

Social Distance and Network Structures

Ryota Iijima and Yuichiro Kamada*

First Version: November 23, 2008;

This version: September 3, 2010

Abstract

This paper analyzes how agents' preferences determine the structures of networks. In our model, agents are endowed with their own multi-dimensional characteristics. We show that, when agents need many similar characteristics in order to prefer being connected with each other, stable networks are cliquish (they exhibit high clustering coefficients) but not closely connected (they have low average path lengths), and vice versa. One implication is that the introduction of new communication technology makes a network closely connected but not cliquish. We relate our model and results to the “strength of weak ties hypothesis,” the “similarity scale,” and the “communication externality.”

JEL Classification Numbers: D85, C72, A14

Keywords: Network formation, heterogeneity, spatial type topologies, clustering, average path length, small worlds, weak-ties

*Iijima: Graduate School of Economics, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, e-mail: iijimaaa@gmail.com; Kamada: Department of Economics, Harvard University, Cambridge, MA 02138, e-mail: ykamada@fas.harvard.edu; We are grateful to Drew Fudenberg, Akihiko Matsui, and Markus Mobius for helpful comments and conversations. Lauren Merrill read through previous versions of this paper and gave us very detailed comments, which significantly improved the presentation of the paper. We also thank Itay Fainmesser, Guido Imbens, Katsuhito Iwai, Kazuya Kamiya, So Kubota, David Miller, Yusuke Narita, Dan Sasaki, Satoru Takahashi and seminar participants at the Fifteenth Decentralization Conference at Tokyo, the 2009 Far East and South Asia Meetings of the Econometric Society, brown bag lunch seminars at Harvard University and the University of Tokyo, Game Theory Workshop 2010 at Kyushu University and Seventeenth Darwin Seminar at Tokyo Institute of Technology.

1 Introduction

The relevance of social networks in the context of economic interactions is increasingly acknowledged.¹ Numerous empirical works have shown that different networks have different structures, where the structures of networks are evaluated by various measures.² Three well-known and well-used measures of network structures are clustering coefficient, average path length, and degree distribution, which represent a network's cliquishness, connectedness, and heterogeneity (of the numbers of neighbors), respectively. This paper focuses on the former two measures, asking the following question: Why do some networks have high clustering coefficient and/or average path length, while others do not?³ The model in this paper provides a possible answer to this question, which gives economic insights about network structures. For example, we explain why the “e-mail network,” in which a link represents an incidence of an email exchange, has a lower clustering coefficient and a lower average path length (Ebel et al., 2002) than the “coauthorship network,” in which a link represents an incidence of coauthorship between two economics scholars (Goyal et al., 2006).⁴ Also, our model can predict how the introduction of new communication technology, such as the Internet, changes the structure of a network.

To answer our motivating question, we construct a model in which we suppose agents are endowed with their own multi-dimensional characteristics. We show that, when agents need many similar characteristics in order to be linked with each other, the stable network is cliquish but not closely connected, under certain regulatory conditions. On the other hand, when it is enough for agents to share a small number of similar characteristics in order to be linked, the stable network is closely connected but not cliquish. We relate our model and results to the “strength of weak ties hypothesis” of Granovetter (1973), the “similarity scale” of Tversky (1977), and the “communication externality” of Rosenblat and Mobius (2004).

A variety of models have been developed to illuminate why different networks have different structures. One of the standard assumptions often used in the economics literature is that people are partitioned into several groups,

¹See Goyal (2005) and Jackson (2008a), among others.

²For references on these empirical works, see, for example, Goyal (2005) and Jackson (2008a).

³We also show that our model accommodates a wide range of observed degree distributions in the data.

⁴Ebel et al. (2002) finds that the “e-mail network” has a clustering coefficient of 0.34 and the average path length of 5.0, while Goyal et al. (2006) finds that the “coauthorship network” has a clustering coefficient of 0.16 and the average path length of 9.5.

and the relationships within a group cost less than the relationships across groups (See Currarini et al. (2009) and Jackson and Rogers (2005)). Such a modeling assumption leads to so-called “small world” property observed in real networks, namely, high clustering coefficient and low average path length. While partitioning of agents implies that an agent belongs to exactly one group, in reality an agent belongs to multiple groups, or more generally, agents might be associated with a level or degree of tastes or attitudes. For example, one might be a student of a graduate school in economics while he likes listening to soft rock.

In modeling situations where agents have multiple aspects of characteristics, or where they belong to multiple groups, it is necessary for agents to have ways to integrate and evaluate the information about the relationships in different dimensions or groups. For example, is it the case that one can make friends with another when *either* the affiliations *or* the tastes matches? Or is it the case that *both* the affiliations *and* tastes have to match? Naturally, different criteria would be relevant in different networks, depending on what types of relationships we deem as “links.” Furthermore, different networks may have different numbers of “relevant dimensions,” as the dimensions that matter might be different across different societies. Some dimension, say religion, might matter a lot in some networks, but much less in others. Even without such cultural issues, it might be the case that the development of communication technology enables us to interact with each other based on new types of interests beyond geographic constraints, which would increase the number of relevant dimensions.

In light of this motivation, this paper models agents’ characteristics, or *types*, as points in a multi-dimensional *type space*, and analyzes how the network structure depends on the notion of distance on the type space. Each coordinate indicates some aspect of agents’ characteristics, such as jobs, locations, tastes, and so forth. Distance in the type space, which we call *social distance*, represents the level or amount of obstacles to their relations, so agents form links with others who are nearby.⁵ We consider a class of notions of distance, *k’th norms*, in which the distance between two points in the type space is the *k’th* smallest distance among *m* dimension-wise distances between them, where *m* denotes the number of dimensions of the type space.⁶ This class of distances is sufficiently tractable to obtain a closed-form solution, yet is complex enough that we can gain relevant economic intuition, for

⁵Akerlof (1997) uses the notion of social distance, too.

⁶Tversky (1977) claims that when similarity relationships are formulated on a multiple dimensional space, they often violate the triangle inequality. Note that our *k’th* norm with $k < m$ does not satisfy the triangle inequality. We will make this point clear in what follows.

example by implementing comparative statics. We first introduce a model based on the benefit and cost of link formation. Our assumptions here are that the benefit of a link is decreasing in the distance between two agents involved, and the cost is increasing linearly with respect to degrees. That is, agents obtain higher benefits from linking to a closer agent in the type space, while they need to pay a fixed cost to maintain a link.⁷ We show that in a unique pairwise stable network each agent has a cutoff on the distance to her neighbors, above which she does not have an incentive to form links.^{8,9} Conversely, for any network generated by a cutoff rule, there exists a pair of benefit and linear cost functions such that the network is a unique pairwise stable one. Based on these results, we then analyze the *cutoff rule model*, in which agents form links if the distance between them is no more than some exogenously given cutoff value. As an approximation of a large network, we focus on the limit of the network as the number of nodes goes to infinity and then the cutoff value goes to zero. It is shown that the limiting values of the clustering coefficient and average path length vary as we vary the value of k and/or m .¹⁰ We also show that a wide range of degree distribution can be obtained by varying the distribution of agents over the type space.

We also consider the case of nonlinear cost functions. We show that a pairwise stable network that is generated by the cutoff rule model always exists, and a strongly stable network always exists and is unique under certain circumstances.¹¹

Although in our model the heterogeneity of agents matters in terms of social distance between agents, there is another way to describe heterogeneity. Fujii and Kamada (2010) introduce a network formation model in which

⁷For example, Selfhout et al. (2009) empirically show that people are likely to be connected if their preferences for music are similar to each other. Although in practice it is sometimes beneficial for people to have links with someone who has very different characteristics, we abstract away from this possibility in this paper.

⁸A pairwise stability requires that no pair of agents would want to form or sever a link between them, with the rest of the network structure fixed.

⁹We do not explicitly model the process by which the pairwise stable network is reached. This is because our focus is not on the details of agents' strategic interactions, but on the question of why different networks have different structures.

¹⁰Watts et al. (2002) consider a model with stochastic link formation. In their model, two agents form a link with a probability that depends, in particular, on the distance between them. Each agent is associated with a vector of characteristics, and the distance measurement corresponds to the case of $k = 1$ in our model. They show that, for plausible range of parameters, networks are "searchable," meaning, in essence, that the average path length of the networks are low. This result is analogous to the part 2 of our Corollary 3 which states that the average path length is increasing in k .

¹¹The notion of strong stability is introduced in Jackson and van den Nouweland (2005). It corresponds to the notion of "core" in cooperative game theory.

each agent has her own intrinsic “sociability,” and two agents form a link if the benefit from the link formation exceeds some (exogenously given) link-formation cost, where the benefit is nondecreasing in their sociabilities.¹² Their model and ours provide complementary approaches to network formation. Indeed, by introducing social distance in their model, or by introducing sociability in our model, we could have a more realistic formation model of social networks. In this paper we do not analyze such a hybrid model in order to highlight the role that social distance plays in network formation.

One of the related themes in the literature is “homophily” in networks.¹³ This theme is closely related to our paper since, in our model, social distance describes the similarity between people’s characteristics. Although the literature on homophily focuses on similarity in terms of one dimension (e.g. ethnicity), our model deals with multiple dimensions. Johnson and Gilles (2000) analyze a related model in which links are formed based on costs that depend on geographical distances between agents. However, they consider only one dimensional spatial model, hence cannot capture the multi-dimensional relationships among agents which could naturally arise in the forementioned example of graduate school and testes for music.¹⁴

The paper is organized as follows. In Section 2, we introduce terminology of networks and the notion of social distance. In Section 3, we present a model of network formation based on the benefit and cost of link formation. In this section, the cost function is assumed to be linear with respect to the degree. It is shown that analyzing the cutoff rule model is enough to understand this model of benefits and costs. In Section 4, the main section, we analyze the cutoff rule model. Clustering coefficients, average path length, and degree distribution are studied. In Section 5, we first discuss the case of nonlinear cost functions, to which the application of the cutoff rule model is not straightforward. Then, we discuss the relationship of our model with the literature, first with the “strength of weak ties hypothesis” proposed by Granovetter (1973, 1995), second with the “similarity scale” proposed by

¹²Their model is a special case of the model proposed by Caldarelli et al. (2002). For a sociability model that takes sociability as a strategic choice variable, see Golub and Livne (2010).

¹³Homophily is a well-observed socio-psychological tendency of people to interact with others similar to oneself. To address this issue, Currarini et al. (2009) use search theoretic network formation model and fit their model to empirical data. Also, Jackson (2008b) formalizes a dynamic model that exhibits homophily and sees its implications on network structures.

¹⁴The literature on “latent space” tries to “embed” agents in given network data to multi-dimensional spaces. Although it also deals with multi-dimensional spaces, as will become clear, its approach differs from ours since it restricts attention only to Euclidean distance. See, for example, Hoff, et al (2002).

Tversky (1977), and finally with the notion of the “communication externality” introduced by Rosenblat and Mobius (2004). Finally we discuss an extension of the cutoff rule model to the case of stochastic link formations. All the proofs are relegated to the Appendix.

2 Definitions

2.1 Terminology

$N = \{1, 2, \dots, n\}$ is a finite set of nodes (or, agents). A *network* g is a set of links between agents in N . A *link* between agents i and j is denoted ij . We say $ij \in g$ if and only if there exists a link between agents i and j . Let $G(N)$ denote the set of all the possible networks defined on the set of agents, N . We focus only on non-directed networks, hence require $ij = ji$. We suppose $ii \notin g$ for all $i \in N$ by convention.

Agent i 's *neighbors* are $j \in N$ with $ij \in g$. Formally, the set of i 's neighbors in g , denoted by $N_i(g)$, is defined as: $N_i(g) = \{j \in N | ij \in g\}$. Agent i 's *degree*, $q_i(g)$, is the number of i 's neighbors, i.e. $q_i(g) = \#N_i(g)$.^{15,16}

A *path* between nodes i and j is a sequence of links $(i_1i_2, i_2i_3, \dots, i_{K-1}i_K)$ such that $i_1 = i$, $i_K = j$, and $i_k \neq i_{k'}$ for all $k \neq k'$. The *path length* between i and j , $PL_{ij}(g)$, is the length of the shortest path between i and j . If there exists no path between i and j , then the path length between i and j is infinite by convention. The *average path length*, $APL(g)$, is the average of $PL_{ij}(g)$'s over all ij 's that have finite path lengths.¹⁷

The *clustering coefficient*, $Cl(g)$, is the average of the probability that a given node's two neighbors are connected to each other. This measure represents the cliquishness of a network. Formally, first define *agent i 's clustering*, $Cl_i(g)$, as follows:

$$Cl_i(g) = \frac{\#\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\}}{\#\{jk | k \neq j, j \in N_i(g), k \in N_i(g)\}},$$

if the denominator is nonzero, and $Cl_i(g) = 0$ otherwise. The denominator in the above expression is $\frac{q_i(g)(q_i(g)-1)}{2}$, the number of possible pairs between i 's neighbors. The numerator is the number of links actually formed among such pairs. The clustering coefficient of a network g is given by $Cl(g) =$

¹⁵Thus, the cardinality of g is $\#g = \frac{1}{2} \sum_{i \in N} q_i(g)$ because $ij = ji$ for all $i, j \in N$.

¹⁶The convention is to denote a degree by d rather than q , but we reserve this notation for the later use when we deal with distances.

¹⁷Thus, strictly speaking, APL is defined only for nonempty networks, $g \neq \emptyset$.

$$\frac{1}{n} \sum_{i \in N} Cl_i(g).^{18}$$

We will suppress each measure's dependence on g when there is no risk of confusion.

2.2 Type Space and Social Distances

Each agent is assumed to be located on a point in $X = [0, 1]^m$, which we call *type space*. Every agent belongs to the type space: Denote by $x_i = (x_{i1}, \dots, x_{im}) \in X$ the point, or *type*, associated with agent $i \in N$.

We assume that x_i 's are independently and identically distributed according to a distribution with a strictly positive and absolutely continuous probability density function f over X . To simplify the analysis, we will assume that f is the uniform distribution except in Subsection 4.4, where it turns out that any of our results does not rely on this assumption.

As mentioned in the Introduction, we will consider various notions of distance (or *social distance* in other words) in the type space X . Specifically, define a class of social distances, which we call the k 'th norm:

Definition 1. For every pair of agents i and j in the type space, the **k 'th norm**, $d^{(k)} : N \times N \rightarrow \mathbb{R}_+$, measures the distance between them as follows:

$$d^{(k)}(i, j) = |x_{il} - x_{jl}| \quad \text{such that}$$

$$\#\{h : |x_{ih} - x_{jh}| \leq |x_{il} - x_{jl}|\} \geq k \quad \text{and} \quad \#\{h : |x_{ih} - x_{jh}| \geq |x_{il} - x_{jl}|\} \geq m - k + 1.$$

Note that this definition boils down to

$$d^{(k)}(i, j) = |x_{il} - x_{jl}| \quad \text{s.t.} \quad \#\{h : |x_{ih} - x_{jh}| < |x_{il} - x_{jl}|\} = k - 1$$

if there is no tie in dimension-wise distances.

To grasp the idea of the definition, suppose, for example, that two agents i and j are located on the type space X with $m = 4$. Their locations are $x_i = (0.3, 0.2, 0.4, 0.6)$ and $x_j = (0.7, 0.7, 0.7, 0.7)$. Then dimension-wise distances are $(0.4, 0.5, 0.3, 0.1)$. If we use 1'st norm, then $d^{(1)}(i, j) = 0.1$; if we use 2'nd norm, then $d^{(2)}(i, j) = 0.3$, and so forth.

¹⁸There is another concept of clustering coefficient, *overall clustering*, that does not average over agents' clusterings but over pairs of neighbors: $Cl(g) = \frac{\sum_i \#\{jk \in g | j \in N_i(g), k \in N_i(g)\}}{\sum_i \#\{jk | j \in N_i(g), k \in N_i(g)\}}$. Clustering coefficient in this paper gives more weights to the clusterings of low-degree nodes than does the overall clustering. The results in this paper do not hinge on the specific choice of the concept of clustering coefficient. Precisely, both concepts give exactly the same set of results in our model.

Considering the situation where agents use social distances when they evaluate the values of relationships with others, the interpretation of the notion of the k 'th norm is that if k is large, agents care about many aspects of others' types, while if k is small, then they care about very few aspects of other's types.^{19,20}

It would be desirable to have a much more complex notion of social distance (such as a weighted average of the k 'th norms over all k 's), but the simple class of distance that we suppose is tractable and yet is enough to obtain relevant economic intuition, for example because it is easy to implement comparative statics. We note that the most important property that drives our results in what follows is the violation of the triangle inequality. We will make this point clearer in the analysis that follows.

We will occasionally restrict our attention to the following special cases of interest, which correspond to $k = m$ and $k = 1$, respectively: the *Max norm*, $d^{max}(i, j) = \max_{1 \leq h \leq m} \{|x_{ih} - x_{jh}|\}$ and the *Min norm*, $d^{min}(i, j) = \min_{1 \leq h \leq m} \{|x_{ih} - x_{jh}|\}$. We sometimes use the notation $d(i, j)$, omitting " (k) ," "*max*," or "*min*," when there is no risk of confusion.

3 The Model and a Preliminary Result

3.1 The Model

An agent's payoff is composed of the benefit and cost associated with his neighbors,

$$u_i(g) = \left(\sum_{j \in N_i(g)} b(d(i, j)) \right) - c(q_i), \quad (1)$$

¹⁹Marmaros and Sacerdote (2006) claim that geographic proximity and race are more important determinants of social interaction than are common interests, majors, and family background. Although our treatment of different dimensions in the type space is symmetric, this symmetry assumption is not crucial to the main results. That is, even if the importance of each dimension differs, our main results remain unchanged: The lengths of different dimensions of the type space can be different.

²⁰One implicit assumption that we employ throughout this paper is that all agents use the same measure of social distance to evaluate the relationships with others. It would be more natural to consider the situation where different agents use different measures, but we abstract away from this possibility because our primary objective is to understand why different networks have different structures, and assuming heterogeneous measures is enough to obtain a meaningful results. Also, it is not obvious what different cutoff levels we should set for agents with different measures, so the results associated with heterogeneous measures would have inevitable arbitrariness. On the other hand, it might be possible that considering heterogeneous measures gives us new insights. We leave this possibility for future research.

where $b(\cdot) > 0$ is a weakly decreasing, left-continuous function and $c(\cdot)$ is a strictly increasing function. Interpretation is that $b(d(i, j))$ denotes the benefit that i obtains from link ij when the distance between i and j is $d(i, j)$, and $c(q_i)$ denotes the cost that i pays to maintain his q_i links.²¹ Let $\Delta c(q) = c(q + 1) - c(q)$ denote the marginal cost of adding one more link. Cost functions are assumed to be homogeneous across all the agents, and are either *linear* (i.e. $\Delta c(q)$ is constant), *concave* (i.e. $\Delta c(q)$ is decreasing), or *convex* (i.e. $\Delta c(q)$ is increasing).²²

We introduce two notions that characterize special classes of networks:

Definition 2. A network g is said to be **efficient** if $\forall g' \in G(N)$, $\sum_{i \in N} u_i(g) \geq \sum_{i \in N} u_i(g')$ holds.

Definition 3. A network g is **pairwise stable** if

$$\begin{aligned} &\forall ij \in g, u_i(g) \geq u_i(g - ij) \text{ and } u_j(g) \geq u_j(g - ij), \text{ and} \\ &\forall ij \notin g, u_i(g) \leq u_i(g + ij) \implies u_j(g) > u_j(g + ij).^{23} \end{aligned}$$

Pairwise stability is the notion that is proposed by Jackson and Wolinsky (1996). In a pairwise stable network, no two agents can mutually weakly (resp. strictly) benefit from adding (resp. deleting) a link between them. We employ this concept to analyze the situation in which each link is formed based on the players' mutual agreement.

In this section, we consider how agents form links in a pairwise stable network. In particular, we will show that a pairwise stable network can be characterized by a class of simple decision rules, or *cutoff rules*. Under such a rule, agents have their own cutoff social distances, above which they do not form links and below which they form links:

Definition 4. g is generated by a **cutoff rule** with $(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_n) \in R_+^n$ if $ij \in g \iff d(i, j) \leq \min\{\hat{d}_i, \hat{d}_j\}$.

We call the above $(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_n)$ a *cutoff value profile*. Note that, given g , a cutoff value profile is not unique in general. For example, if $(\hat{d}_1(g), \hat{d}_2(g), \dots, \hat{d}_n(g))$ is a cutoff value profile for g and $\hat{d}_i \geq \hat{d}_j$ for all $j \in N$, then $(\hat{d}_1(g), \hat{d}_2(g), \dots, \hat{d}_i(g) + \epsilon, \dots, \hat{d}_n(g))$ is also a cutoff value profile where $\epsilon > 0$. We say that a cutoff value profile is *homogeneous* if for all $i, j \in N$, $\hat{d}_i = \hat{d}_j$. Otherwise we say it is *heterogeneous*.

²¹Our specification supposes that cost does not depend on the identity of neighbors. This is without loss of generality so long as the cost is the sum of two terms: the term that is increasing in the number of neighbors and the term that is increasing in the distance.

²²Note that the concavity (resp. convexity) in our notation corresponds to the strict concavity (resp. strict convexity) in usual conventions. We choose this wording just to ease the exposition.

²³By convention, we use " $g + ij$ " for $g \cup \{ij\}$ and " $g - ij$ " for $g \setminus \{ij\}$.

3.2 A Preliminary Result: Linear Cost Functions

In general, a pairwise stable network may be neither unique nor efficient.²⁴ Moreover, a cutoff value profile does not necessarily exist for a pairwise stable network, i.e. a pairwise stable network may not be generated by a cutoff rule.²⁵

The next proposition states that a model with linear cost functions in which we concentrate on pairwise stable networks and the cutoff rule model with a homogeneous cutoff value profile are equivalent, in the sense made clear in the following:

Proposition 1.

1. *Suppose that the cost function is linear, i.e. $c(q) = c_0 + c_1 q$ for some constants c_0 and $c_1 > 0$. Then, a pairwise stable network g exists, and it is unique and efficient. Furthermore, g is generated by a cutoff rule with a homogeneous cutoff value profile.*
2. *Conversely, for any network g that is generated by a cutoff rule with a homogeneous cutoff value profile, there exists a benefit function $b(\cdot)$ and linear cost function $c(\cdot)$ such that g is a unique pairwise stable and efficient network with respect to the pair (b, c) .*

Proof Idea. For Part 1, the proof is constructive. As the marginal cost of any additional link formation is constant, in a pairwise stable network link ij exists if and only if $b(d(i, j))$ is no less than that marginal cost. It is straightforward to see that this type of network is unique, and also that we can use the distance that equates the benefit and the marginal cost as a cutoff value. As the marginal cost is homogeneous and constant, this cutoff value must be homogeneous across agents. Efficiency is also straightforward from the assumption of constant marginal cost. For Part 2, we simply construct a (b, c) pair such that agents would want to form a link with others if and only if they are within the cutoff distance. Uniqueness and efficiency follows directly from Part 1. See Appendix A.1 for details. \square

Thus, in short, we are justified in working with the simpler cutoff rule model instead of working with potentially very complicated benefit-cost functions. Note that the proposition deals with only linear cost functions. The case with nonlinear cost functions is discussed in Subsection 5.4, in which

²⁴See Jackson (2005) for discussions on this issue.

²⁵See Subsection 5.4 for an example of a pairwise stable network in which there is no cutoff value profile.

we show that our main results roughly carry over even in such a setting. Note also that the result is independent of the choice of the notion of social distance (the choice of k and m).²⁶

In the next section, we analyze the cutoff rule model in detail when the number of agents goes to infinity and the cutoff values go to zero. Proposition 1 justifies the results in Propositions 2, 3, and 4 in that section, if we consider situations where the benefit from forming links when the distance is very short decreases fast, and/or the marginal cost of forming an additional link is large.²⁷ In particular, we could directly use these results if we wanted to fit the model to data on real networks to estimate parameters, without a reference to specific benefit and cost functions, provided that we know that the cost function is linear.²⁸

4 Main Section: Cutoff Rule Model

In this section, we present the *cutoff rule model* and analyze how and why different notions of social distances result in different network structures, characterized by clustering coefficient and average path length. Furthermore, we show that a wide range of degree distributions can be obtained by adjusting the agents' distribution over the type space, without changing the results about clustering coefficient and average path length.

As we have shown in the previous section, the cutoff rule model presented here can be interpreted as the model in which agents have benefit and cost functions as in equation (1). The simplicity of the cutoff rule model enables us to obtain results that are expositionally neat and hence appeal to our intuition.

4.1 The Cutoff Rule Model

Each agent i is associated with her own *cutoff value*, denoted by \hat{d}_i . The link between i and j is formed if and only if the distance between them is no more than both i 's cutoff value and j 's. Formally, $ij \in g \iff d(i, j) \leq \min\{\hat{d}_i, \hat{d}_j\}$. In this section, we assume that the cutoff values are common to all the agents. Let the common value be \hat{d} . That is, $\forall i \in N, \hat{d}_i = \hat{d}$.

²⁶This independence is also true when we analyze the case of nonlinear cost functions in Subsection 5.4.

²⁷These situations correspond to taking the limit of $\hat{d} \rightarrow 0$. To see this point, see the proof of Part 2 of Proposition 1.

²⁸Again, note that we discuss nonlinear cost functions in Subsection 5.4.

4.2 Clustering Coefficient

In this subsection, we will analyze how the clustering coefficient depends on the property of the social distance in consideration. We focus on the clustering coefficient in the limit as n tends to infinity and then \hat{d} tends to zero. Formally, we consider the value of Cl^* , defined as follows:

$$Cl^* = \lim_{\hat{d} \rightarrow 0} Cl^{\hat{d}} \quad \text{where} \quad Cl(g) \xrightarrow{n \rightarrow \infty} Cl^{\hat{d}} \quad \text{almost surely.}^{29}$$

Note that the strong law of large numbers ensures that the limit is unique almost surely. Notice the order of the limit. If the order were reversed, this value would be trivially zero for any value of k . For, if we let \hat{d} tend to zero with some fixed n , the set of i 's neighbors would eventually become empty for any $i \in N$ almost surely. Recalling that i 's clustering is zero when i has no more than two neighbors, we would conclude that $\lim_{\hat{d} \rightarrow 0} Cl(g) = 0$, hence $\lim_{n \rightarrow \infty} [\lim_{\hat{d} \rightarrow 0} Cl(g)] = 0$.

We will sometimes use the notation $Cl^*(k, m)$ instead of Cl^* , to make it clear that this value depends on the social distance in consideration.

We can solve for this limit clustering for every pair of k and m .

Proposition 2. *For each m and $k \leq m$,*

$$Cl^*(k, m) = \binom{m}{k}^{-1} \left(\frac{3}{4}\right)^k.$$

Proof Idea. To understand the intuition, consider the case of $m = 2$ and $k = 1$. A typical agent i in the interior of the type space has three classes of neighbors: those who are close to i only with respect to the first dimension, those of only second dimension, and those of both dimensions. As the cutoff goes to zero, the probabilities that agent j being the first, second, and third class, given that he is a neighbor of i , converge to $1/2$, $1/2$, and 0 , respectively. Thus the probability that two of i 's neighbors chosen randomly are of the same class converges to $1/2$. The probability that the two first class neighbors being connected to each other is the probability that two points in a unit interval have a distance no more than 0.5 , which is $3/4$. Hence, in this case, the clustering coefficient converges to $1/2 \cdot 3/4 = 3/8$. See Appendix A.2 for details. \square

To understand the formula given in the above proposition, consider the extreme cases: the Max norm and the Min norm.

²⁹We could have results for $Cl^{\hat{d}}$ for fixed $\hat{d} > 0$, but we consider the only limit value to make the exposition neat and highlight the effects of social distance on network structures.

Corollary 1. *If $1 < m < 9$, $Cl^*(m, m) > Cl^*(1, m)$ holds. That is, Cl^* is higher with the Max norm than with the Min norm.*

The proof follows straightforwardly from Proposition 2, and is relegated to Appendix A.2.

If $m < 9$ holds then the clustering coefficient is higher if social distances are measured by the Max norm ($k = m$) than by the Min norm ($k = 1$). The intuition is that, under the Max norm, the triangle inequality provides an upper bound of the distances between i 's neighbors. Therefore, i 's neighbors are relatively closely located to each other. For the Min norm, however, this is not the case because the triangle inequality is not satisfied: two agents being neighbors of a common agent does not provide an upper bound of their distance, so it is possible that i 's neighbors are quite far away from each other.

The second corollary of Proposition 2 is the following comparative statics:

Corollary 2.

1. $Cl^*(k, m)$ is decreasing in m .
2. $Cl^*(k, m)$ is nonincreasing in k when k is small, reaches its minimum at $k = \lfloor \frac{4}{7}(m + 1) \rfloor$ (provided such k is no more than m), and is nondecreasing when k is large where $\lfloor \cdot \rfloor$ denotes the Gaussian.

Again, the proof is straightforward from Proposition 2, and is relegated to Appendix A.2. We note that in part 2 of this corollary, “nonincreasing” and “nondecreasing” can be replaced with “decreasing” and “increasing,” respectively, except at a possible indifference at the minimum.

According to the first part of Corollary 2, if the number of dimensions of the type space becomes large with fixed k , then the resulting network becomes less cliquish.³⁰ Thus, for example, the introduction of new communication technology, which would increase the number of relevant dimensions, makes a network less cliquish. The second part states that, with fixed m , there is a nonmonotonic relationship between the clustering coefficient and k . A bit more specifically, networks are more cliquish either when agents care about very few aspects of others, or when they care about many aspects of others,

³⁰A similar argument is informally discussed in Chwe (2000). He considers the case of the Max norm in our terminology, and claims that “lower dimension networks have higher transitivity.”

given that the number of relevant dimensions are not too few.^{31,32}

This explains why the “e-mail network” of Ebel et al. (2002) has a lower clustering coefficient than the “coauthorship network” of Goyal et al. (2006): The incidence of e-mail exchanges does not require many similar aspects between the sender of the email and the receiver, while coauthoring needs many similar interests.

4.3 Average Path Length

In this subsection, we solve for the average path length for each k . As in the previous subsection, we focus on the limit value APL^* , formally defined by:

$$APL^* = \lim_{\hat{d} \rightarrow 0} APL^{\hat{d}} \quad \text{where} \quad APL(g) \xrightarrow{n \rightarrow \infty} APL^{\hat{d}} \quad \text{almost surely.}$$

Again, the order of the limit is important. If it were reversed, then it would not be well-defined, as $APL(g)$ is defined as the average of *finite* path lengths, while as the cutoff goes to zero with a fixed number of agents, all the pairs of agents have the path length ∞ almost surely. We also use $APL^*(k, m)$ as before.

The following proposition gives the formula of APL^* for the k 'th norm with $k < m$.

Proposition 3. *Take any k and m such that $k < m$. Then, $APL^*(k, m)$ is $\lfloor \frac{m}{m-k} \rfloor + 1$ if $\frac{m}{m-k}$ is not an integer and $\frac{m}{m-k}$ if it is, where $\lfloor \cdot \rfloor$ is the Gaussian.*

³¹Assuming that the relevant k is not too small, this result seems to be consistent with the empirical result given by Rapoport and Horvath (1961). Rapoport and Horvath (1961) analyze the survey data collected at a junior high school in Ann Arbor area shortly after the beginning of the 1960-1961 school year. In the survey, students are asked to list 10 friends from the first to the tenth. Based on the data, they generate a network with links of the l 'th and the $(l+1)$ 'th friends, for various values of l . They find that θ , a parameter similar to clustering coefficient, decreases with respect to l . Roughly speaking, θ is defined by the fraction of the overlap of the sets of neighbors of two connected agents (The formal definition can be found in Rapoport (1953). It can be shown that θ is monotone in clustering coefficient in (appropriately defined) large networks). Assuming, although arguably, that an agent's closer friends share more aspects that are similar to hers than far friends do, this result is consistent with our result that Cl^* is increasing in k for values of k close to m .

³²As we have explained, the “decreasing” part is quite intuitive. The “increasing” part is due to the “combination” term: $\binom{m}{k}^{-1}$. When m is large and $k(< m)$ is close to m , the change from k to $k+1$ makes the number of possible combinations of dimensions at which two neighbors of agent i are close to each other significantly lower. Thus the probability of i 's two neighbors being connected with each other rises as we move from k to $k+1$.

Proof Idea. To understand the intuition, consider the case of $m = 5$ and $k = 3$, and suppose for a moment (only in this paragraph) that we are dealing with continuum of agents. Almost surely, a pair of agents i and j satisfies $d^{\min}(i, j) > 0$. Let $x_i = (0.3, 0.3, 0.3, 0.3, 0.3)$ and $x_j = (0.7, 0.7, 0.7, 0.7, 0.7)$. Letting $x_1 = (0.7, 0.7, 0.3, 0.3, 0.3)$, $x_2 = (0.7, 0.7, 0.7, 0.7, 0.3)$, we construct a path: $(x_i x_1, x_1 x_2, x_2 x_j)$. In each link, the first $5 - 3$ elements change from 0.3 to 0.7. This change ends in $\lfloor 5/(5 - 3) \rfloor + 1$ steps. The proof is a bit more involved since we deal with a finite number of agents. See Appendix A.3 for details. \square

To understand the proposition, consider the following corollary:

Corollary 3. *Take any k and m such that $k < m$. Then,*

1. $APL^*(k, m)$ is decreasing in m .
2. $APL^*(k, m)$ is increasing in k .

The proof is straightforward from the formula given in Proposition 3, hence is omitted. The average path length in a network, in the limit, tends to be small if the type space is rich (if m is large), and/or if agents do not care about many aspects of the others (if k is small). Thus, the introduction of new communication technology, which would increase the number of relevant dimensions, makes a network closely connected. Also, a network is closely connected if it is enough for people to have a small number of similar aspects for them to be connected. Also, this explains why the “e-mail network” of Ebel et al. (2002) has a lower average path length than the “coauthorship network” of Goyal et al. (2006): The incidence of e-mail exchanges does not require many similar aspects between the sender of the email and the receiver, while coauthoring needs many similar interests.

The result is intuitive: If the number of dimensions at which agents must have similar characteristics to form a link is small relative to the richness of the type space, then each agent has neighbors who have many aspects of characteristics that are different from his ones. Thus, it is easy to have access to agents with very different characteristics through the network.

Proposition 3 rules out the case of the Max norm, where k is exactly equal to m . The next proposition concerns this case.

Proposition 4. $APL^*(m, m) = \infty$.

Proof Idea. Generically two randomly chosen points in the type space have a strictly positive dimension-wise distance for each dimension. Since under the Max norm two agents are linked with each other only if they are within

the cutoff distance with respect to all dimensions, the path length between any randomly chosen agents (generically) goes to infinity as the cutoff goes to zero. See Appendix A.4 for details. \square

Here, we see a striking difference between the k 'th norm with $k < m$ and the Max norm with $k = m$. With any $k < m$, the average path length takes some finite value in the limit as the cutoff goes to zero. But with $k = m$, the average path length goes to infinity. Furthermore, the result in Proposition 4 is true also in the case of Euclidian norm, $d(i, j) = [\sum_{k=1}^m (x_{ik} - x_{jk})^2]^{\frac{1}{2}}$. Notice that the triangle inequality is satisfied by the Max norm (and Euclidian norm), but not by the k 'th norm with $k < m$. Social distances with the triangle inequality describe situations where an agent's neighbors cannot be very far away from each other, which suggests that in such cases path lengths in networks tend to be large.

Propositions 2, 3, and 4 constitute our main results in this paper: The structures of networks, measured by clustering coefficient and average path length, vary with the social distance in consideration. By exploiting the results in Corollaries 2 and 3, one can provide an interpretation for a variety of networks, and fit our model to data by adjusting the parameters in the model, such as k and/or m .

Except for under the Max norm, our model has the “small world” property, i.e. networks have smaller average path lengths compared with lattice networks and larger clustering coefficients compared with randomly generated networks.^{33,34} This is a well-observed property in a variety of networks in reality and has been much studied in the literature. This property is generated in the existing literature by, for example, “rewiring” process (Watts and Strogatz, 1998), hub nodes (Barabasi and Albert 1999), or partitioning of agents into several groups (Jackson and Rogers 2005). Our model gives an alternative explanation for the “small world” property, which depends on the multi-dimensionality of the type space.

4.4 Degree Distribution

So far we have assumed that agents are uniformly distributed over the type space X . But a network generated by such a model has the degree distribu-

³³Precisely, the average path length in a large lattice network is very large if the expected degree is moderate. In a random network in which the probability of link formation between any pair of nodes is p , the clustering coefficient is p . But if the network is large and the expected degree is moderate, p needs to be very small, which results in a very low clustering coefficient.

³⁴In Subsection 5.5 we consider models in which networks with the “small world” property arise even under the Max norm.

tion that converges in distribution to a degenerate point mass distribution, contradicting the empirical evidence about degree distributions.³⁵ As mentioned earlier, however, our results do not rely on the assumption of the uniform distribution. In this subsection, we state this claim formally, and further show that by appropriately changing the distribution over the type space, we can generate a wide range of degree distributions.

Proposition 5. *For any strictly positive and absolutely continuous probability density function f over X , Cl^* and APL^* in Propositions 2, 3, and 4 remain the same.*

Proof Idea. If the distribution over the type space is strictly positive and absolutely continuous over X , we can show that f restricted to the \hat{d} -neighborhood (with respect to any norms) of any agent is close to the uniform distribution as $\hat{d} \rightarrow 0$. The continuity of the clustering coefficient with respect to the distributions of nodes implies the Cl result. The APL result is straightforward since the proofs of Propositions 3 and 4 do not depend on any specific assumptions on the distribution of nodes as long as it is strictly positive. See Appendix A.5 for details. \square

Now, we know that the results we have so far are robust to changes in the distribution over X , as long as the distribution is strictly positive and absolutely continuous. The next proposition shows that a wide range of degree distributions can be attained in the model. Before stating the result, we need a piece of notation:

Definition 5. The **relative degree** of agent i , denoted p_i , is

$$p_i = \frac{q_i}{\max_{j \in N} q_j} \in [0, 1].$$

Thus, the relative degree of agent i is the ratio of his degree to the maximum of the degrees of all the agents. Again, we consider the limit of p_i as n goes to infinity and \hat{d} goes to zero:

$$p_i^* = \lim_{\hat{d} \rightarrow 0} p_i^{\hat{d}} \quad \text{where} \quad p_i \xrightarrow{n \rightarrow \infty} p_i^{\hat{d}} \quad \text{almost surely.}$$

We refer to the minimum of the relative degrees as the “minimum relative degree.”

The next proposition shows that a wide range of relative degree distributions can be obtained by appropriately changing the distribution over X .

³⁵See Barabasi (2002), among others.

Proposition 6. *For any distribution of limit relative degrees with minimum relative degree greater than $\frac{m-k}{m}$, there exists a distribution f that generates it.*

Proof Idea. The simplest way to construct an appropriate distribution over the type space is to deal with the marginal distribution over an arbitrary chosen single dimension, while keeping the marginal distributions on other dimensions uniform. See Appendix A.6 for details. \square

The condition on the minimum relative degree says that if the degree distribution is “not too diverse,” our model can generate that distribution. We note that the condition can be vacuously satisfied for the Max norm, unless the minimum relative degree is exactly equal to zero. Non-degenerate degree distributions are often found empirically. In particular, the scale-free distribution have found extensively in the literature.³⁶ Although many models have been proposed that explain these phenomena, Proposition 6 provides another explanation for them: An agent has high (resp. low) degree if he has many (resp. a few) friends who have similar types to his one.

5 Discussions

In this section, We relate our model to the “strength of weak ties hypothesis” of Granovetter (1973), the “similarity scale” of Tversky (1977), and the “communication externality” of Rosenblat and Mobius (2004).

We also discuss two extensions of our model. In one extension, we consider the possibility that cost functions are nonlinear. Although there might exist multiple pairwise stable networks and a homogeneous cutoff value profile might not exist in this context, we show that analogous results to our main results can be obtained by replacing pairwise stability and a homogeneous cutoff value profile with a stronger notion of strong stability (Jackson and van den Nouweland, 2005) and a heterogeneous cutoff value profile, respectively.

In another extension, we consider the possibility that the link formation is stochastic. In this context, as in Watts and Strogatz (1998), a small stochastic component is enough to make the average path length significantly low while keeping the clustering coefficient almost unchanged.

³⁶The scale-free distribution, which was originally discovered by Pareto (1896), is observed in a variety of networks. Pareto (1896) finds that wealth distribution in Italy had the scale-free feature. Note that the scale-free distribution is often found as a property of the tail of degree distributions, so our condition on the minimum relative degree does not contradict the scale-free distribution observed in the data.

5.1 Information Diffusions and the “Strength of Weak Ties Hypothesis”

Granovetter (1973,1995) observes that “weak ties” bring more useful information than “strong ties”, where the strength of the relationships is measured by frequency of interactions. He also discusses that strong ties typically tend to be informationally redundant because relationships tend to be transitive (e.g. a friend of one’s friend is also his friend) in such cases. On the other hand, weak ties bring new information, since the relationships tend not to be transitive. In our model, link formations with the k ’th norm with small k is closely related to that of weak ties, because with small k , for two agents to be linked with each other, they need not share many similarities. This suggests that varying parameters of social distance in our model has an implication for information dissemination.

To formalize this idea, consider the following stylized model of information dissemination. At period 0, some agent $i \in X$ is randomly selected and obtains a piece of information, which has *value* δ^T for j if $j \in N$ knows it at period $T > 0$ for the first time, where $0 < \delta < 1$. At each period, each agent h meets with his neighbor $h' \in N_h(g)$ with probability $p(D(h, h'))$ and passes the information to her, where $D(h, h')$ is the number of dimensions in which the dimension-wise distance between h and h' is no more than \hat{d} .³⁷ We assume that $p(\cdot)$ is strictly increasing, to capture the idea that if h and h' have many similar characteristics (ex. sharing a workplace or having similar tastes for music), they communicate often. Note that as n goes to ∞ and \hat{d} goes to 0, the fraction of linked pairs that have $D = k$ approaches 1 for almost sure events for any given k . So (with a slight abuse of notation) we simply let $D = k$ and assume that p is strictly increasing in k . This implies that for $P = \{p' | p' = p(k) \text{ for some } k\}$, $p^{-1}(p')$ for $p' \in P$ is strictly increasing. Hence, we expect that the network is generated with high k if the probability of meeting (frequency of interaction) of linked agents is high, that is, ties are strong.

Now, the (ex ante) value of the information in the limit as n goes to infinity and \hat{d} goes to zero is simply δ^{APL^*} almost surely, since agent h can communicate with some agent in any open subset of the set of points within distance \hat{d} from h , if n is sufficiently large. Recalling from Corollary 3 that APL^* is increasing in k , we have that the value of the information is decreasing in k .

Since we have concluded that k is increasing in p (the frequency of in-

³⁷For simplicity, we assume that the informational is conveyed truthfully when information transmission occurs.

tereaction), we can conclude that the value of the information is decreasing in p , that is, when ties are strong, the information is less valuable. Conversely, when ties are weak, the information is more valuable. Therefore, the “strength of weak ties” results.

5.2 The k 'th Norm and Similarity Scale

In this subsection, we discuss the relationship between the k 'th norm and the “similarity scale,” defined in Tversky (1977). Tversky (1977) derives a representation of the “similarity scale” that satisfies several assumptions. The similarity between agent i and agent j is measured by the relationship between their profiles of characteristics, I and J , respectively. For example, we may have $I = \{\text{student, conservative, New York}\}$ and $J = \{\text{student, progressive, London}\}$. The representation of the similarity scale, denoted $s(i, j)$, is as follows:

$$s(i, j) = \theta f(I \cap J) - \alpha f(I \setminus J) - \beta f(J \setminus I),$$

where $\theta, \alpha, \beta \geq 0$. For simplicity, let us assume that $f(A) = \#A$. In order to accommodate this function to our model with type space, let $I = \{x_{i1} \dots, x_{im}\}$ and $J = \{x_{j1} \dots, x_{jm}\}$, and say that $x_{il} \sim x_{jl}$ if and only if $x_{il} - x_{jl} \leq \hat{d}$, where $\hat{d} > 0$. Notice that this binary relation \sim does not satisfy transitivity. We let $I \cap J = \{x_{il} | x_{il} \sim x_{jl}\}$, $I \setminus J = \{x_{il} | x_{il} \not\sim x_{jl}\}$, and $J \setminus I = \{x_{jl} | x_{il} \not\sim x_{jl}\}$.

Then, with the notation $D(i, j)$, which is defined as the number of “common dimensions” (see Subsection 5.1 for the definition), we have:

$$\begin{aligned} s(i, j) &= \theta f(I \cap J) - \alpha f(I \setminus J) - \beta f(J \setminus I) \\ &= \theta D(i, j) - (\alpha + \beta)(m - D(i, j)) \\ &= (\theta + \alpha + \beta)D(i, j) - (\alpha + \beta)m. \end{aligned}$$

This implies that, i and j 's similarity is no less than some threshold value \hat{s} if and only if

$$D(i, j) \geq \frac{\hat{s} + (\alpha + \beta)m}{\theta + \alpha + \beta}.$$

But this is equivalent to saying that i and j are within distance \hat{d} under the k 'th norm where $k = \lfloor (\hat{s} + (\alpha + \beta)m) / (\theta + \alpha + \beta) \rfloor$ and $\lfloor \cdot \rfloor$ is the Gaussian. That is, similarity, as measured in the sense of Tversky (1977), being no less than some threshold implies link formation under the k 'th norm in our model. It would be easy to see that the converse holds, too. That is, link formation under the k 'th norm in our model implies that similarity, as measured in the sense of Tversky (1977), is no less than some threshold.

5.3 Implication for Social Welfare: the Communication Externality

In this subsection, we analyze the implication of our model for social welfare, following a procedure similar to the one in Rosenblat and Mobius (2004, hereafter RM). Following RM, we define the *communication externality* in network g , denoted by $CE(g)$, as follows:

$$CE(g) = T(g) - v(g) + w(g).$$

The three terms represent the communication externality that results from agents' link formations. Concisely, $T(g)$ denotes the benefit of transmitted ideas from other agents, $v(g)$ is the cost generated by the difference in the preferences of agents, denoted by ΔE , and $w(g)$ is the benefit of (informal) institutions serving the needs of specific groups. All of these terms are nonnegative. We explain each of the three terms below. Detailed discussion is available in RM. In what follows, we assume that network g is a limit network with $n \rightarrow \infty$ and $\hat{d} \rightarrow 0$.

Transmission of ideas: $T(g)$ is the benefit of transmitted ideas from other agents. Some examples of these welfare-improving ideas are innovative technologies or information about job opportunities. Seminal empirical studies are Rogers (1995) and Granovetter (1995), respectively. Following the discussion in the previous subsection, let us assume that $T(g)$ is decreasing in APL^* .³⁸

Corollary 3 tells us that APL^* is increasing in k and decreasing in m . Hence we conclude that $T(g)$ is decreasing in k and increasing in m .

Cost of differences: $v(g)$ denotes the cost generated by the difference in the preferences of agents, ΔE . $v(g)$ is increasing in ΔE . For example, more diverse preferences for political policies (hereafter "political types") make collective decision making more difficult.³⁹ To formulate ΔE , we assume, as in RM, that agents are partly influenced by the neighbor's preferences.⁴⁰ For simplicity, suppose that a specific one of m axes of the type space, say axis 1, describes the political type. Let μ denote the mean of the marginal

³⁸RM assume $T(g)$ to be inversely proportional to the degree of *individual separation*, which is defined in their model. They note that the degree of individual separation is closely related to average path length of the network.

³⁹Alesina, Baqir and Easterly (1999) show that the heterogeneity of preferences can reduce the provision of public goods in a community. Alesina and la Ferrara (2000) show that social capital is lower in more heterogeneous communities.

⁴⁰Our formulation of ΔE and $I(g)$ is somewhat different from that of RM. In RM's model, agents are partitioned into two groups, although in this paper the notion of group is not explicitly modeled.

distribution of agents' types with respect to axis 1, i.e. $\mu = E[x_{i1}]$.

Agent i 's "ex post political type" in network g , $\tilde{x}_{i1}(g)$, is assumed to be composed of his original ("ex ante") political type and his neighbors' ex ante political types:

$$\tilde{x}_{i1}(g) = \frac{x_{i1} + \beta \cdot \sum_{j \in N_i(g)} x_{j1}}{1 + \beta \cdot q_i(g)},$$

where $\beta > 0$ represents the degree of social influence. Assume that ΔE is equal to the almost sure limit of the mean distance between randomly chosen two agents' ex post political types as n goes to infinity and then \bar{d} goes to zero.⁴¹ That is,

$$\Delta E = \lim_{\bar{d} \rightarrow 0} \lim_{n \rightarrow a.s. \infty} \frac{1}{n} \sum_{i \in N} |\tilde{x}_{i1}(g) - \mu|.$$

In Appendix A.10, we show that ΔE is decreasing in m and increasing in k . The intuition behind this result is that an agent has more chances to form links with agents who are dissimilar in the political type if the type space is rich (if m is large), and/or if the agents do not care about many aspects of the others (if k is small). Under such circumstances, the diversity of agents' ex post political types is small and hence the network incurs small cost generated by the difference in agents' preferences.

Informal institutions: $w(g)$ represents the benefit of informal institutions. As classical works such as Coleman (1988) observes, a community with high network closure, i.e. Cl , can support cooperative behaviors well. That is, in such a network neighbors can monitor each other and avoid "free riders" by "community enforcement."⁴² From another perspective, Chwe (2000) shows that a network with high Cl facilitates collective actions.⁴³ Hence, we assume that $w(g)$ is decreasing in Cl .

The above decomposition of the communication externality enable us to obtain insights into social welfare. For example, suppose that the type space becomes rich, i.e. m is increased with k being constant, because of advance in communication technology or because of the creation of a new interest group. We can predict three possible external effects on the welfare: First, since this implies smaller APL , we predict a positive effect from transmissions of ideas through the network (higher $T(g)$). Second, a larger m makes neighbors' political preferences less diverse. Hence, the cost of producing public goods and

⁴¹This definition is analogous to the one for ΔE in RM.

⁴²Balmaceda and Escobar (2010) formalize this logic and show that cooperation can be sustained in a network with high clustering.

⁴³In his model, a network with high Cl tends to generate common knowledge among agents so that they can participate in risky collective actions. For more comprehensive discussion on this subject, see Chwe (2001).

of accumulating social capital becomes small (smaller $v(g)$). Third, Corollary 2 shows that a larger m results in a smaller Cl . This makes collective actions of the community harder, so the benefit from informal institutions becomes smaller (smaller $w(g)$).

Hence, the effect of changes in m on the communication externality is not straightforward: It depends on the relative importance of each component, that is, the benefit of transmitted ideas, the cost generated by the difference in preferences, and the benefit of informal institutions. Analogously, it is easy to see that the effect of changes in k is also ambiguous.⁴⁴

5.4 Nonlinear Cost Functions

5.4.1 Nonlinear Cost Functions and Strong Stability

Generally, pairwise stability does not determine a unique network structure. Moreover, a pairwise stable network is not necessarily generated by a cutoff rule. For example, consider the networks depicted in Figure 1.

First, consider the composition of nodes in Figure 1(a): There are four nodes, 1, 2, 3, and 4 located in the type space $X = [0, 1]^2$, with $x_1 = (0.9, 0.1)$, $x_2 = (0.8, 0.95)$, $x_3 = (0.1, 0.25)$, and $x_4 = (0.15, 0.8)$. We consider the case with $k = m = 2$. Calculating the distances, we get $d(1, 2) = 0.85$, $d(1, 3) = 0.8$, $d(1, 4) = 0.75$, $d(2, 3) = 0.7$, $d(2, 4) = 0.65$, and $d(3, 4) = 0.55$. Suppose that $b(d) = \frac{1}{d}$, $c(0) = 0$, $c(1) = 2$, $c(2) = 2.2$, and $c(3) = 2.3$. Notice that the cost function c is concave. In this case, there are three types of pairwise stable network structures, depicted in (a-1), (a-2), and (a-3), respectively. The network in (a-1) is pairwise stable because the cost to form the first link, i.e. $\Delta c(0)$, is so high that no one wants to form a link. The network in (a-2) is pairwise stable because, again, the cost for the node 4 to form the first link is very high that he does not want to form a link even though each of the other three nodes have incentive to form a link with him. There are three other networks of this type, in each of which one agent has degree 0 and other three agents have degree 2. The network in (a-3) is also pairwise stable because the fact that the marginal cost of forming a third link, $\Delta c(2)$ is very low implies that the marginal benefit of deleting a third link is negative.

Next, consider the composition of nodes in Figure 1(b): There are four nodes, 1, 2, 3, and 4 located in the type space $X = [0, 1]^2$, with $x_1 = (0.8, 0.2)$,

⁴⁴Depending on the context, our results on the effects of changes in parameters such as m and k (and even the cost function discussed in Subsection 5.4 in depth) would be suggestive for social planners or “network designers” who, for example, manage social networking services.

$x_2 = (0.75, 0.95)$, $x_3 = (0.4, 0.1)$, and $x_4 = (0.25, 0.8)$. Again, we consider the case with $k = m = 2$. Suppose that $b(d) = \frac{1}{d}$, $c(0) = 0$, $c(1) = 1$, $c(2) = 10$, and $c(3) = 30$. Notice that c is convex. Distances between nodes are $d(1, 2) = 0.75$, $d(1, 3) = 0.4$, $d(1, 4) = 0.6$, $d(2, 3) = 0.85$, $d(2, 4) = 0.5$, and $d(3, 4) = 0.7$. In this case, there are at least two pairwise stable networks, depicted in (b-1) and (b-2), respectively.⁴⁵ Both networks in (b-1) and in (b-2) are pairwise stable because the marginal cost for these nodes to have a second link is very high. But the network in Figure 1(b-2) is not generated by a cutoff rule. For, if it did, the cutoff value of node 1 has to be no less than 0.75 because it is connected to node 2 and $d(1, 2) = 0.75$. The cutoff value of node 3 has to be also no less than 0.7 because it is connected to node 4 and $d(3, 4) = 0.7$. But then, $d(1, 3) = 0.4 < 0.7$ implies that it has to be the case that the link 13 is formed; a contradiction.

Although we have multiplicity of pairwise stable networks in both concave and convex cost functions, the reasons for the multiplicity are quite different. Precisely, in the case of convex cost functions, it is impossible that two networks $g, g' \in G(N)$ are both pairwise stable *and* $g \subsetneq g'$, while it is possible in the case of concave cost functions, as shown in the example in Figure 1(a).

Although multiple pairwise stable networks are possible, a refinement of the concept of pairwise stability, *strong stability* (Jackson and van den Nouweland, 2005), can predict a smaller set (or even a singleton set under certain circumstances) of “stable” networks. By using this stronger notion of stability, we can show that the resulting network, which turns out to exist, can be described by the cutoff rule model analyzed in Section 4.

Before defining strong stability, we need one more definition: We say a network g' is *obtainable from g via deviations by $S \subseteq N$* if

$$\begin{aligned} (ij \in g' \wedge ij \notin g) &\implies i, j \in S \text{ and} \\ (ij \in g \wedge ij \notin g') &\implies \{i, j\} \cap S \neq \emptyset. \end{aligned}$$

That is, g' is obtainable from g via deviations by S if each newly formed link in g' involves the agents only from S , and each deleted link in g' involves at least one agent from S .

Definition 6. A network g is **strongly stable** if for any $S \subseteq N$ and g' that is obtainable from g via deviations by S , $(\exists i \in S \text{ s.t. } u_i(g') > u_i(g))$ implies $(\exists j \in S \text{ s.t. } u_j(g') < u_j(g))$.

This concept requires that a stable network be robust to deviations by any coalitions. For example, the networks in Figure 1(a-3) and Figure 1(b-1) are strongly stable, but other networks in Figure 1 are not. Note that if g is

⁴⁵A network with links $\{14, 23\}$ is also pairwise stable.

strongly stable, then it is also pairwise stable, as the coalition S can be any pair of agents.

The next proposition states that we are assured to have a pairwise stable network that is generated by a cutoff rule, and moreover, that when the cost function is linear or convex, the concept of strong stability selects a unique network, and again it is generated by a cutoff rule.

Proposition 7. *Suppose that the cost function c is linear, concave, or convex. Then, almost surely, there exists a pairwise stable network that is generated by a cutoff rule. Furthermore, if c is linear or convex, there exists a unique strongly stable network, and it is generated by a cutoff rule.*

Proof Idea. We propose an algorithm in which agents make offers to form links with others at each step. The algorithm stops in a finite number of steps, and generates a strongly stable network. A cutoff value profile is given by the one in which each agent's cutoff is the maximum distance among his links with others in the generated network. See Appendix A.7 for details.⁴⁶ \square

A pairwise stable network is not necessarily generated by a cutoff rule if it is not strongly stable. In the example in Figure 1, for instance, the network in (b-2) is pairwise stable, but is not (uniquely) strongly stable. So the fact that it is not generated by a cutoff rule is still consistent with the result in Proposition 7. But it is always the case that there exists a pairwise stable network that is generated by a cutoff rule. Moreover, using the notion of strong stability, we can select a smaller set (or even a singleton set under certain circumstances) of networks in which players form links as if they are using some cutoff values. Note that, as opposed to the case of linear cost functions, the cutoff value profile, if any, in a pairwise stable network under nonlinear cost function is not necessarily homogeneous. An example

⁴⁶Although we were unable to prove the existence of strongly stable network for concave cost functions, there is a sense in which the result seems to hold. For example, in Figure 1(a), the strongly stable network exists, which is simply a Pareto efficient network among the set of pairwise stable networks (Figure 1(a-3)). One difficulty associated with concave cost functions is that, as opposed to the cases of convex or linear cost functions, that an agent does not have an incentive to sever *one* link from a current network does not necessarily imply that she does not want to sever *multiple* links from the network. We leave further investigation of the case of concave cost functions for future research. Also, notice that we do not provide a result on efficiency. This is because efficiency does not hold in general. For example, consider an example in Figure 1(b-1), but suppose that there are only node 1-3, and consider the same k , m , b and c as in the main text, except that now we assume $c(2) = 2.5$. Then, it is straightforward that $g = \{13\}$ is the unique strongly stable network. But this is not efficient: If link 12 were to be added, node 1 obtains $1/0.75 - 1.5 = -1/6$, while node 2 obtains $1/0.75 - 1 = 1/3$, hence the net effect on the total surplus is positive ($-1/6 + 1/3 = 1/6 > 0$).

is the network in Figure 1(a-2), where agents 1-3 and agent 4 cannot have a homogeneous cutoff value profile. Note that this network is not strongly stable, as the network in Figure 1(a-3) is obtainable from the network in Figure 1(a-2) via deviations by $S = \{1, 2, 3, 4\}$ and that all the agents would be better off after such deviations.

A homogeneous cutoff value profile may not exist even in a strongly stable networks: Consider the composition of nodes in Figure 2: There are four nodes, 1, 2, 3, and 4 located in the type space $X = [0, 1]^2$, with $x_1 = (0.7, 0.1)$, $x_2 = (0.8, 0.95)$, $x_3 = (0.2, 0.05)$, and $x_4 = (0.15, 0.4)$. We consider the case with $k = m = 2$. Suppose that $b(d) = \frac{1}{d}$, $c(0) = 0$, $c(1) = 1$, $c(2) = 5$, and $c(3) = 10$. Distances between nodes are $d(1, 2) = 0.85$, $d(1, 3) = 0.5$, $d(1, 4) = 0.55$, $d(2, 3) = 0.9$, $d(2, 4) = 0.65$, and $d(3, 4) = 0.35$. It is straightforward to see that there is a unique pairwise stable network, namely $g = \{12, 34\}$, as in the figure. This is also strongly stable.

Now, because node 1 is connected with node 2, his cutoff value, if any, has to be no less than 0.85. But because node 3 is not connected with node 1, his cutoff value, if any, has to be strictly less than 0.5. This implies that we can not find any homogeneous cutoff value profile. Hence, this example shows that even in a strongly stable network, a homogeneous cutoff value profile may not exist. On the other hand, as Proposition 7 shows, a heterogeneous cutoff value profile must exist. For example, $(\hat{d}_1, \hat{d}_2, \hat{d}_3, \hat{d}_4) = (0.85, 0.85, 0.35, 0.35)$ serves as a heterogeneous cutoff value profile.

The next subsection examines how heterogeneous a cutoff value profile can be, when the number of nodes is very large.

5.4.2 Heterogeneous Cutoff Value Profile

The following proposition shows that the heterogeneity of a cutoff value profile is small when the marginal cost approaches some constant value as the number of agents goes to infinity.

Proposition 8. *Suppose that b is continuous and strictly decreasing, and that for some $c_1 > 0$, $\lim_{d \rightarrow 0} b(d) > c_1$ and $\lim_{q \rightarrow \infty} \Delta c(q) = c_1 > 0$ hold. Then, the cutoff value profile for a strongly stable network, $(\hat{d}_1, \dots, \hat{d}_n)$, is such that*

$$\min_{i \in N} \hat{d}_i \xrightarrow{n \rightarrow \infty} \hat{d} \quad \text{almost surely} \quad \text{and} \quad \max_{i \in N} \hat{d}_i \xrightarrow{n \rightarrow \infty} \hat{d} \quad \text{almost surely,}$$

where $b^{-1}(c_1) = \hat{d} > 0$.

Proof Idea. For each agent, for sufficiently large number of nodes, there are sufficiently many neighbors in his δ -neighborhood. The agent has to be

connected with them in a strongly stable network, for otherwise the network would not be pairwise stable, so it would not be strongly stable, either. This implies that he has a sufficiently large degree, and hence the cost function is almost linear when he decides whether or not to connect with agents outside the δ -neighborhood. Hence, he can be described as if he were using a cutoff that is only slightly different from some fixed cutoff. See Appendix A.8 for details. \square

Next, we analyze networks under a heterogeneous cutoff value profile. We assume that each agent has his own cutoff value, \hat{d}_i , and it is distributed in the interval $[\hat{d} - \epsilon, \hat{d} + \epsilon]$ for some $\epsilon > 0$, according to some (possibly unknown and/or correlated) distribution. That is, agents are using heterogeneous cutoff values, which deviate from \hat{d} by at most ϵ . Define

$$Cl_{hetero}^* = \lim_{\hat{d} \rightarrow 0+} \lim_{\epsilon \rightarrow 0+} Cl^{\hat{d}, \epsilon} \quad \text{where} \quad Cl(g) \xrightarrow{n \rightarrow \infty} Cl^{\hat{d}, \epsilon} \quad \text{almost surely.}$$

$$APL_{hetero}^* = \lim_{\hat{d} \rightarrow 0+} \lim_{\epsilon \rightarrow 0+} APL^{\hat{d}, \epsilon} \quad \text{where} \quad APL(g) \xrightarrow{n \rightarrow \infty} APL^{\hat{d}, \epsilon} \quad \text{almost surely.}$$

Note that the order of the limits implies that we consider the situation where the heterogeneity of the cutoff values is almost negligible relative to the sizes of the cutoff values themselves. Note also that there exists a sequence of (b, c) pair that satisfies the assumptions in Proposition 8 such that the corresponding \hat{d} converges to zero, due to the analogous argument as in Part 2 of Proposition 1, hence the requirement of $\hat{d} \rightarrow 0$ above is not vacuous. The next propositions state that the limit values of the clustering coefficient and the average path length with heterogeneous cutoff values are the same as in the case of homogeneous cutoff values.

Proposition 9. $Cl_{hetero}^* = Cl^*$.

Proof Idea. Given \hat{d} and ϵ , by slightly modifying the calculation in the proof of Proposition 1, we get an upper bound and a lower bound of Cl_{hetero}^* . One can show, for any \hat{d} , that these modified formula approaches Cl^* as ϵ goes to zero. See Appendix A.9 for details. \square

Proposition 10. $APL_{hetero}^* = APL^*$.

The proof strategy runs parallel to that of Proposition 9, hence is omitted.

Summing up, our main results are almost unchanged under the condition that heterogeneity of the cutoff values is almost negligible relative to the cutoff values themselves. Combined with Proposition 8, our results in Section 4 carry over even in the case of nonlinear cost functions provided that they approximate linear functions.

5.5 Extensions of the Cutoff Rule Model to Stochastic Models

In this subsection, we present models that include stochastic components in the cutoff rule model. The results in this subsection apply to the cases of the k 'th norms with any k .

We propose two ways to include stochastic components. We obtain similar results for both cases, but the two models highlight different aspects of possible “randomness” in agents’ relationships.

Stochastic Model 1 (S1): Suppose that each pair of agents i and j uses a cutoff value, $\hat{d} + \epsilon w_{ij}$, where $\epsilon \in [0, 1]$ is a fixed constant and w_{ij} is an independently and identically distributed random term (Thus, this model assumes that the cutoff values are not agent-specific, but pair-specific). Random terms are interpreted as a result of idiosyncratic noises in preferences. We assume that the distribution has a full-support over \mathbb{R}_+ . When $\epsilon > 0$, this model differs from the deterministic cutoff rule model in that we now allow each agent to use various cutoff values. Let the almost sure limit of the average path length of this model as $n \rightarrow \infty$ be APL^{S1} and the almost sure limit of the clustering coefficient of this model as $n \rightarrow \infty$ be Cl^{S1} .

Stochastic Model 2 (S2): Suppose that each agent has the cutoff value of $\hat{d} > 0$ with probability $1 - \epsilon$, and that of D with probability ϵ , where D is a random variable whose distribution has a full-support over $[0, 1]$ and $\epsilon \in [0, 1]$. The interpretation is that fraction ϵ of agents who have different cutoffs are “crazy,” while the remaining fraction $1 - \epsilon$ of agents are “normal.” Assume also that $\Pr(D = 0) = 0$. When $\epsilon > 0$, this model differs from the deterministic cutoff rule model in that agents can differ not only in their types but also in their “sociabilities” (Fujii and Kamada, 2010). For example, an agent with very high D corresponds to a very “social” agent. Let the almost sure limit of the average path length of this model as $n \rightarrow \infty$ be APL^{S2} and the almost sure limit of the clustering coefficient of this model as $n \rightarrow \infty$ be Cl^{S2} .

The first proposition tells us that the prediction about the clustering coefficient is not affected substantially by the introduction of slight stochastic components.

Proposition 11. (i) Cl^{S1} is continuous in ϵ at $\epsilon = 0$; (ii) Cl^{S2} is continuous in ϵ at $\epsilon = 0$.

Proof Idea. In both models, when ϵ is small, most of agents (or pairs) have cutoffs very close to \hat{d} . Also, for each agent i , most of her neighbors consist of the agents with the cutoffs very close to \hat{d} . Since clustering coefficients must take values in $[0, 1]$, we can essentially ignore the agents with cutoffs which

are very different from \hat{d} when we take the limit as $\epsilon \rightarrow 0$. Because the cutoffs of agents who we need to take care of converge to \hat{d} , we can use Proposition 9 to show that the bounds of the clustering coefficient approaches Cl^* from below and above as $\epsilon \rightarrow 0$. \square

The next proposition shows that APL is very small in these stochastic models, irrespective of k , m , and \hat{d} .

Proposition 12. (i) $APL^{S1} \leq 2$; (ii) $APL^{S2} \leq 3$.

Proof Idea. In model S1, any pair of agents i and j can be connected through some agent h who is very close to i and w_{hj} is very high. Such an agent exists for large n almost surely by the strong law of large numbers. Thus the APL in the limit is at most 2. In model S2, any pair of agents i and j can be connected through some pair of agents h and h' , both of whom has very large D , and h is very close to i and h' is very close to j . Again, such a pair of agents exists almost surely for large n by the strong law of large numbers. Thus the APL in the limit is at most 3, unless the APL in the original deterministic model is 2. Since we do not require $\hat{d} \rightarrow 0$ nor $\epsilon \rightarrow 0$, only the upperbounds can be obtained. \square

Proposition 4 tells us that the average path length with the Max norm is very high. However, according to Proposition 12, the existence of stochastic components in the model leads to networks that have surprisingly low APL , regardless of the property of social distance.

Although the models with the k 'th norm with $k < m$ turn out to have so-called ‘‘small world’’ property, i.e. high clustering and low average path length, the model with the Max norm does not. But the model with stochastic components exhibits the small world property, even when we use the Max norm. The reason that we can recover this property is similar to the one in Watts and Strogatz (1998). They construct a model in which with small probability, randomly chosen links in a lattice network are deleted to make other links to some randomly chosen nodes (i.e. the links are ‘‘rewired’’). The resulting networks turn out to have the small world property. By interpreting the network generated by the deterministic part of our model as a lattice network in Watts and Strogatz’s model, the introduction of stochastic components in our model can be seen to have the same spirit as their ‘‘rewiring’’ process.

6 Concluding Remarks

In this paper, we proposed a model that provides an explanation as to why some networks are cliquish (they exhibit high clustering coefficients) and/or closely connected (they have low average path lengths) while others do not. In our model, agents are endowed with their own multi-dimensional characteristics. When agents integrate and evaluate the information about the relationships in different dimensions or groups, we supposed that they measure the “social distance” between themselves and others by using the “ k ’th norm,” in which the distance is the k ’th smallest dimension-wise distances. When k is high, that is, when agents need many similar characteristics in order to be linked with each other, the network is cliquish while it is not closely connected, under certain regulatory conditions. On the other hand, when k is low, that is, when it is enough for agents to share a small number of similar characteristics in order to be linked, the network is not cliquish while it is closely connected. One implication of our result is that the introduction of new communication technology makes a network closely connected but cliquish. We related our model and results to the “strength of weak ties hypothesis” of Granovetter (1973), the “similarity scale” of Tversky (1977), and the “communication externality” of Rosenblat and Mobius (2004). We also showed that the assumption of linear cost function is not essential to our result, by replacing the notion of pairwise stability with that of strong stability. Although the network does not have a small world property in large networks when k is the same as the number of dimensions of the type space, we showed that a stochastic version of the model has the small world property.

Let us suggest possible generalizations of our model. First, the notion of the k ’th norm is tactable and useful to gain economic intuition, but perhaps is too simple a model to fit to data. A generalization of the k ’th norm, for example a weighted average of the k ’th norms over all k ’s, would enable the model to better fit the data. Second, our model supposed that agents’ characteristics are completely determined by their types. One natural way to introduce further heterogeneity is to assume that different agents have different “sociabilities” as in Fujii and Kamada (2010). As mentioned in the Introduction, by introducing social distance in the model of Fujii and Kamada (2010), or by introducing sociability in our model, we could have a more realistic formation model of social networks. Third, some dimensions might not be described appropriately by a continuous variables. If there are such dimensions, we suspect that the average path length becomes smaller, the clustering coefficient becomes larger, and it becomes more difficult to obtain a desired degree distribution, as essentially agents in the “same category”

with respect to some dimension has “distance zero” between them, so even in the limit of letting the cutoff go to zero, these neighbors have nonnegligible effects on clustering and degrees. We leave these possibilities of generalization to future research.

We conclude this paper by explaining how the paper could serve as a basis for future works. First, this paper introduced a model of multi-dimensional type space and various measures of distance that violate the triangle inequality, based on which agents form links. We believe these new ingredients of the model would give us new insights when analyzing situations in which preferences depend on similarity between agents involved. For example, they would be useful in analyzing network formation models, models of matching markets such as marriage or labor markets, voting models, and so forth. Moreover, they would also be useful even in the context of biology literature. For example, Antal et al. (2009) consider an evolutionary model in which individuals cooperate if their opponent is close to oneself in the phenotype space, and show that evolution can favor cooperators. It would be natural to consider a situation where cooperation takes place when some but not all aspects of the individuals’ phenotypes are close to each other. Second, we proved the existence and the uniqueness of a strong stable network under some regulatory conditions. Proving the existence and the uniqueness of a strong stable network is often a hard task, but our result suggests that these are not “impossible results” if we restrict a class of preferences in a tractable manner.

References

- Akerlof, G., Social Distance and Social Decisions, *Econometrica*, **65** pp.1005-1027, 1997
- Alesina, A., Baqir, R., and Easterley, W., Public Goods and Ethnic Divisions, *Quarterly Journal of Economics*, CXIV, pp.1243-1284, 1999
- Alesina, A., and la Ferrara, E., Participation in Heterogeneous Communities, *Quarterly Journal of Economics*, CXV, pp.847-904, 2000
- Antal, T., Ohtsuki, H., Wakeley, J., Taylor, P. and Nowak, M., Evolution of Cooperation by Phenotypic Similarity, *Proceedings of the National Academy of Sciences of the United States of America*, **106**, pp.8597-8600, 2009
- Balmaceda, F. and J. F. Escobar, Cooperation and Trust in Social Networks of Information Transmission, *mimeo*, 2010

- Barabasi, A., *Linked: The New Science of Networks*, Perseus Books, 2003
- Barabasi, A. and Albert R., Emergence of Scaling in Random Networks, *Science*, **286**, pp.509-512, 1999
- Caldarelli, G., Capocci, A., De Los Rios, P. and Munoz, MA., Scale-Free Networks from Varying Vertex Intrinsic Fitness *Physical Review Letters*, **89(25)**: 258702, 2002
- Chwe, M.K., Communication and Coordination in Social Networks, *Review of Economic Studies*, **67**, pp.1-16, 2000
- Chwe, M.K., *Rational Ritual: Culture, Coordination, and Common Knowledge*, Princeton University Press, 2001
- Coleman, JS., Social Capital in the Creation of Human Capital, *American Journal of Sociology*, **94** pp.S95-S120, 1988
- Currarini, S., Jackson, MO., and Pin, P., A Economic Model of Friendship: Homophily, Minorities and Segregation, *Econometrica*, **77**, pp.1003-1045, 2009
- Ebel, H., L-I. Mielsch, S. Bornholdt, Scale-free Topology of E-Mail Networks, *Physical Review E*, **66(3)**, 035103(R), 2002
- Fujii, D. and Kamada, Y., Network Formation Model with Individual Sociability, mimeo, 2010
- Golub, B. and Y. Livne, Strategic Random Networks, mimeo, 2010.
- Goyal, S., *Connections* Princeton University Press, 2005
- Goyal, S., M.J. van der Leij and J. L. Moraga-Gonzalez, Economics: An Emerging Small World, *Journal of Political Economy*, **114(2)**, pp.403-412, 2006
- Granovetter, M., The Strength of Weak Ties, *American Journal of Sociology*, **83(6)** pp.1360-1380, 1973
- Granovetter, M., *Getting a Job* 2nd edition, University of Chicago Press, 1995
- Hoff, P., Raftery, A., and Handcock, M., Latent Space Approaches to Social Network Analysis, *Journal of the American Statistical Association*, **97(460)** 1090-1098, 2002

- Jackson, MO., A survey of network formation models: Stability and efficiency in *Group Formation in Economics: Networks, Clubs and Coalitions*, Cambridge, 2005
- Jackson, M., *Social and Economic Networks* Princeton University Press, 2008a
- Jackson, MO., Average Distance, Diameter, and Clustering in Social Networks with Homophily, *Proceedings of the Workshop in Internet and Network Economics (WINE 2008)*, Springer-Verlag, Berlin Heidelberg, 2008b
- Jackson, MO., and Rogers, B., The Economics of Small Worlds, *Journal of European Economic Association*, **3** pp.617-627, 2005
- Jackson, MO., and van den Nouweland, A., Strongly Stable Networks, *Games and Economic Behavior*, **71**, pp.44-74, 2005
- Jackson, MO., and Wolinsky, A., A Strategic Model of Social and Economic Networks, *Journal of Economic Theory*, **71**, pp.44-74, 1996
- Johnson, C., and Giles, R., Spatial Social Networks, *Review of Economic Design*, **5** pp.273-300, 2000
- Marmaros, D. and Sacerdote, B., How do Friendships Form?, *Quarterly Journal of Economics*, **121(1)**, pp.79-119, 2006
- Pareto, V., *Cours d' Economie Politique*, Droz, 1896
- Rapoport, A., Spread of Information Through a Population with Socio-Structural Bias: II. Various Models with Partial Transitivity, *Bulletin of Mathematical Biophysics*, **15**, pp.535-546, 1953.
- Rapoport, A. and W. J. Horvath, A Study of a Large Sociogram, *Behavioral Science*, **6**, pp.279-291, 1961.
- Rosenblat, T. and Mobius, M., Getting Closer or Drifting Apart, *Quarterly Journal of Economics*, **119(3)**, pp.971-1009, 2004
- Rogers, E.M., *Diffusions of Innovations.*, New York: Free Press
- Selfhout, M. H. M., S. J. T. Branje, T. F. M. ter Bogt, and W. H. J. Meeus, The Role of Music Preferences in Early Adolescents' Friendship Formation and Stability, *Journal of Adolescence*, **32(1)**, pp.95-107, 2009
- Tversky, A., Features of Similarity, *Psychological Review*, **84** pp.327-352, 1977

Watts, D. and Strogatz, S., Collective Dynamics of ‘Small-World’ Networks, *Nature*, **393** pp.440-442, 1998

A Appendix

A.1 Proof of Proposition 1

Proof.

Part 1-1: Existence of a Pairwise Stable Network

Consider the maximum of d 's that satisfies $b(d) - c_1 \geq 0$, and denote it by \hat{d} (The maximum exists because b is nonincreasing and continuous from the left). We have

$$\Delta c(q) = (c_0 + c_1(q + 1)) - (c_0 + c_1q) = c_1 \quad \text{for all } q.$$

g is pairwise stable if and only if (i) there is no link $ij \in g$ such that $u_i(g) < u_i(g - ij)$ and (ii) there is no link $ij \notin g$ such that $u_i(g) \leq u_i(g + ij)$. Now, since $\Delta c(q) = c_1$ for all q , (i) is equivalent to saying that there is no $ij \in g$ such that $0 > b(d(i, j)) - c_1$, and (ii) is equivalent to saying that there is no $ij \notin g$ such that $0 \leq b(d(i, j)) - c_1$. Noting that $b(d(i, j)) - c_1 \geq 0 \iff d(i, j) \leq \hat{d}$, we have that $g = \{ij : d(i, j) \leq \hat{d}\}$ is pairwise stable. Thus, a pairwise stable network exists.

Part 1-2: Uniqueness of the Pairwise Stable Network

Suppose that there are two distinct pairwise stable networks, g and g' . Without loss of generality, there exists a pair of agents $i, j \in N$ such that $ij \in g$ and $ij \notin g'$. But $ij \in g$ and (i) in Part 1 of this proof imply $b(d(i, j)) - c_1 \geq 0$, while $ij \notin g'$ and (ii) in Part 1 of this proof imply $b(d(i, j)) - c_1 < 0$. Contradiction.

Part 1-3: Efficiency of the Pairwise Stable Network

Suppose, to the contrary, that the pairwise stable network g is not efficient. That is, suppose that there is another network g' in which the sum of utilities of all the agents is strictly larger in g' than in g . Let $L_1 = g \setminus g'$ and $L_2 = g' \setminus g$. That is, g' is obtained from g by deleting all the links in L_1 and adding all the links in L_2 . Note that the order of deletion and addition of links doesn't matter for the efficiency from the resulting networks by the definition of efficient networks. Now, for all $ij \in L_1$, we have $b(d(i, j)) - c_1 \geq 0$ from Part 1 of this proof, so the sum of utilities strictly decreases by deletion of links in L_1 unless L_1 consists only of links ij such that $d(i, j) = c_1$. Next,

for all $ij \in L_2$, we have $b(d(i, j)) - c_1 < 0$ from Part 1 of this proof, so the sum of utilities strictly decreases by addition of links in L_2 if L_2 is not empty, and stays constant if it is empty. Hence, the only way that g' be efficient is that L_1 's only elements are the links ij such that $d(i, j) = c_1$, and L_2 is empty. But as deleting the links ij such that $d(i, j) = c_1$ does not change the utility of either i or j and hence it does not change the sum of utilities, g' has the same sum of utilities as g . But this contradicts our starting assumption that g' is such that the sum of utilities of all the agents is strictly larger in g' than in g . This completes the proof.

Part 1-4: Existence of a Homogeneous Cutoff Value Profile

In Parts 1 and 2 of this proof we have shown that the unique pairwise stable network is $g = \{ij : d(i, j) \leq \hat{d}\}$. Let a cutoff value profile be such that $\hat{d}_i = \hat{d}$ for all $i \in N$. This cutoff value profile is homogeneous by definition, and clearly generates network g .

Part 2: Existence of pair (b, c)

Fix a network g that is generated by cutoff rule with a homogeneous cutoff value profile. It suffices to provide one example of (b, c) pair such that g is pairwise stable with respect to the pair (b, c) . Uniqueness and efficiency follows directly from Parts 1-2 and 1-3, respectively.

Let the homogeneous cutoff value be \hat{d} . Consider a pair of functions $b(d) = a \cdot \frac{\hat{d}}{d}$ and $c(q) = a \cdot q$ for some $a > 0$. These functions satisfy the assumptions made in Section 3.1. Notice that the benefit from forming links when the distance is very short decreases fast if a is small, and the marginal cost of forming an additional link is large if a is large.

Now, notice that $ij \in g$ implies $d(i, j) \leq \hat{d}$, which implies $b(d(i, j)) - a = a \cdot \frac{\hat{d}}{d(i, j)} - a \geq 0$, which in turn implies that the marginal benefit for each of agents i and j from link ij is no less than the marginal cost. Also, $ij \notin g$ implies $d(i, j) > \hat{d}$, which implies $b(d(i, j)) - a = a \cdot \frac{\hat{d}}{d(i, j)} - a < 0$, which in turn implies that the marginal benefit for each of agents i and j from link ij is strictly less than the marginal cost. Hence g is pairwise stable. Thus the proof is complete. \square

A.2 Proof of Proposition 2 and its Corollaries

Proof of Proposition 2.

Let the set of points sufficiently away from the boundary be $X(\hat{d}) = \{x_i \in$

$X : 0 < x_{ih} \pm \hat{d} < 1, 0 \leq h \leq m\}$. We have:

$$Cl^* = \lim_{\hat{d} \rightarrow 0} \lim_{n \rightarrow \infty}^{\text{a.s.}} \left[\frac{1}{n} \left(\sum_{x_i \in X(\hat{d})} Cl_i(g) + \sum_{x_j \in X \setminus X(\hat{d})} Cl_j(g) \right) \right] = \lim_{\hat{d} \rightarrow 0} \lim_{n \rightarrow \infty}^{\text{a.s.}} \left[\frac{1}{n} \left(\sum_{i \in X(\hat{d})} Cl_i(g) \right) \right],$$

where $\lim^{\text{a.s.}}$ denotes the almost sure limit, since the volume of $\frac{vol(X(\hat{d}))}{vol(X)} \rightarrow 1$ as $\hat{d} \rightarrow 0$ where $vol(\cdot)$ denotes the volume of a set, and $Cl_j(g)$ takes only a finite value (a value in $[0, 1]$) for any $j \in N$.

Fix k . Take a point x in $X(\hat{d})$ and consider a hypothetical agent situated at the point, named agent i .

We will ignore the possibility of the tie in distances, as it does not occur almost surely, hence does not affect the result.

Now, let $B_{\hat{d}}(x)$ be the \hat{d} -neighborhood of point x .

Consider a randomly chosen $y \in B_{\hat{d}}(x)$ according to the uniform distribution over $B_{\hat{d}}(x)$. Consider a hypothetical agent situated at y and call him j . It is easy to see that $\lim_{\hat{d} \rightarrow 0} Pr(\#\{h | x_{ih} - y_{jh} \leq \hat{d}\} = k) = 1$. So for our result, we consider only the case of $\#\{h | x_{ih} - y_{jh} \leq \hat{d}\} = k$.

Let $Z(x, y) = \{z \in B_{\hat{d}}(x) | \{h | x_{ih} - z_{jh} \leq \hat{d}\} = \{h | x_{ih} - y_{jh} \leq \hat{d}\}\}$. Notice that $\frac{vol(Z(x, y))}{vol(B_{\hat{d}}(x))} \rightarrow \binom{m}{k}^{-1}$ as $\hat{d} \rightarrow 0$.

Now, it is straightforward to see that the probability that $z \in B_{\hat{d}}(x) \setminus Z(x, y)$ is connected to y goes to 0 as \hat{d} goes to 0. Thus we only need to consider z 's in $Z(x, y)$.

Given that y and z are connected, the probability that $\{h | y_{ih} - z_{jh} \leq \hat{d}\} = \{h | x_{ih} - y_{jh} \leq \hat{d}\}$ goes to 1 as $\hat{d} \rightarrow 0$. Hence, the probability that z and y are connected is equal to the probability that the projections of z and y on the restricted space with dimensions in $\{h | x_{ih} - y_{jh} \leq \hat{d}\}$ are within the distance \hat{d} with respect to the Max norm (i.e. the k 'th norm).

This probability is simply:

$$\frac{1}{(\hat{d})^k} \int_0^{\hat{d}} \int_0^{\hat{d}} \cdots \int_0^{\hat{d}} \frac{(2\hat{d} - y_1)(2\hat{d} - y_2) \cdots (2\hat{d} - y_k)}{(2\hat{d})^k} dy_1 dy_2 \cdots dy_k = \left(\frac{3}{4}\right)^k.$$

Hence the desired probability is $\binom{m}{k}^{-1} \cdot \left(\frac{3}{4}\right)^k$, by the strong law of large numbers. This completes the proof. \square

Proof of Corollary 1.

The formula in Proposition 2 implies $Cl^*(m, m) = \left(\frac{3}{4}\right)^m$ and $Cl^*(1, m) = \frac{3}{4m}$. It is straightforward to see that $\left(\frac{3}{4}\right)^m$ is strictly larger than $\frac{3}{4m}$ if and only if $m < 9$, completing the proof. \square

Proof of Corollary 2.

Part 1 is straightforward from the formula in Proposition 2.

We consider Part 2. From the formula in Proposition 2,

$$\begin{aligned} Cl^*(k+1, m) &= \binom{m}{k+1}^{-1} \left(\frac{3}{4}\right)^{k+1} \\ &= \frac{(k+1)!(m-k-1)!}{m!} \left(\frac{3}{4}\right)^{k+1} \\ &= Cl^*(k, m) \frac{3(k+1)}{4(m-k)}. \end{aligned}$$

Taking logs, we get

$$\log(Cl^*(k+1, m)) - \log(Cl^*(k, m)) = \log\left(\frac{3(k+1)}{4(m-k)}\right).$$

Hence, $Cl^*(k+1, m) \geq Cl^*(k, m)$ is equivalent to $\frac{3(k+1)}{4(m-k)} \geq 1$, or $k \geq \frac{4}{7}m - \frac{3}{7}$, completing the proof. \square

A.3 Proof of Proposition 3

Proof.

Take two points in X , x and y . Almost surely, $d^{min}(x, y) > 0$. Hence we restrict attention to the case of $d^{min}(x, y) > 0$. Fix $\hat{d} > 0$ at a value such that $\hat{d} < \frac{1}{2}d^{min}(x, y)$.

Consider a class of sets such that

$$\beta(t) = \{z \in X : |y_l - z_l| < \frac{\hat{d}}{2} \text{ if } l \leq t(m-k), \quad |x_l - z_l| < \frac{\hat{d}}{2} \text{ otherwise}\}$$

for positive integers $t < \frac{m}{m-k}$. Let T be the largest t that satisfies $t < \frac{m}{m-k}$. Also, let $\beta\{0\} = x$. Then by definition, we have $|w_t - w_{t+1}| < \hat{d}$ for all $w_t \in \beta(t)$ and $w_{t+1} \in \beta(t)$ for all $t = 0, \dots, T-1$.

Now, as n goes to infinity, almost surely there is at least one agent in $\beta(t)$ for any t . Thus, almost surely, there exists a path between x and y whose length is no more than $\lfloor \frac{m}{m-k} \rfloor + 1$ if $\frac{m}{m-k}$ is not an integer and $\frac{m}{m-k}$ if it is.

Finally, we show that the path length cannot be less than this value, almost surely. To see this, suppose, to the contrary, that there exists a path with length less than the value above that connects x and y . But such a path has to have a link ww' on it such that $\sharp\{h|w_h - w'_h \leq \hat{d}\} > k$ (the subscripts h denote the index for dimension), so $d(w, w')^{(k)} > \hat{d}$. Contradiction.

Hence, we have that APL^* is exactly $\lfloor \frac{m}{m-k} \rfloor + 1$ if $\frac{m}{m-k}$ is not an integer and $\frac{m}{m-k}$ if it is. \square

A.4 Proof of Proposition 4

Proof. Take any pair of points in X , x and y . Consider a pair of hypothetical nodes, i and j , situated at x and y , respectively. Almost surely, there exists a dimension h such that $|x_{ih} - y_{jh}| > 0$. Write this value as $a > 0$. Then, with cutoff $\hat{d} > 0$, the path length between i and j is bounded below by a/\hat{d} . As $\hat{d} > 0$ goes to zero, this bound goes to infinity. Since this argument holds for all the pairs x and y with $x \neq y$, the proof is completed. \square

A.5 Proof of Proposition 5

Proof.

Fix k . Take any distribution of nodes, f , which is strictly positive and absolutely continuous over X . Consider a point in X and a hypothetical agent i who is situated at that point. Denote his position by x_i . Note that $f(x_i)$ is strictly positive.

Consider a distribution generated by f over the restricted space $B_\delta(x_i) = \{x \in X : d^{(k)}(x_i, x) < \delta\}$. Denote by h the probability density function of this distribution. By definition, we have:

$$h(x) = \frac{f(x)}{\int_{d^{(k)}(x_i, x') < \delta} f(x') dx'}.$$

We will show that h becomes arbitrarily close to the uniform distribution as δ goes to zero.

The continuity of f implies that for all small $\epsilon > 0$, there exists $\delta' > 0$ such that for all x such that $d^{(k)}(x_i, x) < \delta'$, it must be the case that $|f(x_i) - f(x)| < \epsilon$ holds. Hence, we have, for any ϵ and small enough δ ,

$$\frac{f(x_i) - \epsilon}{\int_{d^{(k)}(x_i, x) < \delta} (f(x_i) + \epsilon) dx} \leq h(x) \leq \frac{f(x_i) + \epsilon}{\int_{d^{(k)}(x_i, x) < \delta} (f(x_i) - \epsilon) dx}$$

But the strict positiveness of f implies that both bounds approach the same limit, which proves the claim.

Now, note that Cl^* is continuous in the distribution of nodes. Also, note that the proofs about APL^* do not rely on any specific assumption about f , as long as it is strictly positive over X . Combining this claim and the above claim, we know that we can approximate a clustering coefficient and an average path length with a general distribution by the clustering coefficient and the average path length with the uniform distribution. \square

A.6 Proof of Proposition 6

Proof.

Fix m and k . Let a distribution f over X be a product measure with marginals being uniform over dimensions $2, \dots, m$, and g over the first dimension. We will show that for any relative degree distributions that satisfy our condition on the minimum relative degree, there exists g that generates it.

Since the strong law of large numbers implies that the relative degree at $y \in X$ converges almost surely to

$$\frac{\int_{x \in B_{\hat{d}}(y)} f(x) dx}{\max_{y' \in X} \int_{x \in B_{\hat{d}}(y')} f(x) dx}$$

where $B_{\delta}(x)$ denotes the δ -neighborhood of x , the relative degree in the almost sure limit at y is proportional to:

$$\binom{m-1}{k-1} \cdot g(y) + \left[\binom{m}{k} - \binom{m-1}{k-1} \right] \cdot 1.$$

First we consider the case of $k < m$. Let the first term be $a \cdot g(y)$ and the second term be b . Let the cumulative distribution function of the desired relative degree distribution be K and its “inverse” be $L(y) = \sup x$ s.t. $y \geq K(x)$. Then what we have to show is that there exists a positive constant c such that $L(y_1) = c(a \cdot g(y) + b)$ for all y . Note that this is well-defined, as $g(y)$ depends only on the value of y_1 . Note that since the desired relative degree distribution is bounded away from zero (the minimum relative degree is bounded away from zero), we can find small enough c such that $L(y_1) - cb > 0$ for all y . Take such c . Then, we have:

$$g(y) = \frac{L(y_1) - cb}{ac}.$$

This generates g for any L , so for any K . However, for this g to be well-defined, it is necessary (and sufficient) that the values of g integrate to 1,

so

$$\int_{y \in X} g(y) dy = 1, \quad \text{or} \quad \frac{1}{ac} \left(\int_0^1 L(y_1) dy_1 - b \right) = 1.$$

Notice that the left hand side of the above equation can be unboundedly large by taking $c > 0$ small enough. The infimum of this value can be obtained by setting $L(0) = cb$, hence we need that

$$\frac{1}{a \cdot \frac{L(0)}{b}} \left(\int_0^1 L(y_1) dy_1 - \frac{L(0)}{b} \cdot b \right) < 1 \quad \text{or} \quad \frac{b}{a} \left(\frac{\int_0^1 L(y_1) dy_1}{L(0)} - 1 \right) < 1.$$

Now, a simple algebraic manipulation shows that $\frac{b}{a} = \frac{m-k}{k}$. Also, it is straightforward to see that $\int_0^1 L(y_1) \geq 1$, as K is a cumulative distribution function for relative degrees. Combining, it is sufficient to have that

$$\frac{m-k}{k} \left(\frac{1}{L(0)} - 1 \right) < 1, \quad \text{or} \quad \frac{m-k}{m} < L(0).$$

This completes the proof. \square

A.7 Proof of Proposition 7

Proof.

Fix the types of agents, (x_1, \dots, x_n) . We ignore the possibility that there exist $h, i, j \in N$ such that $d(i, j) = d(i, h)$, or that there exist $i, j \in N$ and $q \in \mathbb{N}$ such that $b(d(i, j)) = \Delta c(q - 1)$, because almost surely such events do not occur. This in particular implies that $N_i(g) \neq N_i(g') \Rightarrow u_i(g) \neq u_i(g')$. We consider the following algorithm that generates a unique network. We will show in the sequel that the algorithm stops in finite steps, the generated network is pairwise stable, and is generated by a cutoff rule. Moreover, we will show that the generated network is strongly stable if the cost function is concave or linear.

Algorithm

step 1

Each player $i \in N(1) := N$ proposes a “request”:

$$r_i(1) = \arg \max_{r'_{i1} \subseteq N(1) \setminus \{i\}} u_i(\{ij | j \in r'_{i1}\}).$$

Generate a network $g' := g(0) \cup \{ij | j \in r_i(1) \text{ and } i \in r_j(1)\} \in G(N)$ where we set $g(0) = \emptyset$. Delete $k'l' = \arg \max_{i \in N, kl \in g'} \{u_i(g' - kl) - u_i(g')\}$ if $u_i(g' - k'l') - u_i(g')$ is positive. Let $g'' = g' \setminus \{k'l'\}$. Then, delete $k''l'' = \arg \max_{i \in N, kl \in g''} \{u_i(g'' - kl) - u_i(g'')\}$ if $u_i(g'' - k''l'') - u_i(g'')$ is positive. Continue

this procedure until the generated network \hat{g} satisfies the property that each link ij satisfies $u_i(\hat{g} - ij) < u_i(\hat{g})$. Let the resulting network be $g(1)$.

step t

Each player $i \in N(t) := N(t-1) \setminus \{j : r_j(t-1) = \emptyset\}$ proposes a “request”:

$$r_i(t) = \arg \max_{r'_{it} \subseteq N(t) \setminus [\{i\} \cup N_i(g(t-1))]} u_i(\{ij | j \in r'_{it}\} \cup g(t-1)).$$

Generate a network $g' := g(t-1) \cup \{ij | j \in r_i(t) \text{ and } i \in r_j(t)\}$. Delete $k'l' = \arg \max_{i \in N, kl \in g'} \{u_i(g' - kl) - u_i(g')\}$ if $u_i(g' - k'l') - u_i(g')$ is positive. Let $g'' = g' \setminus \{k'l'\}$. Then, delete $k''l'' = \arg \max_{i \in N, kl \in g''} \{u_i(g'' - kl) - u_i(g'')\}$ if $u_i(g'' - k''l'') - u_i(g'')$ is positive. Continue this procedure until the generated network \hat{g} satisfies the property that each link $ij \in \hat{g}$ satisfies $u_i(\hat{g} - ij) < u_i(\hat{g})$. Let $g(t)$ be the resulting network.

Let \bar{t} be the first period, if any, such that $N(\bar{t}) = \emptyset$. If such a period does not exist, then denote $\bar{t} = \infty$.

Let us give an intuitive explanation about the algorithm. For each step t , $N(t)$ is the set of “remaining agents.” Each remaining agent makes a request to form links to some of the remaining agents, which would make him better off than the current network if it was accepted by all agents included in it. However, at each step, all the requests are not necessarily satisfied. Instead, we require that only links that are requested by both agents involved are actually formed. Hence, it is possible that some portion of a request is satisfied while the other portion is not satisfied. In such cases, it may be that, after the formation of links based on the requests, some agents have incentives to delete links that currently exist. Such links are deleted in the “deletion procedure” in each step of the algorithm. Step by step, links are gradually formed, and eventually some agents have empty requests. Such agents are removed from the algorithm, and never be made a request to, nor be able to make a request by himself. Eventually, at some step, no agent “remains” in the algorithm, and the algorithm “stops” at such a step.

We prove the following Lemmas to complete the proof of Proposition 7.

Lemma 1. *For every $t \leq \bar{t}$, if $i \in N(t)$, $k \in r_i(t)$, and $l \in N(t) \setminus r_i(t)$, then $d(i, k) < d(i, l)$.*

That is, i 's request $r_i(t)$ is a set of agents who are closer to i than anyone who is in $N(t)$ but is not included in the request.

Lemma 2. *$\bar{t} < \infty$, and $g(\bar{t})$ is unique.*

Hence, the algorithm “stops” in a finite steps, generating a unique network.

Lemma 3. $g(\bar{t})$ is pairwise stable.

Lemma 4. There exists $\hat{d} = (\hat{d}_1, \dots, \hat{d}_n)$ such that $g(\bar{t})$ is generated by a cutoff rule with \hat{d} .

To establish Lemma 4, we first prove the following claim.

Claim 1. Suppose c is convex. If $j \in r_i(t)$, then $\forall t' > t$ such that $i, j \in N(t')$, either $j \in r_i(t')$ or $ij \in g(t' - 1)$ holds.

Claim 1 implies the following.

Claim 2. Let $g = g(\bar{t})$ and suppose $ij \notin g$, and $d(i, j) < \max_{k \in N_i(g)} \{d(i, k)\}$. Then, $u_j(g + ij) < u_j(g)$ holds.

Claim 2 implies Claim 3, which in turn implies Lemma 4.

Claim 3. Let $g = g(\bar{t})$ and suppose $ij \notin g$, and $d(i, j) < \max_{k \in N_i(g)} \{d(i, k)\}$. Then, $d(i, j) > \max_{l \in N_j(g)} \{d(j, l)\}$ holds.

Lemma 5. Suppose c is linear of convex. Then, $g(\bar{t})$ is strongly stable.

Lemma 6. Suppose c is linear of convex. Then, a strongly stable network is unique.

Proof of Lemma 1

Note that $r_i(t)$ maximizes the sum of additional benefits that i obtains minus that of additional costs that he incurs. Separability of u and the definition of r_i imply

$$r_i(t) = \arg \max_{r'_{it} \subseteq N(t) \setminus \{i\} \cup N_i(g(t-1))} \left[\sum_{j \in r'_{it}} b(d(i, j)) - \sum_{s=0}^{\#r'_{it}-1} \Delta c(q_i(g(t-1)) + s) \right].$$

Notice that the second term of the right hand side of the above equality depends only on i 's degree but not on the identities of agents in r'_{it} .

Suppose, to the contrary, that there exist $i, k, l \in N(t)$ such that $d(i, k) > d(i, l)$, $k \in r_i(t)$, and $l \in N(t) \setminus r_i(t)$. Then, depriving $r_i(t)$ of k and adding l to $r_i(t)$ strictly increases i 's additional benefit (the first term of the right hand side of the above equality) with i 's additional cost (the second term) unchanged. This contradicts the assumption that $r_i(t)$ is the maximizer of the right hand side of the above equality. This completes the proof.

Proof of Lemma 2

Since there is no tie in distances, for each t and each $i \in N$, $r_i(t)$ is uniquely determined. Therefore the algorithm generates a unique network, if it ends in finite steps.

Now we prove that the algorithm ends in finite steps. The algorithm can be regarded as a deterministic dynamic process over discrete time $t = 1, 2, \dots$, defined on state space $G(N) \times 2^N$, where the state at t is $(g(t-1), N(t))$. Note that the number of states is finite.

We first show that this process is monotone. To see this, notice that the set $N(t)$ is nonincreasing. Hence it suffices to show that $g(t-1)$ is nondecreasing. To show this, we will prove that no link in $g(t-1)$ is not deleted in the “deletion procedure” at t (i) with a convex or linear cost function, and (ii) with a concave cost function.

First, consider case (i). We show that there is no agent deleting his links in the algorithm, when c is convex or linear. By the definition of the request, for each t , $i \in N(t)$, and $j \in r_i(t)$, we have

$$\begin{aligned} b(d(i, j)) &> \Delta c(q_i(g(t-1)) + \#r_i(t) - 1) \\ &\geq \Delta c(q_i(g(t-1)) + s), \end{aligned}$$

where $0 \leq s < \#r_i(t)$. This ensures that however i 's requested links are actually formed, he cannot become better off by deleting his newly formed links.

Second, consider case (ii). At step 1, the statement trivially holds, since $g(0) = \emptyset$. We have, by the construction of the algorithm, $\Delta c(q_i(g(t)) - 1) < b(d(i, j))$ for all $ij \in g(t)$. Now consider step $t+1$ and suppose that i becomes better off by deleting links in $g(t)$. Let ij be the first link that is deleted from $g(t)$. It must be the case that $\Delta c(q_i(g(t)) + r - 1) > b(d(i, j))$ for some $0 \leq r \leq \#r_i(t)$. But then we would have $\Delta c(q_i(g(t)) + r - 1) > \Delta c(q_i(g(t)) - 1)$, which contradicts the assumption that Δc is decreasing.

Hence, the process is monotone. Therefore, it suffices to show that there does not exist an event in which the process remains in the same state such that $N(t) \neq \emptyset$. This event could happen only if all the remaining agents make nonempty requests, and any of agents' requests are not fulfilled in the step. That is,

$$\forall i \in N(t), [r_i(t) \neq \emptyset] \text{ and } [\forall k \in r_i(t) \ i \notin r_k(t)].$$

Suppose that this is true at step t .

The simplest case is as follows: $N(t) = \{1, 2, 3\}$, $r_1(t) = \{2\}$, $r_2(t) = \{3\}$, and $r_3(t) = \{1\}$. However, Lemma 1 implies that $d(1, 2) < d(1, 3)$, $d(2, 3) < d(2, 1)$, and $d(3, 1) < d(3, 2)$. Contradiction.

Generally, there must exist a sequence of agents $(1, 2, \dots, n')$ in $N(t)$ (with an appropriate renaming) such that $2 \in r_1(t)$, $3 \in r_2(t)$, \dots , $n' \in r_{n'-1}(t)$,

and $1 \in r_{n'}(t)$, while $1 \notin r_2(t)$, $2 \notin r_3(t)$, \dots , $n' - 1 \notin r_{n'}(t)$, and $n' \notin r_1(t)$. By Lemma 1, we have $d(1, 2) < d(1, n')$, $d(2, 3) < d(2, 1)$, \dots , $d(n' - 1, n') < d(n' - 1, n' - 2)$, and $d(n', 1) < d(n', n' - 1)$. Contradiction. This completes the proof.

Proof of Lemma 3

We need to show that in $g(\bar{t})$, (i) no agent has a strict incentive to delete a link, and (ii) no pair has an incentive to add a link.

To show (i), note that we have constructed the network in the way that there is no link to delete at the final step. Moreover, for agents who have left the algorithm in earlier steps, deleting their links does not increase their payoffs. This is because the set of neighbors of each agent who left earlier remains unchanged after the step at which her request was empty, and (just as in the final step,) there is no link for her to delete at that step.

To see (ii), partition the set of agents, (P_1, \dots, P_T) , so that in each cell P_t of the partition, agents contained in it have empty requests at step t . Consider an agent i in a partition P_t . At step t , there exists no agent j in $\bigcup_{l=t}^T P_l$ such that i would be better off by connecting with j at step t . This is because otherwise j 's request would not be empty at step t . After step t , his degree does not change until the algorithm stops, hence i does not have an incentive to form a link with agents in $\bigcup_{l=t}^T P_l$. Suppose that there exists agent $j' \in P_{l'}$ with $l' < t$ such that i has an incentive to form a link with. However, j' does not have an incentive to form a link with agents in $\bigcup_{l=l'}^T P_l$, in particular with $i \in P_t \subseteq \bigcup_{l=l'}^T P_l$. Hence, no agent has an incentive to form a link in the resulting network.

Proof of Claim 1

It suffices to show the statement in the case of $t' = t + 1$. To see this, first suppose that $ij \in g(t + 1)$, given that $j \in r_i(t)$ and $i, j \in N(t + 1)$. Then, this implies $ij \in g(t')$ for every $t' > t$, by the monotonicity of $g(\cdot)$, proved in the proof of Lemma 2. Second, suppose that $j \in r_i(t + 1)$, given that $j \in r_i(t)$ and $i, j \in N(t + 1)$. Then, when $i, j \in N(t + 2)$, we can show that either $j \in r_i(t + 2)$ or $ij \in g(t + 1)$ holds, by repeating exactly the same argument as in the case of $t' = t + 1$, but by replacing t with $t + 1$. We can repeat this argument to show that for any $t' = t + k$ with $k > 0$, the statement of the claim holds.

Now, suppose, to the contrary, that given that $j \in r_i(t)$ and $i, j \in N(t + 1)$, both $j \notin r_i(t + 1)$ and $ij \notin g(t)$ hold. By Lemma 1, $k \in r_i(t + 1)$ implies $k \in r_i(t)$, because of $j \in r_i(t)$ and $j \notin r_i(t + 1)$. That is, we have $r_i(t + 1) \subsetneq r_i(t)$, where the inclusion is strict because of j .

Since the payoff function is separable, $j \in r_i(t)$ implies $\Delta c(q_i(g(t-1)) + \sharp r_i(t) - 1) < b(d(i, j))$. Also, $j \notin r_i(t+1)$, $ij \notin g(t)$, and $j \in N(t+1)$ imply $\Delta c(q_i(g(t)) + \sharp r_i(t+1)) > b(d(i, j))$. Therefore, we have $q_i(g(t-1)) + \sharp r_i(t) \leq q_i(g(t)) + \sharp r_i(t+1)$, because Δc is increasing.

On the other hand, we have $N_i(g(t)) \subseteq N_i(g(t-1)) \cup r_i(t)$, by construction. Together with $r_i(t+1) \subsetneq r_i(t)$, we obtain $N_i(g(t)) \cup r_i(t+1) \subsetneq N_i(g(t-1)) \cup r_i(t)$. This implies that we have $q_i(g(t)) + \sharp r_i(t+1) < q_i(g(t-1)) + \sharp r_i(t)$, because $N_i(g(t)) \cap r_i(t+1) = \emptyset$. But this contradicts our earlier conclusion that $q_i(g(t-1)) + \sharp r_i(t) \leq q_i(g(t)) + \sharp r_i(t+1)$. This completes the proof.

Proof of Claim 2

Denote $k = \arg \max_{k \in N_i(g)} \{d(i, k)\}$, and $l = \arg \max_{l \in N_j(g)} \{d(j, l)\}$.

Suppose, to the contrary, that $u_j(g + ij) > u_j(g)$ holds. But from $ij \notin g$ and the pairwise stability of g , $u_i(g) > u_i(g + ij)$ must hold. That is, we must have $b(d(i, j)) < \Delta c(q_i(g))$. On the other hand, by the pairwise stability of g , we have $u_i(g) > u_i(g - ik)$. That is, $b(d(i, k)) > \Delta c(q_i(g) - 1)$ holds. When c is concave or linear, this contradicts $b(d(i, j)) < \Delta c(q_i(g))$, since $\Delta c(q)$ is nonincreasing and $b(d(i, j)) > b(d(i, k))$.

Consider the case where c is convex. By Lemma 2, $r_i(t') = \emptyset$ for some t' . Since $k \in r_i(t'')$ for some $t'' < t'$, by Lemma 1, $j \in r_i(t'')$ holds. From Claim 1, we have $j \in r_i(t)$ for any $t > t''$ such that $j \in N(t)$. This implies $j \in r_i(t')$, contradicting $r_i(t') = \emptyset$.

Therefore, for c that is either concave, convex, or linear, the statement is proved.

Proof of Claim 3

Suppose, to the contrary, that $d(i, j) < d(j, l)$ holds.

Consider the case in which c is linear or concave. From Claim 2, $u_j(g + ij) < u_j(g)$, it holds that $b(d(i, j)) < \Delta c(q_j(g))$. Since g is pairwise stable, $u_j(g - jl) < u_j(g)$, so that $b(d(j, l)) > \Delta c(q_j(g) - 1)$ holds, where l is defined in the proof of Claim 2 (we define k in the same way as in the proof of Claim 2, too). But this implies $\Delta c(q_j(g) - 1) < \Delta c(q_j(g))$, because of $b(d(i, j)) > b(d(j, l))$. This contradicts that $\Delta c(q)$ is nonincreasing.

Consider the case of convex c . First, note that, as proved in the proof of Lemma 1, there is no agent deleting his links in the algorithm, when c is convex. Then, from $ik \in g$, at some t' , $k \in r_i(t')$ holds. Similarly, by $jl \in g$, at some t'' , $l \in r_j(t'')$ holds. We have $j \in r_i(t')$ and $i \in r_j(t'')$ by Lemma 1 if $j \in N(t')$ and $i \in N(t'')$. Thus, it cannot be the case that $t' = t''$, as it would imply $ij \in g$.

Consider the case of $t' < t''$. Claim 1 implies that $j \in r_i(t)$ for all $t > t'$ whenever $j \in N(t)$. But then $j \in N(t'')$ implies $j \in r_i(t'')$, which would

imply $ij \in g$ as there is no “deletion procedure” in the case of convex c as we have seen already. In a perfectly symmetric manner, we cannot have $t' > t''$. Thus the proof is complete.

Proof of Lemma 4

We claim that

$$\left(\max_{i \in N_1(g)} \{d(1, i)\}, \max_{i \in N_2(g)} \{d(2, i)\}, \dots, \max_{i \in N_n(g)} \{d(n, i)\} \right)$$

is a cutoff value profile $\hat{d} = (\hat{d}_1, \hat{d}_2, \dots, \hat{d}_n)$ generating g , where $g = g(\bar{t})$.

By the definition of the cutoff rule, it suffices to show that we do not have the case in which there exists $ij \notin g$ such that $d(i, j) \leq \min\{\hat{d}_i, \hat{d}_j\}$. Suppose this holds. Then, $ij \notin g$ and $d(i, j) < \max_{k \in N_i(g)} \{d(i, k)\}$ while $d(i, j) < \max_{l \in N_j(g)} \{d(j, l)\}$ hold. This contradicts Claim 3, so that the existence of a cutoff value profile is proved.

Proof of Lemma 5

As shown in Proposition 1, pairwise stable network is unique, and hence Lemma 3 implies that the generated network is the network constructed in the proof of Proposition 1. Due to the separability of the payoff function, it is straightforward to see that the network is also stringly stable. Hence, we constraint attention to the case in which c is convex: We prove that $g = g(\bar{t})$ is strongly stable when c is convex. Take g' that is obtainable from g via deviations by a set of agents $S \subseteq N$. The statement of the lemma is true if

$$[\exists s \in S \ u_s(g') > u_s(g)] \implies [\exists s' \in S, \ u'_s(g') < u'_s(g)].$$

Hence, it suffices to show that it cannot be the case that $u_s(g') > u_s(g)$ for every $s \in S$. Define $D(s) = \{j \in N \mid sj \in g, \ sj \notin g'\}$ and $A(s) = \{j \in N \mid sj \notin g, \ sj \in g'\}$, that is, $D(s)$ (resp. $A(s)$) is a set of agents whose link to $s \in S$ is deleted (resp. added) in the deviations.

We are going to show that, for the profitable deviations by S to be possible, there must exist an infinite sequence of agents, denoted by $s_1, s_2, s_3, \dots \in S$, such that $s_{l+1} \in A(s_l) \setminus \{s_1, s_2, \dots, s_{l-1}\}$ for each s_l . (Since S is finite, this is impossible.) To derive this sequence, we also show that $q_{s_l}(g) \leq q_{s_{l+1}}(g)$ holds, and either $d(s_l, \bar{n}_l) > d(s_{l+1}, \bar{n}_{l+1})$ or $q_{s_l}(g) < q_{s_{l+1}}(g)$ holds, where \bar{n}_i denotes an agent whose distance to s_i is the longest among s_i 's neighbors, i.e. $d(s_i, \bar{n}_i) = \max_{n \in N_{s_i}(g)} \{d(s_i, n)\}$. We prove them by the mathematical induction.

First, take an agent denoted by $s_1 \in S$. Since the rule of the final step of the algorithm and the convexity of c ensures that there is no incentive

to delete links, agents in S cannot be better off by only deleting their links in the deviations, implying $A(s) \neq \emptyset$ for every $s \in S$. There are two cases concerning $A(s_1)$.

- Case 1: $\forall s_i \in A(s_1), u_{s_1}(g) > u_{s_1}(g + s_1 s_i)$.

In this case, if we have $d(s_1, s_i) > d(s_1, \bar{n}_1)$ for every $s_i \in A(s_1)$, then it would be impossible to satisfy $u_{s_1}(g) < u_{s_1}(g')$. To see this, we calculate s_1 's net gain from the deviations as follows. When s_1 's degree increases in the deviations, i.e. $\sharp A(s_1) > \sharp D(s_1)$, his net benefit is

$$\sum_{k \in A(s_1)} b(d(s_1, k)) - \sum_{l \in D(s_1)} b(d(s_1, l)) - \sum_{j=1}^{\sharp A(s_1) - \sharp D(s_1)} \Delta c(q_{s_1}(g) + j - 1).$$

Notice that $\Delta c(q_{s_1}(g)) > b(d(s_1, s_i))$ holds for all $s_i \in A(s_1)$ in this case, and that Δc is increasing. Taking any subset $\bar{A}(s_1) \subset A(s_1)$ such that $\sharp \bar{A}(s_1) = \sharp A(s_1) - \sharp D(s_1)$, the net benefit can be rearranged to

$$\begin{aligned} & \left(\sum_{k \in A(s_1) \setminus \bar{A}(s_1)} b(d(s_1, k)) - \sum_{l \in D(s_1)} b(d(s_1, l)) \right) \\ & + \left(\sum_{k \in \bar{A}(s_1)} b(d(s_1, k)) - \sum_{j=1}^{\sharp A(s_1) - \sharp D(s_1)} \Delta c(q_{s_1}(g) + j - 1) \right), \end{aligned}$$

which is negative because $\forall s_i \in A(s_1), u_{s_1}(g) > u_{s_1}(g + s_1 s_i)$. The same argument carries over to the situation where his degree does not increase in the deviations.

Hence, we can focus on the case where there exists $s_i \in A(s_1)$ such that $d(s_1, s_i) < d(s_1, \bar{n}_1)$ holds. Take such an agent s_i , and denote him by s_2 . The inequality $d(s_1, s_2) < d(s_1, \bar{n}_1)$ and $s_1 s_2 \notin g$ imply, by Claims 2 and 3 above, $u_{s_2}(g) > u_{s_i}(g + s_1 s_2)$ and $d(s_2, \bar{n}_2) < d(s_1, s_2)$.

Notice that we obtained the desired inequality $d(s_2, \bar{n}_2) < d(s_1, \bar{n}_1)$.

Now, we show that $q_{s_1}(g) \leq q_{s_2}(g)$: The pairwise stability of g implies $b(d(s_1, \bar{n}_1)) > \Delta c(q_{s_1}(g) - 1)$, and $u_{s_2}(g) > u_{s_2}(g + s_1 s_2)$ implies $b(d(s_1, s_2)) < \Delta c(q_{s_2}(g))$. By $b(d(s_1, s_2)) \geq b(d(s_1, \bar{n}_1))$, we have that $\Delta c(q_{s_1}(g) - 1) < \Delta c(q_{s_2}(g))$. Hence, we also get inequality $q_{s_1}(g) \leq q_{s_2}(g)$, since Δc is increasing.

- Case 2: $\exists s_i \in A(s_1), u_{s_1}(g) < u_{s_1}(g + s_1 s_i)$.

Take $s_i \in A(s_1)$ such that $u_{s_1}(g) < u_{s_1}(g + s_1 s_i)$, and denote this agent by s_2 . By the pairwise stability of g , we have $u_{s_2}(g) > u_{s_2}(g + s_0 s_1)$. From $\Delta c(q_{s_1}(g)) < b(d(s_1, s_2)) < \Delta c(q_{s_2}(g))$, we get a desired inequality $q_{s_1}(g) < q_{s_2}(g)$.

Hence, we have shown the desired statements for the first step $l = 1$: There exists $s_2 \in A(s_1)$ such that $q_{s_1}(g) \leq q_{s_2}(g)$ holds, and either $d(s_1, \bar{n}_1) > d(s_2, \bar{n}_2)$ (Case 1) or $q_{s_1}(g) < q_{s_2}(g)$ (Case 2) holds.

Next, let us suppose that we have shown the statements up to $l = r$: There exists a sequence (s_1, s_2, \dots, s_r) in S such that $s_{l+1} \in A(s_l) \setminus \{s_1, s_2, \dots, s_{l-1}\}$, and $q_{s_l}(g) \leq q_{s_{l+1}}(g)$ holds, and either $d(s_l, \bar{n}_l) > d(s_{l+1}, \bar{n}_{l+1})$ or $q_{s_l}(g) < q_{s_{l+1}}(g)$ holds for each $l = 1, 2, \dots, r$.

Suppose, to the contrary, that $A(s_{r+1}) \subseteq \{s_1, s_2, \dots, s_r\}$. We show this is impossible, for both cases below.

- Case 1': $\forall s_i \in A(s_{r+1}), u_{s_{r+1}}(g) > u_{s_{r+1}}(g + s_{r+1} s_i)$.

Due to the same discussion as in the Case 1 above, we can focus on the case where there exists $s_i \in A(s_{r+1})$ such that $d(s_{r+1}, s_i) < d(s_{r+1}, \bar{n}_{r+1})$. Take such an agent s_i . As we have derived $d(s_2, \bar{n}_2) < d(s_1, \bar{n}_1)$ and $q_{s_1}(g) \leq q_{s_2}(g)$ in the Case 1 above, we can get $d(s_i, \bar{n}_i) < d(s_{r+1}, \bar{n}_{r+1})$ and $q_{s_{r+1}}(g) \leq q_{s_i}(g)$. But, because $s_i \in \{s_1, s_2, \dots, s_r\}$, at least one of them contradicts the assumed inequalities: $q_{s_l}(g) \leq q_{s_{l+1}}(g)$, and either $d(s_l, \bar{n}_l) > d(s_{l+1}, \bar{n}_{l+1})$ or $q_{s_l}(g) < q_{s_{l+1}}(g)$ for each $l = 1, 2, \dots, r$.

- Case 2': $\exists s_i \in A(s_{r+1}), u_{s_{r+1}}(g) < u_{s_{r+1}}(g + s_{r+1} s_i)$.

Take $s_i \in A(s_{r+1})$ such that $u_{s_{r+1}}(g) < u_{s_{r+1}}(g + s_{r+1} s_i)$. By the same logic with which we have obtained $q_{s_1}(g) < q_{s_2}(g)$ in the Case 2 above, we can get inequality $q_{s_{r+1}}(g) < q_{s_i}(g)$. But, because $s_i \in \{s_1, s_2, \dots, s_r\}$, this contradicts the assumed inequalities: $q_{s_l}(g) \leq q_{s_{l+1}}(g)$ for each $l = 1, 2, \dots, r$.

Hence, for both cases, it must be the case that $A(s_{r+1})$ includes some $s_i \in S \setminus \{s_1, s_2, \dots, s_r\}$, such that $q_{s_{r+1}}(g) \leq q_{s_i}(g)$ holds, and either $d(s_{r+1}, \bar{n}_{r+1}) > d(s_i, \bar{n}_i)$ or $q_{s_{r+1}}(g) < q_{s_i}(g)$ holds. Denote this s_i by s_{r+2} .

Therefore, the mathematical induction is complete.

Since it is impossible for all the elements in the infinite sequence s_1, s_2, \dots are included in finite $S \subseteq N$, we conclude that there is no profitable deviations by any set of agents $S \subseteq N$. Hence, $g = g(\bar{t})$ is strongly stable.

Proof of Lemma 6

Proposition 1 establishes the uniqueness for the case of linear cost functions, so we concentrate on the case of convex cost functions.

Suppose, to the contrary, that a network $g' \neq g = g(\bar{t})$ is also strongly stable. This implies that no pair of agents can profitably deviate from g' . We will show that this contradicts the finiteness of N .

Notice first that there exists some $i \in N$ such that $u_i(g) > u_i(g')$, as otherwise g would not be strongly stable. Take such an agent arbitrarily and call him agent 1. Consider two possible (exhaustive) cases.

- (i) $q_1(g) > q_1(g')$: In this case, pairwise stability of g and the convexity of the cost function imply that there exists some $i \in N_1(g) \setminus N_1(g')$ such that $u_1(g' + 1i) > u_1(g')$.
- (ii) $q_1(g) \leq q_1(g')$: In this case, we can find some $i \in N_1(g) \setminus N_1(g')$ such that there exists $j \in N_1(g')$ such that $d(1, i) < d(1, j)$. Denote this j by 0. To see this, suppose, to the contrary, that for all $i \in N_1(g) \setminus N_1(g')$, for all $j \in N_1(g')$, $d(1, i) > d(1, j)$ holds. Take an arbitrary network $g'' \subset g'$ such that $N_1(g) \cap N_1(g') \subset N_1(g'')$ and $q_1(g) = q_1(g'')$. Such g'' exists because $q_1(g) \leq q_1(g')$. Then we have $u_1(g) \leq u_1(g'') \leq u_1(g')$, where the first inequality holds because we have, when $g \neq g''$, $\forall i \in N_1(g) \setminus N_1(g'')$, $\forall j \in N_1(g'')$, $d(1, i) > d(1, j)$, and the second inequality is due to the pairwise stability of g' and the convexity of c . But this contradicts our earlier conclusion that $u_1(g) > u_1(g')$.

In either case (i) or (ii), we take such i and call him agent 2.

To complete the proof, we construct a sequence of distinct agents, $\{1, 2, \dots\}$, such that $2k \in N_{2k-1}(g) \setminus N_{2k-1}(g')$, $2k + 1 \in N_{2k}(g') \setminus N_{2k}(g)$, $2k + 2 \in N_{2k+1}(g) \setminus N_{2k+1}(g')$, and $d(2k - 1, 2k) > d(2k, 2k + 1) > d(2k + 1, 2k + 2)$ hold for each $k = 1, 2, \dots$. We have considered a portion of the case with $k = 1$ in the previous paragraph. The rest of the first step can be shown to be true by following exactly the same logic as we will have below (by substituting $k = 0$), so we omit its proof.

Now, we start mathematical induction argument to obtain the remaining parts of the infinite sequence and inequalities.

First, suppose we have shown the claims up to step k , and consider step $k + 1$. Then, we must have $u_{2k+2}(g' + (2k + 1)(2k + 2)) < u_{2k+2}(g')$, as otherwise the pair $2k + 1$ and $2k + 2$ could profitably deviate from g' , by adding $(2k + 1)(2k + 2)$ while simultaneously deleting $2k(2k + 1)$. Hence, by the pairwise stability of $g \ni (2k + 1)(2k + 2)$ and the cost convexity, we have $q_{2k+2}(g) \leq q_{2k+2}(g')$. Notice that this implies there is an agent in $N_{2k+2}(g')$ who is not in $N_{2k+2}(g)$, because $(2k + 1)(2k + 2) \in g \setminus g'$. Similarly, we must

have $d(2k+2, i) < d(2k+1, 2k+2)$ satisfied for all $i \in N_{2k+2}(g')$, to ensure that $2k+1$ and $2k+2$ do not profitably deviate from g' .

The two conclusions in the previous paragraph imply that we can find some $i \in N_{2k+2}(g') \setminus N_{2k+2}(g)$ such that $d(2k+2, i) < d(2k+1, 2k+2)$. If $i = 2l-1$ (resp. $2l$) for some $l = 1, 2, \dots, k$, then we would have $d(2l-1, 2k+2) < d(2k+1, 2k+2) < d(2l-1, 2l)$ (resp. $d(2l, 2k+2) < d(2k+1, 2k+2) < d(2l-1, 2l)$) by the inductive supposition. But, this contradicts Claim 3, because we have $(2l-1)2l, (2k+1)(2k+2) \in g$ and $(2l-1)(2k+2) \notin g$ (resp. $2l(2k+2) \notin g$). Hence $i \notin \{1, 2, \dots, 2k+1\}$. Denote this i by $2k+3$.

Since we have $(2k+1)(2k+2) \in g$, $(2k+2)(2k+3) \notin g$ and $d(2k+1, 2k+2) > d(2k+2, 2k+3)$, we can apply Claim 2 to get $u_{2k+3}(g + (2k+2)(2k+3)) < u_{2k+3}(g)$. Then, this implies $q_{2k+3}(g) \geq q_{2k+3}(g')$, due to the cost convexity and the pairwise stability of $g' \ni (2k+2)(2k+3)$. Again, this implies that we can find some $i \in N_{2k+3}(g) \setminus N_{2k+3}(g')$, as $(2k+2)(2k+3) \in g' \setminus g$. By applying Claim 3, we must have $d(2k+3, i) < d(2k+2, 2k+3)$. If $i \in \{2, 3, \dots, 2k+1\}$, then we would have $d(i, 2k+3) < d(2k+2, 2k+3) < d(i, i+1) < d(i-1, i)$ by the inductive supposition. But if i is odd (resp. even), then i and $2k+3$ could profitably deviate from g' by adding $i(2k+3)$ while deleting $(i-1)i$ (resp. $i(i+1)$) and $(2k+2)(2k+3)$, respectively. Also, if $i = 1$, then the profitable deviation by 1 and $2k+3$ from g' is possible. This is because 1 would be better off by adding $(2k+3)1$ (as $(2k+3, 1) < d(2k+2, 2k+3) < d(1, 2)$) in case (i) and by adding $(2k+3)1$ and deleting 01 in case (ii), and $(2k+3)$ would be better off by adding $(2k+3)1$ and deleting $(2k+2)(2k+3)$. Hence, it must be the case that $i \notin \{1, 2, 3, \dots, 2k+1\}$. Denoting such agent i by $2k+4$, we have shown the desired properties for step $k+1$.

We have completed the induction. But since N is finite, it is impossible to have such infinite sequence of distinct agents. This completes the proof. \square

A.8 Proof of Proposition 8

Proof. Consider a point x in the type space X , and a hypothetical agent i who is situated at x , i.e. $x = x_i$.

Let $q(x_i, \delta)$ denote the number of agents in δ -neighborhood of x_i . Then, for any $\delta > 0$ and q' , $q(x_i, \delta) > q'$ holds almost surely as $n \rightarrow \infty$. Also, $\lim_{q \rightarrow \infty} \Delta c(q) = c_1 > 0$ implies that for all $\epsilon > 0$, there exists q' such that for all $q_i > q'$, $|\Delta c(q_i) - c_1| < \epsilon$.

Now, take a small enough ϵ' and $\delta' > 0$ such that $b(\delta') \geq c_1 + \epsilon'$. Such ϵ' and δ' exist since $\lim_{d \rightarrow 0} b(d) > c_1$.

If i is not connected with an agent in his δ' -neighborhood, the resulting network would not be pairwise stable, hence it is not strongly stable. Thus, i

is connected with all the agents in his δ' -neighborhood. Thus, for any $\epsilon > 0$, we have $|\Delta c(q_i) - c_1| < \epsilon$ almost surely as $n \rightarrow \infty$.

Now, consider links with agents outside of the δ' -neighborhood. Since strongly stability implies pairwise stability, $c_1 - \epsilon < \Delta c(q_i)$ (implied by $|\Delta c(q_i) - c_1| < \epsilon$) implies that $ij \notin g$ if $b(d(i, j)) \leq c_1 - \epsilon$, or $\hat{d} + \epsilon' \leq d(i, j)$ for $b^{-1}(c_1) = \hat{d}$ and some $\epsilon' > 0$. Also, for the same reason, $\Delta c(q_i) < c_1 + \epsilon$ (implied by $|\Delta c(q_i) - c_1| < \epsilon$) implies that $ij \in g$ if $c_1 + \epsilon \leq b(d(i, j))$, or $d(i, j) \leq \hat{d} - \epsilon''$ for the same \hat{d} and for some $\epsilon'' > 0$.

Now, for any ϵ' and ϵ'' , there exist agents j and k such that $\hat{d} + \epsilon' < d(i, j) < \hat{d} + 2\epsilon'$ and $\hat{d} - 2\epsilon'' < d(i, k) < \hat{d} - \epsilon''$ almost surely as $n \rightarrow \infty$. Also, these j and k have to satisfy $ij \notin g$ and $ik \in g$ because of the argument in the previous paragraph. Hence, agent i 's cutoff value, denoted by \hat{d}_i , which we know exists from Proposition 7, has to satisfy $\hat{d} - 2\epsilon'' \leq \hat{d}_i < \hat{d} + 2\epsilon'$ almost surely as $n \rightarrow \infty$. Because ϵ' and ϵ'' go to zero as ϵ goes to zero by the continuity and strict decreasingness of b , and because x can be arbitrary, the proof is completed. \square

A.9 Proof of Proposition 9

Proof.

The procedure is almost the same as the Proof for Proposition 1.

We only need to modify the expression in the proof of Proposition 1:

$$\frac{1}{(\hat{d})^k} \int_0^{\hat{d}} \int_0^{\hat{d}} \cdots \int_0^{\hat{d}} \frac{(2\hat{d} - y_1)(2\hat{d} - y_2) \cdots (2\hat{d} - y_k)}{(2\hat{d})^k} dy_1 dy_2 \cdots dy_k$$

to take into account the heterogeneity of the cutoff values.

The expression has lower bound when the node in consideration has the cutoff of $\hat{d} + \epsilon$, where all the other nodes have the cutoffs $\hat{d} - \epsilon$, which is larger than:

$$\begin{aligned} \frac{1}{(\hat{d} + \epsilon)^k} \int_{2\epsilon}^{\hat{d} + \epsilon} \int_{2\epsilon}^{\hat{d} + \epsilon} \cdots \int_{2\epsilon}^{\hat{d} + \epsilon} \frac{(2\hat{d} - y_1)(2\hat{d} - y_2) \cdots (2\hat{d} - y_k)}{(2\hat{d} + 2\epsilon)^k} dy_1 dy_2 \cdots dy_k \\ = \left(\frac{\frac{3}{2}\hat{d}^2 - 2\hat{d}\epsilon - \epsilon + \frac{3}{2}\epsilon^2}{2(\hat{d} + \epsilon)^2} \right)^k. \end{aligned}$$

Also, it has an upper bound when the node in consideration has the cutoff of $\hat{d} - \epsilon$, where all the other nodes have the cutoffs $\hat{d} + \epsilon$, which is smaller than:

$$\frac{1}{(\hat{d} - \epsilon)^k} \int_0^{\hat{d} - \epsilon} \int_0^{\hat{d} - \epsilon} \cdots \int_0^{\hat{d} - \epsilon} \frac{(2\hat{d} - y_1)(2\hat{d} - y_2) \cdots (2\hat{d} - y_k)}{(2\hat{d} - 2\epsilon)^k} dy_1 dy_2 \cdots dy_k$$

$$= \left(\frac{\frac{3}{2}\hat{d} + \frac{1}{2}\epsilon}{2(\hat{d} - \epsilon)} \right)^k.$$

For any $\hat{d} > 0$, both bounds converges to the same desired limit, $\left(\frac{3}{4}\right)^k$ as ϵ goes to zero. This completes the proof. \square

A.10 Proof of a Claim in Subsection 5.3

Proof. We prove that ΔE is decreasing in m and increasing in k . To see this, fix agent i with position x_i . We first calculate the probability that agent i and his neighbor j has a dissimilar political type, i.e. $|x_{i1} - x_{j1}| > \hat{d}$. Assuming that $\hat{d} \leq x_{i1} \leq 1 - \hat{d}$, by the strong law of large numbers, this probability is:

$$\Pr(|x_{i1} - x_{j1}| > \hat{d} \mid j \in N_i(g)) = \frac{\sum_{l=k}^{m-1} \binom{m-1}{k} (2\hat{d})^l (1 - 2\hat{d})^{m-l-1}}{\sum_{l'=k}^m \binom{m}{k} (2\hat{d})^{l'} (1 - 2\hat{d})^{m-l'}}.$$

The limit of this probability as $\hat{d} \rightarrow 0$ is:

$$\lim_{\hat{d} \rightarrow 0} \frac{\binom{m-1}{k} (2\hat{d})^k (1 - 2\hat{d})^{m-k-1}}{\binom{m}{k} (2\hat{d})^k (1 - 2\hat{d})^{m-k}} = \frac{m-k}{m}.$$

Using this limit probability, we now calculate the almost sure limit of \tilde{x}_{i1} as $n \rightarrow \infty$ and then $\hat{d} \rightarrow 0$. First, notice that:

$$\lim_{n \rightarrow a.s. \infty} \tilde{x}_{i1} = \lim_{n \rightarrow a.s. \infty} \frac{x_{i1} + \beta \cdot \sum_{j \in N_i(g)} x_{j1}}{1 + \beta \cdot q_i(g)} = \lim_{n \rightarrow a.s. \infty} \frac{\sum_{j \in N_i(g)} x_{j1}}{q_i(g)}.$$

Note that this converges to x_{i1} as $\hat{d} \rightarrow 0$. Hence,

$$\begin{aligned} & \lim_{\hat{d} \rightarrow 0} \lim_{n \rightarrow a.s. \infty} \tilde{x}_{i1}(g) \\ &= \lim_{\hat{d} \rightarrow 0} \lim_{n \rightarrow a.s. \infty} \Pr(|x_{i1} - x_{j1}| > \hat{d} \mid j \in N_i(g)) \cdot E[x_{j1} \mid |x_{i1} - x_{j1}| > \hat{d}, j \in N_i(g)] \\ & \quad + \lim_{\hat{d} \rightarrow 0} \lim_{n \rightarrow a.s. \infty} \Pr(|x_{i1} - x_{j1}| \leq \hat{d} \mid j \in N_i(g)) \cdot E[x_{j1} \mid |x_{i1} - x_{j1}| \leq \hat{d}, j \in N_i(g)] \\ &= \frac{m-k}{m} \mu + \left(1 - \frac{m-k}{m}\right) x_{i1}. \end{aligned}$$

Now we calculate the value of ΔE . Recalling that as $\hat{d} \rightarrow 0$ the fraction of player i 's who do not satisfy $\hat{d} \leq x_{i1} \leq 1 - \hat{d}$ goes to zero and these agents

have only vanishing effect in the limit because the distance is bounded, we have that by definition,

$$\begin{aligned}\Delta E &= \mathbb{E} \left[\left| \left(\frac{m-k}{m} \mu + \left(1 - \frac{m-k}{m} \right) x_{i1} \right) - \left(\frac{m-k}{m} \mu + \left(1 - \frac{m-k}{m} \right) x_{j1} \right) \right| \right] \\ &= \frac{k}{m} \mathbb{E} [|x_{i1} - x_{j1}|] \\ &= \frac{k}{3m}.\end{aligned}$$

Therefore, ΔE is decreasing in m and increasing in k . \square

A.11 Proof of Proposition 11

Proof. Part (i):

Fix a composition of nodes. Take any $\delta > 0$. Then, there exists $\epsilon > 0$ such that the probability of events in which the cutoff for some arbitrary chosen pair ij is contained in $(\hat{d} - \delta, \hat{d} + \delta)$ is above $1 - \delta$ almost surely as $n \rightarrow \infty$ (We suppress the dependence on n in the sequel in this proof of Proposition 11, to simplify the exposition). Note that the supremum of the possible ϵ tends to zero as δ tends to zero. Thus, Cl^{S1} is a convex combination of clustering coefficients among agents who have cutoffs in $(\hat{d} - \delta, \hat{d} + \delta)$ and some value in $[0, 1]$, with the weight $1 - \delta'$ being placed on the former and the weight δ' on the latter, where $\delta' > 0$ is a constant that tends to zero as δ tends to zero. But we know from the proof of Proposition 9 that the former converges to the clustering coefficient with $\epsilon = 0$ as δ goes to zero (although the proof concerns the limit of clustering coefficient as the cutoff goes to zero, it is straightforward that the proof can be applied to clustering coefficient with any fixed cutoff values). Thus, by taking arbitrary small ϵ , δ goes to zero, and so the convex combination tends to the one with $\epsilon = 0$. This completes the proof.

Part (ii):

Fix a composition of nodes. Cl^{S2} is a convex combination of clustering coefficients among agents who have the cutoff of \hat{d} and some value in $[0, 1]$, with the weight $1 - \delta'$ being placed on the former and the weight δ' on the latter, where $\delta' > 0$ is a constant that tends to zero as ϵ tends to zero. But the former is exactly Cl^{S2} with $\epsilon = 0$. Thus, by taking arbitrary small ϵ , δ' goes to zero, and so the convex combination tends to the one with $\epsilon = 0$. This completes the proof. \square

A.12 Proof of Proposition 12

Proof. Part (i):

Fix the cutoff $\hat{d} > 0$. Consider two points in X , and hypothetical agents i and j who are situated at these points. The probability of the event in which i and at least one of j 's neighbors are connected tends to 1 as n goes to infinity because the random term w_{ij} 's have full support over \mathbb{R}_+ , by the strong law of large numbers. Take such an agent and call him agent h . Now, agents i and h are connected, and agents h and j are connected. Thus, the path length between i and j is at least 2 almost surely as $n \rightarrow \infty$. Since this argument holds for all pairs of points in X , we are done.

Part (ii):

Fix the deterministic part of the cutoff $\hat{d} > 0$. Consider two points in X , and let i and j be hypothetical agents who are situated at these two points. Let their cutoffs be \hat{d}_i and \hat{d}_j . Note that the distance between these points is strictly between 0 and 1 almost surely. Now, for any $\delta > 0$, the probability of the event in which i has at least one neighbor h in the interior of X with $\hat{d}_h > 1 - \delta$ is above $1 - \delta$, for sufficiently large n . Similarly, for any $\delta > 0$, the probability of the event in which j has at least one neighbor h' in the interior of X with $\hat{d}_{h'} > 1 - \delta$ is above $1 - \delta$, for sufficiently large n . Note that for small enough δ , $d(h, h') \leq 1 - \delta$ holds. Now, agents i and h are connected, agents h and h' are connected, agents h' and j are connected. Thus, the path length between i and j is at most 3 almost surely as $n \rightarrow \infty$. This completes the proof. \square

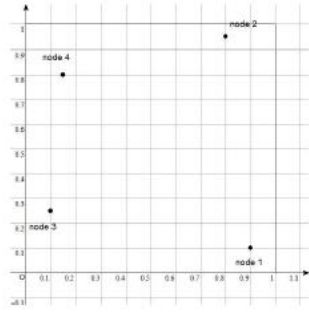


Figure 1(a-1)

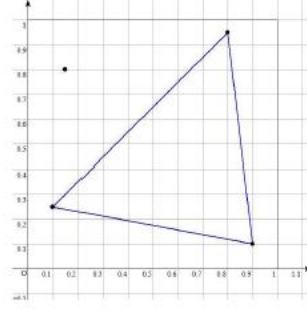


Figure 1(a-2)

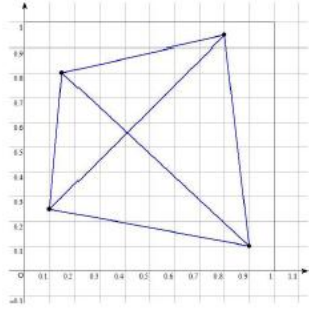


Figure 1(a-3)

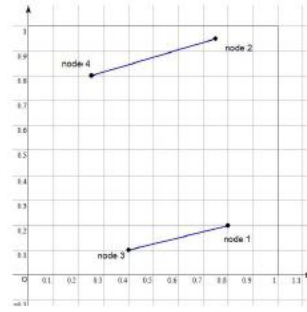


Figure 1(b-1)

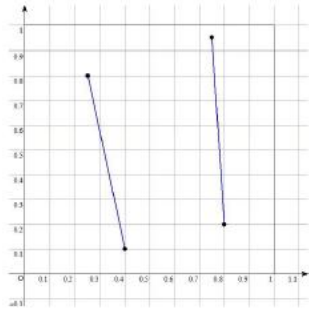


Figure 1(b-2)

Figure 1: (a): $k = m = 2$, $b(d) = 1/d$, $c(0) = 0$, $c(1) = 2$, $c(2) = 2.2$, and $c(3) = 2.3$. All three networks are pairwise stable. A cutoff value profile for the network in (a-2) cannot be homogeneous. The network in (a-3) is strongly stable. (b): $k = m = 2$, $b(d) = 1/d$, $c(0) = 0$, $c(1) = 1$, $c(2) = 10$, and $c(3) = 30$. Both networks are pairwise stable. The network in (b-2) cannot be generated by a cutoff rule, and it is not strongly stable.

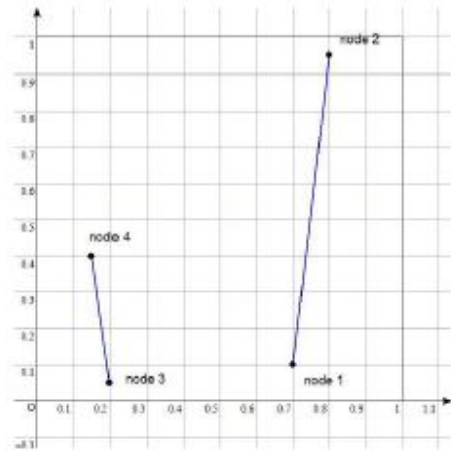


Figure 2: $k = m = 2$, $b(d) = 1/d$, $c(0) = 0$, $c(1) = 1$, $c(2) = 5$, and $c(3) = 10$. The network is strongly stable, but cannot be generated by a cutoff rule with a homogeneous cutoff value profile.