The Structure of Information Networks[1]

Rajgopal Kannan[2] Lydia Ray[3] Sudipta Sarangi[4]

September 2004

[1]We are grateful to Hans Haller, Rob Gilles, Sumit Joshi, Georg Erber, Beth Allen, Mark Machina, Bob Martin, Johanna Francis, Cheryl Long, Raja Kali, Kaz Miyagiwa, Micheal Kosfeld and an anonymous referee for helpful suggestions. We also thank Susanne Maria Schmidt for many useful conversations. The paper has further benefited from the comments of seminar participants at Louisiana State University, University of Arkansas, Colgate University, Midwest Economic Theory Meetings 2002, Game Theory and Applications Conference in Mumbai 2003, Royal Economic Society Meetings 2003, Summer Meetings of the Econometric Society 2003, SED 2004 and GAMES 2004. Sudipta Sarangi gratefully acknowledges the hospitality of *DIW Berlin* where a part of this research was carried out. Rajgopal Kannan acknowledges the support of NSF grants IIS-0329738 and IIS-0312632. This work was also supported in part by DARPA-SenseIT and AFRL under grant number F30602-01-1-0551. All remaining errors are our responsibility.

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

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

[4]*DIW* Berlin and Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA. email: *sarangi@lsu.edu*

Abstract

We develop a strategic model of information acquisition in networks where agents pay for all the pieces of 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 consider two possibilities: (*i*) costs of information increase with distance, and (*ii*) they decrease with distance. We show that there is almost no divergence between the efficient and Nash equilibrium information architectures. We then explore the implications of a spatial model and study the effect of decay in networks where information through longer paths is cheaper. Finally, we also examine a model with costly link formation that combines both types of cost related assumptions.

JEL Classification: D82, D83

Keywords: Information networks, Nash networks, Efficient networks, Decay

1 Introduction

Communication forms one of the major pillars of all societies and economic systems. It results in the dissemination of information, helping well informed agents make better decisions. Often such communication takes place through a network of bilateral links between the participants. This paper develops a strategic model of the formation of information networks where agents choose their own link partners, resulting in different network configurations. The network is used to acquire information, and agents pay for all the information acquired through the network. 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 900B.C. to 612 B.C. to their network of roads, superior organization and discipline. Casson (1994, pg. 53) explains that one of the secrets of the success of the even mightier Persian empire 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. Moreover, the Persian dispatch service was one of the most efficient arms of the state. Governments however, were not the only ones to build these networks. Medieval trading communities often had their own informal networks for gathering 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 about far-flung markets and ensure contract enforcement among trade partners located in distant lands.

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. Railways helped link the two coasts of the continental United States leading to an increase in trade opportunities (see for example Stover (1999)). Rapid technological advances in recent

[1] Amitav 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 with the merchants acting as agents for each other locally.

years such as the Internet and wireless communication have created a global web which facilitates rapid transmission of information in a hitherto unprecedented manner. 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 (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 many extensions can be found in Hedetniemi et al. (1991). The focus of this literature is on the distribution issue and is captured mainly by some aggregate network criterion such as the minimum number of links needed to ensure that the gossip reaches everyone, or the minimum number of rounds required for everyone to hear the gossip. This literature rarely considered the costs of the links or more importantly individual costs and benefits. In other words, strategic interactions are conspicuously absent from this work.

Jackson and Wolinsky (1996) study strategic behavior in networks using a stability concept called *pairwise stability*.[3] Soon after the notion of *Nash networks* was introduced in a paper by Bala and Goyal (2000a; henceforth [BG]).[4] They assume that link formation is unilateral with the initiating player incurring all the costs of establishing the link. A link can only be

[2] The use of networks to study human interaction was pioneered by sociologists in the 1960s. Through an experiment involving farmers in Nebraska and a stockbroker in Boston, Milgram (1967) showed that on average, people have six degrees of separation.

[3] Aumann 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 (2003).

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

broken by the initiating player. In the model, when player i has a link with player j, she can access j's information and the information of all the other players j is linked to, without having to pay for these 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 Nash networks of the Bala and Goyal (2000a) type. It focuses on the structure of information networks where agents pay for all acquired information. We assume that every agent has a unique piece of information with some intrinsic value and would like to gather more information by linking to other agents. Unlike much of the earlier work, we incorporate an element of realism by allowing each agent to have a different endowment of information (see also Galeotti, Goyal and Kamphorst (2004)). Secondly, information seekers have to pay for *all* the information they acquire in the network, i.e., we do not allow for free indirect links. 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. Establishing links however is free. Since the externality is stronger in the case of two-way flow, we focus on this 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 − each providing a different interpretation of network distance. The first interpretation of network distance is a spatial one while in the second instance distance is a proxy for time. 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 like international phone calls being more expensive than domestic ones. It is also true for the historical information networks mentioned at the beginning of this section. With merchant ships serving 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

[5]Distance based cost functions were also used by Johnson and Gilles (2000).

to information delays or involve waiting and hence are cheaper. This cost structure is relevant for some types of electronic networks. Network applications can be characterized by their differing Quality of Service (QoS) requirements, such as real-time video with high bandwidth and low delay requirements, or the opposite where bandwidth requirements are flexible and delay is tolerable. 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)).

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 (see for example, Herzog, Shenker and Estrin (1997)). Network users maximize their payoffs in terms of costs and QoS benefits by selecting appropriate sets of users with whom to share network resources. In particular this can be used to model Available Bit Rate (ABR) traffic 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.

This cost formulation is also of theoretical interest since the same set of incentives govern that behavior in our model can be applied to the Bala and Goyal (2000a) model and its direct extensions for obtaining their results. This assumption also allows us to differentiate between the role of full reliability and two-way flow of information in Nash networks.

Following the analysis of the two benchmark models we explore the implications of a spatial model. We study the effect of decay in networks where information through longer paths is cheaper. An interesting trade-off exists in this model since decay offsets the desire for agents to have longer paths in the network. Finally we examine a model with costly link formation that combines both types of cost related assumptions.

Section 2 describes the basic model. In section 3 we analyze the stability and efficiency properties of networks. Section 4 summarizes the main results and explores alternative formulations that incorporate both cost assumptions. Section 5 concludes.

2 Model

Let $I = \{1, 2, \ldots n\}$ be the set of agents with $n \geq 3$. For ordered pairs $(i, j) \in I \times I$ the shorthand notation ij is used. Agents in the model are information seekers who gain utility from having more information. Each individual $i \in I$ has an information endowment of value $V_i \geq 0$. $V_i \neq V_j$ corresponds to *heterogeneity of agents' endowments*. All agents are aware of the value of the *non-rival* information possessed by other agents. Access to 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 permits two-way flow of information between them.

Strategies. Formally, a strategy of agent $i \in I$ is a vector $g_i = (g_{i1}, g_{i2}, \ldots, g_{ii-1}, g_{ii+1}, \ldots, g_{in})$ where $g_{ij} \in \{0, 1\}$ for each $j \in I \backslash \{i\}$. If i forms a link with j, then $g_{ij} = 1$ allowing information to flow from i to j and from j to i, while if no link exists between i and j then $g_{ij} = 0$, permitting no information flow. Note that only the initiating agent can break a link. We restrict attention to fully reliable links and pure strategies. The set of all strategies of individual i is denoted by $\mathcal{G}_i$. The set of joint strategies (strategy profiles) is denoted by $\mathcal{G} = \times_{i=1}^{n} \mathcal{G}_i$. A strategy profile $g = (g_1, g_2, \ldots, g_n) \in \mathcal{G}$ is equivalent to a (directed) **network**, where each vertex depicts an agent and each link forms an edge with the arrow pointing to the person with whom the link is established.[6] Hence with this identification, $\mathcal{G}$ also represents the set of all possible networks. We now introduce some graph-theoretic definitions for undirected graphs based on West (1996).

A *walk* is a sequences 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**.[7] 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

[6] At 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.

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

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** (g^e) is one where $g_{ij} = 0$ for all pairs ij and a **complete network** (g^c) is a graph in which every player has a direct link to *every* other player. A **center-initiated** star is an acyclic network where only one agent (the central agent) establishes a direct link with all the other $(n-1)$ agents. Similarly, a **periphery-initiated** star is an acyclic network where each of the other $(n-1)$ agents initiate a link with the central agent. A mixed star is a combination of these two types of stars. 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*.

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 $\sum_{i \in I} V_i = \Lambda$.

Costs of Information. Let d_{ij} be the geodesic distance between agents i and j. The cost incurred by agent i to obtain agent j's information is given by $\phi_i(V_j, d_{ij})$.[8] We now list the different properties of the cost function.

1. *Property* **HD:** Information of higher value is more expensive, i.e.,

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

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:

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

[8] Cowan (1997) in her book on the social history of technology provides an interesting historical example of such link costs in the US. Realizing the importance of roads, but lacking revenues, local governments and often the citizens themselves in the early 1800s, made roadwork a condition of property ownership.

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) since we assume costless links. Consequently, 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 to prevent the duplication of links. Such double links are of no strategic importance and can generate uninteresting multiple equilibria. 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.[9] Note that the penalty needs to be imposed on both players as strategies are chosen simultaneously.

[9]Although these duplication cost may seem like an artificial construct, their sole purpose is to rule out strategically uninteresting equilibria. Also, this may be viewed as a penalty for wasting resources to establish unnecessary infrastructure. Link duplication has interesting consequences primarily in the context of reliability problems (see for instance Bala and Goyal (2000b) and Haller and Sarangi (2003)).

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

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

where $i \neq j$ with $\phi(V_j, d_{ij})$ either having the DC *or* the LE property. We also introduce the following functional form for the payoff function that is used to obtain additional insights in the paper.

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

where $i \neq j$.[10] Note that this payoff function satisfies properties HD and DC. A direct link always gives a payoff of zero, while the indirect links yield positive payoffs. 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.

Equilibrium. 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**:** $\Pi_i(g_i \oplus g_{-i}) \geq \Pi_i(g_i' \oplus g_{-i})$ 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 $g = (g_1, \ldots, g_n)$ 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.

Efficiency. The commonly used welfare measure is defined as the sum of the payoffs of all agents. Formally, let $W : \mathcal{G} \to \mathbb{R}$ be defined as**:** $W(g) = \sum_{i \in I} \Pi_i(g)$ 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 1 (*Distance*): We first examine the consequences of incorporating distance in costs assuming $V_i = V$ for all $i \in I$. Let $g_{12} = g_{14} =$

[10]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.

$g_{31} = 1$, and $g_{ij} = 0$ for all other $ij \in I \times I$. This 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 $\phi(V_j, d_{ij}) = V/d_{ij}$, 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, this network is not Nash for the specified payoff function. But the new tree described in the next example 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.*

Example 2 (*Value of Information*): Assume that $V_i = i$, $i = 1, 2, 3, 4$. Let the network g^1 be defined as $g_{12} = g_{14} = g_{34} = 1$, and $g_{ij} = 0$ for all other $ij \in I \times I$. Let the network g^2 be defined as $g_{12} = g_{14} = g_{32} = 1$, and $g_{ij} = 0$ for all other $ij \in I \times I$. First, let the cost of acquiring j's information be $\dfrac{\alpha}{d_{ij}}$ where $\alpha > 0$ is a constant. Now, the cost varies only with 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 is an equilibrium network. Similarly if we apply the cost model of [BG], both these graphs constitute equilibrium configurations as well.

Next 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 away 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. 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 for the two benchmark cost formulations. 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.

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

Proof: See Appendix.

Note that the above theorem holds when all agents have identical endowment values, i.e., the network distance component is the driving force behind this intuitive result. Also 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 of size n there are n^{n-2} trees and the last part of *Theorem 1* clearly indicates that this formula only provides an upper bound on the number of equilibria. Not all trees are Nash as some agent can get higher payoffs by deviating to form another tree.

We next use equation (**2**) to investigate the equilibrium properties of certain popular architectures like the chain and star graphs. Besides, the chain is the largest diameter graph and the star is the smallest diameter graph where agents have positive benefits from indirect links.

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

Insert Figure 1 here.

Proof: See Appendix.

While the above configurations are the only strict Nash networks, the set of chain networks that are Nash includes other configurations as well.

Using the assignment of values described above a chain given by the links $g_{21} = g_{23} = 1$ and $g_{kk+1} = 1$ for all $k \in I \backslash \{1, 2\}$ is also Nash. Clearly, when $V_i = V$ for all i, any chain is Nash and chains where each player makes only one direct link are strict Nash. The formation of these chains can be interpreted in an alternative way $-$ one that allows us to predict the equilibrium outcome in a sequential version of the 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 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. ■

We now analyze the popular star configurations. Their appeal lies in their in simple structure and the fact that they not only arise frequently 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 (i) The center-initiated star network is always Nash. (ii) 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 intuition for why periphery-initiated stars are not Nash for $n \geq 6$ 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, increasing the distance to other players, thereby increasing payoffs. However, for $n < 6$ it may be possible to arrange the agents in such a way (with the lowest value agent at the centre) that a periphery-initiated star is Nash.We show this next through an example.

Example 3: *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. ■

Proposition 2 also states that mixed star networks can only be Nash when conditions for both non-mixed type stars are satisfied. 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.

PROPOSITION 3: *When $V_i = V$ for all i, and equation (2) is satisfied, a Nash network will never contain a periphery-initiated star as a subgraph.*[11]

Proof: See Appendix..

From *Proposition 3*, we know that when $V_i = V$ for all i, Nash networks

[11]Note that a star network must have at least 3 peripheral nodes connected to a central node.

can consist only of center-initiated stars and chains. One such equilibrium network is 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, even with homogeneous endowments 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 2 here.

In Figure 2 player a would be better off by linking to player b. Due to this new link, the number of agents that are at maximal distance from agent a is the largest and the set of agents at minimal distance is the lowest.

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 making the duplication penalty redundant. 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

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

Remark 2: *Assumption DC and payoffs with costly direct links.* Since the proofs are similar, we only state which of our current results hold under this new payoff specification. We find that *Theorem 1(b)* still holds and *Theorem 1(a)* holds as long as $0 < c < \phi(V, 2) - \phi(V, 1)$. Using a modified version of equation (**2**) we find that our result about periphery-initiated star still holds BUT a center-initiated star will no longer be Nash. Our result about chain networks is also valid as long as link formation yields positive net benefits.

Remark 3: *Assumption LE and payoffs with costly direct links.*[12] Longer graph distances affect net benefits negatively via increased operating costs

[12] We thank an anonymous referee for bringing this to our attention.

and due to the additional costs incurred to prevent information decay or erosion of quality of routing/service (Kannan, *et al.* (2004)). In a recent paper on game theoretic routing Fabrikant *et al.* (2003) also have a payoff function where routing through more nodes is costlier. For small direct link costs it is easy to see that complete networks will still be equilibria, while for large link costs one would expect minimally connected networks as in [BG]. For very large link costs the empty network is both Nash and efficient. In the intermediate range a number of possibilities can arise and in the subsequent section we shed more light on this.

3.1.2 Assumption DC and How it Relates to the Literature

We now provide a key insight about Nash networks and discuss its relationship to the rest of the literature.

Remark 4: *Consider any model where the cost of information never exceeds its (positive) 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 access to more players. This ensures that every Nash graph is connected. The two-way flow and (full) link reliability assumptions 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.

It can be argued that these conditions are also applicable to other specifications of the payoff function in the literature, including the one in [BG]. In their model, each agent pays a cost $c > 0$ for each of her direct links, but does not pay anything for knowing others through her direct links. *One way to interpret this is that in [BG] direct links are costly while indirect links are free, i.e., information that comes through a shorter path is more expensive!* Assumption DC may be construed as a general form of this relationship where indirect links also impose positive costs in a specific manner. Moreover, even the agent that does not form the link acquires the information for free, i.e., information is free regardless of the path length. This explains the emergence of minimally connected networks in equilibrium.

Galeotti, Goyal and Kamphorst (2004) extend the basic [BG] framework by introducing endowment heterogeneity as well as different link formation costs. By assuming heterogeneity in endowments but not in costs, they also find minimally connected networks. This is because agents in their model have incentives similar to agents in our model. Heterogeneity in costs leads to a more general result where equilibrium networks are minimal networks in the sense that deletion of link increases the number of components. The intuition for this result is also similar except that in this case heterogeneity in costs can lead to multiple minimal components in equilibrium. Thus their claim that "*... even in settings with considerable heterogeneity, strategic models of network formation yield sharp predictions and equilibrium networks exhibit high centrality and small average distances.*" is not surprising at all.

Next consider Bala and Goyal (2000b) where links can fail with a uniform probability $p \in (0, 1)$. In equilibrium they find super-connected networks when the size of the player set increases and the costs of information are not very high. Here too information obtained through indirect links is free but as the size of the networks increases, the expected value of information coming through shorter paths now decreases (because of the probabilities involved) instead of being constant as in [BG]. Alternatively the implicit cost of information increases with distance, i.e., without additional links expected indirect benefits decrease when the size of society increases. This gives rise to super-connected networks. Consequently, in *Theorem 1* as longer paths always increase the cost of information, we get the limiting case of super-connectedness − the complete network. However in Haller and Sarangi (2003) where different links can have different failure probabilities, such super-connected networks may not arise in equilibrium.[13]

Note that while both the assumptions in *Remark 3* prevent the formation of cycles, they lead to different network formation behaviors. With the undirected flow of benefits, the non-initiating agent is the one who does not have an incentive to add links, while the presence of free indirect benefits provides every agent an incentive to form the least possible number of links. Consequently, recall that in the one-way flow version of [BG], a wheel is the

[13]Note that while decay and link failure probability both lead to cycles, the incentives in both cases are different. For more on this topic see Haller and Sarangi (2003).

equilibrium architecture, i.e., only one of the non-initiating agents in the chain adds a link forming a wheel. Further, as information coming through longer paths is still cheaper every agent forms the lowest number of necessary links. On the other hand in the two-way flow model of [BG] equilibrium outcomes are stars which is also the case in our model.

Later in the paper we retain the two-way flow requirement but relax the assumption that information coming through longer paths is cheaper, i.e., indirect benefits may even have increasing costs. We find that this leads to a particular type of cyclical networks: $k-regular$ networks. In contrast, Caffarelli (2004) finds that in the long run the system converges to a state that consists of starred wheels − wheels in which some agents are like the center of a star. He considers directed flow of benefits with indirect links providing free benefits. However, agents incur a maintenance or usage cost which is a convex function of the number of agents observed, i.e., there is an optimal number of agents each agent wishes to observe. Since it is a one-way flow model with free indirect link benefits, he finds that the equilibrium architecture will be a wheel (and not a regular network) where the number of agents depends on the usage cost. Not surprisingly this wheel also has star like features since this allows agents to limit the number of other players they observe.

3.2 Efficient Networks

We now examine network structures that maximize the sum of the payoffs of all agents and compare them with the Nash networks.

THEOREM 2: *Let the payoff function be given by (1). Then, (i) Under property LE, the (directed) complete graph is the only efficient network. (ii) 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 $\rho(.)$ is an increasing function of V and $\gamma(.)$ is a decreasing function of d_{ij}. Hence the payoff function can be written as

$$\Pi_i(g) = \sum_{j \in g(i)} [V_j - (\rho(V_j)\gamma(d_{ij}))] - \varepsilon \left|\Delta_i^g\right|. \tag{4}$$

Using (**4**) we can obtain a chain which allows us to specify the precise location of the agents in the network.[14]

Proposition 4: *Let the payoff function be given by (4) and assume that player i has value V_i with $V_1 \leq V_2 \leq \cdots \leq V_n$. Then the chain shown below (Fig. 3) is efficient.*

Insert Figure 3 here.

Proof: See Appendix.

Theorem 2(ii) tells us that efficient networks are also chains. The similarity between the strict Nash and efficient architectures stems from the fact that every agent pays for all the information they acquire. The crucial difference with the earlier models is the absence of the externality accruing from the free indirect links. 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. 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 Model Variations

In this section we summarize the main findings of this paper and explore some further modeling ramifications. 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. Only when $V_i = V$ for all $i \in I$ does the Nash network coincide with the efficient one. Second, we find that the properties LE and DC lead to completely opposite equilibrium networks.

Next, the value of the endowment plays no role when costs satisfy the LE property. For the DC cost model, heterogeneous endowments only affect the location of agents in the efficient and stable networks. At best they

[14] The payoff function described in (**2**) is a special case of this.

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. Under certain values of heterogeneous endowments the periphery-initiated star can be sustained as Nash by assigning the role of the central agent to the lowest value agent – a possibility that does not exist in the equal values model. However, as n increases, increasing the total net benefits at stake, agents no longer wish to have all the other agents at a distance of two units. Longer path lengths can lower total costs (by increasing the distance to other high valued agents), leading to a break down of the periphery initiated star.

4.1 Spatial Distance

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 $X = \{x_1, \ldots, x_n\} \subset [0,1]$ 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.

4.2 Information Decay

We now examine the implications of information decay. Clearly decay creates an incentive for the agents to form shorter paths. To study information decay in the DC cost model we now introduce a variation of equation (2). For the sake of simplicity let

$$\Pi_i(g) = \sum_{j \in g(i)} \left(V_j(1 - \delta d_{ij}) - \frac{V_j}{d_{ij}} \right) - \varepsilon \, |\Delta_i^g| \tag{5}$$

where $\delta \in (0,1)$ is the decay parameter. Observe that 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.

Proposition 5: *Let $n \geq 3$ and let the payoff function be given by (5). Then for $\delta \in (1/4, 1]$ the empty network is the unique Nash network. For $\delta \in (0, 1/4]$, every Nash network is connected.*

Proof: See Appendix.

This proposition can also be viewed in another way. A connected network for which $\delta > 1/2D$, where D is the diameter of the graph, cannot be Nash. In contrast to our earlier finding under the DC model, we now show that under equation (**5**) a periphery-initiated star will indeed be Nash over a range of the decay parameter.

Proposition 6: *Let the payoffs be given by (5) and $1/6 < \delta \leq 1/4$. (i) Then periphery-initiated stars are the unique Nash networks. (ii) In case the value of endowment differs across agents, the value is minimal for the central agent.*

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. Based on these two propositions we can make some further observations that charecterize the types of possible Nash networks.

- Since direct links yield negative payoffs, under the decay model a center-sponsored star and mixed star will never arise in equilibrium. In fact, even if direct links yield nonnegative benefits, but the benefits from indirect links are higher, periphery sponsored stars are the more likely outcome.

- Next observe that i's payoff from a link with player j equals zero when she is at a distance $d = \frac{1+\sqrt{1-4\delta}}{2\delta}$ from j. Consequently one $\delta \rightarrow 0$, player i prefers longer paths. Therefore as in the previous section we expect that minimally connected networks will form the Nash networks. The same holds for efficiency.

- Next observe that in this case it is possible to have cyclic networks as well as acyclic networks in equilibrium. For instance with $n = 7$ and $V_i = 1$ for all $i \in N$, the chain network is Nash for $\delta = 0.02$. Further it turns out that $\delta > 0.1355$, the chain is no longer Nash. For example with $\delta = 0.14$ the circle where every player has a payoff of 0.6533 is the Nash network. Furthermore, the circle network is no longer Nash for $\delta \geq 0.194$. In this range all players have negative payoff in the circle and will prefer shorter paths. In fact the periphery-sponsored star is Nash in this range.

To sum up, we find that for very high δ the empty network is stable and for very low δ the outcome is similar to the DC model. Players prefer longer paths and both efficient and Nash networks will tend to be minimally connected. We also find that when $\delta \leq 1/6$, cyclical networks can be stable, but for $\delta > 1/6$, the Nash network again is the minimally connected periphery-sponsored star. Thus Nash networks are trees for both high and low δ and have cycles in the intermediate range. The low δ trees however have large diameters and the high δ trees have low diameters.

Further it can be shown that if $\delta < 1/N^2$ then a Nash network must be acyclic and the chain network discussed above is an example of this. Similarly a Nash network will be acyclic for $\delta < 1/3D^2$ where D is the diameter of the graph.

Finally note that decay is of no consequence under the LE cost model. Under the LE property, information decay reinforces every agent's incentive to establish a direct link to all the other agents leading to the complete network.

4.3 Combining Assumptions LE and DC

We now introduce a model with costly link formation that combines both assumptions LE and DC. The payoff of player i in the network g is given by

$$\Pi_i(g) = \sum_{j \in g(i)} (V_j - \phi(V_j, d_{ij})) - c \sum_{j \neq i} g_{ij} \tag{6}$$

where $c \geq 0$ is the cost of forming a link.

Assumption U: We will now assume that costs of link formation decline up to a distance d^* and then increase. In other words

$$\phi_i(V_j, d_{ij}+1) < \phi_i(V_j, d_{ij}) \text{ for } d_{ij} \leq d^*$$
$$\phi_i(V_j, d_{ij}+1) > \phi_i(V_j, d_{ij}) \text{ for } d_{ij} > d^*.$$

Thus we have a U-shaped cost function with declining costs up to d^*.[15] For simplicity we also assume that $V_i = 1$ for all $i \in N$. Also let $V^* = \sum_{d_{ij}=1}^{d^*} (1 - \phi(1, d_{ij}))$. It is now possible to state the following result.

THEOREM 3: *Let the payoffs be given by (6). Then under assumption U, when $c > V^*$ the empty network is the unique Nash network. When $c \leq V^*$ then every Nash network is connected.*

Proof: We know that for $d_{ij} \leq d^*$, $\sum_{d_{ij} \leq d^*} (1 - \phi(1, d_{ij}))$ is increasing with distance and every player has an incentive to form a minimal number of links. However when $c > V^*$, $g_{ij} = 0$ for all $ij \in I \times I$. Moreover for $d_{ij} > d^*$ no player wants to form a link since $(1 - \phi(1, d^* + 1)) < V^* < c$. Hence the empty network is the unique Nash network. The second part follows from the fact that it is desirable to link to all agents and since the inequality is reversed every agent is willing to form at least one link. ■

When $c \leq V^*$, a fairly large class of networks are still permissible as equilibria. To understand more about equilibrium networks in this range we will consider two different cases.

Case (i): Let $c \to 0$. In this case we get a model that is very similar to the decay model analyzed above. When d^* is very high, i.e., as $d^* \to \infty$, then regardless of the fact that links are free, the equilibrium architecture is a chain network. Similarly, when $d^* \to 1$, we will get the complete networks as the equilibrium Nash architecture. In between this range just as in the decay model equilibrium architectures will depend on the precise value of d^* and both periphery-initiated stars as well cycles could arise in equilibrium.

Case (ii): Intermediate cost range. Formally this implies that $c \leq V^*$, but unlike Case (i) costs are not close to zero, i.e., $c >> 0$. Again when

[15] This section is based on a payoff specification suggested by an anonymous referee and provides a very general model of Nash networks. We are grateful to the referee for this suggestion.

$d^* \rightarrow \infty$, despite the costly nature of link formation, the distance effect dominates and chains are the only Nash networks. Now let $d^* \rightarrow 1$. We know that if links are costless, then g^c is the Nash network. However with costly link formation only a finite number of links will be profitable for each player. Hence as $d^* \rightarrow 1$, we will find $k-$regular networks in equilibrium. In the intermediate range for d^* cyclic networks can be supported as equilibria. These networks can be wheels, or wheels with spokes, or even starred wheels as in Caffarelli (2004) depending on which effect dominates. Yet for some parameter values it is also possible to have periphery-initiated stars in equilibrium. It is difficult to make more detailed remarks for this range without using a more precise formulation for the payoff function.

5 Conclusion

The paper identifies the nature of stable and efficient information networks when agents have to pay for information acquired through indirect links as well. Further the properties of stable networks are investigated by introducing costs of link formation. The range of possible equilibrium networks and the conditions under which they can occur are stated. The paper sheds light on the role of assumptions like two-way information flow, free indirect benefits and full reliability in models of Nash networks.

One interesting extension would be to introduce reliability issues in the model. Another, perhaps more interesting 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.

6 Appendix

1. *Proof of Theorem 1*: (i) Consider a connected graph $g \neq g^c$. Such a g cannot be in equilibrium.[16] Let $g_{ij} = g_{ji} = 0$. Then $d_{ij} > 1$. Without

[16]In part (ii) we show why all equilibrium graphs must be connected.

loss of generality set $g_{ij} = 1$. Under property LE, $\phi(V_j, 1) < \phi(V_j, d_{ij})$. Hence g cannot be an equilibrium network. Further, since this is true for all $d_{ij} > 1$, player i can minimize her costs by establishing direct links with all the players. As a similar reasoning holds for all $i \in I$, g^c will be the equilibrium network. Finally the link duplication penalty ensures that this complete graph is a directed network.

(*ii*) First, consider a disconnected network with k components, i.e., $\mathcal{C}_j$, $j = 1, 2, \ldots, k$. If $|\mathcal{C}_j| < 2$ for all j, a player is indifferent between forming and not forming links. Consequently, g^e will be Nash but not strict Nash. This also holds when $V_i \leq \phi(V_i, 1)$. To prove that it must be a tree, we will first show that the Nash network must be connected. Let g be a disconnected Nash network with components $\mathcal{C}_1$ and $\mathcal{C}_2$ where (at least) one component is at least of size 2. Without loss of generality, let $|\mathcal{C}_2| \geq 2$ and agent $j \in \mathcal{C}_1$. If j links to a player in $\mathcal{C}_2$ then there is some $j' \in \mathcal{C}_2$ who is at least two links away from j. By connecting to $\mathcal{C}_2$, player j will get a positive payoff. Hence g cannot be Nash. 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 $j_1 j_2 j_3 \cdots j_r j_1$. Then either $g_{j_r j_1} = 1$, or $g_{j_1 j_r} = 1$. If $g_{j_r j_1} = 1$, then j_r wants to delete the link since $\phi(V_1, 1) > \phi(V_1, d_{ij})$ for $d_{ij} > 1$. Similarly if $g_{j_1 j_r} = 1$, j_1 will delete the link. Hence a Nash network cannot contain the cycle described above. Finally (direct link) cycles between two agents are ruled out by the duplication penalty. Hence, the Nash network will be a tree.

For the second part the following counter-example suffices. Let $g_{12} = g_{23} = g_{24} = 1$, with no other links existing in the graph. Let $V_2 = 2$, $V_3 = 1$, and $V_4 = 3$. V_1 can take any value. Using equation (**2**), player 1 can get the information of all the other three and has to pay a cost of $(2V_2 + V_3 + V_4)/2 = 4$. If player 1 deletes the current link to 2 and instead links to player 3, then her total benefits will not change but her cost reduces to $V_3 + (V_2/2) + (V_3/3) = 3$. Clearly, the original tree is not in equilibrium. ■

2. *Proof of Proposition 1*: Since the chain is a tree by *Theorem 1* it is a candidate for Nash. Also note that a player who establishes a new link to a second player earns a payoff of zero from this direct link making her indifferent to forming it. The only tree where every player has the smallest number of direct links is the chain. 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 player j, then her costs will be

$[V_j + (V_{j-1} + V_{j+1})/2 + ...] > [V_1 + (V_2/2) + (V_3/3) + ...]$. Hence player k will not gain by changing her strategy. Similarly, any other player $m \in I \backslash \{1, k\}$, 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*: (i) A center-initiated star is a tree where the central agent has a payoff of zero. However, she cannot increase her payoffs by deleting any of her current links. All other agents can only add links since they do not have any links to remove. Forming extra links will lead to cycles that only reduce their current payoff. Thus, no agent can improve his or her payoff by deviating from the current strategy and the center-initiated star is Nash.

(ii) Let agent j's endowment be V_j with $V_1 \leq V_2 \leq \cdots \leq V_n$. First we will argue that in equilibrium the central agent must have the lowest value. Suppose not. Then there exists agent $j \neq 1$ who is the central agent. Since this is a star agent 1 has a link to agent j. Then $k \in I \backslash \{j,1\}$ will receive a higher payoff by linking to player 1. Hence player 1 has to be the central agent. 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 since direct links should always be to the agent with least value. For g^{ps} to be Nash, we need $C_k(g^{ps}) - C_k(g') \leq 0$. This difference can be written as $\{3V_1 + (\Lambda - V_1 - V_k) - 4V_2\}/6 \leq 0$. Now, $(\Lambda - V_1 - V_k) \geq (n-2)V_2$ since V_2 is the smallest value in the set $I \backslash \{1, k\}$. Hence $\{3V_1 - 4V_2 + (n-2)V_2\}/6 \leq 0$, 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. ■

4. *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 $D_1, D_2, \ldots, D_\kappa$. Without loss of generality, we assume $D_1 \leq D_2 \leq \ldots . \leq D_\kappa$. 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. ■

5. *Proof of Theorem 2*: (i) 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 both efficiency and stability require all agents to be as close to each other as possible. (ii) Under property DC, only a connected network can be efficient. Moreover since cycles will raise the cost of acquiring information only trees can be 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 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 to increase 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 efficient graphs. ■

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

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

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	$\cdots$	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 ne 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 given us the chain shown in the figure. ■

7. *Proof of Proposition 5*: Consider player i in the network g, and an arbitrary player k at a geodesic distance $d_{ik} = d$ from player i. If $d = 1$ and player k has no other links, player i is better off deleting this link. Let $d > 1$. Then i's payoff from k is given by $V_k(\frac{d-1}{d} - \delta d)$. This is positive only when $f(d) = \frac{d-1}{d^2} > \delta$. However, $f(d)$ is a decreasing function of d which attains a maximum at 1/4. Hence if $\delta > 1/4$, the empty network is Nash.

For $\delta \leq 1/4$ we show that every Nash network must be connected using contradiction. Consider a non-empty Nash network g with two components $\mathcal{C}_1$ and $\mathcal{C}_2$. Without loss of generality let $|\mathcal{C}_1| \geq |\mathcal{C}_2|$ and let $j \in \mathcal{C}_1$ and let $i \in \mathcal{C}_2$. Since $\delta \leq 1/4$ we know that $\frac{d-1}{d} - \delta d \geq 0$ for $d \geq 2$. Given that $n \geq 3$, $\Pi_j(g + ij) \geq \Pi_j(g)$. Hence g could not have been Nash and for $\delta \leq 1/4$ every Nash network must be connected. This completes the proof of the second part. ■

8. *Proof of Proposition 6*: We will only prove part (*ii*) since (*i*) is easily obtained from this. Let player 1, the lowest value agent be the center of the star. She cannot improve her payoffs by altering her strategy. Player $i \in I\backslash\{1\}$ can either add more links, or sever the link to the center and make one or more links to other nodes. Using (**5**) it is easy to verify that increasing the number of direct links reduces total payoffs. Next consider making exactly one link to a non-central node (say player 2) while deleting the link to player 1. Let g^{ps} be the original star graph and $g' = g^{ps} - g_{k1} + g_{k2}$, $(1 < k < n)$ the modified graph. For g^{ps} to be the Nash we need $\Pi_k(g^{ps}) - \Pi_k(g') \geq 0$. We can write the difference between these 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 $(V_1 - V_2)(2\delta - 1)/2 \geq 0$ for $\delta \leq 1/4$ and the whole expression is positive when $\delta > 1/6$. Thus, a periphery-sponsored star graph will be Nash if $1/6 < \delta \leq 1/4$. This is also true when $V_i = V$ since $\Pi_k(g^{ps}) - \Pi_k(g') = (n-2)(6\delta - 1)V/6$. To show uniqueness, recall that the star architecture is the only possible network with $D < 3$. A center-initiated star or mixed star is never Nash since the central agent has negative payoffs. Hence the periphery-initiated star is the unique Nash equilibrium. ■

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] Caffarelli, F. (2004). Non-Cooperative Network Formation with Network Maintainence Costs, *EUI Working Paper ECO No. 2004/18.*

[6] Casson, L. (1994). Travel in the Ancient World, *Johns Hopkins University Press.*

[7] Cowan, R.S. (1997). A Social History of American Technology, *Oxford University Press.*

[8] Fabrikant, A., A. Luthra, E. Maneva, C. Papadimitrou and S. Schenker (2003). On a Network Creation Game, *PODC 2003*, 347-351.

[9] Galeotti, A., S. Goyal and J. Kamphorst (2004). Network Formation with Heterogeneous Players, Mimeo, *Erasmus University*.

[10] Ghosh, A. (1994). In an Antique Land, *Vintage Books*.

[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 (2003). Nash Networks with Heterogeneous Agents, *LSU Working Paper 2003-06*.

[13] Hedetniemi, S.M., S.T. Hedetniemi and A.L. Liestman (1988). A Survey of Gossiping and Broadcasting in Communication Networks, *Networks*, **18**, 319-349.

[14] 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.

[15] International Telecommunications Union (1999). Challenges to the Network: Internet for Development, *ITU Geneva*.

[16] Jackson, M.O. (2003). A Survey of Models of Network Formation: Stability and Efficiency, Mimeo, *CalTech*.

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

[18] Johnson, C. and R. Gilles (2000). Spatial Social Networks, *Review of Economic Design*, **5**, 273-299.

[19] Kannan, R. S. Sarangi and S.S. Iyengar (2004), Sensor-centric Energy-constrained Reliable Query Routing for Wireless Sensor Networks, *Journal of Parallel and Distributed Computing*, **64**, 839-852.

[20] Litan, R.E. and A.M. Rivlin (2002). Beyond the Dot.coms, *Brookings Institution Press*.

[21] Milgram, S. (1967). The Small World Problem, *Psychology Today*, **2**, 60-67.

[22] Rogers, E and D.L. Kincaid, (1981). Communication Networks: Towards a New Paradigm for Research, *Free Press*.

[23] 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.*

[24] Sarangi, S., L. Ray and R. Kannan (2004). The Structure of Information Networks, *LSU Working Paper 2004-04.*

[25] Slikker, M. and A. van den Nouweland (2001). Social and Economic Networks in Cooperative Game Theory, *Kluwer Academic Publishers.*

[26] Stover, J.F. (1999). Historical Atlas of the American Railroads, *Routledge Publishers.*

[27] Wellman, B. and S. Berkowitz [Eds.] (1988). Social Structures: A Network Approach, *Contemporary Studies in Sociology: Vol. 15*, Cambridge University Press.

[28] West, D.B. (1996). Introduction to Graph Theory, *Prentice Hall Inc.*

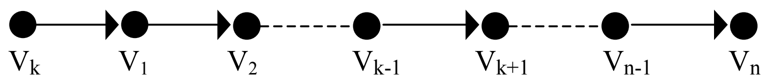

Figure 1: A Chain which is Nash

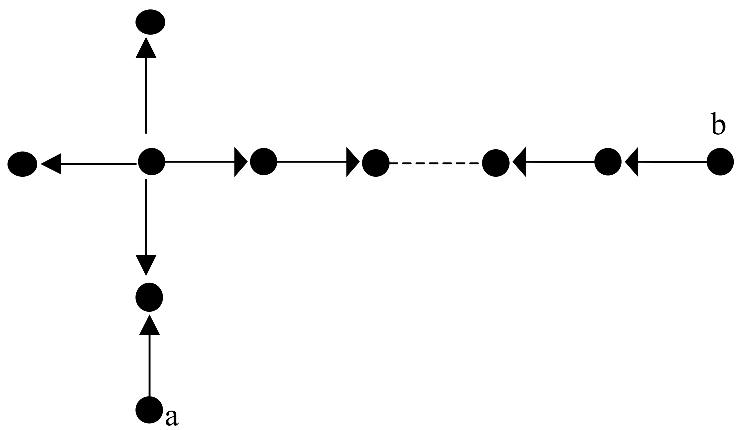

Figure 2: A concatenation of the Center-initiated Star and a Chain which is not in equilibrium

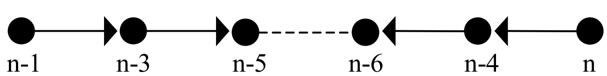

Figure 3: An Efficient Chain