Theoretical Economics 12 (2017), 655–689 1555-7561/20170655

Social distance and network structures

Ryota Iijima
Department of Economics, Yale University

Yuichiro Kamada
Haas School of Business, University of California, Berkeley

This paper proposes a tractable model that allows us to analyze how agents' perception of relationships with others determines the structures of networks. In our model, agents are endowed with their own multidimensional characteristics and their payoffs depend on the *social distance* between them. We characterize the clustering coefficient and average path length in stable networks, and analyze how they are related to the way agents measure social distances. The model predicts the small-world properties under a class of social distance that violates the triangle inequality. Allowing for heterogeneity in link-formation costs, the model also accommodates other well documented empirical patterns of social networks such as skewed degree distributions, positive assortativity of degrees, and clustering-degree correlation.

Keywords. Network formation, heterogeneity, spatial type topologies, clustering, average path length, weak ties.

JEL classification. A14, C72, D85.

1. Introduction

The structure of a social network plays an important role in determining the behavior of agents in various settings, for example, peer effects, opinion formation, and information diffusion.[1] Numerous empirical works have shown that different networks have different structures, where the structures of networks are evaluated by various measures.[2]

Ryota Iijima: riijima@yale.edu
Yuichiro Kamada: y.cam.24@gmail.com

We thank the editor and the referees for their thoughtful and constructive comments. We are grateful to Drew Fudenberg, Ben Golub, Matthew Jackson, Akihiko Matsui, Markus Mobius, and Alvin Roth 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, Hiroaki Kaido, Kazuya Kamiya, Yuichi Kitamura, So Kubota, David Miller, Yusuke Narita, Sheng Qiang, 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, Seventeenth Darwin Seminar at Tokyo Institute of Technology, and the JIMS Workshop of Marketing Dynamics.

[1] See Goyal (2007) and Jackson (2008a), among others, for a survey.

[2] For references on these empirical works, see, for example, Newman (2003).

DOI: 10.3982/TE1873

Two well known and well used measures of network structures are the clustering coefficient and average path length, which represent a network's cliquishness and connectedness, respectively. For example, the "e-mail network," in which a link represents an incidence of an email exchange, has a low clustering coefficient and a low average path length (Ebel et al. 2002). The "co-authorship network," in which a link represents an incidence of co-authorship between two economics scholars, has a high clustering coefficient and a high average path length (Goyal et al. 2006).[3] Why do some networks have low clustering coefficient and/or average path length, while others do not?

To answer our motivating question, let us start from the following casual observation: E-mail exchanges do not require many similar aspects between the sender of the e-mail and the receiver, while co-authoring likely needs both parties to have more similar interests. More generally, in different networks, people have different ways to evaluate their relationships with others. The present paper argues that this difference in how people evaluate relationships explains why different networks have different structures. We construct a model in which agents are endowed with their own multidimensional characteristics. We characterize clustering coefficient and average path length in stable networks, and analyze how they are related to the way agents measure the distances between their characteristics. The model also produces well documented empirical patterns, such as skewed degree distributions, positive assortativity, and negative clustering-degree correlations.

More specifically, this paper models agents' characteristics, or *types*, as points in a multidimensional *type space*, and analyzes how the network structure depends on the notion of distance in the type space. Each coordinate indicates some aspect of agents' characteristics, such as jobs, locations, tastes, political views, and so forth. Distance in the type space, which we call *social distance*, represents the level or amount of obstacles to agents' relations, so agents form links with others who are nearby.[4] So as to compute a distance between two agents, it is necessary for them to have a way to integrate and evaluate the relationships across different dimensions. We consider a class of notions of distance, the kth norms, in which the distance between two points in the type space is the kth smallest distance among m dimension-wise distances between them, where m denotes the number of dimensions of the type space.[5] This class of distances is both sufficiently tractable to obtain closed-form solutions, and also rich enough so that we can gain relevant economic intuition, for example, by implementing comparative statics.

Once we formalize the concept of social distance as above, we postulate a simple network-formation 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

[3] Ebel et al. (2002) find that the e-mail network has a clustering coefficient of 0.034 and an average path length of 5.0, while Goyal et al. (2006) find that the co-authorship network has a clustering coefficient of 0.16 and an average path length of 9.5.

[4] Akerlof (1997), too, uses a notion of social distance to study social decisions such as choices of educational attainment and childbearing.

[5] As we show in Appendix B in the Supplemental Material (available in a supplementary file on the journal website, http://econtheory.org/supp/1873/supplement.pdf), the functional form of the kth norm naturally arises if the benefit of each link is given by the total equilibrium payoff from games (such as a repeated prisoner's dilemma) played at each of m dimensions.

involved, and the cost increases linearly with respect to degrees. That is, agents obtain higher benefits from linking to a closer agent in the type space, while needing to pay a fixed cost to maintain a link.[6] We show that in a unique pairwise stable network, there exists a cutoff on the distance between any pair of agents below which they form a link while above which they do not.[7] 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 networks as the number of agents goes to infinity and then as the cutoff value goes to zero.[8] We analytically derive the limit average path length and clustering coefficient, which vary with k and m, the key parameters that represent how agents perceive the social distance to the other agents.[9] As long as the triangle inequality is violated (i.e., $k < m$), our model predicts the "small world" property, so that average path length is small relative to population size. This result can be viewed as a novel explanation of small worlds in reality that is based on multidimensionality of social distance.

The analytical solutions enable us to study comparative statics. For instance, if $m = 2$, our model predicts lower clustering and lower average path length under the 1st norm than under the 2nd norm. Intuitively, if agents do not care about many aspects, the notion of social distance does not satisfy the triangle inequality: on the one hand, two agents j and k linked to a common agent i need not be close to each other, as they may be linked to i through different aspects of characteristics, leading to a low clustering coefficient (the top panels in Figure 1). On the other hand, this means that the two separate agents j and k can be indirectly connected through i, which leads to a low average path length (the bottom panels in Figure 1). E-mail exchanges do not require many similar aspects between the sender of the e-mail and the receiver, so k can be small relative to m. In such a case, a similar reasoning suggests a short average path length and small clustering coefficient, consistent with the data.

Different networks may be formed based on different sets of "relevant dimensions," whose numbers are captured by the parameter m.[10] Some dimensions, say religion, might matter a lot in some networks, but much less in others. Another instance of variation of relevant dimensions is the introduction of new communication technology. It can create a new dimension through which agents can be linked with each other (such

[6]For 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.

[7]The notion of pairwise stability is introduced in Jackson and Wolinsky (1996). It requires that no pair of agents would want to form a link or no agent strictly wants to sever her own link, with the rest of the network structure fixed.

[8]We elaborate on the interpretations of the large-network limit in Section 6(a).

[9]To supplement our analysis, Appendix E.5 in the Supplemental Material derives bounds on the convergence rates and also provides simulation results under fixed population sizes and cutoffs.

[10]Section 6(b) discusses the interpretation of m in more detail.

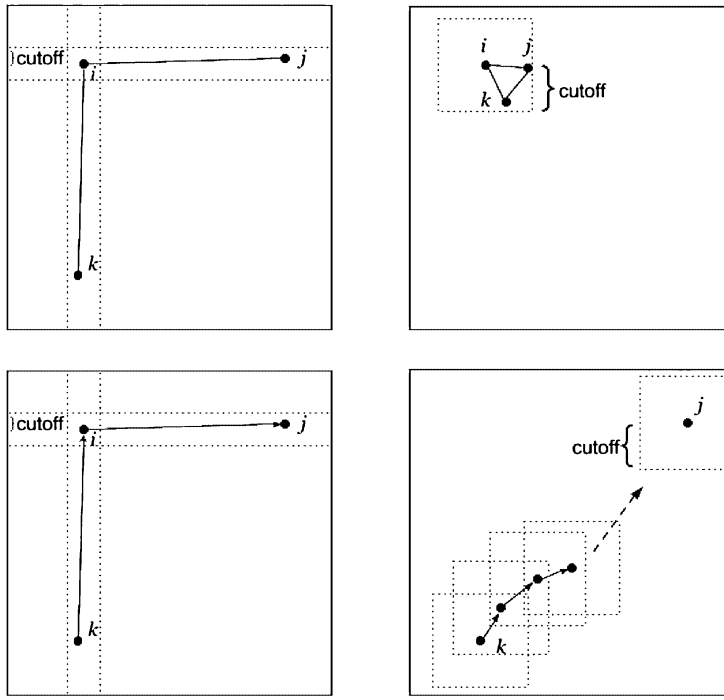

FIGURE 1. The clustering coefficient is low under $k = 1$ (the top left panel), but high under $k = 2$ (the top right panel) with $m = 2$. The average path length is low under $k = 1$ (the bottom left panel), but high under $k = 2$ (the bottom right panel) with $m = 2$

as tastes in music or views on political issues), *even if* they are far away with respect to the geographic dimension. This case can be thought of as increasing m with k being constant.[11] An analogous intuition as above suggests that in such a case both the average path length and the clustering coefficient decrease.[12] This sheds light on a new aspect of the effect of improved communication technology that seems to be absent in the literature. A standard way to model the introduction of new communication technology is to represent it as the decrease of costs for communication or link formation (Gaspar and Glaeser 1998 and Rosenblat and Mobius 2004). We could do the same by varying the cost function (which varies the cutoff); however, our limit result suggests that there can be an effect orthogonal to such a change in linking costs, and it also shows that the effect we identify comes from the violation of the triangle inequality.

Although the main model assumes homogeneous cutoff values, it may be more natural to assume that cutoffs are different across people because of heterogeneity in terms of link formation cost or benefit. We extend the main model by allowing heterogeneous cutoff values in Section 4, and show that the characterizations of clustering coefficients

[11] Chwe (2000) considers a model in which agents' actions depend on the network structure that is formed based on distances in a type space, and conducts comparative statics in which the number of relevant dimensions of the type space varies across networks.

[12] The decrease of the average path length is consistent with the recent research on Facebook networks (Backstrom et al. 2012) that finds very small average path lengths.

and average path lengths and their comparative statics generalize to this case without a substantial change. This extension allows us to predict other well documented empirical patterns of social networks: positive assortativity and degree-clustering correlation. It also approximately accommodates all possible degree distributions.

Our model is not based on a mechanical link-formation process but is utility-based. This feature enables us to obtain welfare implications. We show that the pairwise stable networks are efficient under the linear cost assumption. When the cost function is nonlinear, the existence and uniqueness of pairwise stable networks are not straightforward. Also, there can exist multiple pairwise stable networks. It is shown in Section 5 that a pairwise stable network that is generated by the cutoff rule model always exists. We further show that, under certain conditions, a strongly stable network (a refinement of pairwise stable networks) generated by the cutoff rule model always exists and is unique.[13]

Proofs of our central results are given in Appendix A. Appendixes B–E in the Supplemental Material contain the proofs of the remaining results and also additional results.

1.1 *Related literature*

Violation of the triangle inequality. The main idea behind introducing the kth norm is to formulate a notion of social distance that does not generally satisfy the triangle inequality, which will turn out to be crucial for our results. This is also empirically motivated by observations in cognitive psychology. The seminal paper on this topic by Tversky and Gati (1982) shows that it is often difficult to represent a person's perception of dissimilarities if the standard metrics are used.[14] Our approach is also related to that of multidimensional scaling, a statistical technique widely used in various fields such as cognitive science, information science, and marketing (Richardson 1938 and Torgerson 1958). This method represents the given data on dissimilarities between various objects by embedding them into a multidimensional space.[15] The data often involve dissimilarities that violate the triangle inequality, preventing the analyst from embedding unless the metric used in the target space also violates the triangle inequality. Despite the empirical validity of people's perceptions violating the triangle inequality, network formation models based on such violations are scarce.[16]

Modeling heterogeneity. 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, 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

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

[14]As shown in an earlier version of this paper (Iijima and Kamada 2015), our functional form of the kth norm can be derived from the similarity scale of Tversky (1977).

[15]The literature on "latent space" tries to embed agents in given network data to multidimensional spaces. Although it also deals with multidimensional spaces, its approach differs from ours since it restricts attention only to Euclidean distance. See, for example, Hoff et al. (2002).

[16]One exception is Watts et al. (2002) that we discuss in Section 3.2.

networks with low average path lengths and high clustering coefficients, well observed properties of real world networks. While the partitioning of agents implies that an agent belongs to exactly one group, in reality an agent may well belong to multiple groups (e.g., workplaces or local communities) or, more generally, she might be associated with attitudes or tastes on a variety of aspects (e.g., political/ethical views or tastes in music/sports). Similarly, Johnson and Giles (2000) analyze a one-dimensional spatial model in which links are formed based on costs that depend on distances between agents; their model does not capture the violation of the triangle inequality.

Explaining stylized facts. We predict patterns of network structures consistent with stylized facts in terms of clustering coefficients, average path lengths, degrees, and correlations among these measures. There are other papers that derive similar predictions based on approaches different from ours. For example, Jackson and Rogers (2007) consider a growing network formation model in which each period new agents are randomly connected to old agents while also being able to find connections through local search. König et al. (2014) consider a dynamic myopic adjustment process of network formation and analyze the set of stochastically stable networks.[17] We view our static approach that focuses on agents' heterogeneity to be complementary to those dynamic approaches.

Homophily. We assume that social distance describes the similarity between agents' characteristics, and that similar agents are connected to each other. This assumption is motivated by a well observed sociopsychological tendency called homophily (see Pin and Rogers 2016 for a recent survey on this topic). Our kth norm provides a tractable parameterization of homophily and enables us to analyze how different types of homophily result in different network structures.

Large networks. We take the limit as the number of agents tend to infinity to approximate large networks, which enables us to obtain analytical solutions. We are not the first to use this idea. There are models of growing networks in which agents are added to the population each period (e.g., Price 1976 and Barabási and Albert 1999). There are also static formation models parameterized by the number of agents that the analyst and a static formation model is parameterized by the number of agents that the analyst takes to infinity. Our model is of the latter type. Here we mention two papers in that category in which agents' heterogeneity is emphasized.

First, Jackson (2008b) examines a generalized version of Chung and Lu's (2002) model in which agents are partitioned into groups, so as to understand how network structures are affected by homophily. He shows that under some regulatory conditions, in the limit of large networks, the average path length of networks is determined only by the number of agents and the (second-order) average degree, and thus in particular is unaffected by other details of the model such as the homophily effect. This invariance seems at odds with our result, and we explain in Section 3.2 why such a difference arises.

Second, the way we take the limits is similar to that of Caldarelli et al.'s (2002) network formation model. They fix the set of types and consider the situation with a sufficiently large number of agents in each type.

[17]We note that König et al. (2014) have a different prediction than ours in terms of assortativity: they predict negative assortativity while we predict positive assortativity.

2. The model

2.1 *Network*

The set $N = \{1, 2, \dots, n\}$ 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 nondirected networks; hence we 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., $q_i(g) = \sharp N_i(g)$.

A *path* between nodes i and j is a sequence of links $(i_1 i_2, i_2 i_3, \dots, i_{K-1} i_K)$ 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, $\mathrm{PL}_{ij}(g)$, 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*, $\mathrm{APL}(g)$, is the average of $\mathrm{PL}_{ij}(g)$'s over all ijs that have finite path lengths.[18]

The *clustering coefficient*, $\mathrm{Cl}(g)$, 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*, $\mathrm{Cl}_i(g)$, for each i with $q_i(g) > 1$, by

$$\mathrm{Cl}_i(g) = \frac{\left|\left\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\right\}\right|}{\left|\left\{jk | k \neq j, j \in N_i(g), k \in N_i(g)\right\}\right|}.$$

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 $\mathrm{Cl}(g) = \frac{1}{|\hat{N}|} \sum_{i \in \hat{N}} \mathrm{Cl}_i(g)$, where $\hat{N}$ denotes the set of agents with degrees more than 1.[19] We will suppress each measure's dependence on g when there is no risk of confusion.

2.2 *Type space and social distance*

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 continuous probability density function f over X.[20] To simplify the analysis, we will assume that f is the uniform distribution. In Section 6(b), we confirm that most of our results do not crucially rely on this assumption,

[18] Thus, strictly speaking, APL is defined only for nonempty networks, $g \neq \varnothing$.

[19] There is another concept of clustering coefficient, *overall clustering*, that does not average over agents' clusterings but over pairs of neighbors: $\mathrm{Cl}(g) = \frac{\sum_i |\{jk \in g | j \in N_i(g), k \in N_i(g)\}|}{\sum_i |\{jk | j \in N_i(g), k \in N_i(g)\}|}$, assuming that the denominator is positive. The 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.

[20] Whenever we refer to distributions over X, it is defined with respect to the usual relative topology.

and in Appendix E.4 in the Supplementary Material, we derive qualitatively similar results under a model with a discrete type space.

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

DEFINITION 1. For every pair of agents i and j, the *kth norm*, $d^{(k)} : X \times X \to \mathbb{R}_+$, measures the distance between them as

$$d^{(k)}(x_i, x_j) = |x_{il} - x_{jl}| \quad \text{such that}$$
$$\sharp\{h : |x_{ih} - x_{jh}| \leq |x_{il} - x_{jl}|\} \geq k \quad \text{and} \quad \sharp\{h : |x_{ih} - x_{jh}| \geq |x_{il} - x_{jl}|\} \geq m - k + 1.$$

Note that this definition boils down to

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

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 the 1st norm, then $d^{(1)}(x_i, x_j) = 0.1$; if we use the 2nd norm, then $d^{(2)}(x_i, x_j) = 0.3$, and so forth.

Considering the situation where agents use social distances to evaluate the values of relationships with others, the interpretation of the kth norm is that if k is large, agents care about many aspects of other's types, while if k is small, then they care about very few aspects of others' types.[21]

We note that the formulation of the kth norm treats each dimension symmetrically, so all the dimensions are equally important. But our main results do not depend too much on the symmetry.[22] By conducting analogous proofs, one can show that our main results are robust even when the lengths of sides of the type space (or, equivalently, the units of measurement of distances) are different across dimensions. Although it would be desirable to have a much more complex notion of social distance, the simple class of distance that we suppose is both tractable and also enough to obtain relevant economic intuition because, for example, it is easy to implement comparative statics. The most important property that drives our results in what follows is the violation of the triangle inequality under $k < m$. We will make this point clear in Sections 3 and 6(f).

[21] One 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 homogeneous measures is enough for that purpose. 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.

[22] Marmaros and Sacerdote (2006) claim that geographic proximity and race are more important determinants of social interaction than are common interests, majors, and family background.

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* $d^{\max}(x_i, x_j) = \max_{1\le h\le m}\{|x_{ih} - x_{jh}|\}$ and the *Min norm* $d^{\min}(x_i, x_j) = \min_{1\le h\le m}\{|x_{ih} - x_{jh}|\}$. We sometimes use the notation $d(x_i, x_j)$, omitting (k), max, or min, when there is no risk of confusion.

2.3 *Pairwise stability and cutoff rules*

An agent's payoff depends on the distance from each of his neighbors and the number of his neighbors. We summarize the former component in a benefit term and the latter in a cost term, as

$$u_i(g) = \Bigg(\sum_{j\in N_i(g)} b\big(d(x_i, x_j)\big)\Bigg) - c(q_i),$$

where $b(\cdot) > 0$ is a weakly decreasing, left-continuous function and $c(\cdot)$ is a strictly increasing function. The term $b(d(i, j))$ represents the benefit that i obtains from link ij when the distance between i and j is $d(i, j)$, and $c(q_i)$ represents the cost that i pays to maintain his q_i links.[23] 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).[24]

We introduce two notions that characterize special classes of networks.

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

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

$$\forall ij \in g,\ u_i(g) \ge u_i(g - ij) \text{ and } u_j(g) \ge u_j(g - ij), \text{ and}$$
$$\forall ij \notin g,\ u_i(g) \le u_i(g + ij) \Longrightarrow u_j(g) > u_j(g + ij).^{25}$$

Pairwise stability is the notion that is proposed by Jackson and Wolinsky (1996). In a pairwise stable network, no single agent can strictly benefit from deleting a link, and no two agents can mutually weakly benefit from adding 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

[23] Our specification supposes that the cost of adding a link does not depend on the distance involved. However, we can easily modify the presentation of the model so that the cost depends on both the degree and the distance, by appropriately adding/subtracting the same amount of some distance-dependent portion from both the benefit and the cost terms.

[24] Note 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.

[25] By convention, we use $g + ij$ for $g \cup \{ij\}$ and $g - ij$ for $g\backslash\{ij\}$.

of simple decision rules that we call cutoff rules. Under such a rule, agents have their own cutoff social distances, above which they do not form links and below which they do form links.

DEFINITION 4. A network g is generated by a *cutoff rule* with $(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_n) \in \mathbb{R}^n_+$ if $ij \in g \iff d(x_i, x_j) \leq \min\{\hat{d}_i, \hat{d}_j\}$.

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*. We call a model in which the network is generated by a cutoff rule a *cutoff rule model*.

Under general cost functions, a pairwise stable network may be neither unique nor efficient.[26] 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.[27] However, if we focus on linear cost functions, a pairwise stable network satisfies all these properties. Furthermore, the converse direction is also true, in the sense made clear in the following lemma.

LEMMA 1. *(i) Suppose that the cost function is linear. 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.*

(ii) 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).

For part (i), 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 that we can also 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 (ii), 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 follow directly from part (i).[28]

Lemma 1 implies that 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 func-

[26]See Jackson (2005) for the discussion on inefficiency of pairwise stable networks.

[27]See Appendix E.1 in the Supplementary Material for an example of a pairwise stable network in which there is no cutoff value profile.

[28]The statements hold under the stronger concept of *farsighted stability* (see Chwe 1994, Page et al. 2005, and Jackson (2008a, Chapter 11.5)).

tions is discussed in Section 5, in which we show that our main results in Section 3 roughly carry over even in such a setting. We also note that the proof of the proposition does not depend on any assumption on d except for symmetry ($d(x_i, x_j) = d(x_j, x_i)$). This is why the statement is independent of the choice of k and m.

Appendix B the Supplementary Material shows that a model with the cutoff rule under the kth norm is equivalent to a model in which the benefit of each link is given by the sum of the equilibrium payoffs from m different games, where the payoff structure of each game depends on the distance between the involved players at each dimension. Under this formulation, the value of k is endogenously determined by the primitives, i.e., the way the distances enter into the payoff functions of those games.

3. Clustering coefficient and average path length

In this section, we consider a cutoff rule model with homogeneous cutoff value profile and analyze how and why different notions of social distances result in different network structures, characterized by the clustering coefficient and average path length. More specifically, we consider the limit of the cutoff rule model as the number of agents goes to infinity and the common cutoff value that we denote by $\hat{d}$ goes to zero. The proof of part (ii) of Lemma 1 shows that this corresponds to the limit of large pairwise stable networks where the benefit is scaled down and/or the marginal cost increases while the benefit is greater than the marginal cost for sufficiently short distances. Lemma 1 shows that the results in Theorems 1 and 2 in this section characterize the features of the networks in such a limit.

3.1 *Clustering coefficient*

In this subsection, we will analyze how the clustering coefficient depends on the property of the notion of 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 by

$$\mathrm{Cl}^* = \lim_{\hat{d}\to 0} \mathrm{Cl}^{\hat{d}} \quad \text{where } \mathrm{Cl}(g) \overset{n\to\infty}{\longrightarrow} \mathrm{Cl}^{\hat{d}} \text{ almost surely.}$$

We consider this limit value as an approximation of a large network. This approximation enables us to obtain an analytical formula of the clustering coefficient that facilitates the comparative statics. In Appendix E.5 the Supplementary Material, we examine the tightness of approximation by deriving the order of $|E[\mathrm{Cl}(g)] - \mathrm{Cl}^*|$. We also provide simulation results on the comparative statics of $E[\mathrm{Cl}(g)]$, so as to clarify the extent to which the results that we provide below are economically meaningful. We will sometimes use the notation $\mathrm{Cl}^*(k, m)$ instead of Cl^* to make it clear that this value depends on the notion of social distance in consideration.

We note that the average degree goes to infinity as $n \to \infty$ for fixed $\hat{d} > 0$. The resulting network becomes approximately regular among the agents in $[\hat{d}, 1 - \hat{d}]^m$, that is, for any fixed pair of agents i and j in $[\hat{d}, 1 - \hat{d}]^m$, the fraction $\frac{q_i}{q_j}$ converges to 1 almost surely

as n goes to infinity.[29] In Section 4, we discuss several ways to obtain heterogeneous degree distributions by extending the model and how robust the asymptotic results are under the finite-agent case.

Notice the order of the limits. 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.

In the following theorem, we solve for this limit clustering for every pair of k and m.

THEOREM 1. *For each m and $k \leq m$,*

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

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: (i) those who are close to i only with respect to the first dimension, (ii) those of only second dimension, and (iii) those of both dimensions. As the cutoff goes to zero, the probabilities of agent j being the first, second, and third class, conditional on the event that he is agent i's neighbor, converge to $1/2$, $1/2$, and 0, respectively. This implies that the probability that two randomly chosen neighbors of i are of the same class converges to $1/2$. If two neighbors are in the same class, then the probability that they are connected to each other approximates (as $\hat{d} \to 0$) the probability with which the distance between two randomly chosen points in a unit interval is no more than 0.5. This probability is $3/4$. If two neighbors are in different classes, then they are connected to each other with probability converging to zero as $\hat{d} \to 0$. All in all, the clustering coefficient converges to $1/2 \cdot 3/4 = 3/8$ as $\hat{d} \to 0$.

The first corollary of Theorem 1 is the following comparative statics.

COROLLARY 1. *(i) For any k, $\mathrm{Cl}^*(k,m)$ is decreasing in $m \geq k$.*

(ii) For any $m \geq 2$, $\mathrm{Cl}^*(k+1,m) > \mathrm{Cl}^*(k,m)$ if $k > \frac{4}{7}m - \frac{3}{7}$, and $\mathrm{Cl}^*(k+1,m) < \mathrm{Cl}^*(k,m)$ if $k < \frac{4}{7}m - \frac{3}{7}$.

We note that part (ii) of this corollary implies a possibility for a nonmonotonic pattern of $\mathrm{Cl}^*(k,m)$: it is decreasing in k when k is small, reaches its minimum, and then is increasing when k is large.

According to the first part of Corollary 1, if the number of dimensions of the type space becomes large with fixed k, then the resulting network becomes less cliquish. 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, given that the number of relevant dimensions are not too few. As we

[29] For any $\hat{d} < 1$, the probability that the resulting network is complete goes to zero as $n \to \infty$.

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 $k+1$.

We note that, in general, the comparative statics in k or m does not necessarily mean involving different populations but also comparing different layers of networks for a fixed population. Such multilayered network data are often available and analyzed in practice (Kremer and Miguel 2007, Conley and Udry 2010, Baccara et al. 2012). In particular, when different networks are defined based on different levels of frequency or strength of each interaction, it would be reasonable to think that the set of relevant dimensions is the same. This makes it easier to interpret the comparative statics in k. For example, Rapoport and Horvath (1961) consider comparisons of different layers of networks for a fixed population. They observe that networks with links of stronger friendships involve more overlapping relationships (see footnote 30 for its details). Assuming that k is not too small, the second part of Corollary 1 can also explain this empirical evidence.[30]

The following comparison of two extreme cases — the Max norm and the Min norm — is instructive.

Corollary 2. *If* $1 < m < 9$, $\mathrm{Cl}^*(m,m) > \mathrm{Cl}^*(1,m)$ *holds. That is,* Cl^* *is higher with the Max norm than with the Min norm.*

We omit the proof, which is clear from the formula given in Theorem 1.

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. In Section 6(f), we explicitly use this fact to derive a lower bound of Cl^*. For the Min norm, however, such an argument is not possible because the triangle inequality is not satisfied, so it is possible that i's neighbors are quite far away from each other. Indeed, in Example 4 in Appendix E.3 in the Supplementary Material, we provide a natural extension of the Min norm to show that the clustering coefficient can be arbitrarily small when the triangle inequality is violated. This logic gives us a possible explanation as to why the e-mail network of Ebel et al. (2002) has a low clustering coefficient.

[30] Rapoport and Horvath (1961) analyze the survey data collected at a junior high school in the 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 lth and the $(l+1)$th friends, for various values of l. They find that θ, a parameter similar to the 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 that an agent's closer friends are similar to that agent in terms of more aspects than farther friends are (an admittedly arguable but reasonably natural assumption to make), this result is consistent with our result that Cl^* is increasing in k for values of k close to m.

3.2 *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

$$\mathrm{APL}^* = \lim_{\hat{d}\to 0} \mathrm{APL}^{\hat{d}}, \quad \text{where } \mathrm{APL}(g) \overset{n\to\infty}{\longrightarrow} \mathrm{APL}^{\hat{d}} \text{ almost surely.}$$

Again, the order of the limits 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. As the cutoff goes to zero with the number of agents fixed, all the pairs of agents would have the path length ∞ almost surely. Analogously to the analysis of clustering coefficient, in Appendix E.5 in the Supplementary Material we examine the tightness of approximation. We use APL$^*(k,m)$ as before.

For any $a \in \mathbb{R}$, denote by $\lceil a \rceil$ the minimum integer that does not exceed a. The following result gives the formula of APL* for the kth norm with $k < m$.

THEOREM 2. *(i) For all k and m such that $k < m$,* $\mathrm{APL}^*(k,m) = \lceil \frac{m}{m-k} \rceil$.

(ii) For all m, $\mathrm{APL}^*(m,m) = \infty$.

To see the intuition for part (i), consider the case of $m = 5$ and $k = 3$, and suppose for a moment (only in this paragraph) that we are dealing with a 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 $x_1 = (0.7, 0.7, 0.3, 0.3, 0.3)$ and $x_2 = (0.7, 0.7, 0.7, 0.7, 0.3)$, 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 $\lceil \frac{5}{5-3} \rceil = 3$ steps.

The proof for part (ii) is simple: 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.

There is a striking difference between the kth 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 part (ii) is true also in the case of Euclidean 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 Euclidean norm), but not by the kth 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, suggesting that path lengths in networks tend to be large in such cases.

As a corollary of this theorem, we have the following comparative statics.

COROLLARY 3. *(i) For all k,* $\mathrm{APL}^*(k,m)$ *is weakly decreasing in m.*

(ii) For all m, $\mathrm{APL}^*(k,m)$ *is weakly increasing in k.*

The proof is straightforward from the formula given in Theorem 2, hence, it 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). The result is intuitive: If sharing only a small number of aspects from many is sufficient to form a relationship, then each agent has neighbors who have many aspects of characteristics that are different from his. Thus, it is easy to have access to agents with very different characteristics through the network. The result implies that the introduction of new communication technology, which would increase the number of relevant dimensions (m), makes a network closely connected through indirect paths. Also, since the e-mail network of Ebel et al. (2002) is expected to have small k as explained earlier, the result predicts that it has a low average path length, consistent with data.

This argument shows that the way we can find indirect paths is crucially affected by k and m. This is the reason for the difference from the result in Jackson (2008b) that we mentioned in Section 1.1. His model assumes that some nonvanishing fraction of links among different categories is formed, implying that we can find short indirect paths as in Chung and Lu's (2002) model where types of agents are irrelevant. In our model, however, the probability that two agents are linked can be zero, depending on k and m. Hence, k and m are crucial determinants of the way indirect paths can be found.

Watts et al. (2002) consider a model that shares a similar spirit to ours, studying stochastic network formation and search on networks. They demonstrate by simulations that if agents use only one of many "social dimensions" to assess the relationships to others, the network is more likely to be "searchable." Searchability here means that their search process from a randomly chosen agent to another randomly chosen agent is complete with a high probability as in the small-world experiment of Travers and Milgram (1969). Our focus is different in that we aim to characterize the network structure and to understand how it varies with social distance. For this reason, we adopted a different network formation that allows for rich comparative statics.[31] Since the models are different and we obtain analytical results on network structures while they obtain results on searchability by simulations, it is difficult to compare their results with ours at a formal level.

Except for the case under the Max norm, the networks generated in our model are consistent with empirically observed networks with 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.[32] This is a well

[31] For instance, they show that searchability is nonmonotonic and eventually decreasing in the number of social dimensions. This is seemingly inconsistent with our result that the average path length is monotonically decreasing in m. They claim that the decreasingness is due to the fact that the network formation process becomes more "random," and this aspect of the model is orthogonal to our main argument on the violation of the triangle inequality. By taking the limit as the number of agents goes to infinity, we isolate this "random network" effect, providing an analytical foundation for their reasoning of the nonmonotonicity. The tractability of our model enables us to obtain further results: We characterize the clustering coefficient, and show that our results hold for a wide range of degree distributions. Also, we can conduct comparative statics with respect to k, where it is not clear how one can incorporate the idea of $k > 1$ in the model of Watts et al. (2002). Finally, our model is utility-based, so we can conduct welfare analysis, as we do in Lemma 1.

[32] Precisely, 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, resulting in a very low clustering coefficient.

observed property in a variety of networks in reality. The existing network formation models generating the small-world property in the literature include the "rewiring" process Watts and Strogatz (1998), hub nodes (Barabási and Albert 1999), and partitioning of agents into several groups Jackson and Rogers (2005). Our model can be seen as giving an alternative explanation for the small-world property, which depends on the multidimensionality of the type space.

4. Cutoff heterogeneity and other stylized facts

In the main section, we assume homogeneous cutoff values for simplicity. In this section, we extend the model by allowing cutoffs to be heterogeneous across agents. This allows us to accommodate empirical regularities such as skewed degree distribution, clustering-degree correlation, and positive assortativity. Newman (2003) reviews these three empirical regularities in social networks, and recent studies by Ugander et al. (2011) and Wilson et al. (2012) find them in Facebook networks. Goyal et al. (2006) also find clustering-degree correlations in co-authorship networks.

In this section, as an approximation of a model with a large finite population, we restrict attention to a model with a continuum of agents to avoid technical complications. There is a continuum of agents with measure 1. Each agent i is associated with her type x_i and *relative cutoff value* θ_i. Distributions of x_i and θ_i are independent. The distribution of x_i is uniform as in the main model. The value θ_i follows a cumulative distribution G whose support is a subset of $[0, 1]$ and includes 1. Fix $\hat{d} > 0$. Each agent i has cutoff value $\hat{d}_i = \theta_i \hat{d}$, that is, agents i and j are connected if and only if $d(x_i, x_j) \leq \min\{\theta_i \hat{d}, \theta_j \hat{d}\}$. In this setup, the number of each agent's neighbors is infinite. For this reason, in this section we use the term "degree" to refer to the measure of the set of neighbors rather than its cardinality.

4.1 *Clustering coefficient and average path length*

First we show that the results about clustering coefficient and average path length in Section 3 extend to the case with general distribution of cutoffs.

Fix k and m. Consider two agents i and j such that $\theta_i = \theta_j$ and $x_i, x_j \in [2\hat{d}, 1 - 2\hat{d}]^m$. By symmetry, their clustering coefficients are equal. Let this value be $\mathrm{Cl}^{\hat{d}}(\theta_i, k, m)$. Define

$$\mathrm{Cl}^*(\theta_i, k, m) = \lim_{\hat{d} \to 0} \mathrm{Cl}^{\hat{d}}(\theta_i, k, m).$$

Also, let $\mathrm{APL}^{\hat{d}}(k, m)$ be the average path length given k and m,[33] and let

$$\mathrm{APL}^*(k, m) = \lim_{\hat{d} \to 0} \mathrm{APL}^{\hat{d}}(k, m).$$

[33] Formally, given a tuple $(\theta_i, \theta_j, x_i, x_j)$, and $\hat{d}$, k, and m, the path length between agents i and j given $\hat{d}$, k, and m denoted by $\mathrm{PL}^{\hat{d},k,m}(i, j)$ is uniquely determined. Then, $\mathrm{APL}^{\hat{d}}(k, m)$ is the expectation of $\mathrm{PL}^{\hat{d},k,m}(i, j)$ with respect to two independent draws of θ_i and θ_j from G and two independent draws of x_i and x_j from the uniform distribution. A formal definition of $\mathrm{Cl}^{\hat{d}}(\theta_i, k, m)$ can be given in an analogous manner.

The proof of the following result is based on the same arguments as in Theorems 1 and 2 and thus is omitted.

THEOREM 3. *(i) There exists* $h : \mathrm{supp}(G) \setminus \{0\} \to (0,1)$ *such that, for all* $\theta_i \neq 0$ *and for all* m *and* $k \leq m$,

$$\mathrm{Cl}^*(\theta_i, k, m) = \binom{m}{k}^{-1} \big(h(\theta_i)\big)^k.$$

(ii) For all k *and* m *such that* $k < m$, $\mathrm{APL}^*(k,m) = \lceil \frac{m}{m-k} \rceil$.

(iii) For all m, $\mathrm{APL}^*(m,m) = \infty$.

Notice that the function h does not depend on k or m. The theorem shows the robustness of the results obtained in the analysis of the homogeneous cutoff rule model. In particular, it shows that the comparative-statics results in Corollaries 1, 2, and 3 generalize with minor modifications.

4.2 *Skewed degree distribution*

Skewed degree distributions are often found empirically. In particular, the literature has found scale-free distributions extensively in the real world.[34] Here we show that, under heterogeneous cutoff values, any "relative degree" distribution can be approximated by tuning the distribution of cutoff values.

Let $\hat{X} = [\hat{d}, 1-\hat{d}]^m \subset X$. Take agent i with type $x_i \in \hat{X}$ and relative cutoff value θ_i. Let $D(\theta_i)$ denote the measure of such i's neighbors. Because the type distribution is uniform, D depends only on θ_i and is independent of x_i. We write it as a function of θ by $D(\theta)$. We define the *relative degree* $\hat{D}(\theta_i)$ of agent i with relative cutoff value θ_i by

$$\hat{D}(\theta_i) = \frac{D(\theta_i)}{D(1)}.$$

A *relative degree distribution* H is the cumulative distribution of relative degrees among agents in $\hat{X}$. Here we focus only on the agents in $\hat{X}$ for simplicity. Note that if cutoff values are homogeneous, then H is concentrated on 1.

For example, when θ is uniformly distributed over $[0,1]$, i.e., $G(t) = t$, then the relative degree is computed as $\hat{D}(\theta) = \theta^k(k(1-\theta)+1)$. This is decreasing in k. This means that, when k is higher, there is more degree inequality between agents with higher θ and lower θ.[35]

[34] Scale-free distributions, which were originally discovered by Pareto (1896), are observed in a variety of networks. Pareto (1896) finds that wealth distribution in Italy had the scale-free feature. Note that scale-free distributions are often found as a property of the tail of degree distributions, so our condition on the minimum relative degree (i.e., the requirement that there exists $\underline{r} > 0$ in Proposition 1) does not contradict the scale-free properties.

[35] For instance, this increases the Gini index of the relative degree distribution because it lowers the Lorenz curve pointwise.

The following result shows that any relative degree distribution H can be approximately achieved by some relative cutoff distribution G.

PROPOSITION 1. *Fix any $\epsilon > 0$ and a cumulative distribution function H. There is a distribution G of θ under which $\tilde{H}$ is the relative degree distribution with* $\sup_{r\in[0,1]} |H(r) - \tilde{H}(r)| \leq \epsilon$.

In the proof, we first construct $\tilde{H}$ that has finite support and is no more than ϵ away from H. Then we construct a distribution G of θ under which $\tilde{H}$ is the relative degree distribution. This G has finite support and can be found by an iterative procedure that defines appropriate θs in the support from the smallest to the largest.

4.3 *Positive assortativity*

One stylized fact about social networks is that they exhibit positive assortativity: an agent's neighbors tend to have similar degrees to that of the agent. Here we show that the introduction of cutoff heterogeneity generates a pattern of assortativity that is consistent with such a fact. The intuition is simple. First, degrees are increasing in cutoffs. That is, if agent i has a higher degree than agent j, then i has a higher cutoff than j. To make the argument simple, suppose that i and j are at the same location, x. This means that i's neighbors are a superset of j's, and the additional neighbors are the ones who are sufficiently far away from x so that j cannot be a neighbor. That is, those neighbors have higher cutoffs than the cutoffs of j's neighbors. This means that there is a lower bound on the cutoffs of all additional friends that i has, which suggests that i's neighbors have on average higher cutoffs than j's neighbors, implying higher degrees. The result below formalizes this intuition.

Fix any location $x \in [2\hat{d}, 1-2\hat{d}]^m \subset X$. Let H_d denote the cumulative distribution function of the relative degrees of the neighbors of an agent at x with cutoff $d \leq \hat{d}$.

PROPOSITION 2. *Suppose that G admits a density. Then $H_d(t)$ is nonincreasing in d for all t. In other words, H_d weakly first order stochastically dominates $H_{d'}$ if $d > d'$.*

4.4 *Clustering-degree correlation*

In this subsection we show that the model with heterogenous cutoffs generates a prediction that is consistent with another stylized fact that clusterings are negatively correlated with degrees. Since relative degrees are strictly increasing and continuous in cutoffs for agents in $[2\hat{d}, 1-2\hat{d}]^m$, it suffices to show that $\mathrm{Cl}^*(\theta_i, k, m)$ is decreasing in θ_i.

PROPOSITION 3. *Assume that G is the uniform distribution over $[0,1]$. For all k and m with $k \leq m$, $\mathrm{Cl}^*(\theta_i, k, m)$ is strictly decreasing in θ_i.*

The cases with general distribution of cutoffs are cumbersome and we do not delve into these cases, but summarize key intuitions: There are two effects of an agent i having a high degree or, equivalently, a high cutoff. Consider the additional agents who

can be newly connected to agent i by a marginal increase of i's cutoff. The first effect is that these new agents have the highest expected cutoffs because they are the furthest from i. This suggests that the new agents should contribute to raising i's clustering. The second effect is that the new agents are located at the edge of the set of i's neighbors, which is the most difficult situation for other neighbors of i to be connected to those new agents. This suggests that the new agents should contribute to lowering i's clustering. The result shows that, with the uniform distribution of relative cutoff values and $m = k = 1$, the second effect dominates the first. The reason is that a link can be formed between two agents only if *both* agents' cutoffs are sufficiently high. The first effect pertains only to the cutoff of one agent, so it is not enough to increase agent i's clustering, while the second effect says that one agent is far from others, so it results in a decrease in clustering.

5. Nonlinear cost function

In this section, we consider situations in which the cost functions are not necessarily linear. Nonlinear cost functions may arise under certain situations such as networks on social network services (SNS) in which the additional cost of forming a link may be decreasing or friendship networks in which the additional cost can be increasing because each person has a limited amount of time to spend with friends. We explain the nontriviality associated with such an extension, demonstrate a way to overcome it, and provide sufficient conditions under which our results with linear cost functions are (approximately) robust to this extension.

With a nonlinear cost function, pairwise stability may not determine a unique network structure. Under a nonlinear cost function, i's marginal cost to connect with j depends on the number of links i already has, and there may exist multiple pairwise stable networks. We defer the complete description of concrete examples with multiplicity to Appendix E.1 in the Supplemental Material; intuitively, with concave cost functions, the average cost per link is small when an agent has many links, while she can pay zero cost by not connecting with anyone. For this reason, in an example in the Supplemental Material, an empty network as well as the complete network are both pairwise stable. With convex cost functions, additional cost to connect is high, so there is a bound on the number of links that each agent can have. Depending on to whom an agent is connected, there exist multiple pairwise stable networks. In another example in the Supplemental Material, each agent is involved in exactly one link, and the cost for the second link is so high that no one wants to deviate without severing an existing link. For a similar reason, a pairwise stable network is not necessarily generated by a cutoff rule even with a heterogeneous cutoff value profile.

Despite such difficulty with pairwise stability, 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 3.

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

$(ij \in g' \wedge ij \notin g) \Longrightarrow i, j \in S$ and

$(ij \in g \wedge ij \notin g') \Longrightarrow \{i, j\} \cap S \neq \varnothing$.

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 5. 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 $(\exists j \in S \text{ s.t. } u_j(g') < u_j(g))$.

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

PROPOSITION 4. *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.*

In the proof, we construct 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 pairwise (strong, for the case of convex or linear c) stable network. On such a network, we can find a cutoff value profile, while it may not be homogeneous among the agents.

Now we examine the extent to which cutoff values can be heterogeneous when the number of nodes is very large. 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 5. *Suppose that b is continuous and strictly decreasing, and that for some $c_1 > 0$, $\lim_{d\to 0} b(d) > c_1$ and $\lim_{q\to\infty} \Delta c(q) = c_1 > 0$ hold. Then the cutoff value profile for a pairwise stable network, $(\hat{d}_1, \ldots, \hat{d}_n)$, is such that*

$$\min_{i\in N} \hat{d}_i \overset{n\to\infty}{\longrightarrow} \hat{d} \quad \textit{almost surely} \quad \textit{and} \quad \max_{i\in N} \hat{d}_i \overset{n\to\infty}{\longrightarrow} \hat{d} \quad \textit{almost surely,}$$

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

For each agent, for a sufficiently large number of nodes, there are sufficiently many neighbors in his δ-neighborhood. The agent has to be connected with them in a pairwise stable network. This implies that he has a sufficiently large degree and, hence, the cost function is almost linear when he decides whether 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.

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

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

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

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) pairs that satisfy the assumptions in Proposition 5 such that the corresponding $\hat{d}$ converges to zero, due to the analogous argument as in part (ii) of Lemma 1; hence, the requirement of $\hat{d} \to 0$ 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 6. *The following two equalities hold:* $\mathrm{Cl}^*_{\mathrm{hetero}} = \mathrm{Cl}^*$ *and* $\mathrm{APL}^*_{\mathrm{hetero}} = \mathrm{APL}^*$.

In the proof of clustering coefficient, given $\hat{d}$ and ϵ, we obtain an upper bound and a lower bound of $\mathrm{Cl}^*_{\mathrm{hetero}}$ by slightly modifying the calculation in the proof of Proposition 2. Then we show, for any $\hat{d}$, that these values approach Cl^* as ϵ goes to zero. The proof for the average path length runs parallel to that of clustering; hence, it is omitted.

Summing up, our main results are almost unchanged under the condition that the heterogeneity of the cutoff values is almost negligible relative to the cutoff values themselves. Combined with Proposition 4, our results in Section 3 carry over even in the case of nonlinear cost functions provided that they approximate linear functions.

6. DISCUSSIONS

(a) *Interpretations of the contribution.* There are at least two possible ways to interpret our contribution. The first interpretation is that the analysis gives us insight into how the underlying mechanism of network formation affects prominent measures of network structures, namely the clustering coefficient and the average path length. The closed-form solutions of these measures enabled us to conduct neat comparative statics, but it was obtained in the limit as $n \to \infty$. Thus the theory was obtained at the cost of having very large average degree, which can be at odds with real networks. We note two remarks on this interpretation. First, although the average degree must be very large, we can accommodate differences in the average degrees across networks by having different speeds of $\hat{d}$ going to zero, and also approximately accommodate all relative degree distributions (Section 4.2). Second, in Appendix E.5 in the Supplementary Material, we provide an approximation result that identifies the exact convergence rate of the two measures as $n \to \infty$, and use simulation to further understand the extent to which the comparative statics is applicable for finite n.

The second interpretation is to see our contribution as giving a guide to which parameter in the model is informative of which measure of network structures. Let us be more specific on this alternative interpretation. Note first that, given the population size and type distribution, three variables, k, m, and $\hat{d}$ completely determine the probability distribution over possible network structures. When the population size is large, however, the clustering coefficient and average path length is almost deterministic, and in such a situation we can approximately pin down the values of k and m (Theorems 1 and 2 and the approximation results in Appendix E.5 in the Supplementary Material). That is, by using the data of the clustering coefficient and the average path length, we can solve for the underlying k and m. Then, by using the data of the degree distribution, we can pin down the underlying $\hat{d}$. This procedure gives us the three key variables (k, m, and $\hat{d}$), allowing us to use them to discuss counterfactuals, i.e., what would happen if the type distributions were different. This is especially important because the type distributions may be (at least partially) under the control of the designer of the market.

(b) *Interpretation of the type space.* The parameter m is the number of dimensions that are "relevant" for network formation. We assume that this parameter is a sufficient statistics to capture the difference in any two type spaces, although we allow for the set of relevant dimensions across different networks to be different. Note that this is partly an assumption, but also partly a result. For example, one can imagine that different dimensions can be associated with different cutoffs. Our result shows that as long as all cutoffs are sufficiently small, even if they are heterogeneous across dimensions, the clustering coefficients and the average path lengths are close to their limit values; thus our comparative statics still applies.

In general, it can be difficult to determine the value of m for each situation, which may be viewed as a weakness of our approach. Nonetheless, we choose to include this as a parameter of the model because it helps us understand the nature of social distance and network formations at a conceptual level by conducting comparative statics.[36] For example, in Section 3, we discussed the effect of the introduction of communication technology. Also, variations in m lead to more flexible predictions that are more convenient for fitting empirical data.

(c) *Robustness to non-uniform type distribution.* Here we briefly discuss the effect of distributional changes on the implied limit clustering coefficients and average path lengths.[37] First, it is straightforward to see that the results on the limit average path length do not depend on the uniformity of the distribution.

PROPOSITION 7. *The conclusions of Theorem 2 and Corollary 3 hold for any full-support type distribution over X that has a probability density function f.*

[36]If we interpret m as the number of all conceivable social characteristics and assume that it is fixed across all the situations, then the model exhibits unrealistic descriptions of network formations. For example, political taste, which might be one of relevant aspects in friendship networks, does not seem to play a significant role in co-authorship network formations.

[37]As shown in an earlier version of the paper (Iijima and Kamada 2015), a wide range of relative degree distributions can be attained in the model by allowing for non-uniform type distributions.

The results on the clustering coefficients, however, need not generalize, although the main insights still carry over. First, the exact values of Cl* are no longer valid with a general distribution f. To grasp the intuition, consider the case with $m = 2$ and $k = 1$. For any point x in the interior of X, the set of its neighbors is a union of the set of the agents who are close to x with respect to the first dimension and that of those who are close to x with respect to the second dimension. One implication of the uniform distribution is that, when $\hat{d} > 0$ is sufficiently small, these two sets have approximately the same expected number of agents. With a non-uniform distribution, however, this is not the case, and as a consequence the clustering coefficient becomes higher. The intuition is that if there are more agents in one set than the other, those agents in the first set are likely to form clusters, raising the clustering coefficient. In Appendix E.2 in the Supplementary Material, we provide sufficient conditions under which the comparative-statics results of the limit clustering coefficient under the uniform-distribution model generalize. Specifically, we introduce a measure of asymmetry for each distribution and show that the comparative-statics results are valid when a distribution is not too asymmetric. The upper bounds on asymmetry are easy to compute, and we provide them for various values of m and k.

(d) *Network centrality*. Centrality measures of networks have attracted significant attentions since they can be used to characterize relative importance of agents in various economic settings.[38] Under non-uniform type distributions, our model can be used to examine how centrality interacts with the nature of social distance. While its full analysis is beyond the scope of the current paper, let us provide a simple example as an illustration. Suppose that the types are distributed according to F with $m = 2$, and the cutoff value is $\hat{d} > 0$. Let us employ betweenness as a measure of centrality. Conditional on agent i's type being x_i, her expected centrality depends only on x_i, and we denote it by $C(x_i)$. It is computed as

$$C(x_i) = \mathbb{E}_F\left[\sum_{j\neq i\neq k} \frac{\mathrm{SP}(i; jk)}{\mathrm{SP}(jk)}\Big| x_i\right],$$

where $\mathrm{SP}(i; jk)$ is equal to 1 if i constitutes a shortest path between agents j and k, and 0 otherwise, and $\mathrm{SP}(jk)$ denotes the number of shortest paths between j and k (we used the convention $\frac{0}{0} = 0$). *In this setup, sets of agents having high centralities are different for different k's.* To see this point, suppose that $\hat{d} < 0.6$ and the type distribution F admits two mass points $(0.2, 0.2)$ and $(0.8, 0.8)$ with measure $p < 0.5$, and the remaining agents of $1 - 2p$ are distributed uniformly over the type space. Take agents i and i' with $x_i = (0.5, 0.5)$ and $x_{i'} = (0.8, 0.2)$. Then we can show that there exist $\bar{p} < 0.5$ and $\bar{n} < \infty$ such that for all $p > \bar{p}$ and $n > \bar{n}$, $C(x_i) > C(x_{i'})$ under $k = 2$ but $C(x_i) < C(x_{i'})$ under $k = 1$. This is because, under the Max norm, agent i constitutes shortest paths between agents at $(0.2, 0.2)$ and agents at $(0.8, 0.8)$ but not under the Min norm. The converse property holds for agent i'.

(e) *Finite-agent approximation*. In the main section we focus on characterizing the limit values of the clustering coefficient and average path length as $n \to \infty$ and $\hat{d} \to 0$. In Appendix E.5 in the Supplementary Material, we consider the case of fixed finite n

[38] Zenou (2016) provides a survey on this topic.

and positive $\hat{d}$, and obtain bounds of the deviations of the expected values of clustering coefficient and average path length from the limit values obtained in the main section. We find that the bound of the clustering coefficient does not depend on n, because the probability that i's two neighbors j and k are connected with each other does not depend on the positions of other agents. Although the average path length depends on n in general, we conduct simulations to confirm that our main results on both clustering coefficient and average path length are robust under broad parameter combinations with finite n.

(f) *Discrete type space*. In the main part of this paper, we assumed that agents are distributed over the type space $[0,1]^m$ according to a strictly positive density. In practice, some dimensions may be better described by discrete variables. In Appendix E.4 in the Supplementary Material, we consider the case with a discrete type space and show that our main qualitative results go through in such a model.

(g) *Role of triangle inequality*. Under a more general class of social distance than the kth norm, we can formalize the role of the standard-norm axioms by providing a strictly positive lower bound of the limit clustering coefficient. Specifically, define social distance by $d(x_i, x_j) = \|x_i - x_j\|$, where $\|\cdot\|$ is a standard norm in $\mathbb{R}^m$. That is, for all $\alpha \in \mathbb{R}$ and $y, y' \in \mathbb{R}^m$, it satisfies (i) absolute homogeneity, i.e., $|\alpha|\|y\| = \|\alpha y\|$, (ii) triangle inequality, i.e., $\|y + y'\| \leq \|y\| + \|y'\|$, and (iii) separates points, i.e., if $\|y\| = 0$, then y is the zero vector.

Fix a type x_i of agent i in $(0,1)^m$ and compute a lower bound of the probability that her two neighbors j and k are connected to each other when $\hat{d} \to 0$. Because of the triangle inequality $d(x_j, x_k) \leq d(x_i, x_j) + d(x_i, x_k)$, j and k are connected to each other if $d(x_i, x_j) + d(x_i, x_k) \leq \hat{d}$. To derive a lower bound, we first compute the distribution of the variables $\frac{d(x_i,x_j)}{\hat{d}}$ and $\frac{d(x_i,x_j)}{\hat{d}}$ conditional on the event that $d(x_i, x_j), d(x_i, x_k) \leq \hat{d}$, using absolute homogeneity and separates points. We then derive the distribution of the variable $q := \frac{d(x_i,x_j)+d(x_i,x_k)}{\hat{d}}$ conditional on the same event. This distribution can be used to show that $\int_0^1 mq^{m-1}(1-q)^m\,dq$ is a lower bound of the limit clustering coefficient, which is strictly positive. The detailed proof is given in Appendix E.3 in the Supplementary Material, where we derive a better bound.

However, such a lower bound cannot be established if we relax the standard-norm axioms, and the limit clustering coefficient can be arbitrarily low in general. In Appendix E.3 we generalize the Min norm to a class of social distance. Specifically, for any $w > 0$, we can find social distance $d(\cdot,\cdot)$ in that class such that the limit clustering coefficient is lower than w.

7. 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 multidimensional 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 us-

ing the kth norm, in which the distance is the kth smallest dimension-wise distance. We characterized average path length and clustering coefficient in stable networks, and analyzed how they are related to the way social distances are measured by agents.

One implication of our result is that the introduction of new communication technology makes a network closely connected but cliquish. Our model is utility-based, which allowed us to conduct welfare analyses. We also showed that the assumption of a linear cost function is not essential to our result by replacing the notion of pairwise stability with that of strong stability. The Supplementary Material contains a number of additional results to check the robustness of our prediction, including a model with a discrete type space, a stochastic link-formation model, and an analysis of convergence speeds.

We conclude this paper by explaining how the paper could serve as a basis for future works. First, this paper introduced a model of multidimensional type space and a tractable class of 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 the similarities 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 it is not infeasible if we restrict a class of preferences in a tractable manner.

Appendix A: Proofs of the main results

This appendix contains the proofs of our main results. Proofs of other results can be found in the Supplementary Material.

A.1 *Proof of Theorem 1*

Proof. Let the set of points sufficiently away from the boundary be $X(\hat{d}) = \{x_i \in X : 0 < x_{ih} \pm \hat{d} < 1, h = 1, \ldots, m\}$. Take a point x in $X(\hat{d})$ and consider a hypothetical agent situated at that point, named agent i. We will ignore the possibility of a tie in distances, as it does not occur almost surely; hence, it does not affect the result. Now let $B_{\hat{d}}(x)$ be the (closed) $\hat{d}$-neighborhood of point x with respect to the kth norm.

Fix $\hat{d} > 0$ and consider an increasing sequence of agents $N(n)$ and corresponding networks, $g(n)$, $n = 1, 2, \ldots$, such that $N(n) = \{1, 2, \ldots, n\}$ and $g(n) \subseteq g(n+1)$. For each

n such that $q_i(g(n)) \geq 2$, define a probability measure μ_n over $B_{\hat{d}}(x)$ by

$$\mu_n(\{x\}) = \begin{cases} \dfrac{1}{q_i(g(n))} & \text{if there is an agent (other than } i\text{) whose type is } x, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\nu_n = \mu_n \times \mu_n$, defined over $B_{\hat{d}}(x_i) \times B_{\hat{d}}(x_i)$, denote the product measure. Also, let ν^* denote the uniform distribution over $B_{\hat{d}}(x_i) \times B_{\hat{d}}(x_i)$. Glivenko–Cantelli's theorem implies that ν_n weakly converges to ν^* almost surely.[39]

Note that the clustering of i is determined by ν_n, i.e., this is the probability that two points that are independently chosen (by ν_n) are within $\hat{d}$ distance, given that the two points are different. Thus we have

$$\begin{aligned} \mathrm{Cl}_i(g(n)) &= \frac{q_i(g(n))}{q_i(g(n)) - 1} \nu_n\big(\{(x', x'') | d^{(k)}(x', x'') \leq \hat{d}\}\big) - \frac{1}{q_i(g(n)) - 1} \\ &\overset{\text{a.s.}}{\to} \nu^*\big(\{(x', x'') | d^{(k)}(x', x'') \leq \hat{d}\}\big) \quad (n \to \infty), \end{aligned}$$

where the convergence is ensured by the fact that the set $\{(x', x'') | d^{(k)}(x', x'') \leq \hat{d}\}$ is a ν^*-continuity set. Hence we can compute this limit clustering based on the assumption that there is a continuum of agents uniformly distributed over $B_{\hat{d}}(x_i)$.

Now note that we have

$$\begin{aligned} \mathrm{Cl}^* &= \lim_{\hat{d} \to 0} \lim_{n \overset{\text{a.s.}}{\to} \infty} \left[\frac{1}{n} \left(\sum_{x_i \in X(\hat{d})} \mathrm{Cl}_i(g) + \sum_{x_j \in X \setminus X(\hat{d})} \mathrm{Cl}_j(g) \right) \right] \\ &= \lim_{\hat{d} \to 0} \lim_{n \overset{\text{a.s.}}{\to} \infty} \left[\frac{1}{n} \left(\sum_{x_i \in X(\hat{d})} \mathrm{Cl}_i(g) \right) \right], \end{aligned}$$

since $\frac{\mathrm{vol}(X(\hat{d}))}{\mathrm{vol}(X)} \to 1$ as $\hat{d} \to 0$, where $\mathrm{vol}(\cdot)$ denotes the volume of a set (with respect to the Euclidean norm) and $\mathrm{Cl}_j(g)$ takes only a finite value (a value in $[0, 1]$) for any $j \in N$.

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

Take another agent l, whose type is x'', and let $Z(x, x') = \{x'' \in B_{\hat{d}}(x) | \{h | x_{ih} - x''_{lh} \leq \hat{d}\} = \{h | x_{ih} - x'_{jh} \leq \hat{d}\}\}$. Notice that $\frac{\mathrm{vol}(Z(x,x'))}{\mathrm{vol}(B_{\hat{d}}(x))} \to \binom{m}{k}^{-1}$ as $\hat{d} \to 0$. Now, it is straightforward to see that the probability that $x'' \in B_{\hat{d}}(x) \setminus Z(x, x')$ is connected to x' goes to 0 as $\hat{d}$ goes to 0. Thus we only need to consider x''s in $Z(x, x')$. Hence, the probability that j and l are connected is equal to the probability that the projections of x' and x'' on the restricted k-dimensional space with dimensions in $\{h | x_{ih} - x'_{jh} \leq \hat{d}\}$ are within the distance $\hat{d}$ with respect to the kth norm.

[39] See, for example, Hildenbrand (1974, p. 52).

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 $\binom{m}{k}^{-1}\cdot(\frac{3}{4})^k$. Note that this value is independent of x_i. Hence the desired value (which is the average of Cl_is) takes this value as well almost surely. □

A.2 *Proof of Theorem 2*

Part (i). Take arbitrarily two agents i and j. Almost surely, $d^{\min}(x_i,x_j)>0$. Hence, we restrict attention to the case of $d^{\min}(x_i,x_j)>0$. Fix $\hat{d}>0$ at a value such that $\hat{d}<\frac{1}{2}d^{\min}(x_i,x_j)$. It suffices to show that the almost sure limit of PL_{ij} becomes exactly the same as the formula in the statement of the proposition.

First, we show that the formula is an upper bound of PL_{ij}. Consider a class of sets defined by

$$\beta(t)=\left\{h\in N:|x_{jl}-x_{hl}|<\frac{\hat{d}}{2}\text{ if }l\le t(m-k),\,|x_{il}-x_{hl}|<\frac{\hat{d}}{2}\text{ otherwise}\right\}$$

for positive integers $t<\frac{m}{m-k}$. Let T be the largest t that satisfies $t<\frac{m}{m-k}$. Also let $\beta(0)=\{i\}$ and $\beta(T+1)=\{j\}$. Then, by definition, we have $\iota_t\iota_{t+1}\in g$ for all $\iota_t\in\beta(t)$ and $\iota_{t+1}\in\beta(t+1)$ for all $t=0,\dots,T$.

Now, for any given $\hat{d}>0$, as n goes to infinity, almost surely there is at least one agent in $\beta(t)$ for any $t=1,\dots,T$. This is because, the probability that no agent belongs to the area is p^n for some $p\in(0,1)$, which goes to zero as $n\to\infty$.[40] Thus, almost surely, there exists a path between i and j whose length is no more than $T+1$ as n goes to ∞.

Finally, we show that the formula is a lower bound of PL_{ij}. To see this, suppose, to the contrary, that there exists a path with length less than the value above that connects i and j. But such a path has to have a link ww' on it such that $\sharp\{h|x_{wh}-x_{w'h}\le\hat{d}\}<k$, so $d^{(k)}(w,w')>\hat{d}$. Contradiction.

Hence, we have that APL^* is exactly the minimum integer that is no less than $\frac{m}{m-k}$.

Part (ii). 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. □

A.3 *Proof of Proposition 1*

For convenience, below we abuse notation by treating G (or H) either as a probability measure over $[0,1]$ or as a cumulative distribution over $[0,1]$, depending on the context. First we prove a lemma that deals with the case of H with finite support.

[40] This is also a straightforward implication of the second Borel–Cantelli lemma.

LEMMA 2. *Fix any $\hat{d} \in (0, \frac{1}{2})$. For any relative degree distribution H with finite support that includes* 1*, there exists a distribution of θ whose support includes* 1 *such that the resulting relative degree distribution coincides with H.*

PROOF. Let μ be the measure over X induced by f. Fix any $\hat{d} \in (0, \frac{1}{2})$ and take any relative degree distribution H that has finite support that includes 1. Then there exist $K \in \mathbb{N}$ and vectors $(r^1, \dots, r^K)$ and $(w^1, \dots, w^K)$ with the properties that (i) $0 < r^k < r^{k+1}$ for each $k = 1, \dots, K-1$, (ii) $r^K = 1$, and (iii) $H(\{r^k\}) = w^k > 0$ for each $k = 1, \dots, K$. Since $K = 1$ corresponds to the case with homogeneous cutoff values where G puts all the probability mass on 1, it suffices to consider the case with $K > 1$. For this reason, we now assume $K > 1$.

We construct G with finite support consisting of K points, $(\theta^1, \dots, \theta^K)$, where $0 < \theta^k < \theta^{k+1}$ holds for all $k = 1, \dots, K-1$ and θ^k is assigned weight w^k.

Pick any $\tilde{\theta}^1 \in (0, 1)$. Let $\psi^1(\tilde{\theta}^1) = \mu(\{y \in X | d(x, y) \leq \tilde{\theta}^1 \hat{d}\})$ for $x \in \hat{X}$ (which is independent of x as long as $x \in \hat{X}$). We now compute the sequence $(\tilde{\theta}^2, \dots, \tilde{\theta}^K)$ by the following procedure defined by Steps $k = 2, \dots, K$.

Step k. Conditional on the procedure being not stopped so that $(\tilde{\theta}^1, \dots, \tilde{\theta}^{k-1})$ has been obtained, let

$$\begin{aligned} \psi^k(\theta^k) = &\left(\sum_{l=1}^{k-1} \left(w^l \cdot \mu\left(\left\{ y \in X | d(x, y) \leq \tilde{\theta}^l \hat{d} \right\} \right) \right) \right) \\ &+ \left(1 - \sum_{l'=1}^{k-1} w^{l'} \right) \cdot \mu\left(\left\{ y \in X | d(x, y) \leq \theta^k \hat{d} \right\} \right) \end{aligned} \tag{1}$$

for $x \in \hat{X}$ (which is independent of x as long as $x \in \hat{X}$), where $d(\cdot, \cdot)$ is the given notion of social distance. The function ψ^k is continuous and strictly increasing in θ^k such that $\lim_{\theta^k \to \tilde{\theta}^{k-1}} \psi^k(\theta^k) = \psi^{k-1}(\tilde{\theta}^{k-1})$. If $\psi^k(1) \geq \frac{r^k}{r^1} \psi^1(\tilde{\theta}^1)$, there is a unique value of θ^k such that $\psi^k(\theta^k) = \frac{r^k}{r^1} \psi^1(\tilde{\theta}^1)$. Let such θ^k be $\tilde{\theta}^k$. If $\psi^k(1) < \frac{r^k}{r^1} \psi^1(\tilde{\theta}^1)$, then we stop the procedure.

Step $K+1$. If the procedure is not stopped after any step $k = 2, \dots, K$, then the procedure stops at this step. In this case we say that the procedure completes.

If the procedure completes, it uniquely determines a profile $(\tilde{\theta}^1, \dots, \tilde{\theta}^K)$. Also, if G is such that $G(\{\tilde{\theta}^k\}) = w^k$ for each $k = 1, \dots, K$ and $G([0, 1] \setminus (\bigcup_{k=1}^K \{\tilde{\theta}^k\})) = 0$, then for each $k = 1, \dots, K$, the relative degree of agents with $\tilde{\theta}^k$ is given by r^k.

Next we show that if we pick $\tilde{\theta}^1$ to be sufficiently small, then the above procedure completes every step. To see this, note that (1) implies that $\psi^k(1)$ in each step k is bounded below by

$$w^K \cdot \mu\left(\left\{ y \in X | d(x, y) \leq \hat{d} \right\} \right) > 0.$$

Thus, for each $k = 2, \dots, K$, the inequality $\psi^k(1) > \frac{r^k}{r^1} \psi^1(\tilde{\theta}^1)$ is satisfied for small $\tilde{\theta}^1$ since $\lim_{\tilde{\theta}^1 \to 0} \psi(\tilde{\theta}^1) = 0$.

Now note that by construction there exists $\bar{\theta}^1 \in (0, 1)$ such that the region of $\tilde{\theta}^1$ such that the procedure completes is $(0, \bar{\theta}^1]$. Again by construction, $\psi^k(\tilde{\theta}^k) < \psi^{k+1}(\tilde{\theta}^{k+1})$ for

all $k = 1, \dots, K-1$ and $\psi^K(\tilde{(\theta^K)})$ is strictly increasing and continuous in $\tilde{\theta}^1 \in (0, \bar{\theta}^1]$. Hence, $\psi^K(1) = \frac{r^K}{r^1}\psi^1(\bar{\theta}^1)$ holds. With $\tilde{\theta}_1 = \bar{\theta}_1$, use the procedure defined above to generate a profile $(\tilde{\theta}^1, \dots, \tilde{\theta}^K)$. Note that $\tilde{\theta}^K = 1$. This leads to the desired construction by defining G to be a distribution such that $G(\{\tilde{\theta}^k\}) = w^k$ for each $k = 1, \dots, K$ and $G([0,1] \setminus (\bigcup_{k=1}^K \{\tilde{\theta}^k\})) = 0$. □

When H does not have a finite support, we can find a distribution $\tilde{H}$ whose support is finite and includes 1 such that $\sup_{r\in[0,1]} |H(r) - H'(r)| \le \epsilon$ holds. To do this, fix $\epsilon > 0$ and take large natural number N so that $\epsilon \ge 1/2N$. For each $k = 1, \dots, N-1$, define $r_k := \min\{r \in [0,1] : H(r) \ge k/N\}$. Also let $r_0 := 0$ and $r_N := 1$. Then $(k-1)/N \le H(r) \le k/N$ holds if $r_{k-1} \le r < r_k$ for $k = 1, \dots, N-1$. Construct a function $\tilde{H} : [0,1] \to [0,1]$ by setting (i) for each $r \in [0,1)$, $\tilde{H}(r) = \frac{k-1/2}{N}$, where $k \in \{1, \dots, N\}$ is the unique integer such that $r \in [r_{k-1}, r_k)$, and (ii) $\tilde{H}(1) = 1$. This $\tilde{H}$ is nondecreasing and right-continuous, and thus, a cumulative distribution function. Also, by construction the support of $\tilde{H}$ includes 1. Hence by Lemma 2 there exists a distribution of θ whose support includes 1 such that the resulting relative degree distribution coincides with $\tilde{H}$. Since by constriction of $\tilde{H}$ it must be the case that $|H(r) - \tilde{H}(r)| \le \frac{1}{2N} \le \epsilon$ for each $r \in [0,1]$, this completes the proof of Proposition 1. □

A.4 *Proof of Proposition 2*

Let $\hat{G}$ denote the distribution of cutoff $\theta\hat{d}$, and let $\hat{g}$ be the corresponding density. Fix an agent i at $x \in [2\hat{d}, 1-2\hat{d}]^m \subset X$ with cutoff $d \le \hat{d}$. Let $W(t)$ denote the measure of the mass of agents in x's t-neighborhood (with respect to the kth norm), which is continuous and strictly increasing in $t \in [0, \hat{d}]$.

The measure of agents in i's neighborhood with cutoffs in $[\tau, \tau + d\tau]$ is $W(\tau) \times \hat{g}(\tau)\, d\tau$. So the cumulative distribution function of neighbors' cutoffs, denoted C_d, can be written as

$$C_d(t) = \frac{\displaystyle\int_0^t \hat{g}(\tau)W(\tau)\,d\tau}{\displaystyle\int_0^d \hat{g}(\tau)W(\tau)\,d\tau + \int_d^{\hat{d}} \hat{g}(\tau)W(d)\,d\tau} \quad \text{if } t < d$$

and

$$C_d(t) = \frac{\displaystyle\int_0^d \hat{g}(\tau)W(\tau)\,d\tau + \int_d^t \hat{g}(\tau)W(d)\,d\tau}{\displaystyle\int_0^d \hat{g}(\tau)W(\tau)\,d\tau + \int_d^{\hat{d}} \hat{g}(\tau)W(d)\,d\tau} \quad \text{if } t > d. \tag{2}$$

Now, for $t < d < d'$,

$$\left[\int_0^{d'} \hat{g}(\tau)W(\tau)\,d\tau + \int_{d'}^{\infty} \hat{g}(\tau)W\big(d'\big)\,d\tau\right]$$
$$-\left[\int_0^{d} \hat{g}(\tau)W(\tau)\,d\tau + \int_{d}^{\infty} \hat{g}(\tau)W(d)\,d\tau\right]$$

$$= \int_d^{d'} \hat{g}(\tau)W(\tau)\,d\tau - \int_d^{d'} \hat{g}(\tau)W(d)\,d\tau + \int_{d'}^{\hat{d}} \hat{g}(\tau)\big[W\big(d'\big) - W(d)\big]\,d\tau$$
$$= \int_d^{d'} \hat{g}(\tau)\big[W(\tau) - W(d)\big]\,d\tau + \int_{d'}^{\infty} \hat{g}(\tau)\big[W\big(d'\big) - W(d)\big]\,d\tau \geq 0.$$

This means that the denominator of $H_d(t)$ is nondecreasing in d when $t < d$. Thus, $C_d(t)$ is nonincreasing in d when $t < d$.

Second, we consider the case with $t > d$. Differentiating the numerator of (2) with respect to d, we obtain

$$\hat{g}(d)W(d) - \hat{g}(d)W(d) + \int_d^t \hat{g}(\tau)W'(d)\,d\tau = \int_d^t \hat{g}(\tau)W'(d)\,d\tau.$$

Differentiating the denominator of (2) with respect to d, we get

$$\hat{g}(d)W(d) - \hat{g}(d)W(d) + \int_d^{\hat{d}} \hat{g}(\tau)W'(d)\,d\tau = \int_d^{\hat{d}} \hat{g}(\tau)W'(d)\,d\tau.$$

Thus all we need to show is that

$$\frac{\displaystyle\int_d^t \hat{g}(\tau)W'(d)\,d\tau}{\displaystyle\int_d^{\hat{d}} \hat{g}(\tau)W'(d)\,d\tau}$$

is nonincreasing in d.[41] But this is equal to

$$1 - \frac{\displaystyle\int_t^{\hat{d}} \hat{g}(\tau)\,d\tau}{\displaystyle\int_d^{\hat{d}} \hat{g}(\tau)\,d\tau},$$

which indeed is decreasing in d.

Thus C_d is nonincreasing in d for $t > d$ too. Hence C_d is nonincreasing in $d \in [0, \hat{d}]$. Since the relative degrees are strictly increasing in cutoffs, this shows that $H_d(t)$ is nonincreasing in $d \in [0, \hat{d}]$ for all t. □

A.5 *Proof of Proposition 3*

We show that an agent's clustering is strictly decreasing in her cutoff. As discussed before the statement of the proposition in question, this suffices because degrees are increasing in cutoffs. We show that this is true in the case with $m = k = 1$. This suffices as a proof for general (m, k) pairs because types are independently distributed across dimensions

[41] Sufficiency of this property is elementary. The detailed proof can be obtained from the authors upon request.

and thus the clustering under the general case is monotonic to the clustering under $m = k = 1$.

Fix agent i and his cutoff d. Let $\delta(t)$ be the probability density that i's neighbor's location is $x_i + t$ with $t > 0$. Note that this is decreasing in t. Specifically,

$$\delta(t) = \frac{\hat{d} - t}{2\int_0^d (\hat{d} - t)\, dt} = \frac{\hat{d} - t}{2\hat{d}d - d^2}.$$

This formula follows because the measure of i's neighbors with types in $[x_i + t, x_i + t + dt]$ is $(\hat{d} - t)\, dt$.

Also, the cumulative distribution of cutoffs, $\hat{G}$, can be written as

$$\hat{G}(t) = \frac{t}{\hat{d}}.$$

We consider two agents in the set of neighbors of i, with distance t and s. There are two cases to consider, depending on whether the two agents are on the same side of i or on the other side.

Suppose first that the agents are on the same side. The probability that these two neighbors are connected to each other conditional on such an event can be computed as

$$\begin{aligned} \mathrm{Cl}_i^{\mathrm{same}} = \int_0^{\frac{d}{2}} 2\delta(t) \Bigg(& \int_0^{\frac{t}{2}} 2\delta(s) \frac{1 - \hat{G}(t-s)}{1 - \hat{G}(s)}\, ds \\ & + \int_{\frac{t}{2}}^{2t} 2\delta(s) \cdot 1\, ds + \int_{2t}^{d} 2\delta(s) \frac{1 - \hat{G}(s-t)}{1 - \hat{G}(t)}\, ds \Bigg) dt \\ & + \int_{\frac{d}{2}}^{d} 2\delta(t) \Bigg(\int_0^{\frac{t}{2}} 2\delta(s) \frac{1 - \hat{G}(t-s)}{1 - \hat{G}(s)}\, ds + \int_{\frac{t}{2}}^{d} 2\delta(s) \cdot 1\, ds \Bigg) dt. \end{aligned} \tag{3}$$

In order to understand this formula, let j_t and j_s be the agents with distance t and s, respectively, from agent i. The first term (the integral of t from 0 to $\frac{d}{2}$) corresponds to the case where j_t is less than $\frac{d}{2}$ away from i. There are three subcases for this. The first is when j_s is less than $\frac{t}{2}$ away from i. In such a case, the probability that j_s and j_t are connected to each other is not 1, and such a probability is integrated in the first integral corresponding to this subcase (the integral from 0 to $\frac{t}{2}$). The second subcase is when j_s is more than $\frac{t}{2}$ but less than $2t$ away from i. In this case, it is probability 1 that j_s and j_t are connected to each other. This appears as the integration of 1 from $\frac{t}{2}$ to $2t$. The third subcase corresponds to the situation where j_s is further away, and this subcase is expressed in the integral from $2t$ to d.

The second term (the integral of $\frac{d}{2}$ to d) corresponds to the case where j_t is more than $\frac{d}{2}$ away from i, so j_s is always within the cutoff of j_t. Thus there are only two cases corresponding to the first two subcases in the first case, which is why there are only two integrals for this case.

Suppose next that the agents are on the different sides from each other. The probability that these two neighbors are connected to each other conditioning on such an

event can be computed as

$$\mathrm{Cl}_i^{\text{different}} = \int_0^d 2\delta(t)\left(\int_0^d 2\delta(s)\frac{1-\hat{G}(t+s)}{1-\hat{G}(t)}\frac{1-\hat{G}(t+s)}{1-\hat{G}(s)}\,ds\right)dt. \tag{4}$$

An analogous classification of cases as for the case with $\mathrm{Cl}_i^{\text{same}}$ applies to obtain this formula.

We can use these two numbers to compute agent i's clustering:

$$\mathrm{Cl}_i = \frac{\mathrm{Cl}_i^{\text{same}} + \mathrm{Cl}_i^{\text{different}}}{2}.$$

Let $a = \frac{d}{\hat{d}}$. A straightforward calculation shows that (3) and (4) imply, respectively,

$$\mathrm{Cl}_i^{\text{same}} = \frac{9a^2-28a+24}{6(a-2)^2}, \qquad \mathrm{Cl}_i^{\text{different}} = \begin{cases} \dfrac{2\left(6(a-1)^2+a^2\right)}{3(a-2)^2} & \text{if } 2d \le \hat{d}, \\ \dfrac{1-2(a-1)^4}{3a^2(a-2)^2} & \text{if } 2d > \hat{d}. \end{cases}$$

Thus, for $a \in [0, 1/2]$,

$$\mathrm{Cl}_i = \frac{\dfrac{9a^2-28a+24}{6(a-2)^2} + \dfrac{2\left(6(a-1)^2+a^2\right)}{3(a-2)^2}}{2} = \frac{37a^2-76a+48}{12(a-2)^2}.$$

Differentiating this expression with respect to a, we obtain $-\frac{8(4a-3)}{3(a-2)^3}$, which is strictly negative for $a \in [0, 1/2]$.

Also, for $a \in (1/2, 1]$,

$$\mathrm{Cl}_i = \frac{\dfrac{9a^2-28a+24}{6(a-2)^2} + \dfrac{1-2(a-1)^4}{3a^2(a-2)^2}}{2} = \frac{5a^4-12a^3+16a-2}{12(a-2)^2a^2}.$$

Differentiating this expression with respect to a, we obtain $\frac{4(a-1)(2(a-1)^2-1)}{3(a-2)^2a^3}$, which again is strictly negative for $a \in (1/2, 1]$. Thus, Cl_i is strictly decreasing in $a \in [0, 1]$, so it is strictly decreasing in d, completing the proof. □

References

Akerlof, George A. (1997), "Social distance and social decisions." *Econometrica*, 65, 1005–1027. [656]

Antal, Tibor, Hisashi Ohtsuki, John Wakeley, Peter D. Taylor, and Martin A. Nowak (2009), "Evolution of cooperation by phenotypic similarity." In *Proceedings of the National Academy of Sciences of the United States of America*, 8597–8600, PNAS, Washington, DC. [679]

Baccara, Mariagiovanna, Ayşe İmrohoroğlu, Alistair J. Wilson, and Leeat A. Yariv (2012), "Field study on matching with network externalities." *American Economic Review*, 102, 1773–1804. [667]

Backstrom, Lars, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna (2012), "Four degrees of separation." In *WebSci '12 Proceedings of the 4th Annual ACM web Science Conference*, 33–42, ACM, New York, NY. [658]

Barabási, Albert-László and Réka Albert (1999), "Emergence of scaling in random networks." *Science*, 286, 509–512. [660, 670]

Caldarelli, Guido, Andrea Capocci, Paolo De Los Rios, and Miguel Munoz (2002), "Scale-free networks from varying vertex intrinsic fitness." *Physical Review Letters*, 89, 258702. [660]

Chung, Fan and Linyuan Lu (2002), "The average distances in random graphs with given expected degrees." *Proceedings of the National Academy in Sciences*, 99, 15879–15882. [660, 669]

Chwe, Michael Suk-Young (1994), "Farsighted coalitional stability." *Journal of Economic Theory*, 63, 299–325. [664]

Chwe, Michael Suk-Young (2000), "Communication and coordination in social networks." *Review of Economic Studies*, 67, 1–16. [658]

Conley, Timothy G. and Christopher R. Udry (2010), "Learning about a new technology: Pineapple in Ghana." *American Economic Review*, 100, 35–69. [667]

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin (2009), "An economic model of friendship: Homophily, minorities, and segregation." *Econometrica*, 77, 1003–1045. [659]

Ebel, Holger, Lutz-Ingo Mielsch, and Stefan Bornholdt (2002), "Scale-free topology of E-mail networks." *Physical Review E*, 66, 035103. [656, 667, 669]

Gaspar, Jess and Edward Glaeser (1998), "Information technology and the future of cities." *Journal of Urban Economics*, 43, 136–156. [658]

Goyal, Sanjeev (2007), *Connections*. Princeton University Press, Princeton, New Jersey. [655]

Goyal, Sanjeev, Marco J. van der Leij, and José Luis Moraga-González (2006), "Economics: An emerging small world." *Journal of Political Economy*, 114, 403–412. [656, 670]

Hildenbrand, Werner (1974), *Core and Equilibria of a Large Economy*, volume 5 of Princeton Studies in Mathematical Economics. Princeton University Press, Princeton, New Jersey. [680]

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock (2002), "Latent space approaches to social network analysis." *Journal of the American Statistical Association*, 97, 1090–1098. [659]

Iijima, Ryota and Yuichiro Kamada (2015), "Social distance and network structures." Unpublished paper, Harvard University. [659, 676]

Jackson, Matthew O. (2005), "A survey of network formation models: Stability and efficiency" In *Group Formation in Economics: Networks, Clubs, and Coalitions* (Gabrielle Demange and Myrna Wooders, eds.). Cambridge University Press, New York. [664]

Jackson, Matthew O. (2008a), *Social and Economic Networks*. Princeton University Press, Princeton, New Jersey. [655, 664]

Jackson, Matthew O. (2008b), "Average distance, diameter, and clustering in social networks with homophily." In *Internet and Network Economics* (Christos Papadimitriou and Shuzhong Zhang, eds.), 4–11, Springer, Berlin, Heidelberg. [660, 669]

Jackson, Matthew O. and Brian W. Rogers (2005), "The economics of small worlds." *Journal of European Economic Association*, 3, 617–627. [659, 670]

Jackson, Matthew O. and Brian W. Rogers (2007), "Meeting strangers and friends of friends: How random are social networks?" *American Economic Review*, 97, 890–915. [660]

Jackson, Matthew O. and Anne van den Nouweland (2005), "Strongly stable networks." *Games and Economic Behavior*, 51, 420–444. [659, 673]

Jackson, Matthew O. and Asher Wolinsky (1996), "A strategic model of social and economic networks." *Journal of Economic Theory*, 71, 44–74. [657, 663]

Johnson, Cathleen and Robert P. Giles (2000), "Spatial social networks." *Review of Economic Design*, 5, 273–299. [660]

König, Michael, Claudio J. Tessone, and Yves Zenou (2014), "Nestedness in networks: A theoretical model and some applications." *Theoretical Economics*, 9, 695–752. [660]

Kremer, Michael and Edward Miguel (2007), "The illusion of sustainability." *Quarterly Journal of Economics*, 122, 1007–1065. [667]

Marmaros, David and Bruce Sacerdote (2006), "How do friendships form?" *Quarterly Journal of Economics*, 121, 79–119. [662]

Newman, Mark E. J. (2003), "The structure and function of complex networks." *SIAM Review*, 45, 167–256. [655, 670]

Page, Frank H. Jr., Myrna H. Wooders, and Samir Kamat (2005), "Networks and farsighted stability." *Journal of Economic Theory*, 120, 257–269. [664]

Pareto, Vilfredo (1896), *Cours d' Economie Politique*. F. Rouge. [671]

Pin, Paolo and Brian Rogers (2016), "Stochastic network formation and homophily". In *The Oxford Handbook of the Economics of Networks* (Yann Bramoullé, Andrea Galeotti, and Brian Rogers, eds.). Oxford University Press. [660]

Price, D. D. S. (1976), "A general theory of bibliometric and other cumulative advantage processes." *J. Am. Soc. Inf. Sci.*, 27, 292–306. [660]

Rapoport, Anatol (1953), "Spread of information through a population with socio-structural bias: II. Various models with partial transitivity." *Bulletin of Mathematical Biophysics*, 15, 535–546. [667]

Rapoport, Anatol and William J. Horvath (1961), "A study of a large sociogram." *Systems Research and Behavioral Science*, 6, 279–291. [667]

Richardson, Marion Webster (1938), "Multidimensional psychophysics." *Psychological Bulletin*, 35, 659–660. [659]

Rosenblat, Tanya S. and Markus M. Mobius (2004), "Getting closer or drifting apart?" *Quarterly Journal of Economics*, 119, 971–1009. [658]

Selfhout, Maarten H. W., Susan J. T. Branje, Tom F. M. ter Bogt, and Wim H. J. Meeus (2009), "The role of music preferences in early adolescents' friendship formation and stability." *Journal of Adolescence*, 32, 95–107. [657]

Torgerson, Warren S. (1958), *Theory and Methods of Scaling*. Wiley, Oxford England. [659]

Travers, Jeffrey and Stanley Milgram (1969), "An experimental study of the small world problem." *Sociometry*, 32, 425–443. [669]

Tversky, Amos (1977), "Features of similarity." *Psychological Review*, 84, 327–352. [659]

Tversky, Amos and Itamar Gati (1982), "Similarity, separability, and the triangle inequality." *Psychological Review*, 89, 123–154. [659]

Ugander, Johan, Brian Karrer, Lars Backstrom, and Cameron Marlow (2011), "The anatomy of the Facebook social graph." Unpublished paper. Available at arXiv:1111.4503. [670]

Watts, Duncan J., Peter Sheridan Dodds, and Mark E. J. Newman (2002), "Identity and search in social networks." *Science*, 296, 1302–1305. [659, 669]

Watts, Duncan J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks." *Nature*, 393, 440–442. [670]

Wilson, Christo, Alessandra Sala, Krishna P. N. Puttaswamy, and Ben Y. Zhao (2012), "Beyond social graphs: User interactions in online social networks and their implications." *ACM Transactions on the Web*, 6, Article No. 17. [670]

Zenou, Yves (2016), *Key Players* (Andrea Galeotti, Yann Bramoullé and Brian Rogers, eds.). The Oxford Handbook of the Economics of Networks. Oxford University Press. [677]

Co-editor Nicola Persico handled this manuscript.

Manuscript received 30 May, 2014; final version accepted 13 May, 2016; available online 24 May, 2016.