Efficient Network Structures with Separable Heterogeneous Connection Costs

Heydari, Babak and Mosleh, Mohsen and Dalili, Kia

School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030, Facebook Inc, New York, NY 10003

27 May 2015

Efficient Network Structures with Separable Heterogeneous Connection Costs

Babak Heydari[a] [*], Mohsen Mosleh[a], Kia Dalili[b]

[a]School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030

[b]Facebook Inc, New York, NY 10003

April 28, 2015

Abstract

We introduce a heterogeneous connection model for network formation to capture the effect of cost heterogeneity on the structure of efficient networks. In the proposed model, connection costs are assumed to be separable, which means the total connection cost for each agent is uniquely proportional to its degree. For these sets of networks, we provide the analytical solution for the efficient network and discuss stability implications. We show that the efficient network exhibits a core-periphery structure, and for a given density, we find a lower bound for clustering coefficient of the efficient network.

Keywords: Complex Networks, Connection Model, Efficient networks, Distance-based utility, Core-periphery, Pairwise Stability

JEL Classification Numbers: D85

[*]Address: 1 Castle Point Terrace, Hoboken, NJ 07030, USA, e-mail: *babak.heydari@stevens.edu*, web: http://web.stevens.edu/cens/

1 Introduction

Network formation models are increasingly being used in a variety of economic contexts and other multi-agent systems. These models often study the structural conditions of efficiency, network social welfare, and stability, which is a measure of individual incentives to form, keep or sever links

We build our model based on the *Connection model* proposed by Jackson and Wolinsky (1996)[1] in which agents can benefit from both direct and indirect connections, but only pay for their direct connections. Benefits of indirect connections generally decrease with distance. Jackson and Wolinsky (1996) demonstrated that, for the homogeneous case, the efficient network can only take one of three forms: a complete graph, a star or an empty graph depending on connection cost and benefits. Several models have been proposed to introduce heterogeneity into the connection model, (see for instance: Galeotti et al. (2006); Jackson and Rogers (2005); Persitz (2010); Vandenbossche and Demuynck (2013)); the focus has mainly been on conditions for stability, with few references to efficiency. Finding general analytical solutions for the efficient networks with heterogeneous costs can be intractable, see for example Carayol and Roux (2009).

Here, we focus on finding efficient networks for a particular model of cost heterogeneity that we refer to as the *separable connection cost* model, in which shares of nodes' costs from each connection are heterogeneous, yet fixed and independent of to whom they connect. This is motivated by networks in which heterogeneous agents are each endowed with some resources (time, energy, bandwidth, etc) and the total resource needed to establish and maintain connections for each node can be approximated to be proportional to its degree. We further assume homogeneous benefits decaying with distance. We provide an exact analytical solution for efficient connectivity structures under these assumptions and show

[1] Jackson and Wolinsky (1996) developed their model based on the notion of pairwise stability or *two-sided link formation* where a link is formed upon the "mutual consent" of two agents. There is also another line of literature from Bala and Goyal (1997); Goyal (1993) that studies one-sided and non-cooperative link formation, where agents unilaterally decide to form the links with another agent.

that such networks have at most one connected component, exhibit a core-periphery structure and have diameters no larger than two. We further provide a lower bound for the clustering coefficient and discuss the pairwise stability implications of efficient networks.

2 Model

For n agents, let $b : \{1, ..., n-1\} \to \mathbb{R}$ represent the benefit that an agent receives from (direct or indirect) connections to other agents as function of the distance between them in a graph. Following Jackson and Wolinsky (1996), the (distance-based) utility function of each node, $u_i(g)$, in a graph g and the total utility of the graph, $U(g)$, are as follows:

$$\begin{aligned} u_i(g) &= \sum_{j \neq i: j \in N^{n-1}(g)} b(d_{ij}(g)) - \sum_{j \neq i: j \in N(g)} c_{ij} \\ U(g) &= \sum_{i=1}^{n} u_i(g) \end{aligned} \tag{1}$$

where $d_{ij}(g)$ is the distance between i and j, c_{ij} is the cost that node i pays for connecting to j, and b is the benefit that node i receives from a connection with another node in the network. We assume that $b(k) > b(k+1) > 0$ for any integer $k \geq 1$.

The network $\tilde{g}$ is efficient, if $U(\tilde{g}) \geq U(g')$ for all $g' \subset g^N$, which indicates that $\tilde{g} = \arg\max_g \sum_{i=1}^{n} u_i(g)$. Assuming the *separable* cost model as introduced earlier, connection costs in Equation 1, for a link between i and j can be written as $c_{ij} = c_i$, $c_{ji} = c_j$. We then introduce a connection cost vector, C, and without loss of generality rename nodes such that $c_1 < c_2 < ... < c_n$.

2.1 Efficient Structures under Separable Connection Costs

In Lemma 1, we determine the efficient structure for a connected component. Then in Proposition 1, we determine the structure of the efficient network in general.

Lemma 1 *If the efficient network with separable cost model is connected then it has a "generalized star" structure with the following characteristics: (a) All nodes are connected to node 1 (the node with minimum connection cost). (b) Nodes i and j $(i, j \neq 1)$ are connected iff $b(1) - b(2) > .5(c_i + c_j)$.*

Proof. Let N represents nodes in the connected network. If there exists a subset of nodes $M = \{v_1, ..., v_m\}$ $(M \subset N)$ that are not connected to node 1, we show that the network is not efficient. Since N is connected, there exists a set of links, $L = \{l_1, ..., l_m\}$ where l_i is adjacent to v_i[2]. Suppose l_i connects v_i to w_i ($w_i \neq 1$ by definition). Now, if we remove all l_is and connect all v_is to node 1, we have reduced the total connection cost of the network by $mc_1 - \sum_{k=1}^{m} c_{v_k} < 0$.

Now, to address the benefits, note that we have not changed the number of links; therefore direct benefits remain the same. Furthermore, the diameter of the new network is 2. Therefore every distance that is not 1 is capped at 2, making the total benefit larger than that of the original network. This results in at improvement in the total utility, indicating that the original network was not efficient.

Furthermore, having established that the maximum distance in the efficient network is no larger than two, every node i and j $(i, j \neq 1)$ are connected iif $b(1) - b(2) > .5(c_i + c_j)$.

■

Proposition 1 determines the structure of the efficient network and shows that the efficient network is a spectrum of solutions.

Proposition 1 *In the connection model, for a finite set of agents, $N = \{1, .., n\}$, if $c_{ij} = c_i$ for all $i, j \in N$, where $c_i \in C = \{c_1, c_2, ..., c_n\}$ and assuming, $c_1 < c_2 < \cdots < c_n$, the structure of the efficient network is as follows: Let m be the largest integer between 1 and n such that $2b(1) + 2(m-2)b(2) > (c_m + c_1)$. If $i > m$, then i is isolated. If $i \leq m$, then there*

[2]For any subset $M = \{v_1, \cdots, v_m\}$ of a connected network N $(M \not\equiv N)$, we can show that there are links $L = \{l_1, \cdots, l_m\}$ in N such that v_i is adjacent to l_i.

is exactly one link between i and 1; also there is one link between i and j $(1 < i, j \leq m)$ iff $b(1) - b(2) > .5(c_i + c_j)$.

Proof.

First, we show there is at most one connected component in the efficient network. Next, we find the condition for each node to be in the connected component, which has the generalized star structure according to Lemma 1.

Assume that the efficient network has more than one (e.g. two) connected components with (m_i, ℓ_i) being respectively the number of nodes and links in connected component i. According to Lemma 1, each connected component has a *generalized star* structure. The total benefit of each component is $B_1 = 2\ell_1 b(1) + (m_1(m_1 - 1) - 2\ell_1)b(2)$ and $B_2 = 2\ell_2 b(1) + (m_2(m_2 - 1) - 2\ell_2)b(2)$ respectively. Suppose h and h' are nodes with minimum costs in component 1 and 2 respectively and without loss of generality $c_h < c_{h'}$. If we disconnect all links connected to h' and connect them directly to h, total cost decreases by $(c_{h'} - c_h)$ per link. This also results in the total benefit $B = (\ell_1 + \ell_2)b(1) + ((m_1 + m_2)(m_1 + m_2 - 1) - 2(\ell_1 + \ell_2))b(2)$, which is strictly greater than $(B_1 + B_2)$.

To determine which nodes belong to the connected component, G_C, in the efficient network, we define, for node i, $A_i \triangleq 2b(1) + 2(k-2)b(2) - c_1 - c_i$ where k is the number of nodes in the connected component G_C. We show that i is in G_C *iff* $A_i \geq 0$. First, $A_i > 0$ is the sufficient condition for i to be in G_C. This is because connecting i to node 1 increases the total utility by exactly A_i, as the diameter of G_C is at most 2, according to Lemma 1. Also if $A_i < 0$, then i will be isolated so $A_i \geq 0$ is also the necessary condition. This is because i cannot be only connected to 1 since $A_i < 0$, so the only way for i to be connected is by having more than one link. From Lemma 1, for i to have a link to $j \neq 1$, we must have $c_i + c_j < 2b(1) - 2b(2)$. But: $c_i + c_j > c_1 + c_i > 2b(1) + 2(k-2)b(2) > 2b(1) - 2b(2)$, where we use the fact that $c_j > c_1$ and $A_i < 0$, so i cannot have more than one connection either, thus i will be isolated. Note that $A_i > 0$ also means $A_j > 0$ for all $j < i$ since $c_j < c_i$, thus all lower cost nodes will also be in G_C, so the smallest i for which $A_i < 0$ provides the size of the connected component in the efficient network.

■

2.2 Stability

Assuming heterogeneous, separable connection costs, are there structures that are simultaneously efficient and stable? To investigate this, we use the notion of *pairwise stability*[3]. According to Proposition 1, for a (pairwise) stable network to also be efficient, it has to have a single connected component with a generalized star structure.

We demonstrate that there are pairwise stable networks in the form of a generalized star. Such structures have one connected component with m' nodes, where m' is the largest integer between 1 and n such that $c_{m'} < b(1) + (m' - 2)b(2)$. With N' as the set of every node i where $c_i \leq b(1) - b(2)$, we show that for a stable network with a generalized star structure: 1) N' is a complete graph, 2) Any $i \in N'$ can be the hub, h, and every $j \notin N'$ for which $b(1) - b(2) < c_j < b(1) + (m' - 2)b(2)$ is connected to the hub, h.

Given $c_i < b(1) - b(2)$, any two agents in N' benefit from forming a link, so N' is a complete graph. No agent $i \in N'$ has an incentive to sever any of its links (inside or outside of N') as it pays $c_i < b(1) - b(2)$ per connection. Once all (e.g. m') nodes are connected to node $h \in N'$, severing a link by $j \notin N'$ for which $c_j < b(1) + (m' - 2)b(2)$ decreases its utility. Moreover, there are no two agents j and j' where $j \notin N'$ and $j' \neq h$ who can improve their utilities by agreeing on forming a mutual link given $c_j > b(1) - b(2)$.

The described structure is not the only possible stable structure as one can in general construct stable structures with diameters bigger than 2 [4], nor does this structure ensure efficiency. However for certain cost vectors we can create structures that are simultaneously stable and efficient. For example, let $C = \{c_1, \ldots, c_k, \ldots, c_m, \ldots, c_n\}$. Suppose we have

[3]According to Jackson and Wolinsky (1996), the network g is *pairwise stable* if:

(i) for all $ij \in g$, $u_i(g) \geq u_i(g - ij)$ and $u_j(g) \geq u_j(g - ij)$ and

(ii) for all $ij \in g$, if $u_i(g + ij) \geq u_i(g)$ then $u_j(g + ij) < u_j(g)$.

[4]For instance, if n=4 and for each agent i, $b(1) - b(3) < c_i < b(1)$, a *line* structure is also a stable network

k such that $c_k \leq b(1) - b(2)$ and $c_{k+1} > 2(b(1) - b(2))$. Suppose we have m such that $c_m < (m-1)b(2)$ and $c_{m+1} > b(1) + (m-2)b(2)$. If all $i \leq k$ form a complete graph and every $j(k < j < m)$ is connected to node 1, the resulting network is both stable and efficient.

2.3 Characteristics of Networks with Separable Cost Model

2.3.1 Core-Periphery structure

We show that the efficient network has a Core-periphery structure, a widely observed structure in various social and economic networks (i.e. see for example Zhang et al. (2014); Rombach et al. (2014)). We adopt the formal definition from Bramoullé (2007), which states that a graph g has a core-periphery structure when agents can be partitioned into two sets, the *core* C and the *periphery* P, such that all partnerships are formed within the core and no partnership is formed within the periphery. For an efficient network, let k be the largest integer between 2 and n such that $b(1) - b(2) > .5(c_{k-1} + c_k)$. The efficient network can be partitioned into a set $C = \{1, \ldots, k\}$, which forms a complete subgraph and a set $P = \{k+1, \ldots, n\}$ which can only have connections to the complete subgraph. If $b(1) - b(2) > (c_{k-1} + c_k)$ then k and $k-1$ are connected and there is also a link between every node $i, j(i, j \leq k$ and $c_i, c_j \leq c_k)$, which forms a complete subgraph. Similarly, we can show that for every $i \in P$, which is connected to a node j, $c_j \leq c_k$ and $j \notin P$.

2.3.2 Clustering coefficient

We find the global clustering coefficient of the efficient network and a lower bound for it. For a given density, the following structure results in minimum clustering coefficient for the connected component in the efficient network: Let m be the number of nodes and $\ell(\ell \geq m-1)$ represent the number of links in the connected component. Having nodes sorted according to connection costs and without violating the conditions for efficiency, we connect every two nodes such that minimum number of triangles are made as follows: Starting from node $k = 1$, we establish a link from node k to every node $i > k$. We repeat this process for

$k = 1, \ldots, n$ until the total number of links reaches ℓ. Therefore, we have $\ell = \sum_{k=1}^{p}(n-k)+J$ where $J < n-p+1$. At the k-th iteration, the total number of connected triplets is increased by $\binom{n-k}{2}+\binom{J}{2}$ and the total number of triangles is increased by $(k-1)(n-k)+pJ$. Therefore we have [5] :

$$
\begin{aligned}
C(g) &= \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of nodes}} = \frac{3 \times \{\sum_{k=1}^{p}(k-1)(n-k)+pJ\}}{\sum_{k=1}^{p}\binom{n-k}{2}+\binom{J}{2}} \\
&\geq \frac{3 \times \sum_{k=1}^{p}(k-1)(n-k)}{\sum_{k=1}^{p}\binom{n-k}{2}}
\end{aligned}
\tag{2}
$$

3 Conclusion

Heterogeneous yet separable connection costs cover important classes of real networks. We showed that efficient structures for such networks can be solved exactly and have diameter no larger than two; we also discussed the transitivity and core-periphery nature of such networks. Moreover, we showed the possibility of having simultaneous efficient and stable structures for these classes of network under certain cost vectors and benefit functions. Although benefits are still assumed to be homogeneous, one can easily take into account heterogeneity of direct benefits, as long as the separability assumption is maintained, i.e. cost and direct benefit terms appear together in all analysis and cost terms can capture heterogeneity of direct benefits by embedding them as an off-set in the fixed costs of nodes.

There are cases where the separable cost assumption does not hold, for example when the link cost is the function of how *similar* the two nodes are. These cases in general can be intractable and approximate methods such as *island models* as discussed in Jackson and Rogers (2005) can be used.

[5] For $n \geq 3$ and with some algebra we can show the inequality.

Acknowledgements

This work was supported by DARPA Contract NNA11AB35C. The authors are grateful to Peter Ludlow (Stevens) and Pedram Heydari (UCSD) for insightful comments.

References

Bala, V. and Goyal, S. (1997). Self-organization in communication networks. Technical report, Econometric Institute Research Papers.

Bramoullé, Y. (2007). Anti-coordination and social interactions. *Games and Economic Behavior*, 58(1):30–49.

Carayol, N. and Roux, P. (2009). Knowledge flows and the geography of networks: A strategic model of small world formation. *Journal of Economic Behavior & Organization*, 71(2):414–427.

Galeotti, A., Goyal, S., and Kamphorst, J. (2006). Network formation with heterogeneous players. *Games and Economic Behavior*, 54(2):353–372.

Goyal, S. (1993). *Sustainable communication networks*. Econometric Institute, Erasmus University Rotterdam.

Jackson, M. O. and Rogers, B. W. (2005). The economics of small worlds. *Journal of the European Economic Association*, 3(2-3):617–627.

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

Persitz, D. (2010). Core-periphery r&d collaboration networks. *Working paper*.

Rombach, M. P., Porter, M. A., Fowler, J. H., and Mucha, P. J. (2014). Core-periphery structure in networks. *SIAM Journal on Applied mathematics*, 74(1):167–190.

Vandenbossche, J. and Demuynck, T. (2013). Network formation with heterogeneous agents and absolute friction. *Computational Economics*, 42(1):23–45.

Zhang, X., Martin, T., and Newman, M. (2014). Identification of core-periphery structure in networks. *arXiv preprint arXiv:1409.4813*.