliances are between members of the same cluster. We will characterize the *intra*-cluster alliances formed by agents in $\mathbf{G}$ given $\mathbf{H}_0$. Consider a given cluster $\mathcal{C}_\alpha(\mathbf{H}_0)$, let $\mathcal{I}_\alpha = \{i_1^\alpha, i_2^\alpha, ..., i_n^\alpha\}$ denote the set of agents arranged in increasing order of their index who belong to this cluster. Let $|\mathcal{C}_\alpha(\mathbf{H}_0)|$ denote the size of this cluster. For ease of exposition, let us assume without loss of generality that:

$$\delta_{i_1^\alpha} \leq \delta_{i_2^\alpha} \leq \delta_{i_3^\alpha} \cdots \leq \delta_{i_n^\alpha} \tag{5}$$

with at least one strictly inequality. After obtaining our characterization result (Proposition 3), we discuss the case where all agents face the same hostility level, i.e., we have equality throughout in (5). Let $\delta_1 < \delta_2 < \cdots < \delta_s$, $s \leq n$, denote the distinct values of friendship measures in the cluster $\mathcal{C}_\alpha(\mathbf{H}_0)$.

**Definition (Friendship class and partition):** A *friendship class*, $\Delta_r^\alpha(\mathbf{H}_0)$, is the set of all agents in the cluster $\mathcal{C}_\alpha(\mathbf{H}_0)$ with friendship measure $\delta_r$:

$$\Delta_r^\alpha(\mathbf{H}_0) = \left\{ i_k^\alpha \in \mathcal{I}_\alpha : \delta_{i_k^\alpha} = \delta_r \right\}$$

The *friendship partition*, $\Delta^{\alpha}\mathcal{C}_{\alpha}(\mathbf{H}_0) = \{\Delta_1^{\alpha}(\mathbf{H}_0), .., \Delta_r^{\alpha}(\mathbf{H}_0), .., \Delta_s^{\alpha}(\mathbf{H}_0)\}$, is the collection of friendship classes in $\mathcal{C}_{\alpha}(\mathbf{H}_0)$.

Friendship classes will play an important role in our characterization result. All agents within the same friendship class face the same level of hostility. Agents belonging to a lower-index friendship class face greater hostility than agents belonging to higher-index friendship class. Thus, for example, $i_i^{\alpha} \in \Delta_1^{\alpha}(\mathbf{H}_0)$ and $i_n^{\alpha} \in \Delta_s^{\alpha}(\mathbf{H}_0)$. Recalling the definition of a degree partition, let $\mathcal{D}^{\alpha}(\mathbf{G}^*) = \{D_0^{\alpha}(\mathbf{G}^*), D_1^{\alpha}(\mathbf{G}^*), ..., D_m^{\alpha}(\mathbf{G}^*)\}$ denote the *degree partition* of agents belonging to $\mathcal{C}_{\alpha}(\mathbf{H}_0)$ in the limit network $\mathbf{G}^* \equiv \mathbf{G}^*(\mathbf{H}_0)$. We will now examine how agents are distributed across this degree partition as a function of their friendship measure, i.e., their hostility level. Specifically, we will connect $\mathcal{D}^{\alpha}(\mathbf{G}^*)$ to the friendship partition $\Delta^{\alpha}\mathcal{C}_{\alpha}(\mathbf{H}_0)$.

We will begin by elaborating on how link formation proceeds according to our dynamic game. Recall that (active) agents proceed in increasing order of their index and can only propose alliances with those who belong to their cluster. Therefore, we can consider how links are formed within any given cluster, say $\mathcal{C}_{\alpha}(\mathbf{H}_0)$. The *first* active agent in $\mathcal{C}_{\alpha}(\mathbf{H}_0)$ to propose an alliance will be $i_1^{\alpha}$. Note that at this stage, say $\mathbf{G}_0(\kappa)$, no alliances have been formed and thus $\eta_{i^{\alpha}}(\mathbf{G}_0(\kappa)) = 0$ for all $i^{\alpha} \in \mathcal{C}_{\alpha}(\mathbf{H}_0)$. Of course, if $i_1^{\alpha}$ is agent 1 in cluster $\mathcal{C}_1(\mathbf{H}_0)$ who initiates the game, then $\mathbf{G}_0(\kappa) = \mathbf{G}^e$. The incremental payoff to agent $i_1^{\alpha}$ from proposing an alliance with agent $i_k^{\alpha}$ is $\Delta\nu(0) + w(1) - c\left(\delta_{i_k^{\alpha}}\right)$. Since $c\left(\delta_{i_n^{\alpha}}\right) \leq c\left(\delta_{i_k^{\alpha}}\right)$ for all $i_k^{\alpha} \in \mathcal{I}_{\alpha} \backslash \{i_1^{\alpha}, i_k^{\alpha}\}$, it follows that the most profitable alliance is with agent $i_n^{\alpha}$. However, if this alliance yields a negative payoff to at least one agent, then the link will not be formed. We can of course have isolated agents in a cluster, i.e., agents who have no alliances. We now offer a sufficient condition under which there will be no isolated agents, i.e., $D_0^{\alpha}(\mathbf{G}^*) = \emptyset$. Note that agents characterized by relatively high levels of hostility (i.e., low $\delta$-values) are unattractive as allies because they expose potential partners to a high cost of defending an alliance. Thus, low $\delta$-value agents are most likely to be isolated. By the same logic, an agent with a low level of hostility is an attractive ally permitting partners to harness economies of scale at a relatively lower cost of defense. Thus, the first step in ensuring that there is no isolated agent is that there exists at least one agent willing to reciprocate an alliance with someone facing the most hostility. The agents facing the highest and lowest hostility in $\mathcal{C}_{\alpha}(\mathbf{H}_0)$ are respectively $i_1^{\alpha} \in \Delta_1^{\alpha}(\mathbf{H}_0)$ and $i_n^{\alpha} \in \Delta_s^{\alpha}(\mathbf{H}_0)$. Therefore, suppose that:

$$\Delta\nu(0) + w(1) \geq c\left(\delta_{i_1^{\alpha}}\right) \tag{6}$$

Since $c\left(\delta_{i_1^{\alpha}}\right) > c\left(\delta_{i_n^{\alpha}}\right)$, it follows from (6), that $i_1^{\alpha}$ has a strict incentive to propose a link to $i_n^{\alpha}$, and $i_n^{\alpha}$ has a weak incentive to reciprocate the link. Therefore, the alliance will be formed, and $g_{i_1^{\alpha}, i_n^{\alpha}} = 1$ will be the first link formed in the cluster $\mathcal{C}_{\alpha}(\mathbf{H}_0)$ and $i_1^{\alpha} \notin D_0^{\alpha}(\mathbf{G}^*)$. The next step is to extend this argument to all agents $i_k^{\alpha} \in \mathcal{I}_{\alpha} \backslash \{i_1^{\alpha}, i_n^{\alpha}\}$ and thus prove that $D_0^{\alpha}(\mathbf{G}^*) = \emptyset$. The following characterization result will be useful in this regard.

**Proposition 1** (***Preferential attachment***) *Suppose agents $i$ and $j$ in the same cluster have a mutually profitable link in a network $\mathbf{G}$. Consider an agent $k \neq i, j$ in the same cluster such that $\delta_k \geq \delta_i$ and $\eta_k(\mathbf{G}') \geq \eta_i(\mathbf{G})$ where $\mathbf{G} \subseteq \mathbf{G}'$. Then agents $k$ and $j$ have a mutually profitable link in $\mathbf{G}'$.*

Proposition 1 states that if an agent, say $j$, has a profitable alliance with some agent $i$, then $j$ will also have mutually profitable links with all agents who have a friendship measure in $\mathbf{H}_0$ and a degree in $\mathbf{G}$ at least as large as that of agent $i$. Further, this incentive is strengthened as the density of $\mathbf{G}$ increases. We will now employ Proposition 1 to show by induction that since all agents $i_k^\alpha \in \mathcal{I}_\alpha \backslash \{i_1^\alpha, i_n^\alpha\}$ face lower hostility than $i_1^\alpha$, they will also have a mutually profitable link with $i_n^\alpha$. Since this is true for agent $i_1^\alpha$, now suppose it is true for agents $\{i_1^\alpha, i_2^\alpha, .., i_k^\alpha\}$. In particular, $i_k^\alpha$ formed the link with $i_n^\alpha$ when it was the active agent in stage $\mathbf{G}_0(\kappa)$. We will show that it is true for agent $i_{k+1}^\alpha$ when it is the active agent in stage $\mathbf{G}_0(\kappa')$, $\kappa' > \kappa$. Note that:

$$\eta_{i_n^\alpha}\left(\mathbf{G}_0\left(\kappa'\right)\right) = k > \eta_{i_1^\alpha}\left(\mathbf{G}_0\left(\kappa'\right)\right) = \cdots = \eta_{i_k^\alpha}\left(\mathbf{G}_0\left(\kappa'\right)\right) = 1 > \eta_{i_{k+1}^\alpha}\left(\mathbf{G}_0\left(\kappa'\right)\right) = \cdots = \eta_{i_{n-1}^\alpha}\left(\mathbf{G}_0\left(\kappa'\right)\right) = 0$$

Therefore, using (5), it follows that:

$$\Delta v(0) + w\left(\eta_{i_n^\alpha}\left(\mathbf{G}_0\left(\kappa'\right)\right) + 1\right) - c\left(\delta_{i_n^\alpha}\right) \geq \Delta v(0) + w\left(\eta_{i_l^\alpha}\left(\mathbf{G}_0\left(\kappa'\right)\right) + 1\right) - c\left(\delta_{i_l^\alpha}\right)$$

for all $i_l^\alpha \in \mathcal{I}_\alpha \backslash \{i_{k+1}^\alpha, i_n^\alpha\}$ from A.2(a) and A.3(a). Thus, agent $i_n^\alpha$ is the most profitable ally for $i_{k+1}^\alpha$. Further, $\eta_{i_k^\alpha}(\mathbf{G}_0(\kappa)) = \eta_{i_{k+1}^\alpha}(\mathbf{G}_0(\kappa')) = 0$ and $i_k^\alpha$ had a mutually profitable link with $i_n^\alpha$ in $\mathbf{G}_0(\kappa)$. It follows from Proposition 1 that $i_{k+1}^\alpha$will also have a mutually profitable link with $i_n^\alpha$ in $\mathbf{G}_0(\kappa')$. We have shown that (6) is sufficient to ensure that the cluster will be connected, i.e., agents participate in at least one alliance and $D_0^\alpha(\mathbf{G}^*) = \emptyset$. Note that this argument also implies that agent $i_n^\alpha$ is the most connected in the cluster, i.e., $i_n^\alpha \in D_m^\alpha(\mathbf{G}^*)$. In fact, we can prove more generally that all agents in $\Delta_1^\alpha(\mathbf{H}_0)$ have the fewest number of alliances in the cluster while those in $\Delta_s^\alpha(\mathbf{H}_0)$ have the most.

**Proposition 2** *Suppose (6) is satisfied. Then* $\Delta_1^\alpha(\mathbf{H}_0) \subseteq D_1^\alpha(\mathbf{G}^*)$ *and* $\Delta_s^\alpha(\mathbf{H}_0) \subseteq D_m^\alpha(\mathbf{G}^*)$.

We now characterize agents facing intermediate levels of hostility. We show that agents facing lower hostility have more alliances. Further, we show that the pattern of alliances assumes the form of a nested split graph. Combining these two results, it follows that the allies of an agent facing less hostility will include the allies of an agent facing greater hostility. This permits us to connect the degree partition $\mathcal{D}^\alpha(\mathbf{G}^*)$ in the alliance network to the friendship partition $\Delta^\alpha \mathcal{C}_\alpha(\mathbf{H}_0)$ in the affinity network.

**Proposition 3** *The degree partition* $\mathcal{D}^\alpha(\mathbf{G}^*)$ *of a cluster* $\mathcal{C}_\alpha(\mathbf{H}_0)$ *exhibits a NSG architecture. In particular, if* $i_k^\alpha, i_l^\alpha \in \mathcal{C}_\alpha(\mathbf{H}_0)$ *such that* $\delta_{i_k^\alpha} \leq \delta_{i_l^\alpha}$, *then* $\mathbf{N}_{i_k^\alpha}(\mathbf{G}^*) \subseteq \mathbf{N}_{i_l^\alpha}(\mathbf{G}^*)$.

Figure 4 illustrates Proposition 3. Note that the agent with the lowest hostility is allied to all, while the agents with the highest hostility are allied only to the lowest hostility agent. In-between, the number of alliances is increasing as hostility decreases. Further, the neighborhoods have a nested structure. The intuition is as follows. An agent with lower hostility poses lower cost in a defensive alliance. Therefore, a potential partner when choosing an ally between two agents with different hostility levels will ceteris paribus choose the one with lower enmity. Thus, at each stage of the link formation game, preferential attachment implies that an agent facing lower hostility will have at least as many alliances as an agent facing greater hostility.

Another way to visualize the intra-cluster NSG architecture is as a "core-periphery" subnetwork composed of a hierarchal order of agents according to their degree. The peripheral agents are those who are connected only to the core agents but not among themselves; the core agents are connected to all other core agents and differ only with respect to the peripheral agents they are connected to. Recall the degree partition $\mathcal{D}^{\alpha}(\mathbf{G}^{*}) = \{D_0^{\alpha}(\mathbf{G}^{*}), D_1^{\alpha}(\mathbf{G}^{*}), ..., D_m^{\alpha}(\mathbf{G}^{*})\}$ within the cluster $\mathcal{C}_{\alpha}(\mathbf{H}_0)$ and let $\lfloor x \rfloor$ denote the largest integer smaller than or equal to $x$. The peripheral agents arranged in increasing number of alliances and their set of allies are as follows:

Table 1: Peripheral Agents in $\mathcal{C}_{\alpha}(\mathbf{H}_0)$

| **Peripheral agents** | **Set of allies** |
|---|---|
| $D_1^{\alpha}(\mathbf{G}^{*})$ | $D_m^{\alpha}(\mathbf{G}^{*})$ |
| $D_2^{\alpha}(\mathbf{G}^{*})$ | $D_m^{\alpha}(\mathbf{G}^{*}) \cup D_{m-1}^{\alpha}(\mathbf{G}^{*})$ |
| $D_3^{\alpha}(\mathbf{G}^{*})$ | $D_m^{\alpha}(\mathbf{G}^{*}) \cup D_{m-1}^{\alpha}(\mathbf{G}^{*}) \cup D_{m-2}^{\alpha}(\mathbf{G}^{*})$ |
| $\cdots$ | $\cdots$ |
| $D_{\lfloor \frac{m}{2} \rfloor}^{\alpha}(\mathbf{G}^{*})$ | $D_m^{\alpha}(\mathbf{G}^{*}) \cup D_{m-1}^{\alpha}(\mathbf{G}^{*}) \cup D_{m-2}^{\alpha}(\mathbf{G}^{*}) \cdot \cdot \cup D_{\lfloor \frac{m}{2} \rfloor+1}^{\alpha}(\mathbf{G}^{*})$ |

The smallest set of peripheral agents are those in $D_1^{\alpha}(\mathbf{G}^{*})$ connected only to the core agents in $D_m^{\alpha}(\mathbf{G}^{*})$ while the largest set of peripheral agents are those in $D_{\lfloor \frac{m}{2} \rfloor}^{\alpha}(\mathbf{G}^{*})$ who are connected to all core agents. The core agents arranged in decreasing number of alliances are as follows:

Table 2: Core Agents in $\mathcal{C}_{\alpha}(\mathbf{H}_0)$

| **Core agents** | **Set of allies** |
|---|---|
| $D_m^{\alpha}(\mathbf{G}^{*})$ | $D_1^{\alpha}(\mathbf{G}^{*}) \cup D_2^{\alpha}(\mathbf{G}^{*}) \cup D_3^{\alpha}(\mathbf{G}^{*}) \cup D_4^{\alpha}(\mathbf{G}^{*}) \cup \cdots \cup D_m^{\alpha}(\mathbf{G}^{*})$ |
| $D_{m-1}^{\alpha}(\mathbf{G}^{*})$ | $D_2^{\alpha}(\mathbf{G}^{*}) \cup D_3^{\alpha}(\mathbf{G}^{*}) \cup D_4^{\alpha}(\mathbf{G}^{*}) \cup \cdots \cup D_m^{\alpha}(\mathbf{G}^{*})$ |
| $D_{m-2}^{\alpha}(\mathbf{G}^{*})$ | $D_3^{\alpha}(\mathbf{G}^{*}) \cup D_4^{\alpha}(\mathbf{G}^{*}) \cup \cdots \cup D_m^{\alpha}(\mathbf{G}^{*})$ |
| $\cdots$ | $\cdots$ |
| $D_{\lfloor \frac{m}{2} \rfloor+1}^{\alpha}(\mathbf{G}^{*})$ | $D_{\lfloor \frac{m}{2} \rfloor}^{\alpha}(\mathbf{G}^{*}) \cup D_{\lfloor \frac{m}{2} \rfloor+1}^{\alpha}(\mathbf{G}^{*}) \cup \cdots \cup D_m^{\alpha}(\mathbf{G}^{*})$ |

The agents in $D_m^{\alpha}(\mathbf{G}^{*})$ are core agents with the largest number of allies and they are connected to all agents – whether peripheral or core – in their cluster. Agents in $D_{\lfloor \frac{m}{2} \rfloor+1}^{\alpha}(\mathbf{G}^{*})$ are core agents with the fewest number of allies and, while being connected to all core agents, are only allied with peripheral agents in the set $D_{\lfloor \frac{m}{2} \rfloor}^{\alpha}(\mathbf{G}^{*})$.

We now identify an interesting *symmetry* property of a friendship class: all agents in the same friendship class have the same allies in $\mathbf{G}^*$. In particular, all of them have the same degree. This is once again a consequence of preferential attachment. Suppose to the contrary that $i, j \in \Delta_r^\alpha(\mathbf{H}_0)$ but $\eta_i(\mathbf{G}^*) > \eta_j(\mathbf{G}^*)$. Let $\mathbf{G}_0(\kappa)$ be the first stage in the link formation game when an agent $k$ forms a link with agent $i$. Thus, in $\mathbf{G}_0(\kappa)$, both $i$ and $j$ have the same degree $\eta'$. Further, by hypothesis, agents $i$ and $k$ have a profitable link and thus:

$$\min\left\{\Delta v\left(\eta'\right) + w\left(\eta_k\left(\mathbf{G}_0(\kappa)\right)+1\right) - c\left(\delta_k\right), \Delta v\left(\eta_k\left(\mathbf{G}_0(\kappa)\right)\right) + w\left(\eta'+1\right) - c\left(\delta_i\right)\right\} \geq 0 \tag{7}$$

Also, by hypothesis, since $k \notin \mathbf{N}_j(\mathbf{G}^*)$, it must be true that:

$$\min\left\{\Delta v\left(\eta_j\left(\mathbf{G}^*\right)\right) + w\left(\eta_k\left(\mathbf{G}^*\right)+1\right) - c\left(\delta_k\right), \Delta v\left(\eta_k\left(\mathbf{G}^*\right)\right) + w\left(\eta_j\left(\mathbf{G}^*\right)+1\right) - c\left(\delta_j\right)\right\} < 0 \tag{8}$$

However, $\eta_j(\mathbf{G}^*) \geq \eta'$, $\eta_k(\mathbf{G}^*) \geq \eta_k(\mathbf{G}_0(\kappa))$, and $\delta_i = \delta_j$. Therefore, the LHS of (8) is at least as great as the that of (7), a contradiction. This establishes the symmetry property.

The symmetry property permits us to answer the question posed at the beginning of this subsection, namely what if (5) holds as an equality throughout in a cluster $\mathcal{C}_\alpha(\mathbf{H}_0)$. In this case, all agents in the cluster belong to the same friendship class. Thus, all agents will have the same allies in $\mathbf{G}^*$. This implies that all agents in the cluster will form a *dominant group* subnetwork in which they are mutually interconnected.

**Remarks:** We now offer two comments, one with respect to the link formation game and the other with respect to the characterization result.

**a.** (*Single versus multiple link proposal*) Our basic game assumes that the active agent proposes at most a *single* link to a passive agent. This allows us to draw out the role of preferential attachment in a transparent manner. Note, however, that we can permit the active agent to propose *multiple* links. This will decrease the number of rounds in the basic game but under our assumptions, particularly (6), will generate the same qualitative results. To illustrate, since agent $i_1^\alpha$ found it profitable to ally with agent $i_n^\alpha \in \Delta_s^\alpha(\mathbf{H}_0)$ in the basic game, in the multiple link case $i_1^\alpha$ will find it profitable to propose links with *all* agents in $\Delta_s^\alpha(\mathbf{H}_0)$ by virtue of Proposition 1. Since (6) holds, each agent in $\Delta_s^\alpha(\mathbf{H}_0)$ will accept $i_1^\alpha$'s proposal. Thus, $\Delta_s^\alpha(\mathbf{H}_0) \subseteq \mathbf{N}_{i_1^\alpha}(\mathbf{G}^*)$ as in the single link case.

**b.** (*Robustness of the NSG characterization*) The pairwise-stability criterion is relatively weak and thus admits a large number of stable networks. The basic game picks one alliance network from this set and shows that it possesses an NSG architecture within each cluster. This begs the question of whether the NSG characterization is a property of all pairwise-stable networks. This is difficult to show since the set of pairwise-stable networks can be large. However, we demonstrate in Section 6.2 that a subset of pairwise-stable networks – the class of *strongly stable* networks – display the NSG property.

## 4 Variable Affinity Network

So far we have kept the affinity network as fixed and examined its influence on the alliance network. However, it is possible that after harnessing sufficient economies of scale from their alliances, highly connected agents in different clusters may look past their differences to transform their hostile or neutral relationship to a friendly one in the affinity network. In other words, changing a relationship in the affinity network is a prelude to forming connections in the alliance network. Thus, if agent $i$ proposes to transform a relationship with agent $j$ in network $\mathbf{H}$, then it is with an eye towards forming a subsequent alliance with $j$ in network $\mathbf{G}$. We now assume that agents have a sufficiently long horizon to commit costly resources to build the necessary trust with an erstwhile enemy. Thus, we now take a *long run* view of alliance formation that accommodates a two-way interaction between the affinity and alliance layers in the multilayer network.

Henceforth, we will denote the incremental utility to $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ from a change in relationship with agent $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$ in $\mathbf{H}$ and the subsequent alliance with $j$ in $\mathbf{G}$ as:

$$\pi_i(\mathbf{G}+g_{ij},\mathbf{H}\oplus h_{ij})-\pi_i(\mathbf{G},\mathbf{H})=[\Delta v(\eta_i(\mathbf{G}))+\Delta c_0(\delta_i(\mathbf{H}))]+\left[\Theta_{ij}w\left(\eta_j(\mathbf{G})+1\right)-c(\delta_j(\mathbf{H})+1)-\tau\right]$$

If agents $k \in \mathcal{C}_\alpha(\mathbf{H}_0)$ and $l \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$ have already transformed their relationship in $\mathbf{H}$, then there exists a path in $\mathbf{H}\oplus h_{kl}$ through which each pair of agents drawn from the two clusters are distant friends. Let $(i,j)\neq(k,l)$ denote that either $i\neq k$, or $j\neq l$, or both. Thus, if any pair $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ and $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$ intend to form an alliance, where $(i,j)\neq(k,l)$, then these agents can form an alliance without having to incur the cost $\tau$ of first revising their affinity relationship. The incremental utility to agent $i$ (and a corresponding expression holds for agent $j$) is given by:

$$\pi_i(\mathbf{G}+g_{ij},\mathbf{H}\oplus h_{kl})-\pi_i(\mathbf{G},\mathbf{H})=\begin{cases}\Delta v(\eta_i(\mathbf{G}))+\left[\Theta_{ij}w\left(\eta_j(\mathbf{G})+1\right)-c(\delta_j(\mathbf{H}))\right], & j\neq l\\ \Delta v(\eta_i(\mathbf{G}))+\left[\Theta_{ij}w\left(\eta_j(\mathbf{G})+1\right)-c(\delta_j(\mathbf{H})+1)\right], & j=l\end{cases}$$

We now define pairwise stability for a multilayer network which is an extension of the monolayer case.

**Definition (Pairwise-stability for multilayer networks):** The multilayer network $(\mathbf{G}^*,\mathbf{H}^*)$ is *pairwise-stable* if:

1. Given the network $\mathbf{H}^*$:

   - No agent $i$ has an incentive to delete an existing link with some agent $j$ in $\mathbf{G}^*$, i.e., $\pi_i(\mathbf{G}^*,\mathbf{H}^*)-\pi_i(\mathbf{G}^*-g_{ij},\mathbf{H}^*)\geq 0$.
   - No pair of agents $i$ and $j$ who are unlinked in $\mathbf{G}^*$ should have an incentive to form a link, i.e., $\pi_i(\mathbf{G}^*+g_{ij},\mathbf{H}^*)-\pi_i(\mathbf{G}^*,\mathbf{H}^*)>0$ implies $\pi_j(\mathbf{G}^*+g_{ij},\mathbf{H}^*)-\pi_j(\mathbf{G}^*,\mathbf{H}^*)<0$.

2. Given the network $\mathbf{G}^*$:

   - No pair of agents $i$ and $j$ for whom $\mathcal{F}_{ij}(\mathbf{H}^*)=-1$ have an incentive to transform $h_{ij}\in\{-1,0\}$ to $h_{ij}=+1$, i.e., $\pi_i(\mathbf{G}^*+g_{ij},\mathbf{H}^*\oplus h_{ij})-\pi_i(\mathbf{G}^*,\mathbf{H}^*)>0$ implies $\pi_j(\mathbf{G}^*+g_{ij},\mathbf{H}^*\oplus h_{ij})-\pi_j(\mathbf{G}^*,\mathbf{H}^*)<0$.

## 4.1 The Augmented Link Formation Game

We now consider an augmented sequential link formation game that accommodates changes in both the affinity and alliance networks.

- Given a non-empty primitive network $\mathbf{H}_0$, link formation starts in the alliance network $\mathbf{G}$ starting from an empty network. The sequential process of link formation on this layer culminates in a limit network that we now denote as $\mathbf{G}_1 \equiv \mathbf{G}(\mathbf{H}_0)$.

- The game now shifts to the network $\mathbf{H}$. Players once again move sequentially in the order of their index starting from the state $\left(\mathbf{H}_0^{(0)} = \mathbf{H}_0, \mathbf{G}(\mathbf{H}_0), 1\right)$. The action set of the active agent in $\mathbf{H}$ is different from that in $\mathbf{G}$. *First*, no links in $\mathbf{H}$ can be deleted. This is in accordance with our assumption that affinity relationships have matured bilaterally over a period of time and thus cannot be expunged unilaterally. *Second*, there is no incentive to convert a friend into an enemy because this makes an agent relatively unattractive as an ally to a potential partner. *Third*, by virtue of assumption (A.4), there is no incentive to revise a relationship within a cluster. An agent $i$ can change an existing neutral relationship with agent $j$ within the cluster to one of direct friends and the consequent increase in the friendship measure bestows a gain in own costs equal to $c_0(\delta_i(\mathbf{H}_0)) - c_0(\delta_i(\mathbf{H}_0) + 1) < \tau$. This is consonant with our formulation that any transformation of affinity links is a precursor to forging alliances in the alliance network, and two agents within the same cluster do not have to resort to this intermediate step in order to connect in $\mathbf{G}$. Therefore, the only choice we allow an active agent is to commit resources to convert a hostile or neutral relation *outside* the cluster into a friendly one.

- Suppose agent $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ is the active agent. The active agent $i$ can propose to an agent $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$, $\alpha \neq \alpha'$, with whom $h_{ij} \in \{-1, 0\}$ to change the relationship to a friend (i.e., to $h_{ij} = +1$). Note that if another agent $k \in \mathcal{C}_\alpha(\mathbf{H}_0)$ had prior to $i$'s move already established a friendly relation with some agent, say $l$, in $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$, then $i$ has no incentive to make an overture to $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$. This is because a friendly path between clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ has already been created in the affinity network through $h_{kl} = +1$. Thus, agent $i$ can free ride on this link to form alliances in $\mathbf{G}$ with members of $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ without having to first transform an affinity link with $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$.

- Suppose, therefore, that when agent $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ is the active agent and proposes to agent $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$, then there is no friendly path connectiong clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$. This new relationship in the affinity network imposes a cost of $\tau > 0$ but increases $i$'s friendship measure (lowers hostility level) to $\delta_i(\mathbf{H}_0) + 1$. This increase in the friendship measure confers two benefits to agent $i$. *First*, by virtue of assumption A.3(a), it decreases $i$'s own costs:

$$\Delta c_0(\delta_i(\mathbf{H}_0)) = c_0(\delta_i(\mathbf{H}_0)) - c_0(\delta_i(\mathbf{H}_0) + 1) > 0 \tag{9}$$

  *Second*, by creating a bridge to an agent $j$ in another cluster, it permits the formation of an alliance with $j$ in $\mathbf{G}$ with net benefit:

$$\Delta\nu(\eta_i(\mathbf{G}_1)) + \left[\Theta_{ij} w\left(\eta_j(\mathbf{G}_1) + 1\right) - c(\delta_j(\mathbf{H}_0) + 1) - \tau\right] \tag{10}$$

The active agent $i$ chooses a potential partner $j$ from another cluster with whom the sum of (9) and (10) is positive and maximum. If no such agent in another cluster exists, then the active agent does not propose any revision of links in the affinity network.

- The action set of the passive agent $j$ is to accept or reject the overture to transform the link in the affinity network. Agent $j$ will reciprocate if its incremental payoff is non-negative.

$$\left[\Delta\nu\left(\eta_j\left(\mathbf{G}_1\right)\right)+\Delta c_0\left(\delta_j\left(\mathbf{H}_0\right)+1\right)\right]+\left[\Theta_{ij}w\left(\eta_i\left(\mathbf{G}_1\right)+1\right)-c\left(\delta_i\left(\mathbf{H}_0\right)+1\right)-\tau\right]\geq 0$$

- If the passive agent rejects, then the relationship continues to be hostile, while if the passive agent accepts, then the relationship becomes friendly. This change in relationship is observed by all agents.

- The game continues sequentially with the next active agent and continues until no pair of agents have no incentive to revise affinity relationships. Let $\mathbf{H}_1 \equiv \mathbf{H}\left(\mathbf{G}_1\right)$ denote the limit network. In $\mathbf{H}_1$, for each state $\{(\mathbf{H}(\mathbf{G}_1), \mathbf{G}_1, i)\,; i \in \mathcal{N}\}$, no active agent $i$ and passive agent $j$ have an incentive to transform a relationship from enemy or neutral to friend. Note that cycles are ruled out because, having transformed a hostile or neutral relationship into friendly by committing resources and realizing the attendant benefits, there is no incentive to retransform the relationship back into one of hostility.

- Link formation now moves back to the alliance network. Players move sequentially in the order of their index starting from the state $(\mathbf{G}_1(0), \mathbf{H}_1, 1)$. Suppose $i$ is the active agent. If the active agent $i$ had transformed relationships in $\mathbf{H}_1$ with agents $\{j_1, j_2, .., j_l\}$, then $i$ will first propose an alliance with $j_s \in \{j_1, j_2, .., j_l\}$ with whom (9) plus (10) is the maximum (with ties broken by choosing the passive agent with the highest index) and which $j_s$ will acquiesce.[15] In subsequent stages of link formation, when $i$ is once again the active agent, then it will choose an ally from $\{j_1, j_2, .., j_l\}\backslash\{j_s\}$ choosing the same principle. If $i$ has exhausted all new alliances with members from $\{j_1, j_2, .., j_l\}$, or $i$ had not been involved in transforming any affinity relations in $\mathbf{H}_1$, then $i$ will choose an ally from outside or within own cluster from whom it derives the highest payoff. The passive player will accept or reject the overture leading to the next active agent until all profitable opportunities for alliances are exhausted.

- The process continues to alternate between the two networks $\mathbf{G}_r \equiv \mathbf{G}\left(\mathbf{H}_{r-1}\right)$ and $\mathbf{H}_r \equiv \mathbf{H}\left(\mathbf{G}_r\right)$, $r \geq 2$, in the manner described above. We will prove in Theorem 2 that the process terminates in a set of networks $\left(\mathbf{G}^* \equiv \mathbf{G}\left(\mathbf{H}^*\right), \mathbf{H}^* \equiv \mathbf{H}\left(\mathbf{G}^*\right)\right)$ such that there are no profitable opportunities to amend the architecture on either network. Specifically, given $\mathbf{H}^*$ there are no mutually profitable links to form or unprofitable links to delete in network $\mathbf{G}^*$, and given $\mathbf{G}^*$ there is no hostile or neutral relationship in network $\mathbf{H}^*$ that two agents would like to transform into a friendly one. In other words, it converges to a pairwise-stable multilayer network. The logic follows the argument of Theorem 1 and we now record this formally. It establishes the existence of a pairwise-stable multilayer network.

[15] Recall that agent $i$ had chosen $j_s$ to transform their affinity link because $j_s$ would yield the highest positive payoff from an alliance in $\mathbf{G}$. In turn, $j_s$ had accepted to transform the affinity relationship because the alliance with $i$ yielded a non-negative payoff from an alliance in $\mathbf{G}$. Thus, when $i$ and $j_s$ have the opportunity to cement this alliance during link formation in $\mathbf{G}$, then they will follow through on this alliance.

**Theorem 2** *The augmented link formation game converges to a limit* $(\mathbf{G}^* \equiv \mathbf{G}(\mathbf{H}^*), \mathbf{H}^* \equiv \mathbf{H}(\mathbf{G}^*))$ *which is pairwise-stable.*

### 4.2 The Pairwise-Stable multilayer Network

We now characterize the pairwise-stable multilayer network $(\mathbf{G}^*, \mathbf{H}^*)$. Consider the augmented link transformation game when it moves from layer $\mathbf{G}_1$ to layer $\mathbf{H}_1$. Consider any two clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$. Let $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ and $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$ be the most connected agents in their respective clusters (with the highest index agent chosen in case of a tie). Note from Proposition 2 that degree correlates positively with friendship, and thus these two agents are also the ones facing the lowest hostility in their respective clusters. Thus, as the following lemma indicates, these agents are the most likely candidates to transform their relationship to a friendly one since their realize the highest incremental utilities within their cluster from such a transformation in $\mathbf{H}$ and a subsequent alliance in $\mathbf{G}$. Let $\mathcal{D}^\alpha(\mathbf{G}_1) = \{D_0^\alpha(\mathbf{G}_1), D_1^\alpha(\mathbf{G}_1), ..., D_{m(\alpha)}^\alpha(\mathbf{G}_1)\}$ denote the degree partition of agents belonging to $\mathcal{C}_\alpha(\mathbf{H}_0)$ in the network $\mathbf{G}_1$ and define $\mathcal{D}^{\alpha'}(\mathbf{G}_1)$ analogously. Also, let $\Delta_{s(\alpha)}^\alpha(\mathbf{H}_0)$ (respectively, $\Delta_{s(\alpha')}^{\alpha'}(\mathbf{H}_0)$) denote the highest friendship class in $\mathcal{C}_\alpha(\mathbf{H}_0)$ (respectively, $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$). From Proposition 2 we know that $\Delta_{s(\alpha)}^\alpha(\mathbf{H}_0) \subseteq D_{m(\alpha)}^\alpha(\mathbf{G}_1)$ and $\Delta_{s(\alpha')}^{\alpha'}(\mathbf{H}_0) \subseteq D_{m(\alpha')}^{\alpha'}(\mathbf{G}_1)$.

**Lemma 1** *Consider any two clusters* $\mathcal{C}_\alpha(\mathbf{H}_0)$ *and* $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ *and let* $i \in \mathcal{C}_\alpha(\mathbf{H}_0) \cap \Delta_s^\alpha(\mathbf{H}_0)$ *and* $j \in \mathcal{C}_{\alpha'}(\mathbf{H}_0) \cap \Delta_s^{\alpha'}(\mathbf{H}_0)$*. For any* $k \in \mathcal{C}_\alpha(\mathbf{H}_0)$ *and* $l \in \mathcal{C}_{\alpha'}(\mathbf{H}_0)$ *:*

$$\begin{aligned} \pi_k(\mathbf{G}_1 + g_{kl}, \mathbf{H}_0 \oplus h_{kl}) - \pi_k(\mathbf{G}_1, \mathbf{H}_0) &\leq \pi_i(\mathbf{G}_1 + g_{ij}, \mathbf{H}_0 \oplus h_{ij}) - \pi_i(\mathbf{G}_1, \mathbf{H}_0) \\ \pi_l(\mathbf{G}_1 + g_{kl}, \mathbf{H}_0 \oplus h_{kl}) - \pi_l(\mathbf{G}_1, \mathbf{H}_0) &\leq \pi_j(\mathbf{G}_1 + g_{ij}, \mathbf{H}_0 \oplus h_{ij}) - \pi_j(\mathbf{G}_1, \mathbf{H}_0) \end{aligned}$$

Let $\overline{\eta}^\alpha$ and $\overline{\eta}^{\alpha'}$ denote the respective degrees in $\mathbf{G}_1$ of the maximally connected agents belonging to $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$, and $\delta_i(\mathbf{H}_0) \equiv \overline{\delta}^\alpha$ and $\delta_j(\mathbf{H}_0) \equiv \overline{\delta}^{\alpha'}$ denote their respective friendship levels. Further, let:

$$\Omega_\alpha(\mathbf{G}_1, \mathbf{H}_0) \equiv \left[\Delta v(\overline{\eta}^\alpha) + \Delta c_0\left(\overline{\delta}^\alpha\right)\right] \tag{11}$$

$$\Omega_{\alpha'}(\mathbf{G}_1, \mathbf{H}_0) \equiv \left[\Delta v\left(\overline{\eta}^{\alpha'}\right) + \Delta c_0\left(\overline{\delta}^{\alpha'}\right)\right] \tag{12}$$

Then the maximally connected agent in $\mathcal{C}_\alpha(\mathbf{H}_0)$ will propose to transform a neutral or hostile relationship with the maximally connected agent in $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ if:

$$\Omega_\alpha(\mathbf{G}_1, \mathbf{H}_0) + \Theta_{ij} w\left(\overline{\eta}^{\alpha'} + 1\right) - c\left(\overline{\delta}^{\alpha'} + 1\right) - \tau > 0 \tag{13}$$

and the other agent will reciprocate if:

$$\Omega_{\alpha'}(\mathbf{G}_1, \mathbf{H}_0) + \Theta_{ij} w(\overline{\eta}^\alpha + 1) - c\left(\overline{\delta}^\alpha + 1\right) - \tau \geq 0 \tag{14}$$

Now, recalling that $w(\eta+1)>0$ for all $\eta\geq 0$ by assumption A.2, let us define:

$$\Theta_{\alpha\alpha'}(\mathbf{G}_1,\mathbf{H}_0)\equiv\min\left\{\frac{\tau-\Omega_{\alpha}(\mathbf{G}_1,\mathbf{H}_0)+c\left(\overline{\delta}^{\alpha'}+1\right)}{w\left(\overline{\eta}^{\alpha'}+1\right)},\frac{\tau-\Omega_{\alpha'}(\mathbf{G}_1,\mathbf{H}_0)+c\left(\overline{\delta}^{\alpha}+1\right)}{w\left(\overline{\eta}^{\alpha}+1\right)}\right\} \tag{15}$$

$\Theta_{\alpha\alpha'}(\mathbf{G}_1,\mathbf{H}_0)$ is the threshold value of discord at which at least one agent, given their current degree in $\mathbf{G}_1$ and friendship in $\mathbf{H}_0$, is indifferent towards transforming a link in $\mathbf{H}_0$. An examination of (15) shows that, ceteris paribus, two agents have an incentive to transform their relationship if its cost $\tau$ is low, the respective hostility levels they face is low (i..e, their $\delta$-values are high), and link formation in $\mathbf{G}_1$ has conferred a high enough degree on each to make it attractive to overcome any hurdle posed by their mutual discord. If $\Theta_{ij}=\Theta_{ji}<\Theta_{\alpha\alpha'}(\mathbf{G}_1,\mathbf{H}_0)$, then at least one agent will get a negative payoff from revising their affinity relationship and will either not make such an overture (if it is the active agent) or will reject the overture (if it is the passive agent). From Lemma 1, this is also true for all pairs of agents drawn from the two clusters. Thus the existing affinity relationships in $\mathbf{H}_0$ between the two clusters will continue to remain hostile. Recalling (1), we have the following result:

**Proposition 4** *Consider the affinity network* $\mathbf{H}_0$ *and suppose that for each pair of clusters* $\mathcal{C}_{\alpha}(\mathbf{H}_0)$ *and* $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ *the divergence in their core norms satisfies*:

$$|\theta_{\alpha}(\mathbf{H}_0)-\theta_{\alpha'}(\mathbf{H}_0)|>\frac{1}{\Theta_{\alpha\alpha'}(\mathbf{G}_1,\mathbf{H}_0)}-1$$

*Then* $(\mathbf{G}^*=\mathbf{G}_1,\mathbf{H}^*=\mathbf{H}_0)$ *is the pairwise-stable multilayer network.*

If the dissonance in core beliefs is sufficiently large between each pair of clusters, then agents within a cluster have no incentive to change their cross-cluster affinity relationships in $\mathbf{H}_0$. Thus the architecture of $\mathbf{H}_0$ remains unchanged. Consequently, the friendship levels of agents continue to remain the same as in $\mathbf{H}_0$. When link formation returns to the alliance layer, then the strategic incentives to form alliances remain the same as when link formation first started in $\mathbf{G}$. Since all profitable opportunities to form alliances had already been exhausted in $\mathbf{G}_1$, the architecture of the alliance network remains unchanged from $\mathbf{G}_1$. Thus, all alliances that are forged continue to be within clusters and we do not observe any alliances spanning disparate clusters. Despite the high degrees of potential partners in $\mathbf{G}_1$, and the accompanying economies of scale, all clusters continue to remain hostile and isolated in both the affinity and alliance layers.

Once again consider $i\in\mathcal{C}_{\alpha}(\mathbf{H}_0)$ and $j\in\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ who are maximally connected in $\mathbf{G}_1$ within their respective clusters. Now suppose $\Theta_{ij}=\Theta_{ji}>\Theta_{\alpha\alpha'}(\mathbf{G}_1,\mathbf{H}_0)$, i.e.,

$$|\theta_{\alpha}(\mathbf{H}_0)-\theta_{\alpha'}(\mathbf{H}_0)|<\frac{1}{\Theta_{\alpha\alpha'}(\mathbf{G}_1,\mathbf{H}_0)}-1$$

Then, because the difference in their norms is relatively small, the two agents $i$ and $j$ have an incentive to transform their affinity relationship. Therefore, there exists at least one agent pair in the two clusters who will effect a change in their affinity relationship. Recall that *at most* one link between two clusters will be transformed into a positive one, since other agents in the two clusters can free ride on this "friendly" link to connect to others in the opposite cluster. Therefore, the pair transforming their link generate positive externalities for all other agents in the two clusters. Note that $\mathbf{H}_1 \neq \mathbf{H}_0$ because at least one neutral or hostile link in $\mathbf{H}_0$ has been transformed to a friendly one. Since this transformation is predicated on the mutual profitability of an alliance in $\mathbf{G}$, it follows that $\mathbf{G}_2 \neq \mathbf{G}_1$. Note an important difference now from link formation in the very first iteration on $\mathbf{G}_0$. In $\mathbf{G}_0$, links could only be proposed to an agent *within* the cluster; in $\mathbf{G}_1$, an active agent can now propose links to agent *outside* the cluster as well who are distant friends thus leading to cross-cluster alliances in $\mathbf{G}_2$. Since degrees and friendships have now changed, accounting for these changes in the expressions for (11), (12) and (15) generates a new threshold value of $\Theta_{\alpha\alpha'}(\mathbf{G}_2, \mathbf{H}_1)$. As before, if the difference in norms exceeds $\frac{1}{\Theta_{\alpha\alpha'}(\mathbf{G}_2,\mathbf{H}_1)} - 1$ for two separate clusters in $\mathbf{H}_1$, then they will continue to remain disconnected in the affinity network; if the inequality is strict in the reverse direction, then at least one agent pair from the two clusters will change their relationship to a positive one thus engendering further changes in both the affinity and alliance networks. This process converges to a limit pairwise-stable multilayer network $(\mathbf{G}^*, \mathbf{H}^*)$ and we now turn to its characterization.

We begin with $\mathbf{G}^*$. Recall that in $\mathbf{G}_1$ all alliances were formed *within* a cluster that assumed an NSG structure. Starting from $\mathbf{G}_2$, agents can also form alliances *outside* the cluster with distant friends. We will thus be interested in the architecture of both *intra-cluster* alliances as well as *inter-cluster* alliances in $\mathbf{G}^*$. With respect to the latter, we will need the following definition.

**Definition (Clique):** A set of agents $\{i_1, i_2, .., i_p\}$ and $\{j_1, j_2, .., j_r\}$ from respectively two distinct clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ form an *inter-cluster clique*, or simply an $(\alpha, \alpha')$-*clique*, in $\mathbf{G}$ if $g_{ij} = 1$ for each $i \in \{i_1, i_2, .., i_p\}$ and $j \in \{j_1, j_2, .., j_r\}$. In a clique, all agents are mutually connected in the alliance network.

**Proposition 5** *Consider the limit network $\mathbf{G}^*$ in the pairwise-stable multilayer network $(\mathbf{G}^*, \mathbf{H}^*)$.*
(**a**) *The* ***intra-cluster*** *degree partition in $\mathbf{G}^*$ is a nested split graph.*
(**b**) *Let $\Delta_l^\alpha(\mathbf{H}^*)$ and $\Delta_{l'}^{\alpha'}(\mathbf{H}^*)$ denote the lowest indexed friendship classes in distinct clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ such that $g_{ij} = 1$ for $i \in \Delta_l^\alpha(\mathbf{H}^*)$ and $j \in \Delta_{l'}^{\alpha'}(\mathbf{H}^*)$. Then the set of agents $\Delta_l^\alpha(\mathbf{H}^*) \cup \Delta_{l+1}^\alpha(\mathbf{H}^*) \cup \cdot\cdot \cup \Delta_{s(\alpha)}^\alpha(\mathbf{H}^*)$ in $\mathcal{C}_\alpha(\mathbf{H}_0)$ and the set of agents $\Delta_{l'}^{\alpha'}(\mathbf{H}^*) \cup \Delta_{l'+1}^{\alpha'}(\mathbf{H}^*) \cup \cdot\cdot \cup \Delta_{s(\alpha')}^{\alpha'}(\mathbf{H}^*)$ in $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ form an* ***inter-cluster*** $(\alpha, \alpha')$-clique *in $\mathbf{G}^*$.*

With regard to part (a) of Proposition 5, we know that intra-cluster alliances in $\mathbf{G}_1$ had an NSG topology. We have also seen that in each subsequent iteration $\mathbf{G}_{r+1}$ first order dominates $\mathbf{G}_r$. An important aspect of this first order domination is that if an agent $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ belonging to a friendship class, say $\Delta_l^\alpha(\mathbf{H}^*)$, experiences a strict increase in degree (during the transition from $\mathbf{G}_r$ to $\mathbf{G}_{r+1}$) due to new alliances with a set of agents $\{i_1, i_2, .., i_s\} \subset \mathcal{C}_\alpha(\mathbf{H}_0)$, then all agents in the same friendship class $\Delta_l^\alpha(\mathbf{H}^*)$ will also form mutually profitable alliances with the set $\{i_1, i_2, .., i_s\}$. Thus, the nested structure of neighborhoods is inherited in each successive iteration of the alliance network and, therefore, by the limit $\mathbf{G}^*$.

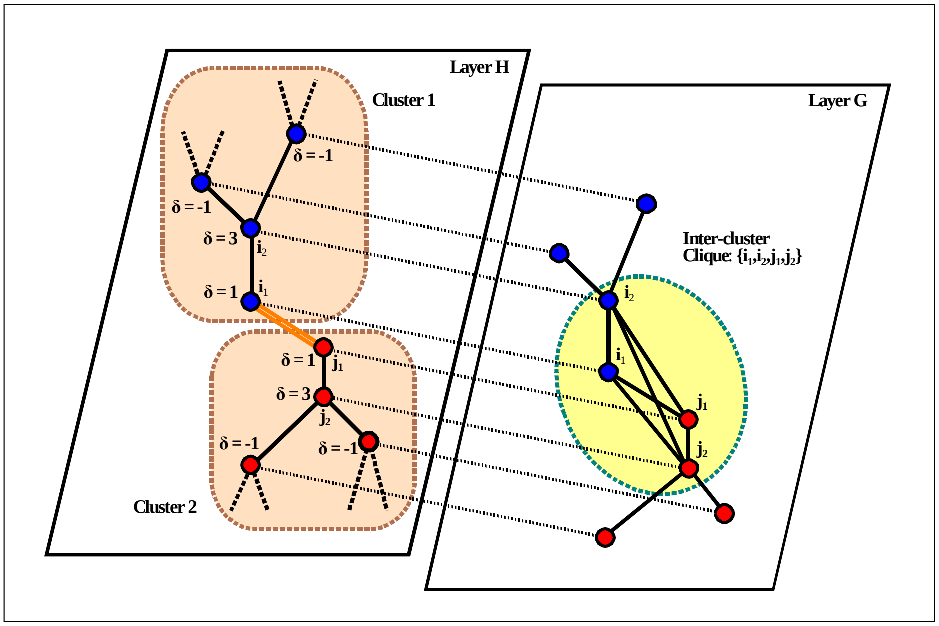


Figure 5: Inter-Cluster Clique in the Alliance Network

Part (b) of Proposition 5 states that if two agents from different clusters have formed an alliance, then all agents in these two clusters with greater degree and greater friendship measure will end up interconnecting with each other. The incentives can best be explained with reference to Figure 5. Agent $i_1$ in cluster 1 and agent $j_1$ in cluster 2 have transformed their neutral relationship to friendly in the affinity network which is indicated by the double line connecting the two agents. Note that $\delta_{i_1} = \delta_{j_1} = 1$ and $\eta_{i_1} = \eta_{j_1} = 1$ prior to transforming their relationship. Suppose $i_1$ was the active agent and $j_1$ was the passive agent when this relationship was transformed. Therefore, $i_1$'s incremental payoff is:

$$[\Delta v(1) + \Delta c_0(1)] + [\Theta_{12} w(2) - c(1)] - \tau \tag{16}$$

An identical expression holds for agent $j_1$. Now consider agent $i_2$ in cluster 1 who belongs to a higher friendship class than $i_1$ and also has greater degree. Then $i_2$ will also have a mutually profitable link with $j_1$ in the alliance network. The incremental payoff to $i_2$ from forming an alliance with $j_1$ is:

$$\Delta v(3) + [\Theta_{12} w(3) - c(2)] \tag{17}$$

Agent $i_2$ does not incur the cost $\tau$ since it can free ride on the friendship link between $i_1$ and $j_1$.[16] We can now compare term-wise the incremental payoffs of $i_1$ and $i_2$. Note that $\Delta v(3) > \Delta v(1)$ from A.2(b), and $\Theta_{12} w(3) - c(2) > \Theta_{12} w(2) - c(1)$ by virtue of A.2(a) and A.3(a). Further, since $\Delta c_0(1) < \tau$ from (A.4), it follows that (17) strictly exceeds (16). An identical argument establishes that $j_1$ will reciprocate the link with $i_2$, and that mutually profitable links will also form between $i_1$ and $j_2$, and between $i_2$ and $j_2$. Hence,

[16] Even if $i_2$ had to incur the cost $\tau$, it would still form an alliance with $j_1$ because its incremental payoff given by (17) is greater than (16) corresponding to $i_1$. The possibility of free riding only strengthens this incentive.

the transformed link $h_{i_1 j_1} = +1$ in the affinity network spurs the creation of an inter-cluster clique in the alliance network composed of $\{i_1, i_2, j_1, j_2\}$. More generally, suppose $\underline{\delta}_\alpha$ (respectively, $\underline{\delta}_{\alpha'}$) be the agent with the lowest hostility level in $\mathcal{C}_\alpha(\mathbf{H}_0)$ (respectively, $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$) who are willing to form an alliance with each other. Then all agents in $\mathcal{C}_\alpha(\mathbf{H}_0)$ with $\delta$-value exceeding $\underline{\delta}_\alpha$, and all agents in $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ with $\delta$-value exceeding $\underline{\delta}_{\alpha'}$, will also have a mutually profitable alliance. Thus, we have an $(\alpha, \alpha')$-*clique* forming across two distinct clusters.

We now turn to the characterization of $\mathbf{H}^*$. For each pair of clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$, define $\Theta_{\alpha\alpha'}(\mathbf{G}^*, \mathbf{H}^*)$ as in (15) but with respect to $(\mathbf{G}^*, \mathbf{H}^*)$. We will let $\mathbf{H}^* \backslash \mathbf{H}_0$ denote the new friendly relations that have been created and which did not exist in $\mathbf{H}_0$.

**Proposition 6** *Consider the limit network* $\mathbf{H}^*$ *in the pairwise-stable multilayer network* $(\mathbf{G}^*, \mathbf{H}^*)$.
(**a**) *Two clusters* $\mathcal{C}_\alpha(\mathbf{H}_0)$ *and* $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ *are connected via a transformed friendly link in* $\mathbf{H}^* \backslash \mathbf{H}_0$ *if*:

$$|\theta_\alpha(\mathbf{H}_0) - \theta_{\alpha'}(\mathbf{H}_0)| < \frac{1}{\Theta_{\alpha\alpha'}(\mathbf{G}^*, \mathbf{H}^*)} - 1 \tag{18}$$

(**b**) *Suppose two clusters* $\mathcal{C}_\alpha(\mathbf{H}_0)$ *and* $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ *have a mutually friendly link in* $\mathbf{H}^* \backslash \mathbf{H}_0$ *and that (without loss of generality)* $\theta_\alpha(\mathbf{H}_0) < \theta_{\alpha'}(\mathbf{H}_0)$. *If there exists an "intermediate" cluster* $\mathcal{C}_{\alpha''}(\mathbf{H}_0)$ *such that*:

$$\theta_\alpha(\mathbf{H}_0) < \theta_{\alpha''}(\mathbf{H}_0) < \theta_{\alpha'}(\mathbf{H}_0) \tag{19}$$

*then both* $\mathcal{C}_\alpha(\mathbf{H}_0)$ *and* $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ *have a friendly link with* $\mathcal{C}_{\alpha''}(\mathbf{H}_0)$ *in* $\mathbf{H}^* \backslash \mathbf{H}_0$.
(**c**) *Two clusters* $\mathcal{C}_\alpha(\mathbf{H}_0)$ *and* $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ *are disconnected in* $\mathbf{H}^*$ *if*:

$$|\theta_\alpha(\mathbf{H}_0) - \theta_{\alpha'}(\mathbf{H}_0)| > \frac{1}{\Theta_{\alpha\alpha'}(\mathbf{G}^*, \mathbf{H}^*)} - 1 \tag{20}$$

Parts (a) and (c) of Proposition 6 formally reiterates the result that agents in two disparate clusters will be involved in an alliance if their norms are sufficiently close and remain separate if their norms are sufficiently divergent. Part (c) draws out the role of "*bridges*" in the affinity network. A cluster $\mathcal{C}_{\alpha''}(\mathbf{H}_0)$ whose norm is intermediate between two clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$ may form a bridge between these two "extremes" that would otherwise not have an incentive to connect given their level of discord.

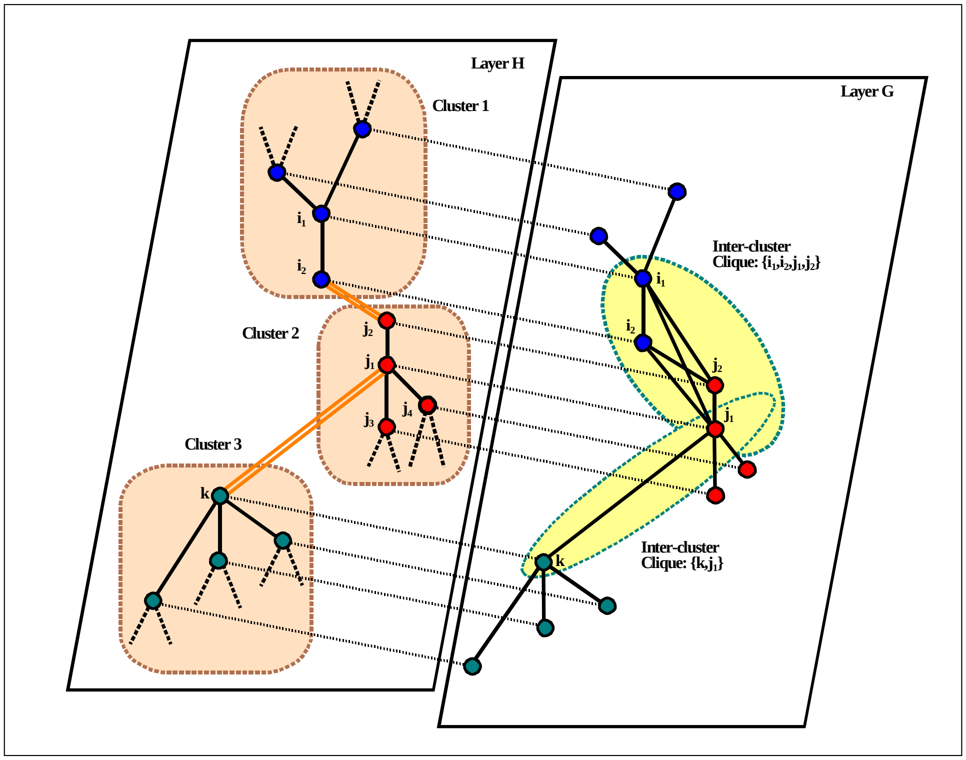


Figure 6: Overlapping Cliques in the Alliance Network

We now draw out the role of *bridge agents* who facilitate alliances across clusters. Consider Figure 6 which assumes that the relationship given by (18) holds between clusters 1 and 2, and between clusters 2 and 3. Further, cluster 2 is intermediate between the other two in the sense of (19). Finally, the relationship between clusters 1 and 3 is characterized by (20). The transformed relationship $h_{i_2 j_2} = +1$ (shown by the double line) in the affinity network connects clusters 1 and 2 and prompts the creation of the inter-cluster clique $\{i_1, i_2, j_1, j_2\}$ with agents $i_1$ and $j_1$ free riding on the friendly path created between the two clusters by $h_{i_2 j_2} = +1$. Likewise, the transformed relationship $h_{j_1 k} = +1$ precipitates the creation of the inter-cluster clique $\{k, j_1\}$; in this particular example, we are assuming that the degree and friendship measure of $j_2$ is not sufficient for an alliance with $k$ and the inclusion of $j_2$ in the clique. Also, despite a friendly path now existing between clusters 1 and 3, the divergence in their core beliefs dissuades agents $k$ and $i_1$ from forming a link in the alliance network. Thus the two inter-cluster cliques overlap with agent $j_1$ at its intersection. Also note that in the affinity layer, there is now a friendly path created between the three clusters with agents $j_1$ and $j_2$ serving as the bridge between clusters 1 and 3. Therefore, it is the sufficiently well-connected and low hostility level agents such as $j_1$ and $j_2$ in the intermediate clusters who serve as conduits – or bridge agents – connecting agents from more extreme clusters. Thus, a bridge agent connecting two extreme clusters is characterized by three features: (i) a norm that is intermediate between the norm of the two extreme clusters; (ii) a high friendship measure, i.e., low hostility, within own cluster in the affinity network; and (iii) a high degree in the alliance network. It is important to note that all three features are needed in order for an agent to qualify as a bridge agent. For example, agents $j_3$ and $j_4$

are also in the intermediate cluster but are not members of the inter-cluster cliques due to high hostility in the affinity network and low degree in the alliance network.

The more interesting case is when, despite the relationship between clusters 1 and 3 characterized by (20), the agents in these two clusters end up forming an alliance through the aegis of agents in cluster 2 who serve as bridge agents. Due to condition (20), agents such as $k$ and $i_1$ are sufficiently divergent in terms of their norms such that their incremental payoff does not cover the cost $\tau$ of transforming their relationship. However, with the friendly path that is now created through cluster 2, these agents can eschew the tranformation cost of $\tau$ and free ride to a mutually profitable alliance in the affinity network. Therefore, the links $h_{i_2 j_2} = +1$ and $h_{j_1 k} = +1$ confer positive externalities within clusters as well as across clusters permitting the formation of alliances between disparate agents who otherwise would not have an incentive to ally with each other. Therefore, through the bridge provided by agents $j_1$ and $j_2$, we have an inter-cluster clique $\{k, i_1, i_2, j_1, j_2\}$ that spans three clusters.

# 5 Motivating Examples

We now provide a set of real world examples to substantiate our main results. To illustrate the *emergence of an NSG alliance structure among agents belonging to the same cluster in the affinity network with high centrality* (*degree*) *corresponding to greater friendship measure*, we look to the Western Pacific and its overlapping relationships in **G** in Figure 7. These relationships run largely through the United States in an NSG type configuration. At the north end of the figure, the NATO security alliance forms a connected alliance that runs dominantly through the United States to form paths to other states and alliances. Prior to 2023, the graph can be partitioned into a set of cliques (NATO, AUKUS, FIVE EYES) and an independent set (Japan, Republic of Korea, India, Taiwan, and Micronesia) forming a partial star architecture with the United States at its center. The United States' position as an economic partner, its cultural ties, and its role as a democratic security guarantor, imparts to it the highest aggregate ranking of friendship (lowest hostility). Thus, in accordance with our result, the United States has an exceptionally high centrality in the affinity network and is by far the highest degree node in the alliance network.

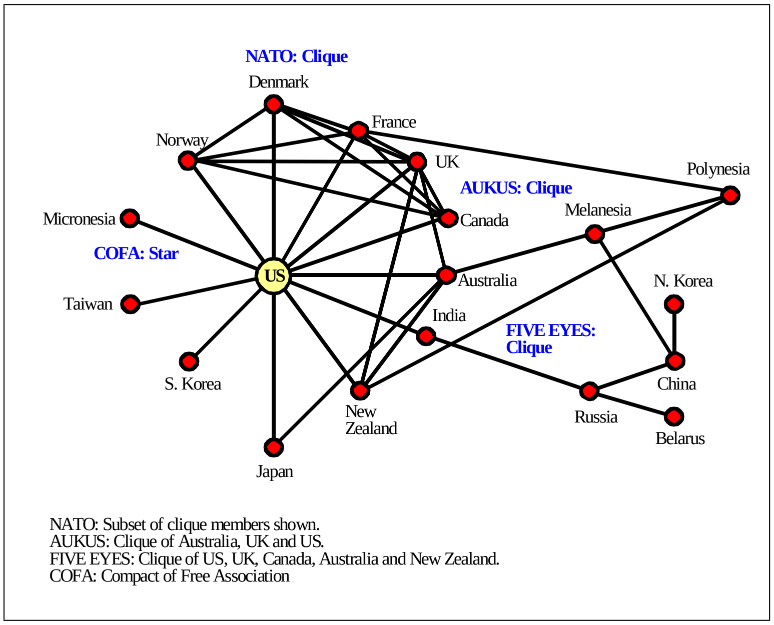


Figure 7: Alliance Network in the Pacific

Our model indicates that *when two agents belong to separate clusters in the affinity network, then as a precursor to forming a link in the alliance network this pair of agents have to establish a bridge in the affinity network*. In the post-pandemic years (2022-23), Japan made moves in both the **H** and **G** layers. *First*, Japan and Australia signed a security cooperation agreement (a link in **G**) that built off years of increasing economic ties (establishing a bridge in **H**). *Second*, Japan and South Korea are making significant diplomatic and economic investments at the encouragement of the United States in flipping their negative relationship into a positive one (establishing a bridge in **H**) as a precursor to security agreements (connecting in **G**). These moves are changing the existing star network in the Pacific where the United States underwrote security for all states (the graph depicted in figure 3(a)) into a more interconnected web of alliances (figure 3(d)).

A real world application of our result that there is *alliance formation among enemies when their disagreement over a norm is relatively small* is provided by Balkan conflict from 1992-95 (Becker et al 2023). In this case, there were three (singleton) clusters, with each cluster corresponding to an ethnic group – Bosniak, Croat and Serb. Figure 5 depicts the multilayered links between the three ethnic groups indexed by layer. The period of the conflict (1992-95) can be divided into three periods, indexed by $t \in \{1, 2, 3\}$, and the alliance structure that prevailed during these three periods is also indicated in figure 5. At the multilayer network's most base *affinity* layer, all relationships were negative since each group held deep and lasting animosities towards the others which predated the war by centuries. The other layers indicate the strategic dimensions that emerged in different periods of the conflict and prompted pairs of otherwise hostile ethnic groups to forge military alliances based on congruence in their interests in a layer.

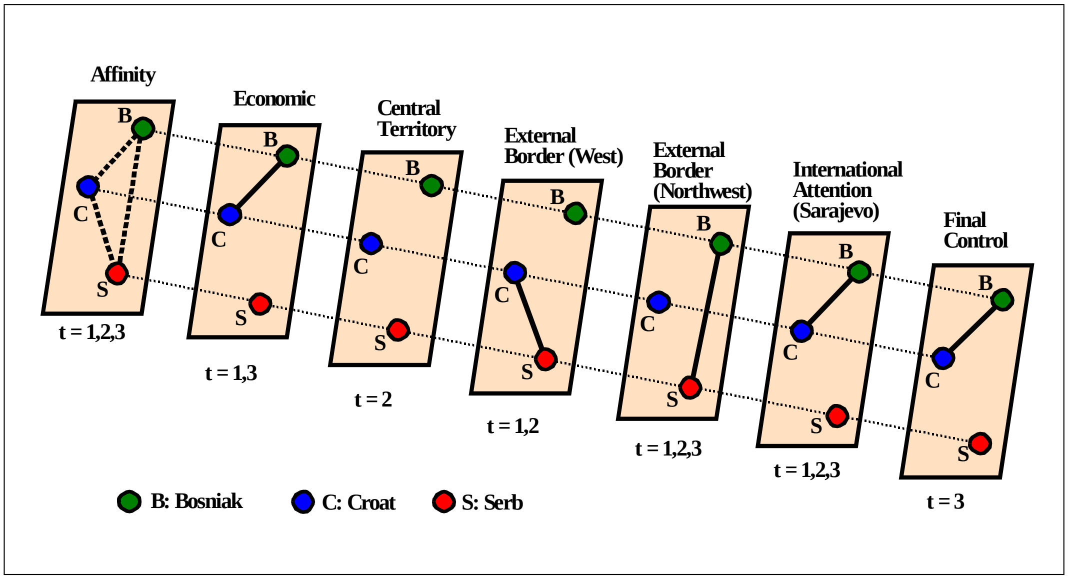


Figure 8: The Balkan Conflict (1992-1995)

At the outset of the war, the Serbs held nearly all pre-existing military hardware and ammunition, while the Bosniaks had essentially none. This forced the Bosniaks to form a relationship with the Croats in the *economic* layer to import (smuggle) weapons through the Croatian coast in contravention of international arms embargos to the war zone. This tacit partnership quickly dissolved as the Bosniaks and Croats started competing for the villages they shared in central Bosnia that began to rapidly ethnically segregate. Thus, the *central territory* layer was an empty network as the Bosniaks and Croats engaged in an internecine campaign while the Serbs disengaged to build their own forces and reserves against the other two. In the *external border* (*west*) layer, the Croats formed local partnerships with the proximate Serbs in an effort to define the Serbian external border with neighboring Croatia and expel Bosniaks. In the *external border* (*northwest*) layer, the Serbs were attempting to unite with a breakaway Serbian region between Croatia and Bosnia centered around the city of Bihac. Here a small "dissenter" Bosniak group allied with the Serbs against the Croats. The unique exterior border incentives in and around Bihac allowed the two most hostile ethnic groups – the Serbs and a small subset of Bosniaks – to form a breakaway alliance. The layer labeled *international attention* (*Saravejo*) shows the alliance between the local Bosniak and Croat forces in the Bosnian capital of Sarajevo against the besieging Serbs despite the ongoing bitter fighting between Bosniak and Croat forces in the majority of the country at that time. This was because in Sarajevo, the international community had daily broadcasts of the siege in a recognizable European capital that had just held an Olympic games (1984). Neither the Bosniaks or Croats of Sarajevo could afford to break their partnership and lose the city to the Serbs because the struggle against the Serbs for the city perimeter kept international attention and support focused on their plight. The same logic precipitated a strong military alliance between the Bosniaks and Croats for *final control* of Bosnia against the more powerful Serbs. As the Bosniaks and Croats degraded their forces fighting each other, they realized the Serbs were becoming a relatively much stronger and more pressing threat to them both. This "enemy of my enemy is my friend" logic became existential, and they formed a unified fighting force that quickly pressed the Serbs to sue for peace and thus end the war. The war in Bosnia-Herzegovina was a uniquely complex ethnic conflict within the dissolution of communist Yugoslavia that highlights the role that multilayered links between competing groups can play in forming stacked alliances starting at the base affinity layer.

Now consider our result that one of the qualities of a bridge agent is that *it is a moderating agent in the affinity network that connects more extreme clusters.* We often observe this free riding on a "friendly" link in geopolitics and geoeconomics. In. 2023, Japan and South Korea worked through the United States to establish "future-oriented" relations. While roughly 60% of South Koreans oppose a close relationship with Japan, the two American allies have been able to form collective military links through the United States' positive affinity network links.[17] The Swiss perform the role of an interlocutor prominently in diplomatic relations. While they occupy a positive space within a western Europe affinity network, they can branch more hostile relationship such as between the United States and Iran. This is precisely the reason Geneva and Vienna host so many international conferences and negotiations, allowing hostile nations in the affinity network to connect through the bridge set up by the Swiss and Austrians.

Another telling experiment of this dynamic is China's diplomatic efforts to be a multinational peace broker in 2023. At the Munich defense forum, China proposed a peace plan to end the Russian invaision of Ukraine. However, this proposal was dead on arrival in Ukraine and its western backers. Conversely, only a few weeks earlier, China brokered a thawing of tensions between Saudi Arabia and Iran, something the United States led western coalition had failed to do for a generation. Why the counterfactual? As our result on the role of intermediate agents serving as structural bridges in the affinity network indicates, China was not an intermediate cluster between Russia and Ukraine. This was made obvious by the Chinese Ambassador to France's remark, "Even these ex-Soviet countries don't have an effective status in international law because there was no international agreement to materialize their status as sovereign countries." Telling Ukraine, and nearly all of Eastern Europe, that they are not nation-states is an even more extreme position than Russia's. Therefore, in accordance with our result on intermediate agents serving as bridges, China could not serve as a bridging cluster and its peace plan failed. Conversely, China was an intermediate nation to both Iran and Saudi Arabia. China holds pseudo or outright positive relations with both making it a moderating presence between the two extremes, thereby providing incentives to both to connect through China in the alliance network.

# 6 Discussion

In this section we elaborate upon specific aspects of the main model.

[17] Indeed the authors expect, at the time of writing, this dynamic to be repeated with the United States conversely piggy-backing through South Korea to establish military relationships with the smaller South China Sea nations who are establishing economic ties with South Korea for their defense industry post Russian invasion of Ukraine (and Chinese aggression in their Exclusive Economic Zones).

## 6.1 Conflict Models

We will show how the reduced form payoff function given by (4) can be deduced from an explicit conflict model. The conflict game is drawn from Baliga and Sjöstrom (2012). Consider agent $i$ with neighborhood $\mathbf{N}_i(\mathbf{G})$. Then agent $i$ can find itself paired in pairwise conflict games – one with any of its own enemies, and one with any of the enemies of its allies. Therefore, agent $i$ can be involved pairwise in $\eta_i(\mathbf{G})$ conflict games. In each conflict game, the paired agents move simultaneously to choose either an aggressive action of *hawk* or a peaceful action of *dove*. The conflict game, with $i$ playing rows and $j$ playing columns, is given in strategic form by:

| $i \setminus j$ | *Hawk* | *Dove* |
|---|---|---|
| *Hawk* | $-\widetilde{c}_i\left(\delta', \eta'\right), -\widetilde{c}_j\left(\delta'', \eta''\right)$ | $\mu - \widetilde{c}_i\left(\delta', \eta'\right), -\rho$ |
| *Dove* | $-\rho, \mu - \widetilde{c}_j\left(\delta'', \eta''\right)$ | $0, 0$ |

where $\mu > 0$ and $\rho > 0$. Recalling (3), agent $i$ incurs a cost $\widetilde{c}_i\left(\delta', \eta'\right)$ by choosing hawk and no cost by choosing dove. When $j$ is a direct enemy of $i$, then $\widetilde{c}_i\left(\delta', \eta'\right) = \widetilde{c}_i(\delta_i, \eta_i)$, i.e., $i$ incurs a net cost based on its own degree and friendship levels. When $j$ is not a direct enemy but an enemy of one of $i$'s allies, say agent $k \in \mathbf{N}_i(\mathbf{G})$, then $\widetilde{c}_i\left(\delta', \eta'\right) = \widetilde{c}_i(\delta_k, \eta_k)$; it is the net cost that $i$ incurs by defending the alliance with $k$. In a hawk-dove interaction, the agent choosing hawk wins (a value of $\mu$) and dove loses (a value of $-\rho$). A hawk-hawk or dove-dove interaction results in a stalemate (a value of 0); thus, hawk-hawk only imposes costs while dove-dove yields 0 to both agents.

Following Baliga and Sjöstrom (2012), an agent $i$ is *coordinating* if hawk (dove) is a best response to hawk (dove). In other words, an aggressive action is met with agression, while a peaceful action is reciprocated with a peaceful one. We will accordingly assume that for each agent $i \in \mathcal{N}$ and all $(\delta, \eta) \in \mathbb{Z} \times \mathbb{Z}_+$:

$$\mu < \widetilde{c}_i(\delta, \eta) < \rho$$

There are two pure strategy Nash equilibria – hawk-hawk and dove-dove – and one mixed strategy Nash equilibrium. In the dove-dove Nash equilibrium, there would be no incentive to form alliances and thus the alliance network will be empty, a trivial case of the NSG architecture. In the hawk-hawk Nash equilibrium, agent $i$ will receive a total payoff (the sum of payoffs from each conflict game it is engaged in) equal to:

$$\pi_i(\mathbf{G}, \mathbf{H}_0) = -\sum_{k \in \mathbf{N}_i(\mathbf{G}) \cup \{i\}} \widetilde{c}_i(\delta_k, \eta_k) \tag{21}$$

while in the mixed strategy Nash equilibrium, which can be expected to be focal, the total payoff of agent $i$ is equal to:

$$\pi_i(\mathbf{G}, \mathbf{H}_0) = \sum_{k \in \mathbf{N}_i(\mathbf{G}) \cup \{i\}} \left[\frac{\rho\mu}{\rho - \mu} - \frac{\rho}{\rho - \mu}\widetilde{c}_i(\delta_k, \eta_k)\right] \tag{22}$$

Recalling (3), maximizing payoffs in either the hawk-hawk or mixed strategy Nash equilibrium is equivalent to maximizing $\pi_i(\mathbf{G}, \mathbf{H}_0)$ given by (4).[18] Therefore, our payoff specification is grounded in a proper conflict game.

**Remark:** (*Other parametric restrictions*) Baliga and Sjöstrom (2012) classify agents as *opportunistic* if hawk (dove) is a best response to dove (hawk), and *hawk-dominant* (*dove-dominant*) if hawk (dove) is a dominant strategy. The dove-dominant case once again yields an empty affinity network. The hawk-dominant case, and the pure strategy Nash equilibria of the opportunistic case yield total payoff given by (21). Thus, once again our reduced form payoff applies. The mixed strategy Nash equilibrium of the opportunistic case yields an anomalous result where agents would seek more hostilities and less allies.

### 6.2 Static Link Formation Game

We will explore a static alliance formation game adapted from Dutta et al. (1998) and show that the set of strongly stable equilibrium networks display the NSG architecture within each cluster. Thus, we provide an alternative approach to link formation that is distinct from the dynamic game as well as demonstrate the robustness of the NSG architecture. Each agent makes an announcement of intended alliances. An *announcement* by agent $i$ is of the form $s_i = (a_{ij})_{j \neq i}$. The intended alliance $a_{ij} \in \{0, 1\}$, where $a_{ij} = 1$ means that $i$ intends to form an alliance with $j$, while $a_{ij} = 0$ means that $i$ intends no such alliance. Let $S_i$ denote the set of announcements, or strategies, of agent $i$. An alliance between agents $i$ and $j$ is formed if and only if $a_{ij} = a_{ji} = 1$. We denote the formed link by $g_{ij} = 1$ and the absence of a link by $g_{ij} = 0$. A strategy profile $s = \{s_1, s_2, ..., s_n\}$, consisting of a strategy for each agent, induces a network $\mathbf{G}(s)$. To simplify the notation we shall often omit the dependence of the network on the underlying strategy profile.

A strategy profile $s^* = \{s_1^*, s_2^*, ..., s_n^*\}$ is *Nash* if and only if $\pi_i(\mathbf{G}(s_i^*, s_{-i}^*), \mathbf{H}_0) \geq \pi_i(\mathbf{G}(s_i, s_{-i}^*), \mathbf{H}_0)$, for all $s_i \in S_i$ and for all $i \in \mathcal{N}$, where $s_{-i}$ is the strategy profile of all agents other than $i$. The corresponding network is referred to as a *Nash network*. The Nash criterion is, however, not discriminating enough. For this purpose we will employ a *strong stability* property to refine the Nash equilibrium. Let $S \subset \mathcal{N}$ denote a coalition of agents. A network $\mathbf{G}'$ can be obtained from a network $\mathbf{G}$ through deviations by a coalition $S \subset \mathcal{N}$ if:

1. $g_{ij} = 1$ in $g'$ and $g_{ij} = 0$ in $\mathbf{G}$ implies that $i, j \in S$. In words, any new alliances added in the movement from $\mathbf{G}$ to $\mathbf{G}'$ can only be formed by agents in the coalition $S$.
2. $g_{ij} = 1$ in $g$ and $g_{ij} = 0$ in $\mathbf{G}'$ implies that $\{i, j\} \cap S \neq \varnothing$. In words, if any links are deleted in the movement from $\mathbf{G}$ to $\mathbf{G}'$, then at least one of the agents severing the alliance should be from the coalition $S$.

[18] We would expect the agents to coordinate on the same Nash equilibrium in each of their conflict games. However, even if agents play different Nash equilibria in different conflict games, the total payoff will be a linear combination of (21) and (22), and maximizing these total payoffs would be equivalent to maximizing (4).

**Definition (Strong Stability):** A network $\mathbf{G}$ is said to be *strongly stable* if for any coalition $S$ and any $\mathbf{G}'$ that can be obtained from $\mathbf{G}$ through deviations by $S$, $\pi_i(\mathbf{G}', \mathbf{H}_0) > \pi_i(\mathbf{G}, \mathbf{H}_0)$ for some $i \in S$ implies that $\pi_j(\mathbf{G}', \mathbf{H}_0) \leq \pi_j(\mathbf{G}, \mathbf{H}_0)$ for some $j \in S$.

The definition of strong stability that we employ is due to Dutta and Mutuswami (1997). According to their definition, if a network $\mathbf{G}$ is *not* strongly stable, then there exists a coalition $S$ that can deviate to some network $\mathbf{G}'$ in which *all* members of $S$ are strictly better off.

**Definition (Equilibrium Network):** A network $\mathbf{G}$ is an *equilibrium network* if there is a Nash strategy profile supporting $\mathbf{G}$, and the network $\mathbf{G}$ is strongly stable.

In our network setting, the only *unilateral* decision that an agent has is to sever alliances. The first property of an equilibrium network is, therefore, that no agent should have an incentive to delete any subset of its alliances. Note that forming an alliance is a *bilateral* decision requiring agreement by both agents. The second property of an equilibrium network states that, for any coalition, the member agents have no incentive to bilaterally form alliances that did not exist in the equilibrium network. The second property permits a refinement of the set of Nash networks that satisfy the first property. The next result shows that all equilibrium networks display an intra-cluster NSG structure in which the neighborhood of an agent with a lower friendship measure is nested within the neighborhood of an agent with a higher friendship measure.

**Proposition 7** *An equilibrium network exists. In an equilibrium network* $\mathbf{G}$*, all agents belonging to the same cluster form an alliance with an NSG architecture such that if* $\delta_j \geq \delta_i$*, then* $\mathbf{N}_i(\mathbf{G}) \subseteq \mathbf{N}_j(\mathbf{G}) \cup \{j\}$*.*

## 6.3 Non-Separable Benefits and Costs

We had assumed additively separable benefit and cost functions. This permitted us to avoid interaction between degrees of agents, or between degree and hostility. However, our results would continue to hold under a more general non-separable specification with suitable restrictions on the interaction terms. We now spell out the precise set of restrictions that are needed. Suppose the gross benefit to agent $i$ from a link with agent $j$ is more specified as $\psi(\eta_i, \eta_j)$. the function $\psi : \mathbb{Z}_+^2 \to \mathbb{R}_+$ is assumed to satisfy the following conditions:

**Assumption (A.2)*:** *For all* $i, j \in \mathcal{N}$:
**(a)** $\psi(\eta_i + 1, \eta_j) \geq \psi(\eta_i, \eta_j)$, and $\psi(\eta_i, \eta_j + 1) \geq \psi(\eta_i, \eta_j)$, for $1 \leq \eta_i, \eta_j < N - 1$.
**(b)** $\psi(\eta_i + 2, \eta_j) - \psi(\eta_i + 1, \eta_j) \geq \psi(\eta_i + 1, \eta_j) - \psi(\eta_i, \eta_j)$, for $1 \leq \eta_i < N - 2$ and $1 \leq \eta_j \leq N - 1$.
**(c)** $\psi(\eta_i + 1, \eta_j) - \psi(\eta_i, \eta_j) \geq \psi(\eta_j + 1, \eta_i) - \psi(\eta_j, \eta_i)$, for $1 \leq \eta_j \leq \eta_i < N - 1$.
**(d)** $\psi(\eta_i + 1, \eta_j + 1) - \psi(\eta_i, \eta_j + 1) \geq \psi(\eta_i + 1, \eta_j) - \psi(\eta_i, \eta_j)$, for $1 \leq \eta_i, \eta_j < N - 1$.

Thus, (a) gross returns are increasing in the degrees of the partners involved in the alliance; (b) the gross benefit shows increasing returns with respect to own degree; (c) the gross benefit to an agent $i$ with a higher degree from linking to an agent $j$ with a lower degree is at least as great as the gross benefit to $j$ from linking with $i$; and (d) gross benefits display the property of increasing differences in degrees $\left(\eta_i, \eta_j\right)$. Turning to costs, we will impose the following assumptions on the cost function $c(\delta, \eta)$.

**Assumption (A.3)*:** *For all* $i \in \mathcal{N}$, $\delta_i \in \mathbb{Z}$, and $1 \leq \eta_i < N-1$
(**a**) $c(\delta_i+1, \eta_i) < c(\delta_i, \eta_i)$ and $c(\delta_i, \eta_i+1) < c(\delta_i, \eta_i)$.
(**b**) $c(\delta_i, \eta_i) - c(\delta_i+1, \eta_i) \geq c(\delta_i-1, \eta_i) - c(\delta_i, \eta_i)$.
(**c**) $c(\delta_i, \eta_i) - c(\delta_i, \eta_i+1) \geq c(\delta_i, \eta_i-1) - c(\delta_i, \eta_i)$.
(**d**) $c(\delta_i, \eta_i+1) - c(\delta_i+1, \eta_i+1) \geq c(\delta_i, \eta_i) - c(\delta_i+1, \eta_i)$.

Therefore, (a) the cost to an agent $i$ is lower when it faces less hostility in $\mathbf{H}$, i.e., it has relatively more friends than enemies. Further, for the same level of hostility, the cost is less if agent $i$ has more allies; (b) the incremental cost is decreasing in the friendship measure given degree; (c) the incremental cost is decreasing in degree given the friendship measure; and (d) the cost function displays increasing differences in $(\delta_i, \eta_i)$.

The payoff of agent $i \in \mathcal{C}_\alpha(\mathbf{H}_0)$ is now given by:

$$\begin{aligned}\pi_i(\mathbf{G},\mathbf{H}) &= \sum_{j \in \mathbf{N}_i(\mathbf{G}) \cap \mathcal{C}_\alpha(\mathbf{H}_0)} \left[\psi\left(\eta_i(\mathbf{G}), \eta_j(\mathbf{G})\right) - c\left(\delta_j(\mathbf{H}), \eta_j(\mathbf{G})\right)\right] \\ &+ \sum_{j \in \mathbf{N}_i(\mathbf{G}) \cap \mathcal{C}_{\alpha'}(\mathbf{H}_0); \mathcal{F}_{ij}(\mathbf{H})=1} \left[\Theta_{ij}\psi\left(\eta_i(\mathbf{G}), \eta_j(\mathbf{G})\right) - c\left(\delta_j(\mathbf{H}), \eta_j(\mathbf{G})\right) - \tau h_{ij}^{+}\right] \end{aligned} \quad (23)$$

The NSG structure, and the consequent results, follow from Proposition 1 on preferential attachment. We will show that the non-separable case satisfies a stronger version of this proposition.

**Proposition 8** *Suppose agents $i$ and $j$ in the same cluster have a mutually profitable link in a network* $\mathbf{G}$. *Consider an agent $k \neq i, j$ in the same cluster such that $\delta_k \geq \delta_i$ and $\mathbf{N}_i(\mathbf{G}) \subseteq \mathbf{N}_k(\mathbf{G}) \cup \{k\}$. Then agents $k$ and $j$ have a mutually profitable link in $\mathbf{G}' \supseteq \mathbf{G}$.*

This proposition generates an NSG architecture within each cluster with degree positively correlated with friendship. For example, consider agent $i$ who forms the first link with agent $j$ when $\mathbf{G}$ is empty. Since trivially $\mathbf{N}_i(\mathbf{G}^e) \subseteq \mathbf{N}_k(\mathbf{G}^e) = \emptyset$, it follows that agent $j$ will subsequently form an alliance with agent $k$ satisfying $\delta_k \geq \delta_i$. The argument applies to each successive alliance formed by agent $i$. Thus, an agent $k$ with a higher friendship measure than agent $i$ measure will have greater degree than $i$ and a neighborhood nesting that of $i$. This nested neighborhood result along with Lemma 1 generates the characterization result on interlocking cliques. We show in the appendix that Lemma 1 also holds in the non-separable case under A.2* and A.3*.

### 6.4 Endogenous Affinity Network and Norms

We have assumed that in the initial position the affinity network is given. As a first step towards a microfounded affinity network, we can assume that the affinity network is initially empty, and draw upon the definitive analysis of Hiller (2017) to augment our link formation game with the prior formation of an affinity network. Adapting Hiller, we can craft an *affinity formation game* as follows. Let $\{\theta_1(\rho), \theta_2(\rho), ..., \theta_K(\rho)\}$ denotes the distribution of norms that separates agents. The parameter $\rho$ captures the dimension on which the norm is based, i.e., $\rho \in \{\text{culture, ideology, politics, security}\}$. Assume that $n_k$ denotes the number of agents who subscribe to the norm $\theta_k(\rho)$ such that $n_k \geq 1$, $n_k \neq n_{k+1}$, and $\sum_{k=1}^{K} n_k = N$. Each agent is endowed with a given intrinsic level of strength that is normalized to unity. An agent can augment this strength through positive connections in the affinity network with agents *who share the same norm*. Formally, the strength gained by an agent $i$ from establishing a positive connection with agent $j$ in the affinity network is equal to 1 if $\theta_i(\rho) = \theta_j(\rho)$ and 0 otherwise.

Each agent simultaneously proposes positive (friendship) or negative (enemy) links to other agents in the affinity layer. A negative offer is interpreted as an initiation of hostile relations and imposes a cost on the proposer. A positive connection between two agents is established if both agents extend a positive link to each other, while a negative connection is formed if at least one agent extended a negative link. A positive connection does not confer any direct benefit (the payoff is zero from a positive link) but imparts indirect benefits by increasing the strength of an agent, with aggregate strength in an affinity network equal to the sum of own strength and the strengths of friends who share the same norm. The formation of a negative connection imposes a "conflict cost" on both agents and engages them in a zero-sum game in which the winner extracts rents from the loser that are increasing (respectively, decreasing) in own (respectively, enemy's) aggregate strength.

We can now draw upon Hiller (2017) to characterize the Nash equilibrium of the affinity announcement game. Agents will have an incentive to offer a positive link only to those who share the same norm. Thus all agents who share the same norm, say $\theta_k(\rho)$, will be mutual friends and comprise a cluster $\mathcal{C}_k$ whose aggregate strength is equal to $n_k$. Our assumption that $n_k \neq n_{k+1}$ ensures that the clusters are asymmetric with respect to size and thus unequal with respect to aggregate strength. Suppose without loss of generality that $n_1 < n_2 < \cdot\cdot < n_K$. Then cluster $\mathcal{C}_{k+1}$ has higher aggregate strength than $\mathcal{C}_k$. Agents in clusters with higher aggegate strength extend negative links to those in clusters with lower aggregate strength to extract rents. Consequently, the affinity network, say $\mathbf{H}_0$, corresponding to the Nash equilibrium is an endogenously formed balanced layer.

The norm can change due to events, or shocks, that change $\rho$ and can engender changes in the affinity network. For example, suppose that after the formation of $\mathbf{H}_0$, the new norm is $\theta_k(\overline{\rho})$ where $\overline{\rho}$ reflects a cluster's concern over external threats to its security. Let $\Phi_{kk'}$ denote the total number of negative links connecting clusters $\mathcal{C}_k(\mathbf{H}_0)$ and $\mathcal{C}_{k'}(\mathbf{H}_0)$ and:

$$\theta_k(\overline{\rho}) = \sum_{k' \neq k} \xi_k(k')\, \Phi_{kk'} \tag{24}$$

where $\xi_k(k')$ is the weight placed by agents in $\mathcal{C}_k(\mathbf{H}_0)$ on the threat posed by those in $\mathcal{C}_{k'}(\mathbf{H}_0)$. Thus agents in lower-indexed clusters (and thus lower aggregate strength) have relatively higher values of the norm (greater perception of the external threat) than higher-indexed clusters. Now suppose that one such cluster, say $\mathcal{C}_2(\mathbf{H}_0)$, faces increased hostility from a higher-indexed cluster, say $\mathcal{C}_K(\mathbf{H}_0)$. Or, the level of hostility is the same but the threat perception of $\mathcal{C}_2(\mathbf{H}_0)$, as given by $\xi_2(K)$, increases. Either way, this will raise the value of $\theta_2(\overline{\rho})$ and draw it closer to $\theta_1(\overline{\rho})$, the norm for $\mathcal{C}_1(\mathbf{H}_0)$. Now that their norms are sufficiently close, there may be an incentive for bridge agents in the two clusters to transform their relationship to positive and use this to leverage the formation of security pacts in the alliance network. The case of Philippines discussed in the introduction corresponds to this case. The confrontation with China increased the security threat to Philippines in the South China Sea and forced it to reconsider its affinities and alliances. In the period following this confrontation, Philippines has signed over 18 security agreements with other countries.[19]

The norm could also change due to exogenous shocks. An example is the epoch-making fall of the Berlin Wall, where clusters were defined by opposing political ideologies. The liberal Western and former Eastern bloc countries had animosity across and friendship within their clusters. When the Berlin Wall fell, the demarcation line was redrawn, and the Eastern European nations became part of the liberal world, and the liberal cluster enlarged. Consequently, we witnessed the formation of alliances between former enemies with Poland, Hungary, Bulgaria, Romania and the Czech Republic joining NATO. How about shocks that do not alter the memberships within clusters but increase intra-cluster friendships within the affinity network. Under these circumstances, hostility declines, or the costs of forming a link in the alliance network declines. As a result, a denser intra-cluster NSG becomes possible.

## 7 Conclusion

This paper aims to show that the defensive alliances between nations can be modeled as a multilayer network of signed relationships. But unlike previous economic studies, we identify a multilayer of relationships and order them with a foundational affinity network ($\mathbf{H}$) that, through non-cooperative gameplay, results in a higher-level defense alliance layer ($\mathbf{G}$). The implications of our findings are broad and show how nations can build security alliances based on their position in an affinity network connecting all nations. We hope that future research will extend these results from two layers to multiple ordered layers of relations with specific parameters tied to economic, cultural, education, geographic, and political layers. Further, while we have offered a preliminary discussion of a microfounded affinity network and endogenously changing norms, a systematic analysis incorporating the co-evolution of affinities, norms and alliances is a productive avenue for further research.

[19] The Washington Post ("As tensions rise with China at sea, Philippines strikes deals", Monday, March 11, 2024) explicitly refers to the new defense pacts of the Philippines as a "network of alliances" to deter Chinese aggression.

# 8 Appendix

**Proof of Theorem 1:** We will show that cycles cannot emerge along an improving path in $\mathbf{G}$. Let $g_{ij} = 1$ be the *first* link that is deleted along an improving path when the state is $(\mathbf{G}_q(\kappa), \mathbf{H}_0, i)$. The active agent $i$ will delete the link with $j$ only if:

$$\Delta\nu\left(\eta_i\left(\mathbf{G}_q(\kappa)\right) - 1\right) + \left[w\left(\eta_j\left(\mathbf{G}_q(\kappa)\right)\right) - c\left(\delta_j\right)\right] < 0 \tag{25}$$

Suppose $(\mathbf{G}_q(\kappa'), \mathbf{H}_0, k \in \{i,j\})$, $\kappa' < \kappa$, was the state at an earlier stage when the link $g_{ij} = 1$ was established with either $i$ or $j$ as the active agent. Since the link between $i$ and $j$ is the first link that is deleted, it must be true that $\eta_k\left(\mathbf{G}_q(\kappa)\right) \geq \eta_k\left(\mathbf{G}_q(\kappa')\right)$ for all $k \in \mathcal{N}$ and $\eta_k\left(\mathbf{G}_q(\kappa)\right) > \eta_k\left(\mathbf{G}_q(\kappa')\right)$ for $k \in \{i,j\}$. From A.2(a) and A.2(b) respectively:

$$w\left(\eta_j\left(\mathbf{G}_q(\kappa)\right)\right) > w\left(\eta_j\left(\mathbf{G}_q(\kappa')\right)\right), \quad \Delta\nu\left(\eta_i\left(\mathbf{G}_q(\kappa)\right) - 1\right) \geq \Delta\nu\left(\eta_i\left(\mathbf{G}_q(\kappa')\right)\right) \tag{26}$$

Since agent $i$ had accepted the link in state $(\mathbf{G}_q(\kappa'), \mathbf{H}_0, k \in \{i,j\})$:

$$\Delta\nu\left(\eta_i\left(\mathbf{G}_q(\kappa')\right) - 1\right) + \left[w\left(\eta_j\left(\mathbf{G}_q(\kappa')\right)\right) - c\left(\delta_j\right)\right] \geq 0 \tag{27}$$

It follows from (25)-(27) that:

$$0 > \Delta\nu\left(\eta_i\left(\mathbf{G}_q(\kappa)\right) - 1\right) + \left[w\left(\eta_j\left(\mathbf{G}_q(\kappa)\right)\right) - c\left(\delta_j\right)\right] > \Delta\nu\left(\eta_i\left(\mathbf{G}_q(\kappa')\right) - 1\right) + \left[w\left(\eta_j\left(\mathbf{G}_q(\kappa')\right)\right) - c\left(\delta_j\right)\right] \geq 0$$

which is a contradiction. Therefore, no links will be deleted along an improving path and thus no cycles will emerge in $\mathbf{G}$. It follows that the link formation game will converge where the limit will satisfy the conditions of pairwise stability. ■

**Proof of Proposition 1:** Since agents $i$ and $j$ have a mutually profitable link in $\mathbf{G}$:

$$\begin{aligned}
\left[v\left(\eta_i\left(\mathbf{G} + g_{ij}\right)\right) - v\left(\eta_i\left(\mathbf{G}\right)\right)\right] + \left[w\left(\eta_j\left(\mathbf{G} + g_{ij}\right)\right) - c\left(\delta_j\right)\right] &\geq 0 \\
\left[v\left(\eta_j\left(\mathbf{G} + g_{ij}\right)\right) - v\left(\eta_j\left(\mathbf{G}\right)\right)\right] + \left[w\left(\eta_i\left(\mathbf{G} + g_{ij}\right)\right) - c\left(\delta_i\right)\right] &\geq 0
\end{aligned}$$

and at least one inequality is strict. Since $\eta_k\left(\mathbf{G}'\right) \geq \eta_i\left(\mathbf{G}\right)$, $\eta_j\left(\mathbf{G}'\right) \geq \eta_j\left(\mathbf{G}\right)$, and $\delta_k \geq \delta_i$, it follows for agent $k$ from A.2(b) and A.2(a) respectively that:

$$\begin{aligned}
&\left[v\left(\eta_k\left(\mathbf{G}' + g_{kj}\right)\right) - v\left(\eta_k\left(\mathbf{G}'\right)\right)\right] + \left[w\left(\eta_j\left(\mathbf{G}' + g_{kj}\right)\right) - c\left(\delta_j\right)\right] \\
&\qquad \geq \left[v\left(\eta_i\left(\mathbf{G} + g_{ij}\right)\right) - v\left(\eta_i\left(\mathbf{G}\right)\right)\right] + \left[w\left(\eta_j\left(\mathbf{G} + g_{ij}\right)\right) - c\left(\delta_j\right)\right] \geq 0
\end{aligned}$$

while for agent $j$ it follows from A.2(b), A.2(a) and A.3(a) that:

$$\left[v\left(\eta_j\left(\mathbf{G}'+g_{kj}\right)\right)-v\left(\eta_j\left(\mathbf{G}'\right)\right)\right]+\left[w\left(\eta_k\left(\mathbf{G}'+g_{kj}\right)\right)-c\left(\delta_k\right)\right]$$
$$\geq\left[v\left(\eta_j\left(\mathbf{G}+g_{ij}\right)\right)-v\left(\eta_j\left(\mathbf{G}\right)\right)\right]+\left[w\left(\eta_i\left(\mathbf{G}+g_{ij}\right)\right)-c\left(\delta_i\right)\right]\geq 0$$

and at least one LHS is strictly positive. Therefore, agents $k$ and $j$ have a mutually profitable link in $\mathbf{G}'$. ■

**Proof of Proposition 2:** We will first prove that $\Delta_1^\alpha\left(\mathbf{H}_0\right)\subseteq D_1^\alpha(\mathbf{G}_1)$. Suppose to the contrary that $i\in\Delta_1^\alpha\left(\mathbf{H}_0\right)\cap D_l^\alpha(\mathbf{G}_1)$ for $l\geq 2$. Thus, $\eta_i\left(\mathbf{G}_1\right)>\eta_j\left(\mathbf{G}_1\right)$ for $j\in D_1^\alpha(\mathbf{G}_1)$. Let $k\in\mathbf{N}_i\left(\mathbf{G}_1\right)\backslash\mathbf{N}_j\left(\mathbf{G}_1\right)$ denote the agent with whom $i$ formed a link when it had $\eta_j\left(\mathbf{G}_1\right)$ number of links, i.e, the same number of links as $j$. Let $\mathbf{G}_0\left(\kappa\right)$ denote the stage along the improving path when this link was formed, and so $\eta_j\left(\mathbf{G}_1\right)=\eta_i\left(\mathbf{G}_0\left(\kappa\right)\right)$. Therefore:

$$\Delta v\left(\eta_i\left(\mathbf{G}_0\left(\kappa\right)\right)\right)+\left[w\left(\eta_k\left(\mathbf{G}_0\left(\kappa\right)\right)+1\right)-c\left(\delta_k\right)\right]\geq 0 \tag{28}$$
$$\Delta v\left(\eta_k\left(\mathbf{G}_0\left(\kappa\right)\right)\right)+\left[w\left(\eta_i\left(\mathbf{G}_0\left(\kappa\right)\right)+1\right)-c\left(\delta_i\right)\right]\geq 0 \tag{29}$$

and at least one inequality is strict. Since $k\notin\mathbf{N}_j\left(\mathbf{G}_1\right)$ in the limit network $\mathbf{G}_1$, it must be true that agents $j$ and $k$ do not have a mutually profitable link in $\mathbf{G}_1$:

$$\min\left\{\Delta v\left(\eta_k\left(\mathbf{G}_1\right)\right)+\left[w\left(\eta_j\left(\mathbf{G}_1\right)+1\right)-c\left(\delta_j\right)\right],\Delta v\left(\eta_j\left(\mathbf{G}_1\right)\right)+\left[w\left(\eta_k\left(\mathbf{G}_1\right)+1\right)-c\left(\delta_k\right)\right]\right\}<0 \tag{30}$$

However, since $\delta_j\geq\delta_i$ (given that $i\in\Delta_1^\alpha$) and $\mathbf{G}_0\left(\kappa\right)\subseteq\mathbf{G}_1$, it follows from $\eta_j\left(\mathbf{G}_1\right)=\eta_i\left(\mathbf{G}_0\left(\kappa\right)\right)$ and A.3(a) that:

$$\Delta v\left(\eta_j\left(\mathbf{G}_1\right)\right)+\left[w\left(\eta_k\left(\mathbf{G}_1\right)+1\right)-c\left(\delta_k\right)\right]\geq\Delta v\left(\eta_i\left(\mathbf{G}_0\left(\kappa\right)\right)\right)+\left[w\left(\eta_k\left(\mathbf{G}_0\left(\kappa\right)\right)+1\right)-c\left(\delta_k\right)\right]\geq 0$$
$$\Delta v\left(\eta_k\left(\mathbf{G}_1\right)\right)+\left[w\left(\eta_j\left(\mathbf{G}_1\right)+1\right)-c\left(\delta_j\right)\right]\geq\Delta v\left(\eta_k\left(\mathbf{G}_0\left(\kappa\right)\right)\right)+\left[w\left(\eta_i\left(\mathbf{G}_0\left(\kappa\right)\right)+1\right)-c\left(\delta_i\right)\right]\geq 0$$

which contradicts (30). Thus, $\eta_i\left(\mathbf{G}_1\right)\leq\eta_j\left(\mathbf{G}_1\right)$ for all $j\in D_1^\alpha(\mathbf{G}_1)$ and $i\in\Delta_1^\alpha\left(\mathbf{H}_0\right)$, and hence $\Delta_1^\alpha\left(\mathbf{H}_0\right)\subseteq D_1^\alpha(\mathbf{G}_1)$. We now prove that $\Delta_s^\alpha\left(\mathbf{H}_0\right)\subseteq D_m^\alpha(\mathbf{G}_1)$. We have already shown in the main text that $i_n\in\Delta_s^\alpha\left(\mathbf{H}_0\right)$ and $i_n\in D_m^\alpha(\mathbf{G}_1)$. The same argument can be repeated for each member of $\Delta_s^\alpha\left(\mathbf{H}_0\right)$ in declining order of index to show that $\Delta_s^\alpha\left(\mathbf{H}_0\right)\subseteq D_m^\alpha(\mathbf{G}_1)$. ■

**Proof of Proposition 3:** Suppose $i_k,i_l\in\mathcal{C}_\alpha\left(\mathbf{H}_0\right)$ such that $\delta_{i_k}\leq\delta_{i_l}$. Let $i^{(1)}\in\mathbf{N}_{i_k}\left(\mathbf{G}^*\right)$ denote the first partner for $i_k$ and $\mathbf{G}_0\left(\kappa^{(1)}\right)$ be the stage along the improving path when this link was established. There are two cases to consider: (i) Suppose $i_k$ was the active player when the link was formed. Note that $\eta_{i_k}\left(\mathbf{G}_0\left(\kappa^{(1)}\right)\right)=0\leq\eta_{i_l}\left(\mathbf{G}_0\left(\kappa^{(1)}\right)\right)$ since $i^{(1)}$ is the first partner for $i_k$. Consider a subsequent stage $\mathbf{G}_0\left(\kappa^{(1)\prime}\right)$ when $i_l$ is the active agent. Then, from Proposition 1, $i^{(1)}$ and $i_l$ also have a mutually profitable link and thus $i^{(1)}\in\mathbf{N}_{i_l}\left(\mathbf{G}^*\right)$. (ii) Suppose $i^{(1)}$ is the active player. Along an improving path, $\eta_{i_k}\left(\mathbf{G}_0\left(\kappa^{(1)}\right)\right)\leq\eta_{i_l}\left(\mathbf{G}_0\left(\kappa^{(1)}\right)\right)$. Thus, the payoff from linking with $i_l$ is weakly greater than linking with $i_k$, and from the construction of the link formation game, agent $i^{(1)}$ would have approached $i_l$ before $i_k$.

Thus $i^{(1)} \in \mathbf{N}_{i_l}(\mathbf{G}^*)$.

Now suppose this property is true for agents $i^{(1)}, i^{(2)}, ..., i^{(r)} \in \mathcal{C}_\alpha$, i.e., these agents are the first $r$ partners of $i_k$ and belong to $\mathbf{N}_{i_k}(\mathbf{G}^*) \cap \mathbf{N}_{i_l}(\mathbf{G}^*)$. Consider the next partner $i^{(r+1)}$ of agent $i_k$ and suppose this link was formed in stage $\mathbf{G}_0\left(\kappa^{(r)}\right)$ along the improving path. Suppose $i_k$ was the active player when this link was formed. Since $\eta_{i_k}\left(\mathbf{G}_0\left(\kappa^{(r)}\right)\right) = r \leq \eta_{i_l}\left(\mathbf{G}_0\left(\kappa^{(r)}\right)\right)$, it follows from Proposition 1 that $i^{(r+1)}$ and $i_l$ also have a mutually profitable link when $i_l$ is the active agent. Now suppose $i^{(r+1)}$ was the active agent when the link with $i_k$ was formed. Then, similar to the reasoning with $i^{(1)}$, agent $i^{(r+1)}$ would have first formed this link with $i_l$. Therefore, $i^{(r+1)} \in \mathbf{N}_{i_l}(\mathbf{G}^*)$. This completes the induction step and proves the nestedness property. ■

**Proof of Theorem 2:** In the augmented link formation game, as each iteration of link formation occurs in the alliance network $\mathbf{G}_r$, $r \geq 1$, potentially new alliances are added but none of the existing links are deleted. Therefore, letting the vector $\mathbf{d}(\mathbf{G}) = \{d_1(\mathbf{G}), d_2(\mathbf{G}), ..., d_N(\mathbf{G})\}$ denote the *degree distribution* of agents in $\mathbf{G}$, it follows that $\mathbf{d}(\mathbf{G}_r)$ first order dominates $\mathbf{d}(\mathbf{G}_{r-1})$, i.e., $d_i(\mathbf{G}_r) \geq d_i(\mathbf{G}_{r-1})$ for all agents $i \in \mathcal{N}$ and $r \geq 1$. We will write this formally as $\mathbf{G}_r \succeq \mathbf{G}_{r-1}$. Now consider the affinity network and let the vector $\boldsymbol{\delta}(\mathbf{H}) = \{\delta_1(\mathbf{H}), \delta_2(\mathbf{H}), ..., \delta_N(\mathbf{H})\}$ denote the *friendship distribution* of agents in $\mathbf{H}$. In each iteration of affinity transformation that occurs in the affinity network $\mathbf{H}_r$, $r \geq 1$, hostile or neutral links are converted into friendly ones but no existing relationships can be obviated. Consequently, $\boldsymbol{\delta}(\mathbf{H}_r)$ first order dominates $\boldsymbol{\delta}(\mathbf{H}_{r-1})$ for all $r \geq 1$. We will write this formally as $\mathbf{H}_r \succeq \mathbf{H}_{r-1}$. Extending to the multilayer, we will say that $(\mathbf{G}_r, \mathbf{H}_r) \succeq (\mathbf{G}_{r-1}, \mathbf{H}_{r-1})$. Thus, the sequence $\{(\mathbf{G}_r, \mathbf{H}_r); r \in \mathbb{Z}_+\}$ is monotonically increasing in the sense of first order dominance. Since the set of multilayer networks is finite, this sequence will converge and the limit is a pairwise-stable multilayer network. ■

**Proof of Lemma 1:** Note that $\pi_i(\mathbf{G}_1 + g_{ij}, \mathbf{H}_0 \oplus h_{ij}) - \pi_i(\mathbf{G}_1, \mathbf{H}_0)$ is equal to:

$$[v(\eta_i(\mathbf{G}_1 + g_{ij})) - v(\eta_i(\mathbf{G}_1))] + [c_0(\delta_i(\mathbf{H}_0)) - c_0(\delta_i(\mathbf{H}_0) + 1)] + \left[\Theta_{ij} w\left(\eta_j(\mathbf{G}_1) + 1\right) - c(\delta_j(\mathbf{H}_0))\right]$$

Since $\eta_i(\mathbf{G}_1) \geq \eta_k(\mathbf{G}_1)$, it follows from A.2(b) that:

$$v(\eta_i(\mathbf{G}_1 + g_{ij})) - v(\eta_i(\mathbf{G}_1)) \geq v(\eta_k(\mathbf{G}_1 + g_{kl})) - v(\eta_k(\mathbf{G}_1))$$

Since $i \in \mathcal{C}_\alpha(\mathbf{H}_0) \cap \Delta_s^\alpha(\mathbf{H}_0)$, it follows that $\delta_i(\mathbf{H}_0) \geq \delta_k(\mathbf{H}_0)$. Therefore, from A.3(b):

$$c_0(\delta_i(\mathbf{H}_0)) - c_0(\delta_i(\mathbf{H}_0) + 1) \geq c_0(\delta_k(\mathbf{H}_0)) - c_0(\delta_k(\mathbf{H}_0) + 1)$$

Finally, since $\eta_j(\mathbf{G}_1) \geq \eta_l(\mathbf{G}_1)$, $\delta_j(\mathbf{H}_0) \geq \delta_l(\mathbf{H}_0)$ and $\Theta_{ij} = \Theta_{kl}$, it follows from A.2(a) and A.3(a) that:

$$\Theta_{ij} w\left(\eta_j(\mathbf{G}_1) + 1\right) - c(\delta_j(\mathbf{H}_0)) \geq \Theta_{kl} w(\eta_l(\mathbf{G}_1) + 1) - c(\delta_l(\mathbf{H}_0))$$

Therefore:

$$\pi_i(\mathbf{G}_1 + g_{ij}, \mathbf{H}_0 \oplus h_{ij}) - \pi_i(\mathbf{G}_1, \mathbf{H}_0) \geq \pi_k(\mathbf{G}_1 + g_{kl}, \mathbf{H}_0 \oplus h_{kl}) - \pi_k(\mathbf{G}_1, \mathbf{H}_0)$$

The verification for agents $j$ and $l$ is identical. ■

**Proof of Proposition 4:** The proof is provided in the main text. ■

**Proof of Proposition 5:**

**a**. We have proved in Proposition 3 that $\mathbf{G}_1$ has an NSG architecture in each cluster. Now suppose this is true for $\mathbf{G}_r$, $r \geq 2$. We will prove it for $\mathbf{G}_{r+1}$ by contradiction. Suppose there exists a cluster $\mathcal{C}_\alpha(\mathbf{H}_0)$ with agents $i$ and $j$ such that $\eta_i(\mathbf{G}_{r+1}) \leq \eta_j(\mathbf{G}_{r+1})$ but $\mathbf{N}_i(\mathbf{G}_{r+1}) \nsubseteq \mathbf{N}_j(\mathbf{G}_{r+1})$. In particular, there exists an agent $k \in \mathcal{C}_\alpha(\mathbf{H}_0)$ such that $k \in \mathbf{N}_i(\mathbf{G}_{r+1}) \backslash \mathbf{N}_j(\mathbf{G}_{r+1})$. Since $\mathcal{C}_\alpha(\mathbf{H}_0)$ has an NSG structure in $\mathbf{G}_r$, and $\mathbf{G}_r \subseteq \mathbf{G}_{r+1}$, the link $g_{ik} = 1$ must have been added when link formation was occurring in $\mathbf{G}_{r+1}$. Thus, $\eta_i(\mathbf{G}_r) < \eta_j(\mathbf{G}_r)$. Recalling Proposition 3 which demonstrated that degree is positively correlated with friendship, it follows that $\delta_i \leq \delta_j$. Now suppose the network is $\mathbf{G}_{r+1}(\kappa)$ when the link $g_{ik} = 1$ is formed in $\mathbf{G}_{r+1}$. There are two possible cases.

**Case I:** Suppose $i$ was the active agent and $k$ acquiesced as the passive agent. Then, in some subsequent state $(\mathbf{G}_{r+1}(\kappa'), \mathbf{H}_r, k)$, i.e., when $k$ is the active agent, then $k$ will have a mutually profitable link with $j$.

$$\pi_k\left(\mathbf{G}_{r+1}\left(\kappa'\right) + g_{kj}, \mathbf{H}_r\right) - \pi_k\left(\mathbf{G}_{r+1}\left(\kappa'\right), \mathbf{H}_r\right) = \Delta v\left(\eta_k\left(\mathbf{G}_{r+1}\left(\kappa'\right)\right)\right) + \left[w\left(\eta_j\left(\mathbf{G}_{r+1}\left(\kappa'\right)\right) + 1\right) - c\left(\delta_j\right)\right]$$

From A.2(b):

$$\Delta v\left(\eta_k\left(\mathbf{G}_{r+1}\left(\kappa'\right)\right)\right) \geq \Delta v\left(\eta_k\left(\mathbf{G}_{r+1}(\kappa)\right)\right)$$

and, since $\eta_i(\mathbf{G}_r) < \eta_j(\mathbf{G}_{r+1}(\kappa'))$, from A.2(a) and A.3(a):

$$w\left(\eta_j\left(\mathbf{G}_{r+1}\left(\kappa'\right)\right) + 1\right) - c(\delta_j) > w\left(\eta_i(\mathbf{G}_r) + 1\right) - c(\delta_i)$$

Therefore:

$$\pi_k\left(\mathbf{G}_{r+1}\left(\kappa'\right) + g_{kj}, \mathbf{H}_r\right) - \pi_k\left(\mathbf{G}_{r+1}\left(\kappa'\right), \mathbf{H}_r\right) > \Delta v\left(\eta_k\left(\mathbf{G}_{r+1}(\kappa)\right)\right) + w\left(\eta_i(\mathbf{G}_r) + 1\right) - c(\delta_i) \geq 0$$

where the second strict inequality follows from the fact that agent $k$ had acquiesced to a link with $i$ when the network was $\mathbf{G}_{r+1}(\kappa)$. Agent $j$ will reciprocate because:

$$\begin{aligned}
\pi_j\left(\mathbf{G}_{r+1}\left(\kappa'\right) + g_{kj}, \mathbf{H}_r\right) - \pi_j\left(\mathbf{G}_{r+1}\left(\kappa'\right), \mathbf{H}_r\right) &= \Delta v\left(\eta_j\left(\mathbf{G}_{r+1}\left(\kappa'\right)\right)\right) + \left[w\left(\eta_k\left(\mathbf{G}_{r+1}\left(\kappa'\right)\right) + 1\right) - c(\delta_k)\right] \\
&> \Delta v\left(\eta_i(\mathbf{G}_r)\right) + \left[w\left(\eta_k\left(\mathbf{G}_{r+1}(\kappa)\right) + 1\right) - c(\delta_k)\right] > 0
\end{aligned}$$

where the last strict inequality follows since $i$ had proposed a link to $k$ in $\mathbf{G}_{r+1}(\kappa)$. Therefore, it cannot be the case that when all profitable opportunities have been exhausted in $\mathbf{G}_{r+1}$ then agents $k$ and $j$ will remain unlinked.

**Case II:** Suppose $k$ was the active agent when the network was $\mathbf{G}_{r+1}(\kappa)$. Then, according to the link formation protocol, $k$ would have proposed a link with agent $j$ rather than $i$ because:

$$\begin{aligned}\pi_k\left(\mathbf{G}_{r+1}(\kappa)+g_{kj},\mathbf{H}_r\right)-\pi_k\left(\mathbf{G}_{r+1}(\kappa),\mathbf{H}_r\right) &= \Delta v\left(\eta_k\left(\mathbf{G}_{r+1}(\kappa)\right)\right)+\left[w\left(\eta_j\left(\mathbf{G}_{r+1}(\kappa)\right)+1\right)-c\left(\delta_j\right)\right] \\ &> \Delta v\left(\eta_k\left(\mathbf{G}_{r+1}(\kappa)\right)\right)+\left[w\left(\eta_i\left(\mathbf{G}_r\right)+1\right)-c\left(\delta_i\right)\right]\end{aligned}$$

and once again agent $j$ will accept the proposal. Therefore, once again we have a contradiction.

It follows that each intra-cluster architecture in $\mathbf{G}_{r+1}$ will have an NSG architecture. Since $\mathbf{G}^*$ is reached in a finite number of steps, it follows that the result also holds for $\mathbf{G}^*$.

**b**. Let $i\in\mathcal{C}_\alpha(\mathbf{H}_0)\cap\Delta_l^\alpha(\mathbf{H}^*)$ and $j\in\mathcal{C}_{\alpha'}(\mathbf{H}_0)\cap\Delta_{l'}^{\alpha'}(\mathbf{H}^*)$. Let $\mathbf{H}^*_{-\alpha,\alpha'}$ denote the affinity network $\mathbf{H}^*$ in which there is no friendly link between clusters $\mathcal{C}_\alpha(\mathbf{H}_0)$ and $\mathcal{C}_{\alpha'}(\mathbf{H}_0)$. There are two possible cases:

**Case I:** Suppose $i$ and $j$ incurred the cost $\tau$ of transforming their affinity relationship allowing all other agents in the two clusters to free ride on the friendly path they have created. Following the same argument as Lemma 1, for any $k_1\in\mathcal{C}_\alpha(\mathbf{H}_0)\setminus\{i\}$ and $k_2\in\mathcal{C}_{\alpha'}(\mathbf{H}_0)\setminus\{j\}$:

$$\begin{aligned}\pi_{k_1}\left(\mathbf{G}^*,\mathbf{H}^*_{-\alpha,\alpha'}\oplus h_{k_1k_2}\right)-\pi_{k_1}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*_{-\alpha,\alpha'}\right) &\geq \pi_i\left(\mathbf{G}^*,\mathbf{H}^*_{-\alpha,\alpha'}\oplus h_{ij}\right)-\pi_i\left(\mathbf{G}^*-g_{ij},\mathbf{H}^*_{-\alpha,\alpha'}\right)\geq 0 \\ \pi_{k_2}\left(\mathbf{G}^*,\mathbf{H}^*_{-\alpha,\alpha'}\oplus h_{k_1k_2}\right)-\pi_{k_2}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*_{-\alpha,\alpha'}\right) &\geq \pi_j\left(\mathbf{G}^*,\mathbf{H}^*_{-\alpha,\alpha'}\oplus h_{ij}\right)-\pi_j\left(\mathbf{G}^*-g_{ij},\mathbf{H}^*_{-\alpha,\alpha'}\right)\geq 0\end{aligned}$$

where at least one of the last inequality in each case is strictly positive. Since $k_1$ and $k_2$ free ride, it follows that:

$$\begin{aligned}\pi_{k_1}\left(\mathbf{G}^*,\mathbf{H}^*\right)-\pi_{k_1}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*\right) &> \pi_{k_1}\left(\mathbf{G}^*,\mathbf{H}^*_{-\alpha,\alpha'}\oplus h_{k_1k_2}\right)-\pi_{k_1}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*_{-\alpha,\alpha'}\right)>0 \\ \pi_{k_2}\left(\mathbf{G}^*,\mathbf{H}^*\right)-\pi_{k_2}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*\right) &> \pi_{k_2}\left(\mathbf{G}^*,\mathbf{H}^*_{-\alpha,\alpha'}\oplus h_{k_1k_2}\right)-\pi_{k_2}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*_{-\alpha,\alpha'}\right)>0\end{aligned}$$

and the result follows.

**Case II:** Suppose a pair of agents, where at least one agent differs from $i$ or $j$, were the ones transforming their affinity relationship. Call this pair of agents transforming their affinity relationship as $\widetilde{i}\in\mathcal{C}_\alpha(\mathbf{H}_0)\cap\Delta_s^\alpha(\mathbf{H}^*)$ and $\widetilde{j}\in\mathcal{C}_\alpha(\mathbf{H}_0)\cap\Delta_{s'}^{\alpha'}(\mathbf{H}^*)$, where $s\geq l$ and $s'\geq l'$. Following the same argument as that in Case I, all agents with friendship measures greater than or equal to those of $\widetilde{i}$ and $\widetilde{j}$ will also have an incentive to form an alliance. Now consider agents $i$ and $j$ from the statement of the proposition. These two agents will free ride on the friendly link created by $\widetilde{i}$ and $\widetilde{j}$ and have a profitable alliance by hypothesis. Thus, for any two agents $k_1$ and $k_2$ whose friendship measures are greater than or equal to those of $i$ and $j$ and who also free ride, we have:

$$\begin{aligned}\pi_{k_1}\left(\mathbf{G}^*,\mathbf{H}^*\right)-\pi_{k_1}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*\right) &\geq \pi_i\left(\mathbf{G}^*,\mathbf{H}^*\right)-\pi_i\left(\mathbf{G}^*-g_{ij},\mathbf{H}^*\right)\geq 0 \\ \pi_{k_2}\left(\mathbf{G}^*,\mathbf{H}^*\right)-\pi_{k_2}\left(\mathbf{G}^*-g_{k_1k_2},\mathbf{H}^*\right) &\geq \pi_j\left(\mathbf{G}^*,\mathbf{H}^*\right)-\pi_j\left(\mathbf{G}^*-g_{ij},\mathbf{H}^*\right)\geq 0\end{aligned}$$

where at least one of the last inequality in each case is strictly positive. This proves the result. ■

**Proof of Proposition 6:** The proof follows from the definition of the threshold value, $\Theta_{\alpha\alpha'}(\mathbf{G}^*, \mathbf{H}^*)$. ■

**Proof of Proposition 7:** To save space, we will suppress reference to $\mathbf{H}_0$.
(*Existence*): We first establish existence. Recall that all alliances are formed within clusters. Consider the network in which each cluster $\mathcal{C}_\alpha$ is complete, i.e., all agents in each cluster are mutually interconnected. Denote this network as $\mathbf{G}^c$. If it is an equilibrium, then we are done. Otherwise, there exists a coalition $S'$ and a network $\mathbf{G}'$ that can be obtained from $\mathbf{G}^c$ by $S'$ such that $\pi_i(\mathbf{G}') > \pi_i(\mathbf{G}^c)$ for all $i \in S'$. Since all alliances are intra-cluster, it implies that $S' \subset \mathcal{C}_\alpha$ for some cluster $\mathcal{C}_\alpha$. Specifically:

$$\pi_i(\mathbf{G}') = \left[v\left(\eta_i\left(\mathbf{G}'\right)\right) - c_0\left(\delta_i\right)\right] + \sum_{j \in \mathbf{N}_i(\mathbf{G}')} \left[w\left(\eta_j\left(\mathbf{G}'\right)\right) - c\left(\delta_j\right)\right] > \pi_i(\mathbf{G}^c), \quad i \in S'$$

Since no new links could be added in $\mathbf{G}^c$, the deviation must involve members in $S'$ deleting their links. This implies in particular that in the cluster $\mathcal{C}_\alpha$:

$$\Delta v\left(|\mathcal{C}_\alpha| - 2\right) + \left[w\left(|\mathcal{C}_\alpha| - 1\right) - c\left(\delta_j\right)\right] < 0, \quad i \in S', \quad j \in \mathbf{N}_i(\mathbf{G}^c)\backslash\mathbf{N}_i(\mathbf{G}') \tag{31}$$

If $\mathbf{G}'$ is an equilibrium, then we are done. Otherwise, there exists a coalition $S''$ that can obtain a network $\mathbf{G}''$ in which each member is strictly better off. We claim that this movement from $\mathbf{G}'$ to $\mathbf{G}''$ can only involve a deletion of links. Suppose to the contrary that the movement from $\mathbf{G}'$ to $\mathbf{G}''$ involves addition of links and let $S' \cap S''$ denote the non-empty subset of agents who are involved in forming alliances, either among themselves or with others in $S''\backslash S'$ in the move from $\mathbf{G}'$ to $\mathbf{G}''$. Note that this intersection cannot be empty because firms in $\mathcal{C}_\alpha \backslash S'$ are completely connected among themselves; thus a member of $S'$ has to be involved if new links are created starting from $\mathbf{G}'$. Consider any $i \in S' \cap S''$. Since $i$ was completely connected in $\mathbf{G}^c$, and deleted links in the move to $\mathbf{G}'$, any new alliance that it forms in the move to $\mathbf{G}''$ must be with some agent $j \in \mathbf{N}_i(\mathbf{G}^c)\backslash\mathbf{N}_i(\mathbf{G}')$ with whom it earlier dissolved an alliance. Since the deviation to $\mathbf{G}''$ is strictly profitable:

$$\Delta v\left(\eta_i\left(\mathbf{G}''\right) - 1\right) + \left[\eta_j\left(\mathbf{G}''\right) - c\left(\delta_j\right)\right] > 0 \tag{32}$$

However, $\eta_i\left(\mathbf{G}''\right) - 1 \leq |\mathcal{C}_\alpha| - 2$ and $\eta_j\left(\mathbf{G}''\right) \leq |\mathcal{C}_\alpha| - 1$. Therefore, using A.2(a), (31) and (32):

$$0 < \Delta v\left(\eta_i\left(\mathbf{G}''\right) - 1\right) + \left[w\left(\eta_j\left(\mathbf{G}''\right)\right) - c\left(\delta_j\right)\right] \leq \Delta v\left(|\mathcal{C}_\alpha| - 2\right) + \left[w\left(|\mathcal{C}_\alpha| - 1\right) - c\left(\delta_j\right)\right] < 0$$

a contradiction. Thus, the move from $\mathbf{G}'$ to $\mathbf{G}''$ involves only deletion of links by agents. If $\mathbf{G}''$ is an equilibrium then we are done, otherwise another coalition could profitably deviate by further deleting links. Since the number of networks are finite, this process of deletion will eventually converge to some $\mathbf{G} \neq \mathbf{G}^e$ or to $\mathbf{G}^e$ from which no coalition can gain through additional deletions. The same argument as the one above establishes that no new links will be formed either. Thus this limit network is an equilibrium network.

(*Proof that $\delta_j \geq \delta_i$ implies $\mathbf{N}_i(\mathbf{G}) \subseteq \mathbf{N}_j(\mathbf{G})$*) Suppose to the contrary that $\mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G}) \neq \emptyset$ in an equilibrium network $\mathbf{G}$. Index the agents such that:

$$\begin{aligned} 1,2,...,L &\in \mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G}) \\ L+1,L+2,...L' &\in \mathbf{N}_i(\mathbf{G})\cap\mathbf{N}_j(\mathbf{G}) \\ L'+1,L'+2,...,L'' &\in \mathbf{N}_j(\mathbf{G})\backslash\mathbf{N}_i(\mathbf{G}) \end{aligned}$$

Let $\mathbf{G}' = \mathbf{G} - \sum_{l=1}^{L} g_{il}$ denote the network in which agent $i$ has deleted all the links in $\mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G})$. Since agent $i$ has no incentive to delete any subset of links:

$$\pi_i(\mathbf{G})-\pi_i(\mathbf{G}') = \left[v\left(\eta_i\left(\mathbf{G}\right)\right) - v\left(\eta_i\left(\mathbf{G}'\right)\right)\right] + \sum_{l=1}^{L}\left[w\left(\eta_l\left(\mathbf{G}\right)\right) - c\left(\delta_l\right)\right] + \sum_{l=L+1}^{L'}\left[w\left(\eta_l\left(\mathbf{G}\right)\right) - w\left(\eta_l\left(\mathbf{G}'\right)\right)\right] \geq 0$$

Now consider the coalition $S = \{j\}\cup\mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G})$ and let $\mathbf{G}'' = \mathbf{G}' + \sum_{l=1}^{L} g_{jl}$ denote the network in which each agent $l \in \mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G})$ deletes its alliance with $i$ and forms an alliance with $j$. Note that $\eta_l\left(\mathbf{G}''\right) = \eta_l\left(\mathbf{G}\right) = \eta_l\left(\mathbf{G}'\right)+1$ for $l \in \mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G})$. Further, $\eta_i\left(\mathbf{G}'\right) = \eta_j\left(\mathbf{G}\right)$ and $\eta_i\left(\mathbf{G}\right) = \eta_j\left(\mathbf{G}''\right)$. For agent $j$:

$$\begin{aligned} \pi_j(\mathbf{G}'') - \pi_j(\mathbf{G}) &= \left[v\left(\eta_j\left(\mathbf{G}''\right)\right) - v\left(\eta_j\left(\mathbf{G}\right)\right)\right] + \sum_{l=1}^{L}\left[w\left(\eta_l\left(\mathbf{G}\right)\right) - c\left(\delta_l\right)\right] \\ &+ \sum_{l=L+1}^{L'}\left[w\left(\eta_l\left(\mathbf{G}''\right)\right) - w\left(\eta_l\left(\mathbf{G}\right)\right)\right] + \sum_{l=L'+1}^{L''}\left[w\left(\eta_l\left(\mathbf{G}''\right)\right) - w\left(\eta_l\left(\mathbf{G}\right)\right)\right] \end{aligned}$$

Comparing terms, it follows that $\pi_j(\mathbf{G}'') - \pi_j(\mathbf{G}) > \pi_i(\mathbf{G}) - \pi_i(\mathbf{G}') \geq 0$. Therefore $j$ has a strict incentive to form links with all agents in $\mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G})$ and move from network $\mathbf{G}$ to $\mathbf{G}''$. We now show that each agent $l$ in $\mathbf{N}_i(\mathbf{G})\backslash\mathbf{N}_j(\mathbf{G})$ has a strict incentive to reciprocate the alliance with $j$. From the equilibrium property of $\mathbf{G}$, $l$ will not delete the link with $i$:

$$\pi_l(\mathbf{G}) - \pi_l(\mathbf{G} - g_{il}) = \left[v\left(\eta_l\left(\mathbf{G}\right)\right) - v\left(\eta_l\left(\mathbf{G} - g_{il}\right)\right)\right] + \left[w\left(\eta_i\left(\mathbf{G}\right)\right) - c\left(\delta_i\right)\right] \geq 0$$

By forming an alliance with $j$ by joining the coalition $S$:

$$\pi_l(\mathbf{G}'') - \pi_l(\mathbf{G}) = \left[v\left(\eta_l\left(\mathbf{G}''\right)\right) - v\left(\eta_l\left(\mathbf{G}\right)\right)\right] + \left[w\left(\eta_j\left(\mathbf{G}''\right)\right) - c\left(\delta_j\right)\right]$$

From A.2(b), $v\left(\eta_l\left(\mathbf{G}''\right)\right) - v\left(\eta_l\left(\mathbf{G}\right)\right) > \left(\eta_l\left(\mathbf{G}\right)\right) - v\left(\eta_l\left(\mathbf{G}'\right)\right)$. Further, from A.3(a), $c\left(\delta_j\right) \leq c\left(\delta_i\right)$ since $\delta_j \geq \delta_i$. Finally, $w\left(\eta_i\left(\mathbf{G}\right)\right) = w\left(\eta_j\left(\mathbf{G}''\right)\right)$ since $\eta_i\left(\mathbf{G}\right) = \eta_j\left(\mathbf{G}''\right)$. Therefore, $\pi_l(\mathbf{G}'') - \pi_l(\mathbf{G}) > 0$. Therefore, given the network $\mathbf{G}$, we have identified a coalition $S$, and a network $\mathbf{G}''$ that can be obtained from $\mathbf{G}$, such that all agents in $S$ are strictly better off. This contradicts the starting hypothesis that $\mathbf{G}$ is an equilibrium network.

(*NSG characterization*) The method of proof differs from that in Propositions 2 and 3 which depended on the specifics of the link formation game. Dropping reference to $\mathbf{H}_0$, consider the cluster $\mathcal{C}_\alpha$. To avoid trivialities, assume no agent is isolated in $\mathbf{G}$. Let $\Delta^{\alpha}\mathcal{C}_{\alpha} = \{\Delta_1^{\alpha}, .., \Delta_r^{\alpha}, .., \Delta_s^{\alpha}\}$ denote the friendship partition and $\mathcal{D}^{\alpha}(\mathbf{G}) = \{D_0^{\alpha}(\mathbf{G}), D_1^{\alpha}(\mathbf{G}), ..., D_m^{\alpha}(\mathbf{G})\}$ denote the degree partition of agents belonging to $\mathcal{C}_{\alpha}$. Consider $i \in \Delta_s^{\alpha}$ and note that $\delta_i \geq \delta_j$ for all $j \in \mathcal{C}_{\alpha}\backslash\{i\}$. Thus, $\mathbf{N}_j(\mathbf{G}) \subseteq \mathbf{N}_i(\mathbf{G})$ for all $j \in \mathcal{C}_{\alpha}\backslash\{i\}$ implies $\cup_{j\neq i}\mathbf{N}_j(\mathbf{G}) \subseteq \mathbf{N}_i(\mathbf{G})$. Since there are no isolated agents, it follows that agents in $\Delta_s^{\alpha}$ are connected to all agents in $\mathcal{C}_{\alpha}$ and thus $\Delta_s^{\alpha} \subseteq D_m^{\alpha}(\mathbf{G})$. Now consider an agent $i \in \Delta_1^{\alpha}$. Since $\delta_i \leq \delta_j$ for all $j \in \mathcal{C}_{\alpha}\backslash\{i\}$, it follows that $\mathbf{N}_i(\mathbf{G}) \subseteq \mathbf{N}_j(\mathbf{G})$. We now claim that $\mathbf{N}_i(\mathbf{G}) = D_m^{\alpha}(\mathbf{G})$. Suppose $k \notin D_m^{\alpha}(\mathbf{G})$ but $k \in \mathbf{N}_i(\mathbf{G})$. Then, $k \in \mathbf{N}_i(\mathbf{G}) \subseteq \mathbf{N}_j(\mathbf{G})$ for all $j \in \mathcal{C}_{\alpha}\backslash\{i\}$. Therefore, $k \in D_m^{\alpha}(\mathbf{G})$, a contradiction. It also follows that $\Delta_1^{\alpha} \subseteq D_1^{\alpha}(\mathbf{G})$ establishing that the characterization of Proposition 2 holds.

We now turn to the intermediate elements of the degree partition $\mathcal{D}^{\alpha}(\mathbf{G})$. Let $l_1, l_2, ..., l_m$ denote agents from $D_1^{\alpha}(\mathbf{G}), D_2^{\alpha}(\mathbf{G}), ..., D_m^{\alpha}(\mathbf{G})$ respectively who have the lowest friendship measure in their set. Consider $l_2 \in D_2^{\alpha}(\mathbf{G})$ and note that $D_m^{\alpha}(\mathbf{G}) \subset \mathbf{N}_{l_2}(\mathbf{G})$. It is a *proper* subset because we have already established that $D_m^{\alpha}(\mathbf{G})$ is also the neighborhood for agents in $D_1^{\alpha}(\mathbf{G})$ and $l_2$ has strictly more alliances than those in $D_1^{\alpha}(\mathbf{G})$. We now argue that the additional alliances of $l_2$ must be with agents in $D_{m-1}^{\alpha}(\mathbf{G})$. Suppose to the contrary that $k \in \mathbf{N}_{l_2}(\mathbf{G})$ but $k \notin D_{m-1}^{\alpha}(\mathbf{G}) \cup D_m^{\alpha}(\mathbf{G})$. Then, $\eta_k(\mathbf{G}) < \eta_{l_{m-1}}(\mathbf{G})$. Since $\delta_{l_2} \leq \delta_{l_3} \leq \cdot\cdot \leq \delta_{l_{m-1}} \leq \delta_{l_m}$, we have $k \in \mathbf{N}_{l_2}(\mathbf{G}) \subseteq \mathbf{N}_{l_3}(\mathbf{G}) \subseteq \cdots \subseteq \mathbf{N}_{l_m}(\mathbf{G})$. Therefore, $\mathbf{N}_k(\mathbf{G}) = D_2^{\alpha}(\mathbf{G}) \cup D_3^{\alpha}(\mathbf{G}) \cup \cdot\cdot \cup D_m^{\alpha}(\mathbf{G})$. We have already proved that $\mathbf{N}_{l_{m-1}}(\mathbf{G}) \cap D_1^{\alpha}(\mathbf{G}) = \emptyset$ and, therefore, $\mathbf{N}_{l_{m-1}}(\mathbf{G}) \subseteq D_2^{\alpha}(\mathbf{G}) \cup D_3^{\alpha}(\mathbf{G}) \cup \cdot\cdot \cup D_m^{\alpha}(\mathbf{G}) = \mathbf{N}_k(\mathbf{G})$. Thus, $\eta_k(\mathbf{G}) \geq \eta_{l_{m-1}}(\mathbf{G})$, a contradiction to $\eta_k(\mathbf{G}) < \eta_{l_{m-1}}(\mathbf{G})$. $n_{l_{m-1}}(g) \leq n_k(g)$, a contradiction. Therefore, $\mathbf{N}_{l_2}(\mathbf{G}) = D_{m-1}^{\alpha}(\mathbf{G}) \cup D_m^{\alpha}(\mathbf{G})$ for $l_2 \in D_2^{\alpha}(\mathbf{G})$. Continuing inductively, for any $i \in D_{x+1}^{\alpha}(\mathbf{G})$, $1 \leq x < \lfloor \frac{m}{2} \rfloor$, $\mathbf{N}_i(\mathbf{G}) = D_{m-x}^{\alpha}(\mathbf{G}) \cup D_{m-x+1}^{\alpha}(\mathbf{G}) \cup \cdots \cup D_{m-1}^{\alpha}(\mathbf{G}) \cup D_m^{\alpha}(\mathbf{G})$, and for any $j \in D_{m-x}^{\alpha}(\mathbf{G})$, $\mathbf{N}_j(\mathbf{G}) = D_{x+1}^{\alpha}(\mathbf{G}) \cup D_{x+2}^{\alpha}(\mathbf{G}) \cup \cdots \cup D_{m-1}^{\alpha}(\mathbf{G}) \cup D_m^{\alpha}(\mathbf{G})$. This proves the NSG property. ■

**Proof of Proposition 8:** It will be convenient to let:

$$\Delta\psi(\eta, \eta') = \psi(\eta + 1, \eta') - \psi(\eta, \eta')$$

Agents $i$ and $j$ have a mutually profitable link in $\mathbf{G}$. Noting that $j \notin \mathbf{N}_i(\mathbf{G})$, and dropping reference to $\mathbf{G}$ to simplify the notation:

$$\sum_{l \in \mathbf{N}_i(\mathbf{G})} \Delta\psi(\eta_i, \eta_l) + \left[\psi(\eta_i + 1, \eta_j + 1) - c(\delta_j, \eta_j + 1)\right] \geq 0 \quad (33)$$

$$\sum_{l \in \mathbf{N}_j(\mathbf{G})} \Delta\psi(\eta_j, \eta_l) + \left[\psi(\eta_j + 1, \eta_i + 1) - c(\delta_i, \eta_i + 1)\right] \geq 0 \quad (34)$$

and at least one of the inequalities is strict. Letting $\eta_l'$ denote the degree of agent $l$ in $\mathbf{G}' \supseteq \mathbf{G}$ (where once again reference to the network $\mathbf{G}'$ is dropped), $\eta_l' \geq \eta_l$ for all $l \in \mathcal{N}$. The incremental payoffs to agents $k$

and $j$ respectively from forming a link in $\mathbf{G}'$ is equal to:

$$\sum_{l\in\mathbf{N}_k(\mathbf{G}')} \Delta\psi\left(\eta'_k,\eta'_l\right) + \left[\psi\left(\eta'_k+1,\eta'_j+1\right) - c\left(\delta_j,\eta'_j+1\right)\right] \tag{35}$$

$$\sum_{l\in\mathbf{N}_j(\mathbf{G}')} \Delta\psi\left(\eta'_j,\eta'_l\right) + \left[\psi\left(\eta'_j+1,\eta'_k+1\right) - c\left(\delta_k,\eta'_k+1\right)\right] \tag{36}$$

Note that $\mathbf{N}_i(\mathbf{G}) \subseteq \mathbf{N}_k(\mathbf{G}')$ and $\mathbf{N}_j(\mathbf{G}) \subseteq \mathbf{N}_j(\mathbf{G}')$. For all $l \in \mathbf{N}_i(\mathbf{G})$, since $\eta'_k \geq \eta_i$, it follows from respectively parts (b) and (d) of A.2* that $\Delta\psi(\eta'_k,\eta'_l) \geq \Delta\psi(\eta_i,\eta'_l) \geq \Delta\psi(\eta_i,\eta_l)$. Further, from A.2*(a), $\psi\left(\eta'_k+1,\eta'_j+1\right) \geq \psi\left(\eta_i+1,\eta'_j+1\right) \geq \psi\left(\eta_i+1,\eta_j+1\right)$. Further, since $\eta'_j \geq \eta_j$, from A.3*(a), $c\left(\delta_j,\eta'_j+1\right) \leq c\left(\delta_j,\eta_j+1\right)$. Therefore, each term in (35) dominates the corresponding term in (33). Likewise, noting that $\delta_k \geq \delta_i$, each term in (36) dominates the corresponding term in (34). This proves the result. ■

**Proof of Lemma 1 for the Non-separable Case:** Dropping reference to $\mathbf{G}_1$, we will let $\eta_i = \eta_i(\mathbf{G}_1)$ and $\eta_i + 1 = \eta_i(\mathbf{G}_1 + g_{ij})$. Then, $\pi_i(\mathbf{G}_1 + g_{ij}, \mathbf{H}_0 \oplus h_{ij}) - \pi_i(\mathbf{G}_1, \mathbf{H}_0)$ is equal to:

$$\sum_{h\in\mathbf{N}_i(\mathbf{G}_1)} \Delta\psi(\eta_i,\eta_h) + \left[c(\delta_i,\eta_i) - c(\delta_i+1,\eta_i+1)\right] + \left[\Theta_{ij}\psi\left(\eta_i+1,\eta_j+1\right) - c\left(\delta_j,\eta_j+1\right)\right]$$

Similarly, $\pi_k(\mathbf{G}_1 + g_{kl}, \mathbf{H}_0 \oplus h_{kl}) - \pi_k(\mathbf{G}_1, \mathbf{H}_0)$ is equal to:

$$\sum_{h\in\mathbf{N}_k(\mathbf{G}_1)} \Delta\psi(\eta_k,\eta_h) + \left[c(\delta_k,\eta_k) - c(\delta_k+1,\eta_k+1)\right] + \left[\Theta_{kl}\psi(\eta_k+1,\eta_l+1) - c(\delta_l,\eta_l+1)\right]$$

Since there is an NSG structure within each cluster with degree positively related to friendship, $\mathbf{N}_k(\mathbf{G}_1) \subseteq \mathbf{N}_i(\mathbf{G}_1)$ and thus $\eta_i \geq \eta_k$. It follows from parts (b) and (c) respectively of A.2* that:

$$\sum_{h\in\mathbf{N}_i(\mathbf{G}_1)\backslash\{k\}} \Delta\psi(\eta_i,\eta_h) + \Delta\psi(\eta_i,\eta_k) \geq \sum_{h\in\mathbf{N}_k(\mathbf{G}_1)\backslash\{i\}} \Delta\psi(\eta_k,\eta_h) + \Delta\psi(\eta_k,\eta_i)$$

Since $i \in \mathcal{C}_\alpha(\mathbf{H}_0) \cap \Delta_s^\alpha(\mathbf{H}_0)$, it is true that $\delta_i \geq \delta_k$. Combining with $\eta_i \geq \eta_k$, it follows that:

$$c(\delta_k+1,\eta_i+1) - c(\delta_i+1,\eta_i+1) \geq c(\delta_k+1,\eta_i) - c(\delta_i+1,\eta_i) \geq c(\delta_k,\eta_i) - c(\delta_i,\eta_i)$$

where the first and second inequalities follow respectively from parts (d) and (b) of A.3*. Rearranging the terms:

$$c(\delta_i,\eta_i) - c(\delta_i+1,\eta_i+1) \geq c(\delta_k,\eta_i) - c(\delta_k+1,\eta_i+1) \tag{37}$$

Further, note that:

$$c(\delta_k+1,\eta_k+1) - c(\delta_k+1,\eta_i+1) \geq c(\delta_k,\eta_k+1) - c(\delta_k,\eta_i+1) \geq c(\delta_k,\eta_k) - c(\delta_k,\eta_i)$$

where the first and second inequalities follow respectively from parts (d) and (c) of A.3*. Rearranging the terms:

$$c(\delta_k, \eta_i) - c(\delta_k + 1, \eta_i + 1) \geq c(\delta_k, \eta_k) - c(\delta_k + 1, \eta_k + 1) \tag{38}$$

From (37) and (38), it follows that:

$$c(\delta_i, \eta_i) - c(\delta_i + 1, \eta_i + 1) \geq c(\delta_k, \eta_k) - c(\delta_k + 1, \eta_k + 1)$$

Finally, note that $\Theta_{kl} = \Theta_{ij}$, $\eta_j \geq \eta_l$ and $\delta_j \geq \delta_l$. Therefore, from A.2*(a):

$$\psi\left(\eta_i + 1, \eta_j + 1\right) \geq \psi\left(\eta_k + 1, \eta_j + 1\right) \geq \psi\left(\eta_k + 1, \eta_l + 1\right)$$

and from A.3*(a):

$$c\left(\delta_j, \eta_j + 1\right) \leq c\left(\delta_l, \eta_j + 1\right) \leq c\left(\delta_l, \eta_l + 1\right)$$

It follows that:

$$\pi_i\left(\mathbf{G}_1 + g_{ij}, \mathbf{H}_0 \oplus h_{ij}\right) - \pi_i\left(\mathbf{G}_1, \mathbf{H}_0\right) \geq \pi_k\left(\mathbf{G}_1 + g_{kl}, \mathbf{H}_0 \oplus h_{kl}\right) - \pi_k\left(\mathbf{G}_1, \mathbf{H}_0\right)$$

The verification for agents $j$ and $l$ is identical. ■

# References


[1] R. Aumann and R. Myerson, 1988. Endogenous Formation of Links between Players and Coalitions: An Application of the Shapley Value. In The Shapley Value, (A. Roth, editor), 175-191, Cambridge University Press.

[2] S. Baliga and T. Sjöstrom, 2012. The Strategy of Manipulating Conflict, *American Economic Review*, 102(6), 2897-2922.

[3] Becker C., P. Devine, H. Dogo and E. Margolin, 2022. Marking Territory: Modeling the Spread of Ethnic Conflict in Bosnia and Herzegovina, 1992-1995 SSRN: https://ssrn.com/abstract=3160015.

[4] F. Bloch, 2012. Endogenous Formation of Alliances in Conflict. In The Oxford Handbook of the Economics of Peace and Conflict (M.R. Garfinkel and S. Skaperdas, editors), 473–502, Oxford University Press.

[5] D. Cartwright and F. Harary, 1956. Structural Balance: A Generalization of Heider's Theory, *Psychological Review*, 63(5), 277–293.

[6] J.A. Davis, 1967. Clustering and Structural Balance in Graphs, *Human Relations*, 20(2), 181–187.

[7] B. Dutta, A. van den Nouweland and S. Tijs, 1998. Link Formation in Cooperative Situations, *International Journal of Game Theory*, 27, 245–256.

[8] B. Dutta and S. Mutuswami, 1997. Stable networks, *Journal of Economic Theory* 76, 322–344.

[9] D. Easley and J. Kleinberg, 2010. Networks, Crowds and Markets: Reasoning About a Highly Connected World. Cambridge University Press, New York.

[10] J. Franke and T. Öztürk, 2015. Conflict Networks, *Journal of Public Economics*, 126,104-113.

[11] S. Goyal, A. Vigier, and M. Dziubinski, 2016. Conflict and Networks. In The Oxford Handbook of the Economics of Networks (Y. Bramoullé, A. Galeotti, and B. Rogers, editors), Oxford University Press.

[12] T. Hiller, 2017. Friends and Enemies: A Model of Signed Network Formation, *Theoretical Economics*, 12, 1057-1087.

[13] K. Huremovic, 2015. A Noncooperative Model of Contest Network Formation, *Journal of Public Economic Theory*, 23(5). DO1:10.1111/jpet.12475.

[14] M.O. Jackson and A. Wolinsky, 1996. A Strategic Model of Social and Economic Networks, *Journal of Economic Theory*, 71, 44-74.

[15] M.O. Jackson and A. Watts, 2002. The Evolution of Social and Economic Networks, *Journal of Economic Theory*, 106, 265-295.

[16] M.O. Jackson and S. Nei, 2015. Networks of Military Alliances, Wars and International Trade. Proceedings of the National Academy of Sciences of the United States of America, 112, 15277–15284.

[17] S. Joshi, A.S. Mahmud and S. Sarangi, 2020. Network Formation with Multigraphs and Strategic Complementarities, *Journal of Economic Theory*, 180, 105033, SSN 0022-0531, https://doi.org/10.1016/j.jet.2020.105033.

[18] S. Joshi, A.S. Mahmud and H. Tzavellas, 2023. multilayer Formation. SSRN: https://ssrn.com/abstract=4499938.

[19] M.D. König, D. Rohner, M. Thoenig and F. Zilibotti, 2017. Networks in Conflict: Theory and Evidence from the Great War of Africa, *Econometrica*, 85(4), 1093-1132.

[20] R.P. Kundu and S.G. Pandey, 2023. Cooperation and Balance in Signed Networks: A Model of multilayer Formation, SSRN: https://ssrn.com/abstract=4499176.