

The Structure of Information Networks¹

Sudipta Sarangi² Rajgopal Kannan³ Lydia Ray⁴

January 2003

¹We are grateful to Rob Gilles, Hans Haller, Georg Erber, Bob Martin, Johanna Francis, Cheryl Long and Micheal Kosfeld for helpful suggestions. We also thank Susanne Maria Schmidt for many useful conversations. The paper has also benefitted from the comments of seminar participants at Louisiana State University, University of Arkansas, Colgate University, Midwest Economic Theory Meetings and the Game Theory and Applications Conference in Mumbai. Sudipta Sarangi gratefully acknowledges the hospitality of DIW Berlin where a part of this research was carried out. All remaining errors are our responsibility.

²Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA. email: *sarangi@lsu.edu*

³Department of Computer Science, Louisiana State University, Baton Rouge, LA 70803, USA. email: *rkannan@csc.lsu.edu*

⁴Department of Computer Science, Louisiana State University, Baton Rouge, LA 70803, USA. email: *lydia@csc.lsu.edu*

Abstract

We develop a model of information acquisition in a network where agents pay for all the information they acquire including those through indirect links. The cost of information depends on the value of the information itself and the distance it traverses in the network. We find that when the costs of information increase with distance, the complete network is the only Nash network. When cost of information decreases with distance or delay, all equilibrium information networks are minimally connected, though not all trees are Nash. We analyze the popular star and chain networks and identify strict Nash networks. We show that there is almost no divergence between the efficient and equilibrium information networks. We explore the implications of a spatial model and information decay and discuss the relationship with experimental evidence.

JEL Classification: D82, D83

Keywords: Information networks, Nash networks, Efficient networks

1 Introduction

Communication forms one of the major pillars of all societies and economic systems. It leads to the dissemination of information, helping well informed agents make better decisions. The problem of constructing a rational economic order as Hayek (1945) says: “If we possess all the relevant information, . . . , the problem which remains is purely one of logic.” Often such communication takes place through a network of bilateral relationships between the participants. This paper develops a model of the formation of information networks where agents pay for all the information acquired through the network. The actions of agents consists of choosing their link partners resulting in different network configurations. Network structure determines the payoffs and hence the set of stable and efficient networks.

Road building in the ancient empires provides the earliest example of information networks. The legendary Roman roads were intended for the quick movement of troops, couriers carrying information and government officials. Casson (1994) attributes the rise of the Assyrians from 900 to 612 B.C. to their network of roads, superior organization and discipline. The Persian empire which was founded approximately half a century later greatly surpassed the Assyrian kingdom in its geographic extant. Casson (1994, pg. 53) explains that one of the secrets of their success was swift and sure communication between the capital and the most distant centers. Their ‘royal road’, maintained primarily for government couriers but open to all, ran from Sardis near the east coast of the Mediterranean, some 1600 miles to Susa the Persian capital, not far from the head of the Persian Gulf. The Persian dispatch service was one of the most efficient arms of the state. The mounts for the service were supplied by the king, its riders formed an elite unit, and the administrative head of the service was one of the government’s high dignitaries.

Governments however, were not the only ones to build these networks. Medieval trading communities often had their own informal networks for the dissemination of information. The Cairo *geniza* documents have provided a wealth of information about such networks.¹ Greif (1993) shows that the Maghribi traders relied on such a network to obtain information

¹Amitav Ghosh in his book *In an Antique Land* (1994) traces the life of a slave through letters between merchants in Cairo and Mangalore (India) via Yemen. The bulk of information in such letters was about prices, profits and execution of orders as the merchants acted as agents for each other.

about far-flung markets and ensure contract enforcement among trade partners located in distant lands. The traders behaved as a coalition that meted out collective punishment for dishonesty by an agent.

The advent of new means of communication has always resulted in the growth of economic activity. Railways helped link the two coasts of the continental United States leading to an increase in trade opportunities (see for example Stover (1999)). The interlinked system of national highways dramatically increased the mobility of goods and services in the post-World War II United States. In her book on the social history of American technology, Cowan (1997) shows how economic activity received a boost with the development of the telegraph, telephone, wireless and fax.

Indeed rapid technological advances in recent years such as the Internet and wireless communication have created a global web which facilitates rapid transmission of information in a manner that is hitherto unprecedented. The Internet can be used to make payments, trade stocks, check the weather, express opinions, listen to music and buy or sell a wide range of products. Law enforcement officers use it to catch criminals while members of churches, clubs and community groups keep in touch with each other via email. The biggest impact on the economy can be attributed to the efficiency gains from B2B (business to business) and B2C (business to consumers) transactions. In a recent study Litan and Rivlin (2002) find that it has a significant impact (0.25-0.5% annually) on US productivity growth while other studies have shown strong synergies between education, life expectancy, income and such information networks (International Telecommunications Union, 1999). Very little of the research in this area however, takes the motivations of the individual participants in to account. Our paper examines the formation of information networks by focusing on the incentives of individual agents in the network.

Information networks were first analyzed by graph theorists in the context of *gossip* and *broadcast* problems.² In a gossip network every individual posses a unique piece of gossip which needs to be communicated to the others

²The use of networks as an institution to study human interaction was pioneered by sociologists in the 1960s. Milgram (1967) showed that on average people, regardless of whether we consider farmers in Nebraska and stockbrokers in Boston, have six degrees of separation. Since then such models have gained wide acceptance and have been used to study diverse social phenomena ranging from job-referral networks (Granovetter (1974) and Montgomery (1991)) to power within groups (Markovsky, Willer and Patton, (1988)).

(Baker and Shostak (1972)). In the broadcast version on the other hand, one person wishes to communicate information to all others in the network. A survey of this literature, including the basic problem and its extensions can be found in Hedetniemi et al. (1991). The focus of these models is mainly on some aggregate network criteria such as the minimum number of links needed to ensure that the gossip reaches everyone, or the minimum number of rounds required before all members of the network hear the gossip. Similar considerations predominate the broadcast models. This literature rarely considered the costs of the links or individual payoffs. In other words, strategic interactions are conspicuously absent from this work.

Jackson and Wolinsky (1996) study strategic behavior in networks using an equilibrium concept called pairwise stability.³ Soon after the notion of Nash networks was introduced in a paper by Bala and Goyal (2000a; henceforth [BG]). The model investigates both static as well as dynamic network formation. The basic structure is similar to the connections model of Jackson and Wolinsky (1996), but relies on Nash equilibrium as the stability concept.⁴ Link formation is unilateral with the initiating player incurring all the costs of establishing the link. Further, a link can only be broken by the initiating player. As in the previous model, if player i has a link with player j , then i can access j 's information as well as the information of all the other players j is linked to without having to pay for the indirect links. The paper analyzes information flow in a directed and undirected network both in the presence and absence of information decay.

Our paper builds on the work on Nash networks by Bala and Goyal (2000a) and focuses on the structure of an information network with agents paying for all acquired information. We assume that every agent has a unique piece of information with some intrinsic value and would like to gather as much information as possible by linking to other agents. Our other modeling assumptions however are quite different from those in the models mentioned above. Unlike much of the earlier work, we incorporate an element of realism by allowing each agent to have a different endowment

³Aumann and Myerson (1988) is perhaps the first to introduce a strategic version of the problem but does not provide a complete characterization of the solution. An excellent survey of the pairwise stability literature can be found in Jackson (2001). Also, Watts (2001) and Jackson and Watts (1998) extend the basic model by studying dynamic network formation.

⁴A third strand of the literature looks at network formation using cooperative game theory. For a comprehensive survey see Slikker and van den Nouweland (2001).

of information. Secondly, information seekers have to pay for *all* the information they acquire through the links. We believe that many of the results in the earlier papers are due to the fact that information obtained through indirect links is free. In order to understand the implications of this externality, we require agents to pay for all the information they receive, including those acquired through indirect links. Since the externality is stronger in the case of two-way flow, we analyze the bi-directional information flow model. Third, instead of using an exogenously given link cost, we develop an alternative cost formulation with two components. The first component requires that costs be in direct proportion to the value of information. In other words, information of higher value costs more. The second component allows costs of information to vary with the distance it travels in the network.⁵

We consider two possible cases of relating costs with distance. We now discuss each of these and explain under what situations these possibilities can arise. One allows information coming from a greater distance to be more expensive. This is applicable to information networks where physical distance is relevant for exchanging information. For example, international phone calls cost more than domestic ones. It is also applicable to the examples of historical information networks mentioned at the beginning of this section. Given that merchant ships served as the means of communication between trading partners, greater distance meant greater costs. The alternative cost formulation allows for the cost of information to vary inversely with distance. This implies longer paths in the network lead to information delays or waiting time and hence are cheaper. This cost formulation is especially appropriate for some types of electronic networks. Network applications can be characterized by their differing Quality of Service (QoS) requirements from the underlying network architecture, for example Constant Bit Rate (CBR) traffic streams or Variable Bit Rate (VBR) applications such as real-time video with high bandwidth and low delay requirements. While the technological aspects of this problem have been extensively studied, there is a growing literature on pricing in computer networks (see for instance Sairamesh et al. (1995)). Recent emphasis has been placed on the examination of cost formulations that enable users to select from a range of differentiated services. With the advent of cost based service provisioning network users are expected to pay for or at least share the cost of each network resource they utilize. For example, Herzog, Shenker and Estrin

⁵Distance based cost functions are also used in a paper by Johnson and Gilles (2000) who analyze the connections model of Jackson and Wolinsky (1996) in a spatial context.

(1997) use cooperative game theory to suggest a cost based pricing strategy for the Internet. Network users can be viewed as independent agents who are maximizing their payoffs in terms of costs and QoS benefits by selecting appropriate sets of users with whom to share network resources. This situation is modeled by our second cost formulation where network users depend on strategic reasoning to form an equilibrium network that maximizes their payoff. In particular this can be used to determine the equilibrium network architecture for users subscribing to data transfer services. Data traffic accounts for a significant proportion of commercial traffic on the Internet. The vast majority of this traffic is classified as Available Bit Rate (ABR) whose bandwidth requirements are *elastic* and suitable for applications like email, file transfer or web browsing. These applications cannot tolerate any packet (or information) loss but have flexible delay or bandwidth requirements. Therefore, as in the model, users will accept the routing of their data via longer paths in exchange for lower costs. The paper investigates the effect of this property on network architecture.

We find that when information costs increase with distance the only Nash network is the complete network. But when information costs decrease with distance, we find that equilibrium information networks have a minimally connected structure. We investigate these equilibrium trees further using a specific payoff formulation. We find that the center-initiated star continues to be Nash along with different types of chains. Chain networks are strict Nash and the only subgame perfect equilibria. Interestingly, when all agents have identical values a Nash network will never contain a periphery-initiated star as a subgraph, although this is possible with heterogeneity in the agents' information endowment. With heterogeneous information endowments a periphery-initiated star is in equilibrium only when the player set is small. We also identify the general conditions under which all Nash networks will be minimally-connected. We find that in the absence of indirect link externalities, Nash and efficient information networks have the same architecture, though the two networks may not coincide.

Section 2 describes the basic model. In section 3 we analyze the stability and efficiency properties of networks. Section 4 summarizes the main results, explores alternative formulations and has comments on recent experimental evidence on network formation.

2 Model

Let $I = \{1, 2, \dots, n\}$ be the set of agents with $n \geq 3$. For ordered pairs $(i, j) \in I \times I$ the shorthand notation ij is used. Individuals in the model are information seekers who gain utility from having more information. Each individual $i \in I$ has an information endowment of value $V_i \geq 0$, and $V_i \neq V_j$. The possibility that agents can have different values of information is motivated by the idea that not all information has the same value. It describes the *heterogeneity of agents' endowments*.⁶ All agents are aware of the value of the *non-rival* information possessed by other agents. Access to the information possessed by other agents can be gained by forming links with them and through links established by other connected agents. Agents simultaneously form links with other individuals resulting in a network that allows for the two-way flow of information between them. Thus each agent's strategy consists of deciding the set of agents with whom to link.

Strategies. Formally, a strategy of agent $i \in I$ is a vector $g_i = (g_{i1}, g_{i2}, \dots, g_{ii-1}, g_{ii+1}, \dots, g_{in})$ where $g_{ij} \in \{0, 1\}$ for each $j \in I \setminus \{i\}$. If i forms a link with j , then $g_{ij} = 1$, and if no link exists between i and j then $g_{ij} = 0$. Links allow information to flow both ways, i.e., if $g_{ij} = 1$ then information can flow from i to j as well as from j to i .⁷ If there is no link between two agents, then information does not flow between them. In this paper we restrict attention only to pure strategies and fully reliable links.

The set of all strategies of individual i is denoted by \mathcal{G}_i with $g_i \in \mathcal{G}_i$ being the generic member. The set of all possible networks is denoted by $\mathcal{G} = \times_{i=1}^n \mathcal{G}_i$. A strategy profile represented as $g = (g_1, g_2, \dots, g_n) \in \mathcal{G}$ is equivalent to a (directed) **network**, where each vertex depicts an individual and each link forms an edge with the direction pointing towards the person with whom the link was established.⁸ We now introduce some graph-theoretic definitions used in the paper.

The graph-theoretic concepts defined here are based on West (1996) and

⁶This is contrast to *heterogeneity in information*. Note that when $V_i = V$ for all i , making the endowments equal in value, each agent's information is still unique and therefore heterogeneous.

⁷Similarly information also flows both ways when $g_{ji} = 1$. A directed edge indicates the initiating agent for a particular link and not the direction of information flow as in some directed network models.

⁸At the risk of abusing notation we will use g to denote both the directed and associated undirected graph and use other labels to distinguish between different directed graphs.

are defined on the undirected graph. A *walk* is a sequence of vertices and edges in a graph such that each vertex belongs only to the preceding and succeeding edge. In a directed graph this must follow the direction of the arrows. A (open) walk with no repeated vertices is called a **path**. A network g is said to be *connected* if there is a path between any two agents ij in the network. We use $g(i)$ to denote the connected subgraph to which player i belongs. A connected graph with *no cycles* (or loops) is called a **tree**.⁹ A **leaf** node is a terminal vertex. Let l_{ij} denote the distance from i to j . Then the *diameter* D of a graph is the maximum distance l_{ij} over vertex pairs ij . A network is said to be **super-connected** if it is still connected after the deletion of *any* link. We now define some of the common types of networks that arise in our paper.

An **empty network** is one where $g_{ij} = 0$ for all pairs ij and a **complete network** is a graph in which every player has a direct link to *every* other player. A **center-initiated** star is an acyclic network where one agent (the central agent) establishes the direct link with all the other $(n - 1)$ agents. Similarly, a **periphery-initiated** star is the acyclic network where all the other $(n - 1)$ agents initiate exactly one link – with the central agent, while a mixed star is a combination of these two types of stars with the central agent initiating some links and the peripheral agents initiating the others. A connected acyclic network with exactly one path is called a **chain**. Finally, two networks g and g' are equivalent if g' is obtained as a permutation of the strategies of the agents in g . The equivalence relation partitions \mathcal{G} into classes and each class is referred to as an *architecture*. We next define the payoff function of each player with players paying for all information – those obtained through direct as well as indirect links.

Benefits. The benefits of player i are given by the total information that she can access from the connected component of the network to which she belongs, i.e., $B_i(g) = \sum_{j \in g(i)} V_j$. The process of information transmission is assumed to be frictionless – information does not decay as it travels through the network. Also let $\sum_{i=1}^n V_i = \Lambda$, where Λ is finite.

Costs of Information. Let d_{ij} be the geodesic distance between agent i and j , i.e., the length of the shortest path between i and j . In other words,

⁹Note that our definition of a tree coincides with the notion of a *minimally connected* network described in [BG] for their two-way information flow model.

distance between agents is measured in terms of the graph itself, and not in terms of any externally defined space.¹⁰ The cost paid by agent i to obtain the information possessed by agent j is given by $\phi_i(V_j, d_{ij})$.¹¹ We now list the different properties of the cost function.

1. *Property HD:* We assume information of higher value is more expensive, i.e.,

$$\phi_i(V_j, d_{ij}) > \phi_i(V'_j, d_{ij}) \quad \text{where } V_j > V'_j.$$

We allow distance to influence costs in two possible ways. These two properties describe alternative notions of distance that are appropriate in different types of situations.

2. *Property LE:* Information that comes through longer paths in the network is more expensive.

$$\phi_i(V_j, d_{ij} + 1) > \phi_i(V_j, d_{ij}).$$

3. *Property DC:* Information coming through shorter paths in the network is more expensive.

$$\phi_i(V_j, d_{ij} + 1) < \phi_i(V_j, d_{ij}).$$

Property LE is more appropriate in the context of physical distance, where it is usually more expensive to obtain information from places that are further away. International phone calls and mail are usually more expensive than their domestic counterparts. Property DC serves as a proxy for situations where distance is correlated with delay. Information that comes through longer paths involves a longer waiting time and hence is cheaper. This property is appropriate for information obtained through electronic networks with ABR traffic. Alternatively it describes the fact that ordinary mail which is cheaper than express mail is often routed through longer routes or tickets involving longer travel routes are cheaper.

Our formulation of the costs of information differs significantly from the one used in [BG] and Jackson and Wolinsky (1996). Here agents incur a separate cost for each piece of information they acquire. This cost depends on

¹⁰See Johnson and Gilles (2000) for spatial or geometric distance considerations.

¹¹Cowan (1997) provides an interesting historical example of such link costs in the US. Realizing the importance of roads, while lacking revenues, local governments and often the citizens themselves in the early 1800s, made roadwork a condition of property ownership.

the value of the information as well as the distance between agents exchanging information, i.e., on the use of the network by the agents. In the two papers mentioned above, each direct link has a given (identical) cost that is independent of the value of information and network distance. Also note that in our model links themselves are costless. Consequently, an agent's strategy has does not have the same implications in our model. Different network structures affect the cost of each piece of information by altering the geodesic distance between agents. A strategy in our context thus may be interpreted as the act of establishing the infrastructure for information networks and may provide a better interpretation of examples like phone calls.

Duplication Costs. Since links themselves are costless in the model, there is no constraint on the duplication of links. These additional links do not help in the transmission of information. Moreover, they are of no strategic importance and can generate uninteresting multiple equilibria. Link duplication will result in undirected architectures that will just be a replica of the directed equilibrium networks. Consequently, to eliminate situations where players establish bidirectional links with each other, we impose a penalty for duplication. Let $\Delta_i^g = \{j : g_{ij} = 1 \text{ and } g_{ji} = 1\}$ be the set of agents with whom player i has double links in network g . Each double link imposes a small positive penalty $\varepsilon > 0$ ($\varepsilon \ll \min\{V_i\}$) on both players.¹² Note that the penalty needs to be imposed on both players as strategies are chosen simultaneously.

Payoffs. The payoffs to player from the network g are given by

$$\Pi_i(g) = \sum_{j \in g(i)} (V_j - \phi(V_j, d_{ij})) - \varepsilon |\Delta_i^g| \quad (1)$$

where $i \neq j$. This payoff will of course depend on how the cost function varies with distance, i.e., whether it has the DC or LE property. We also introduce a functional form for the payoff function that will be used to obtain some results and insights in the paper. The payoff to player i from the network g according to this simple specification will be given by

¹²Although the duplication cost may seem like an artificial construct, their sole purpose is to rule out equilibria that are strategically uninteresting. Also given our interpretation of strategies, this may be viewed as penalty for wasting resources to establish infrastructure that remains unused. Link duplication has interesting consequences primarily in the context of reliability problems (see for instance Bala and Goyal (2000b) and Haller and Sarangi (2001)).

$$\Pi_i(g) = \sum_{j \in I} \left(V_j - \frac{V_j}{d_{ij}} \right) - \varepsilon |\Delta_i^g| \quad (2)$$

where $i \neq j$.¹³ Note that this payoff function satisfies properties HD and DC. In fact, according to this specification a direct link always gives a payoff of zero, while the indirect links yield positive payoffs.¹⁴ It is this property of the above payoff function that allow us to focus directly on the significance of paying for information obtained through indirect links. Observe also that the cost of information never exceeds its value. This assumption is maintained throughout the paper. Finally, note that when the network is connected every individual is able to acquire the information possessed by all the other players.

Example 1: The following example based on equation (2) explains the cost model clearly. Let agent i have information of value V_i , $i = 1, 2, 3$.

Insert Figure 1 here.

The payoffs of the three players are computed as follows:

$$\Pi_1(g) = V_2 - \frac{V_2}{1} + V_3 - \frac{V_3}{2} = \frac{V_3}{2}.$$

Similarly, $\Pi_2(g) = V_3 - \frac{V_3}{1} + V_1 - \frac{V_1}{1} = 0$ and the payoff of player 3 can be computed as $\Pi_3(g) = V_1 - \frac{V_1}{2} + V_2 - \frac{V_2}{1} = \frac{V_1}{2}$. ■

Equilibrium and Efficiency. Given a network g , let g_{-i} denote the network that remains when all of agent i 's links have been removed. Let $g = g_i \oplus g_{-i}$ where the symbol \oplus indicates that g is formed by the union of links in g_i and g_{-i} . A strategy g_i is said to be a **best response** of agent i to g_{-i} if

¹³Of course, when $g_{ij} = 0$ (or $g_{ji} = 0$), then no link exists between players i and j and their benefits and costs are both zero.

¹⁴Allowing the direct links to have a small positive payoff here does not alter the qualitative nature of our results. The crucial assumption here is the fact that shorter paths cost more. The role of this assumption is discussed in more detail later in the paper.

$$\Pi_i(g_i \oplus g_{-i}) \geq \Pi_i(g'_i \oplus g_{-i}) \text{ for all } g'_i \in G_i.$$

Let $BR_i(g_{-i})$ denote the set of agent i 's best response to g_{-i} . A network $g = (g_1, \dots, g_n)$ is said to be a **Nash network** if $g_i \in BR_i(g_{-i})$ for each i , i.e., agents are playing a Nash equilibrium. A *strict* Nash network is one where agents are playing strict best responses. Also recall that in a Nash network only the initiating agent can break a link.

The commonly used welfare measure is defined as the sum of the payoffs of all agents. Formally, let $W : \mathcal{G} \rightarrow \mathbb{R}$ be defined as

$$W(g) = \sum_{i=1}^n \Pi_i(g) \text{ for } g \in \mathcal{G} .$$

A network g is **efficient** if $W(g) \geq W(g')$ for all $g' \in \mathcal{G}$. We now illustrate the implications of our cost function through two examples by comparing the equilibrium outcomes to those of [BG]. Both examples use equation (2) as the payoff function.

Example 2 (Distance): We first examine the consequences of incorporating distance in costs. Hence in this example there is no endowment heterogeneity or $V_i = V$ for all i .

Insert Figure 2 here.

The above figure is a Nash network in the [BG] formulation where each agent pays a cost $0 < c \leq V = 1$ only for her direct links. However, when $\phi(V_j, d_{ij}) = \frac{V}{d_{ij}}$, player 3 would be able to minimize her cost by removing the current link, and choosing either $g_{34} = 1$ or $g_{32} = 1$, which would keep one of them at a distance of 3 units. Thus, the above graph is not Nash for the specified payoff function. But the new tree shown below (Figure 3) where agent 3 links to one endpoint, is an equilibrium under equation (2) as well as in [BG]. *Thus, even when the value of information is the same for*

all players distance or delay alters the cost of information acquisition and plays a crucial role in determining the equilibrium. ■

Insert Figure 3(a) and 3(b) here.

Example 3 (*Value of Information*): In figure 3 assume that $V_i = i$, $i = 1, 2, 3, 4$. We denote the first network by g^1 and the second by g^2 . First, let the cost of acquiring j 's information be $\alpha \frac{1}{d_{ij}}$ where $\alpha > 0$ is a constant. So, the cost varies only with the distance and not with the value of information. Under this cost function all players get the same payoff from both g^1 and g^2 and each of these graphs denotes an equilibrium network. Similarly if we apply the cost model of [BG], both these graphs constitute equilibrium configurations as well.

Now consider equation (2) where higher valued information is costlier. With this cost function, each player will want to keep the players with larger values of information as far as possible. The payoff of agent 3 is now higher in the network g^2 where it links to player 2 instead of player 4 because of the lower costs and the network g^1 is no longer an equilibrium, i.e., the fact that more information costs more is critical. Thus it is easy to see that if the costs are a function of the value of information and distance traveled by the information, in equilibrium, the players will have to be careful about their selection of direct link partners. ■

3 Equilibrium and Efficiency

This section contains our results. We begin by describing Nash networks. This is followed by the characterization of efficient networks.

3.1 Equilibrium Outcomes

Our first result pertains to the basic architecture of an equilibrium information network. This is followed by an examination of the chain and star

networks. To obtain a better understanding of the role of value of information in network formation we also present results for the case when all agents have identical endowments of information. Finally we analyze efficient networks. Throughout this section it is assumed that the cost function satisfies Property HD.

THEOREM 1: *Let the payoff function be given by (1). Then, (a) Under property LE, the (directed) complete graph is the only Nash equilibrium. (b) Under property DC, every Nash network is either empty or a tree. However, not every tree is Nash.*

Proof: See Appendix.

First, note that the above theorem holds when all agents have identical values, i.e., the network distance component is the driving force behind this intuitive result. Since longer paths are costlier, under property LC, every agent wants a direct link with every other agent and the result follows. Similarly, under property DC, direct links and cycles lead to lower payoffs. So every agent minimizes the number of links she forms which gives us the tree architecture. Since every Nash network is a tree we know from Cayley's formula (West (1996)) that there are a finite number of equilibria. Cayley's formula states that for a vertex set size n there are n^{n-2} trees and the last part of the theorem clearly indicates that this formula only provides an upper bound on the number of equilibria. Not all trees are Nash because some agent can get higher payoffs by deviating to form another tree. Hence we now use equation (2) to investigate the equilibrium properties of certain popular architectures like the chain and star graphs.

PROPOSITION 1: *Let the payoffs be given by equation (2) and assume that player i has value V_i with $V_1 \leq V_2 \leq \dots \leq V_n$. Then the following chains are the only strict Nash networks for $k = 1, 2, \dots, n$.*

Insert Figure 4 here.

Proof: See Appendix.

While the above configurations are the only strict Nash networks, the set of chain networks that are Nash includes some other configurations as well. Consider for instance the assignment of values described above. Then a chain given by the links $g_{21} = g_{23} = 1$ and $g_{kk+1} = 1$ for all $k \in I \setminus \{1, 2\}$ is also a Nash. Clearly, when $V_i = V$ for all i , any chain is Nash but only the chain where each player makes one direct link is strict Nash. The formation of these chains can be interpreted in an alternative way which allows us to predict the equilibrium outcome in a sequential version of this game.

Remark 1: *In a sequential version of this game, (a) When $V_i = V$ for all i , the chain is the only subgame perfect equilibrium, (b) Under endowment heterogeneity, the chain configurations shown above (Fig. 4) are the only subgame perfect equilibria.*

Sketch of Proof: Since the paper does not develop the sequential game formally, we provide only a sketch of the proof. For (a) consider a player k for whom $|g(k)| = n - 1$. By Theorem 1 the last player to join the network will not add any links. Next let $|g(k)| < n - 1$. If k belongs to the connected component, then by Theorem 1 player $g_{ik} = 0$ for $i \in g(k)$. Recall forming more than one link reduces payoff and since all players have the same value, she can form this single link with any $j \notin g(k)$. Finally, let $g(k) = \emptyset$. In this case player k is better off forming a link to any player in the connected component of the network. We now argue that player k will not choose to link to a player in the middle of the network. Let player k establish a link to a node j somewhere in the middle of the chain instead of either terminal node. The chain is now divided into two parts around node j which we denote by g^A and g^B . If k severs the link to j and instead connects to the node at the extreme end of g^A (without loss of generality), then the distance to all the players in g^B will be increased while the distance to those in g^A will remain the same. Since, all the values are equal, this change will increase the total payoff of k and players will always be better off by removing a branch and forming a bigger chain. Hence the result in (a) follows. The proof of part (b) is similar. ■

Chain graphs of the above kind are not the only Nash trees possible. We now analyze the popular star configuration. Their appeal lies in their in simple structure and fact that not only do they frequently arise in information networks, but are also a commonly observed social phenomenon (see for instance Rogers and Kincaid (1981) and Wellman and Berkowitz (1988)).

PROPOSITION 2: *Let the payoffs be given by equation (2). Then (a) The center-initiated star network is always Nash. (b) The periphery-initiated star can never be Nash for $n \geq 6$.*

Proof: See Appendix.

It is easy to see that the center-initiated star is Nash even if all agents possess the same value of information. However, it is not strict Nash since the central agent gets a payoff of zero in equilibrium. The above proposition also identifies a sufficient condition under which periphery-initiated stars cannot be sustained as equilibrium networks. The intuition for this is fairly straight-forward. When the number of players in the game increases, the amount of information at stake also increases and players in the network now wish to increase their distance from other players in the network. This creates an incentive to access the center of the star through indirect links, which increases distance to the other players, thereby increasing payoffs. However, for $n < 6$ it is easy to verify that periphery-initiated stars can indeed be Nash. We show this next through an example.

Example 4: *When $n < 6$, periphery-initiated stars can be Nash.* Consider a periphery-initiated star with $n = 4$. Let $V_1 = 1$, $V_2 = V_3 = V_4 = 6$. Player 1 who is the central agent will not add any links. But each of the peripheral agents can delete their current link and form a link to some other peripheral agent. Since each peripheral agent has the same value, they will all behave identically. In the periphery-initiated star, each of them has a cost of $1 + 2\frac{6}{2} = 7$. If one of them deviates to the alternative strategy of linking to another peripheral agent, then the new cost will be $6 + \frac{1}{2} + \frac{6}{3} = 8\frac{1}{2}$ which is higher than the previous one. Hence this particular star is in equilibrium. ■

From Proposition 2 it is also clear that mixed stars networks can only be Nash when both conditions (a) and (b) of the above proposition are satisfied simultaneously. Consequently, the agent with the lowest value must always be the central agent in such a network and the number of peripheral agents initiating links must not exceed five. We now examine the formation of periphery initiated stars in a network where all agents have identical values.

PROPOSITION 3: *When $V_i = V$ for all i , and equation (2) is satisfied, a Nash network will never contain a periphery-initiated star (Fig. 5) as a subgraph.¹⁵*

¹⁵We consider a graph to be a star if there are at least 3 peripheral nodes connected to

Insert Figure 5 here.

Proof: See Appendix.

From *Proposition 3*, we know that $V_i = V$ for all i , Nash networks can consist only of center-initiated stars and chains. One such equilibrium network is a the caterpillar. A **caterpillar** is a tree in which each vertex has at most two non-leaf neighbors. Note that under endowment heterogeneity, the chain subgraph of the caterpillar will have to satisfy the conditions identified in *Proposition 1*. However, not all concatenations of center-initiated stars and chains will be in equilibrium. The network shown below is composed of chains and center-initiated star and yet is not in equilibrium.

Insert Figure 6 here.

In Figure 6 player a would be better off by linking to player b . Due to this new link, the number of agents that are at maximum distance from agent a is maximized and the of number agents at minimum distance is minimized.

3.1.1 Costly Direct Links

Suppose we now assume that each direct link has cost $c > 0$ along with property HD and LE or DC. The resulting model will be a merger of Bala and Goyal and our proposed model. Let $\mu_i^d(g)$ be the set of agents with whom player i has a direct link. Then the payoff to player i from the network g is given by

$$\Pi_i(g) = \sum_{j \in g(i)} (V_j - \phi(V_j, d_{ij})) - c |\mu_i^d(g)| \quad (3)$$

a central node.

We can write a similar version of equation (2) with direct link costs as

$$\Pi_i(g) = \sum_{j \in I} \left(V_j - \frac{V_j}{d_{ij}} \right) - c \left| \mu_i^d(g) \right|. \quad (4)$$

Remark 2: *Relationship between equations (1-2) and payoffs with costly direct links.* Note that the payoff functions above do not require a duplication penalty since each direct link is now costly. Since the proofs are similar, we only state which of our current results hold under the new payoff specification. We find that Theorem 1(b) still holds and Theorem 1(a) holds as long as $0 < c < \phi(V, 2) - \phi(V, 1)$.¹⁶ Our result about periphery-initiated star still holds while a center-initiated star will no longer be Nash. Similarly as long as links formation is beneficial our result about chain networks is also valid. Next we identify general properties of networks under which a Nash network will always be a tree.

Remark 3: *Consider any model where the cost of information never exceeds its value and links are fully reliable. The (undirected) Nash network will always be a tree if (1) the information flow is two-way, and (2) the cost of information is more if it comes via a shorter path.* Since the cost of information never exceeds its value, a player's total payoff never decreases as she gets more information. This ensures that every Nash graph is connected. The two-flow and full link reliability ensure that no links need be duplicated. Next suppose that the Nash network g is a tree but the cost of information is less if it comes through a shorter path. In this case it is easy to see that every node will form a direct link with every other node leading to a complete network in which no player can be worse off. Clearly, the network g could not have been Nash. Thus a Nash network will not contain any cycles. Hence every Nash network satisfying the above two conditions will be a tree.

While these conditions apply to our model, they also hold for other specifications of the payoff function including the one in [BG]. In their model direct links are costly while indirect links are free, i.e., information that comes through the shorter path is more expensive. Note that the class of Nash networks in [BG] would be much more restricted if the costs of direct links would vary proportionately with different values of information. On the other hand in Bala and Goyal (2000b) we find super-connected networks when the size of the player set increases and the costs of information

¹⁶Thus the minimally connected nature of Nash networks shown in [BG] seems to be a fairly robust phenomenon.

are not very high. Here too information obtained through indirect links is free, but links are prone to failure with a constant probability. As the size of the networks increases, the value of information coming through shorter paths now decreases instead of being constant as in [BG]. This leads to the super-connected networks. Therefore it is also not surprising that in Haller and Sarangi (2001) where links can have different failure probabilities, such super-connected networks may not arise. Alternatively as the cost of information increases with distance, i.e., expected indirect benefits decrease when the size of society increases as in Bala and Goyal (2000b), super-connected networks will emerge. Hence in Theorem 1 as longer paths always increase the cost of information, we get the limiting case of super-connectedness – the complete network.

3.2 Efficient Networks

We now examine network structures that maximize the sum of the payoffs of all agents.

THEOREM 2: *Let the payoff function be given by (1). Then, (a) Under property LE, the (directed) complete graph is the only efficient network. (b) Under property DC, the chain is the only efficient network.*

Proof: See Appendix.

The above statements are also true for the case when $V_i = V$ for all $i \in I$. In order to better characterize the chains obtained under Theorem 2(b), we now introduce a multiplicative cost function. Let $\phi(V_j, d_{ij}) = \rho(V_j)\gamma(d_{ij})$, where $\rho(\cdot)$ is an increasing function of V and $\gamma(\cdot)$ is a decreasing function of d_{ij} . Hence the payoff function can be written as

$$\Pi_i(g) = \sum_{j \in g(i)} [V_j - (\rho(V_j)\gamma(d_{ij}))] - \varepsilon |\Delta_i^g|. \quad (5)$$

Clearly, the payoff function described in (2) is a special case of this. Using (5) we can obtain a chain which allows us to specify the precise location of the agents in the network.

PROPOSITION 4: *Let the payoff function be given by (5) and assume that player i has value V_i with $V_1 \leq V_2 \leq \dots \leq V_n$. Then the chain shown below (Fig. 7) is efficient.*

Insert Figure 7 here.

Proof: See Appendix.

Theorem 2(b) tells us that efficient networks are also chains. The similarity between the strict Nash and efficient architectures stems from the fact that in our model every agent pays for all the information they acquire. The crucial difference with the earlier models is the fact that the externality derived from the free indirect links is absent in our formulation. Therefore depending on the cost function we find that efficient networks, like Nash networks, are either complete or minimally connected. Further, when all agents have identical values of information, the set of Nash networks is the same as the efficient networks and when agents have different values of information, the two sets do not always coincide because the value of information affects the location of agents.

4 Discussion

In this section we summarize the main findings of this paper and explore some further modeling ramifications. We also compare our model with some recent network experiments.

Our first conclusion is that even in the absence of free indirect benefits there can be a conflict between stability and efficiency. Although the chain architecture is both stable and efficient, due to the heterogeneity of agent endowments, the two networks are not identical. Hence when all agents have the same value of information the Nash network coincides with the efficient one. Second, we find that the properties LE and DC lead to completely opposite equilibrium networks. Hence we examine an alternative distance metric by introducing a spatial formulation.

Spatial Considerations. Following Johnson and Gilles (2000) we will now assume that the agents have a *fixed* location on \mathbb{R} . Player $i \in I$ is located at x_i and the set $X = \{x_1, \dots, x_n\} \subset [0, 1]$ with $x_1 = 0$ and $x_n = 1$ represents the spatial distribution of players. Without loss of generality assume that $x_i < x_j$ if $i < j$. This implies that for all $i, j \in I$ the *distance* between players i and j is given by $d_{ij} = |x_i - x_j| \leq 1$. Thus, instead of a network based distance metric, we now have costs dependent on the spatial distance between players. The results obtained under property DC remain

unaltered since incentives of agents do not change. Next consider the cost function with the LE property. Since the agents are arranged on \mathbb{R} , for any three agents $i < j < k$, we have $d_{ij} + d_{jk} \geq d_{ik}$. Consequently, the complete graph will still be Nash but other networks like the chain can now also be supported as Nash. The large set of stable networks in Johnson and Gilles can therefore be attributed to the indirect link externality. We next discuss the role of agent heterogeneity for the model.

Endowment Heterogeneity. The value of the endowment clearly plays no role when costs satisfy the LE property. For the DC cost model, when the set of players is large, it turns out that there is no qualitative difference between assuming heterogeneous endowments or $V_i = V$ for all $i \in I$. Heterogenous endowments only affect the location of agents in the efficient and stable networks and at best can eliminate the co-ordination problem associated with the equal value of information model. However, when the player set is small, under the DC cost model the value of endowments seems to play a role. When $V_i = V$, we find that a periphery initiated star can never be Nash, but with heterogenous endowments it is a possibility as shown in Example 4. This is because under certain values of heterogeneous endowments it is possible to sustain the periphery-initiated star as a Nash by assigning the role of the central agent to the lowest value agent, a possibility that does not exist in the context of the equal values model.¹⁷ However, as the player set increases, increasing the total value of net benefits at stake, agents no longer wish to have all the other agents at a distance of two units. Longer path lengths can be used to reduce total costs (by increasing the distance to other high valued agents), leading to a break down of the periphery initiated star. This is similar to the argument in Bala and Goyal (2000b) where in the presence of uniform probabilities for link failure equilibrium networks become super-connected with an increase in the player set. The intuition there is that as the value of information at stake increases, agents try to insure themselves against link failure by forming alternate paths.

Information Decay. We now examine the implications of information decay, i.e., the value of information decreases as it traverses through the network. Clearly, this will create an incentive for the agents to form shorter paths. Under the LE property, information decay reinforces every agent's incentive to establish a direct link to all the other agents leading to the

¹⁷Note that Example 4 is also not valid for $V < \frac{3}{2}$.

complete network. To study the impact of information decay for the DC cost model we now introduce a variation of the payoff function in given by equation (2). Let

$$\Pi_i(g) = \sum_{j \in I} \left(V_j(1 - \delta d_{ij}) - \frac{V_j}{d_{ij}} \right) - \varepsilon |\Delta_i^g| \quad (6)$$

where $\delta \in (0, 1)$ is the decay parameter. A direct link now yields a payoff of $-\delta V_i$ and as before this functional form enables us to focus on the indirect links.¹⁸ In contrast to our earlier finding we now show that under equation (6) a periphery-initiated star will indeed be a Nash over a range of the decay parameter.

PROPOSITION 5: *Let the payoffs be given by (6). Then the periphery-initiated star with the central agent having minimum value is Nash for $1/6 < \delta \leq 1/4$.*

Proof: See Appendix.

In the absence of decay, the periphery-initiated star fails to be Nash when the player set increases, since agents wish to increase their distance from other high endowment agents. However, decay acts as a countervailing force to this cost based incentive. We find that since longer paths decrease the benefit obtained from the information itself, for certain values of the decay parameter agents continue to access other high information agents through the central agent. Thus while decay is of no consequence under the LE cost model, it can lead to different equilibrium networks under the DC cost model.

Network Experiments. In recent years there have been a number of experiments that take into account strategic behavior in the formation of networks (see Kosfeld (2002)). Here we will focus on two experiments that are directly relevant to our model. First, the ubiquitous nature of the center-initiated star requires some comment especially in the context of recent experimental evidence (see Falk and Kosfeld (2002)). They find that despite being the (strict) Nash, subjects do not favor centre-sponsored star of the [BG] model. They claim that this is because of the inequality inherent in this architecture since the central agent bears all the costs of

¹⁸This payoff function is purely for illustrative purposes and a complete investigation of the decay model is outside the scope of this paper.

network formation. In our model while the links themselves are costless for this network, the central agent still has the lowest payoff. When all values are equal to V , the central agent gets zero while all the others get $(n - 2)\frac{V}{2}$ and in the [BG] model the corresponding payoffs are $(n - 1)(V - c)$ and $(n - 1)V$. Thus relative inequality in our model is lower compared to [BG] when $c > \frac{V}{2}$. Further, if relative inequality is indeed the correct explanation (and we believe so) it would be easy to verify this by introducing heterogeneous endowments in the model. If the central agent is given a substantially higher endowment than the other players, it would be possible to form a center-initiated star where all agents have the same payoff after the network is formed.

The Deck and Johnson (2002) experiment tests the Johnson and Gilles (2000) model using Nash equilibrium instead of pairwise stability.¹⁹ Different cost benefit parametrization are tested using three institutions. The first two called ‘Split’ and ‘Primary’ do not allow for indirect benefits. In ‘Split’ the cost of a link is split equally between the two agents and under ‘Primary’ subjects can submit a bid for some proportion of the cost of the link. Not surprisingly subjects are able to achieve the Nash and efficient network, i.e., the complete network most of the times. This is similar to the LE cost model considered here where indirect links are more expensive. In the third institution called ‘Secondary’ subjects are allowed to submit bids on links in which they are not directly involved. This is a complex institution that permits agents to cross-subsidize their indirect links. Two main findings emerge from this experiment. First subjects are quite intolerant to free-riding. This supports the relative inequality idea suggested in Falk and Kosfeld (2002). As before we suggest that using differential endowments can provide conclusive evidence to support this hypothesis. The second finding is that under ‘Secondary’ subjects often tend to cross-subsidize indirect links. Deck and Johnson (2002) suggest that this is an attempt to insure against the failure of indirect links. We believe an experiment that includes both costly indirect links like ours, and free indirect links with the possibility of cross-subsidization, can be used to verify the indirect link insurance hypothesis.

Finally, while reliability and detailed examination of decay may possible extensions of the basic model, we believe another even more interesting

¹⁹The experimental design is quite fascinating with subjects being labelled station managers who must bid on the construction costs of railway tracks.

problem would be to impose the restriction that each agent could only form a limited number of links. Such a link formation capacity constraint would lead to interesting insights on network formation because informationally advantaged agents will now form links only with other such agents. While it will lead to minimal network architectures, more importantly, it can lead to social stratification based on the information endowment of agents.

5 Appendix

1. *Proof of Theorem 1:* (a) Consider a graph g that is not complete. We show that this cannot be in equilibrium. Let i and j be two agents such that $g_{ij} = g_{ji} = 0$. Then $d_{ij} > 1$. Without loss of generality let agent i form a link with j , i.e., $g_{ij} = 1$. Now under property LE, we have $\phi(V_j, 1) < \phi(V_j, d_{ij})$. Hence g cannot be an equilibrium network. Further, as this is true for all $d_{ij} > 1$, player i can minimize her costs by establishing direct links with all the players. Since a similar reasoning holds for all players $i \in I$, the complete network will be the equilibrium network. Finally the link duplication penalty ensures that this complete graph is a directed network.

(b) First, consider a disconnected network with k components, denoted as \mathcal{C}_j , $j = 1, 2, \dots, k$. If $|\mathcal{C}_j| < 2$ for all j , a player is indifferent between forming and not forming links. Consequently, the empty network will be an equilibrium. Of course this holds when $V_i \leq \phi(V_i, 1)$. However, the empty network will not be strict Nash.

In order to prove that it must be a tree, we will first show that the Nash network must be connected. Assume that there exists a disconnected Nash network with at least one component of size at least 2. Consider any two components \mathcal{C}_1 and \mathcal{C}_2 . Without loss of generality, let $|\mathcal{C}_2| \geq 2$ and agent j belongs to \mathcal{C}_1 . If j establishes a link to a player in \mathcal{C}_2 then there will be at least one agent in \mathcal{C}_2 who is at least 2 links away from j . By connecting to \mathcal{C}_2 , player j will get a positive payoff. Hence the Nash network must be connected. Next, we show that a connected Nash network will contain no cycles, i.e., it must be minimally connected. Suppose not. Then there exists a Nash network with at least one cycle. Consider the cycle $j_1 j_2 j_3 \dots j_r j_1$. Then either $g_{j_r j_1} = 1$ or $g_{j_1 j_r} = 1$. If $g_{j_r j_1} = 1$, then agent j_r wants to delete the link since $\phi(V_1, 1) > \phi(V_1, d_{ij})$ for $d_{ij} > 1$. When $g_{j_r j_1} = 0$, we know that $d_{ij} > 1$. By a similar reasoning when $g_{j_1 j_r} = 1$, agent j_1 will delete the link.

Hence a Nash network cannot contain the cycle described above. Finally we need to rule out cycles between two agents, i.e., cases when $g_{ij} = 1$ and $g_{ji} = 1$. Since both agents will incur a cost $\varepsilon > 0$ in this case, such cycles cannot be a part of an equilibrium network. Hence, the Nash network will be a tree.

To prove the second part, we present the following counter-example. Let $g_{12} = 1$, $g_{23} = g_{24} = 1$ with no other links existing in the graph. Assign $V_2 = 2$, $V_3 = 1$, $V_4 = 3$. V_1 can be assigned any value. Assume that the payoff function is given by equation (2). In the current network player 1 can get the information of all the other three nodes and has to pay a cost of $V_2 + \frac{V_3 + V_4}{2} = 4$. If player 1 deletes the current link to player 2 and instead links to player 3, then his total benefit will not change but the cost is reduced to $V_3 + \frac{V_2}{2} + \frac{V_4}{3} = 3$. Clearly, this tree is not in equilibrium. ■

2. Proof of Proposition 1: We first show that a chain satisfies the condition of Theorem 1 and hence is a candidate for Nash. We then use (2) to show that the networks shown in the above figure is strict Nash. Let $\{1, 2, \dots, n\}$ be the sequence of players forming the chain. By Theorem 1 we know that $g_{1n} = g_{n1} = 0$. Also, if $g_{ij} = 1$, player i will not break the link since it will destroy connectivity in the network. If $g_{ij} = 1$, then $g_{ji} = 0$ because of the duplication penalty. A player does not establish a new link to a second player since it will lead to a payoff of zero from this direct link. The only tree that has this property is the chain and hence a strict Nash network must be a chain. Now we show that in the above graph each player is playing their unique best response. First consider player k . If player k deletes the link with player 1 and links to node j , then the cost will be $\left[V_j + \frac{1}{2}(V_{j-1} + V_{j+1}) + \dots \right]$ which is bigger than the previous cost $\left[V_1 + \frac{V_2}{2} + \frac{V_3}{3} + \dots \right]$. Hence player k will not gain by changing her strategy. Similarly, any other player $m \in I \setminus \{1, k\}$, will incur a larger cost by deviating from the current strategy. Thus, this particular group of chains is in equilibrium. ■

3. Proof of Proposition 2: (a) In a center-initiated star the central agent in the star forms direct links to all the other agents. This graph is connected and has no cycles and therefore is a tree. Clearly, the central agent has a total payoff of zero. However, removing any of its current links will not lead to a higher payoff. All other agents can only add links since they do not

have any links to remove. Forming extra links would lead to cycles that will only reduce the current payoff. Thus, no agent can improve his or her payoff by deviating from the current strategy and the center-initiated star is Nash.

(b) Let the distribution of values be such that agent j has value V_j and $V_1 \leq V_2 \leq \dots \leq V_n$. First observe that for an equilibrium the central agent must have the least value. Otherwise, suppose agent $j \neq 1$ occupies the center. Since this is a star agent 1 has link to agent j . In this situation, all the other players except j will get a higher payoff by linking to player 1. Hence player 1 must be the center of the star. Let the initial periphery-initiated star with agent 1 in the center be denoted by g^{ps} . Let agent k , $2 < k \leq n$, alter her strategy by deleting the link to 1 and establishing a link to player 2. Recall that direct links should always be to the agent with least value, and since agent 1 is ruled out, player k can get the highest payoff only by linking to player 2. We denote the resultant network by $g' = g^{ps} - g_{k1} + g_{k2}$. Hence $C_k(g^{ps}) = V_1 + \sum_{j \neq 1, k} \frac{V_j}{2} = V_1 + \frac{(\Lambda - V_1 - V_k)}{2}$ and $C_k(g') = \frac{V_1}{2} + V_2 + \sum_{j \neq 1, 2, k} \frac{V_j}{3} = V_2 + \frac{V_1}{2} + \frac{(\Lambda - V_1 - V_2 - V_k)}{3}$. For g^{ps} to be the Nash, it must have the lower cost, i.e., we need $C_k(g^{ps}) - C_k(g') \leq 0$. We can rewrite this difference as $\frac{V_1}{2} + \frac{(\Lambda - V_1 - V_k)}{6} - 2\frac{V_2}{3} \leq 0$. Now, $(\Lambda - V_1 - V_k) \geq (n-2)V_2$ since V_2 is the smallest value in the set $I/\{1, k\}$. Hence $\frac{V_1}{2} - 2\frac{V_2}{3} + (n-2)\frac{V_2}{6} \leq 0$, or $(n-2)V_2 \leq 4V_2 - 3V_1$. This is inconsistent if $(n-2) \geq 4$ i.e., for $n \geq 6$. Hence, the result. ■

3. Proof of Proposition 3: First, we prove that a periphery-initiated star itself is never Nash when all values are equal. The central agent in such a star has payoff zero and the acyclicity principle prevents it from forming any additional links. Each peripheral agent has a cost $V(1 + (n-2)\frac{1}{2})$. This peripheral agent can initiate a link to another peripheral agent only after deleting the current link to the center. In that case, her cost will be $V(1 + \frac{1}{2} + (n-3)\frac{1}{3})$ which is less than the previous one.

When periphery-initiated star forms a subgraph of some tree, then one or more of its edges will be connected to some other component of the tree (see Figure 5). Let the central agent have κ branches excluding the branch associated with agent a with diameters $D_1, D_2, \dots, D_\kappa$. Without loss of generality, we assume $D_1 \leq D_2 \leq \dots \leq D_\kappa$. Let $\mu(j, D_i)$ denote the

number of players that are D_i links away from node j . In the above graph $\mu(a, 1) \geq 1$ and $\mu(a, 2) \geq k$. Also, for any $D_i > D_\kappa$, we set $\mu(a, D_i) = 0$. Let player a now delete its current link and form a link to the player at farthest end of the longest diameter branch. Denote the new network by g' . In g' , $\mu(a, 1)$ is the same as before while $\mu(a, 2)$ is reduced by $(k - 1)$. Also, now, $\mu(a, D) \geq 1$ for any D such that $D_\kappa \leq D \leq (D_1 + D_\kappa)$. Since the total number of players is fixed and the distance from some of the players has increased whereas the number of nodes that are 2 links away has decreased, the cost component for player a has decreased. Thus, a periphery-initiated star, where a peripheral agent has a choice of making a link to one end of the whole tree, can never form a component of a Nash network. ■

4. Proof of Theorem 2: Under property LE, there is no conflict between efficiency and stability. The network that minimizes the costs of all agents also minimizes the overall cost since. Under property DC, only a connected network can be efficient. Further, it must contain no cycles since cycles will raise the cost of acquiring information for all agents in the cycle. Hence we consider only trees as potential efficient graphs. Next we argue that only the chain is efficient. Consider a tree with diameter k ($0 < k < n - 1$). If $k = n - 1$, then we have a chain. Let $d_{ij} = k \neq 1$. Then there exists a star network (center-initiated or periphery-initiated or mixed) somewhere in the graph in the path between i and j since at least one vertex has two edges emanating from it. Without loss of generality, let such a star be at a distance m ($0 < m < k$) from j . Since the star has at least two arms one of these links can be rearranged in such a way that it becomes a part of the diameter. The diameter now becomes $k + 1$ and irrespective of the agent who forms this new link the total cost is lowered for the entire tree. This eliminates all other trees except chains as potential graphs. ■

5. Proof of Proposition 4: Consider an arbitrary chain given by j_1, \dots, j_n . Then using equation (3) we can write the total cost of this network as

$$\sum_{i=1}^n \sum_{k=1}^n \{\rho(V_{j_k})\gamma(|k - i|) + \rho(V_{j_i})\gamma(|i - k|)\} \quad (7)$$

where $|k - i|$ denotes the distance between agents k and i . We can compute the total costs using the table shown below, where $d_{j_i j_k}$ measures the geodesic distance between agents j_i and j_k .

$d_{j_i j_k}$	j_1	j_2	j_3	\dots	j_n	Total
j_1	0	1	2		$n-1$	$H(j_1)$
j_2	1	0	1		$n-2$	$H(j_2)$
\vdots						
j_n	$n-1$	$n-2$	$n-3$		0	$H(j_n)$

Note that $H(j_i)$ is the sum of the path lengths from agent j_i to all the other agents or $H(j_i) = \sum_{k=1}^n \gamma(d_{j_i j_k})$ for all $i \in I$. Clearly for the chain shown we have $H(j_i) = H(j_{n-i})$. We can rewrite the total cost of this chain as $\sum_{i=1}^{n/2} \{\rho(V_{j_i}) + \rho(V_{j_{n-i+1}})\} H(j_i)$. This sum will be minimized if for increasing order of $H(j_i)$, the order of its coefficients, i.e., $\{\rho(V_{j_i}) + \rho(V_{j_{n-i+1}})\}$ is decreasing. This implies that the two nodes with the highest value of information must be farthest apart and those with the smallest should be the least apart. This gives us the chain shown in the figure. ■

6. Proof of Proposition 5: Let player 1, the agent with the lowest information endowment be the center of the star. She cannot break any links and the duplication penalty ensures that she will not add any links. Player $i \in I \setminus \{1\}$ can either increase the number of links, or sever the link to the center and make one or more links to other nodes. The first possibility can be ruled out since a direct link always yields negative payoffs and does not create any *new* shortest paths to any other agent. Thus increasing number of direct links will reduce the total payoffs. In the second case, making exactly one link to some other node will produce a shortest path of length one to the directly connected node, a shortest path of length 2 to the central node and shortest paths of length 3 to the other indirectly connected nodes. As before adding more than one direct link will not improve upon the *existing* shortest paths. Next let g^{ps} be the original star graph. The graph g' is created when player k , $1 < k < n$, severs the link to player 1 and creates a link to the player with the next smallest value. Without loss of generality let this be player 2 or $g' = g^{ps} - g_{k1} + g_{k2}$. Hence $\Pi_k(g^{ps}) = -\delta V_1 + \sum_{j=2, j \neq k}^n V_j(1 - 2\delta - \frac{1}{2})$ and $\Pi_k(g') = -\delta V_2 + V_1(1 - 2\delta - \frac{1}{2}) + \sum_{j=3, j \neq k}^n V_j(1 - 3\delta - \frac{1}{3})$. For g^{ps} to be the Nash, it must have yielded a higher payoff, i.e., we need $\Pi_k(g^{ps}) - \Pi_k(g') \geq 0$. We can rewrite the difference between the two payoffs as $(V_1 - V_2)(\delta - \frac{1}{2}) + \sum_{j=3, j \neq k}^n V_j(\delta - \frac{1}{6})$. Note that the diameter of the underlying undirected

graph is 2. Using this it is easy to verify that for $\delta > 1/4$ the graph itself is not connected. This provides the upper bound on δ . Next $(V_1 - V_2)(\delta - \frac{1}{2})$ is non-negative for $\delta \leq 1/4$ and the whole expression is positive when $\delta > 1/6$. Thus, a periphery-initiated star graph will be Nash if $1/6 < \delta \leq 1/4$. ■

References

- [1] Aumann, R. and R. Myerson (1988) Endogenous Formation of Links Between Players and Coalitions: An Application of the Shapley Value. In The Shapley Value Ed. Alvin Roth, *Cambridge University Press*, 175-191.
- [2] Baker, B. and R. Shostak (1972) Gossips and Telephones, *Discrete Math*, **2**, 191-193.
- [3] Bala, V. and S. Goyal (2000a) A Non-Cooperative Model of Network Formation, *Econometrica*, **68**, 1181-1230.
- [4] Bala, V. and S. Goyal (2000b), "A Strategic Analysis of Network Reliability", *Review of Economic Design*, **5**, 205-228.
- [5] Casson, L. (1994) Travel in the Ancient World, *Johns Hopkins University Press*.
- [6] Cowan, R.S. (1997) A Social History of American Technology, *Oxford University Press*.
- [7] Deck, C. and C. Johnson (2002) Link Bidding in a Laboratory Experiment, Mimeo, *University of Arkansas*.
- [8] Falk, A. and M. Kosfeld (2002) An Experimental Investigation of Network Formation, Mimeo, *University of Zurich*.
- [9] Ghosh, A. (1994) In an Antique Land, *Vintage Books*.
- [10] Granovetter, M. (1974), Getting a Job, *University of Chicago Press*.
- [11] Greif, A. (1993) Contract Enforcement and Economic Institutions in Early Trade: The Maghribi Traders' Coalition, *American Economic Review*, **83**, 525-548.

- [12] Haller, H. and S. Sarangi (2001) Nash Networks with Heterogeneous Agents, Mimeo, *Virginia Tech*.
- [13] Hayek, F.A. (1945) The Use of Knowledge in Society, *American Economic Review*, **35**, 519-530.
- [14] Hedetniemi, S.M., S.T. Hedetniemi and A.L. Liestman, (1988) A Survey of Gossiping and Broadcasting in Communication Networks, *Networks*, **18**, 319-349.
- [15] Herzog, S., S. Shenker and D. Estrin (1997) Sharing the “Cost” of Multicast Trees: An Axiomatic Analysis, *IEEE\ACM Transactions on Networking*, **5**, 847-860.
- [16] International Telecommunications Union (1999), Challenges to the Network: Internet for Development, *ITU Geneva*.
- [17] Jackson, M.O. (2001) The Stability and Efficiency of Social and Economic Networks, forthcoming. In *Advances in Economic Design*, Ed. Murat Sertel.
- [18] Jackson, M.O. and A. Watts (1998), The Evolution of Social and Economic Networks, forthcoming, *Games and Economic Behavior*.
- [19] Jackson, M.O. and A. Wolinsky (1996) A Strategic Model of Social and Economic Networks, *Journal of Economic Theory*, **71**, 44-74.
- [20] Johnson, C. and R. Gilles (2000) Spatial Social Networks, *Review of Economic Design*, **5**, 273-299.
- [21] Kosfeld, M. (2002) Network Experiments, Mimeo, *University of Zurich*.
- [22] Litan, R.E. and A.M. Rivlin (2002) Beyond the Dot.coms, *Brookings Institution Press*.
- [23] Markovsky, B., D. Willar and T. Patton (1988), Power Relations in Exchange Networks, *American Sociological Review*, **53**, 220-236.
- [24] Milgram, S. (1967), The Small World Problem, *Psychology Today*, **2**, 60-67.
- [25] Montgomery, J. (1991), Social Networks and Labor Outcomes: Towards and Economic Analysis, *American Economic Review*, **81**, 1408, 1418.

- [26] Rogers, E and D.L. Kincaid, (1981), *Communication Networks: Towards a New Paradigm for Research*, *Free Press*.
- [27] Sairamesh, J., D.F. Ferguson and Y. Yemini (1995), An Approach to Pricing, Optimal Allocations and Quality of Service Provisioning in High-packet Networks, *Proceeding of INFOCOM 1995*.
- [28] Slikker, M. and A. van den Nouweland (2001) *Social and Economic Networks in Cooperative Game Theory*, *Kluwer Academic Publishers*.
- [29] Stover, J.F. (1999) *Historical Atlas of the American Railroads*, *Routledge Publishers*.
- [30] Watts, (2001), A Dynamic Model of Network Formation, *Games and Economic Behavior*, **34**, 331-341.
- [31] Wellman, B. and S. Berkowitz [Eds.] (1988) *Social Structures: A Network Approach*, *Contemporary Studies in Sociology: Vol. 15*, Cambridge University Press.
- [32] West, D.B. (1996) *Introduction to Graph Theory*, *Prentice Hall Inc*.



