MACROECON & INT'L FINANCE WORKSHOP
presented by: Laura Veldkamp
FRIDAY, Dec. 3, 2010
3:30 pm – 5:00 pm, **Room: HOH-302**

Germs, Social Networks and Growth

Alessandra Fogli and Laura Veldkamp*

First version: July 2010
This version: November 2010

Abstract

A country's social institutions undoubtedly affects their economy. But how does this effect operate and how much does it matter for economic development? Economists often overlook social institutions because they are difficult to observe, to describe formally and to quantify. This paper uses tools from network analysis to explore how different social structures might affect a country's rate of technological progress. The network model also explains why societies with a high prevalence of contagious disease might adopt growth-inhibiting institutions and how small differences in disease prevalence can produce large differences in incomes. Empirical work uses differences in disease characteristics to identify an effect of social structure on technology diffusion.

How people organize themselves as a society undoubtedly affects economic activity and countries' average level of income. But how does this effect operate and how much does it matter for development? Economists typically overlook findings of sociologists and anthropologists because social characteristics are difficult to observe, to describe formally and to quantify.[1] This paper uses tools from network analysis to explore how different social structures might affect a country's rate of technological progress. The network model also explains why societies might adopt growth-inhibiting institutions and how small differences in disease prevalence can produce large differences in incomes. Our empirical work identifies an effect of social structure on technology diffusion.

There is a long history in the networks literature of measuring the speed of information or technology diffusion within various kinds of networks (Jackson (2008), Granovetter (2005)). Given these findings, a simple way to explain the effect of social structure on GDP is to show that some types of social networks disseminate new technologies more efficiently than others and append a

*Corresponding author: afogli@umn.edu, Department of Economics, University of Minnesota, 90 Hennepin Ave. Minneapolis, MN 55405. lveldkam@stern.nyu.edu, 44 West Fourth St., suite 7-77, New York, NY 10012. We thank participants at the 2010 SED meetings for their comments and suggestions. We thank Corey Fincher and Damian Murray for help with the pathogen data. Laura Veldkamp thanks the Hoover Institution for their hospitality and financial support through the national fellows program. Keywords: growth, development, technology diffusion, economic networks, social networks, pathogens, disease. JEL codes: E02, O1, O33, I1.

[1] Of course, there is a small economics literature and a much more extensive sociology literature on the effects of social institutions on income. See e.g. Greif (1994) for economics and Granovetter (2005) for a review of the sociology literature.

production economy where the average technology level is related to output and income. There are two problems with this explanation. First, if social contacts are something people can choose, then choosing a social structure that inhibits growth seems sub-optimal. Second, this explanation is difficult to quantify or test. While researchers have mapped social networks in schools or on-line communities (Jackson, 2008), measuring the social network structure for an entire economy is not feasible.

Our theory for why some societies choose growth-inhibiting social structures revolves around the idea that communicable diseases and technologies spread in similar ways - through human contact. When choosing a social structure, people are balancing the advantages of rapid technology diffusion against the risk of infection. In countries where communicable diseases are inherently more prevalent, a social structure that inhibits the spread of disease and technology will be optimal.

The idea that disease prevalence and social structure are related can help to isolate the effect of social structure on GDP. First, we show that, to protect themselves from disease, people should form economic networks with the property that your contacts are likely to be direct contacts with each other. These kinds of structures with mutual friends or contacts are often called "triples." When relationships have many triples, each group of friends have few links with the outside world. This limited connectivity reduces the risk of an infection entering the community, but it also restricts the community's exposure to new technologies. In contrast, having a dispersed community brings the benefit faster technology diffusion, at the cost of a higher rate of mortality.

Networks with many triples are common in collectivist societies. In order to enforce community norms, communities need to be dense networks of people who know each other. The existence of friends-in-common is what allows the group to punish deviators and what allows the collectivist norms to be enforced (Coleman, 1988). These societies are typically characterized by high levels of within-group trust, much less trust in strangers, and more localized economic activity. In individualistic societies, where people interact primarily through large markets, they tend to come in contact with people who do not know each other. In other words, the network in an individualistic society has few triples. Therefore, collectivist societies are more effective at protecting their members from the spread of disease. Since our theory predicts that societies with high disease prevalence should be more likely to adopt a collectivist structure, we use disease prevalence to help empirically identify social structure.

Specifically, we compare an individualistic to a collectivist society modeled in the following way: In the collectivist society, relationship triples are abundant, meaning that the country is populated by communities of people who mostly all know one another, and know each other's friends. In the individualistic society, relationships are dispersed. Agents interact, socially or economically, with others who do not know each other. This would be the case if most transactions took place in large,

anonymous markets. Section 1 model the epidemiology of disease and technology in each society and show that when the initial prevalence of infectious disease is higher, people prefer collectivist social networks, in order to reduce their chance of infection. This also reduces the growth rate of productivity.

Using historical pathogen prevalence data from the Gideon database and measures of a society's individualism from Hofstede (2001), section 3 tests the model's predictions for the relationship between disease prevalence and social structure. We adopt a number of different strategies to distinguish the effect of disease on social structure from the reverse effect of from some jointly causal factor. Finally, we quantify how much of the cross-country difference in technology diffusion this mechanism can explain.

The last section of the paper tackles the following question: If societies became collectivist when diseases were prevalent in the 1930's, why didn't their social structure change in the latter 20th century when vaccinations and modern medicine eradicated many of these diseases? The answer lies in the fact that networks often have multiple equilibria. When most people in the network are in relationships with many triples, having the same kind of relationships can be optimal. So while the society may have originally become collectivist to protect itself from disease, it may continue to be collectivist, even if social welfare would be improved from being individualistic, because it is stuck in a collectivist equilibrium, from which no individual has an incentive to deviate.

Related literature The paper contributes to four growing literatures. A closely related literature is one that considers the effects of social structure on economic outcomes. Most of this literature considers particular firms, industries or innovations and how they were affected by the social structure in place (e.g., see Granovetter (2005) or Rauch and Casella (2001)). In contrast, this paper takes a more macro approach and studies the types of social networks that are adopted throughout a country's economy and how those affect technology diffusion economy-wide.

Thus in its scope, the paper is much more related to work on technology diffusion. But what sets this paper apart from that body of work is its insights about why societies adopt networks that do not facilitate the exchange of ideas and its links to empirical measures of social structure.

The literature on culture and its effects on national income is similarly macro in scope. For example, work by Tabellini (2005) and Algan and Cahuc (2007) examine the relationship between cultural characteristics and economic outcomes. Work by Bisin and Verdier (2001), Bisin and Verdier (2000) and Fernández and Fogli (2005) examines the transmission of culture. Cole, Mailath, and Postlewaite (1992) investigate how social norms affect savings choices, and in turn growth. But this literature typically regards culture as an aspect of preferences. We look at social structure, which characterizes the set of relationships people have. Greif (1994) argues that culture is an

important determinant of a society's social structure. While this is undoubtedly true, we examine a different determinant of social structure that is more easily measurable, pathogen prevalence. Our approach lends itself better to quantifying the effects of social structure on economic outcomes.

Likewise, the work on the importance of political institutions by Acemoglu, Johnson, and Robinson (2002) and Acemoglu and Johnson (2005) is similar in its objectives and the approach of using pathogen prevalence to identify variation in endogenous institutions. But instead of examining political institutions, we study an equally important but distinct type of institution, social structure.

1 A Network Diffusion Model

Our model serves two purposes. First, it is meant to fix ideas. The concept of social structure is a fungible one. We want to pick a particular aspect of social structure, the social network, to anchor our analysis on. In doing this, we do not exclude the possibility that other aspects of social or cultural institutions are important for technology diffusion and income. But we will try to make the case that this aspect is important.

The second role of the model is that it helps us answer the following question: The richest countries have income and productivity levels that are 100 times higher than the poorest countries. How can small differences of a few percentage points in disease prevalence explain such large income disparities? To answer this kind of question requires model and some reasonably calibrated parameter values. The next section takes up this quantitative exercise.

A key feature of our model linking social structure to technological progress is that technologies spread by human contact. This is not obvious since one might think new ideas could be just as easily spread by print or electronic media. However, economists and sociologists have long noted the importance of human contact. In his 1969 presidential address, Kenneth Arrow remarks, "While mass media play a major role in alerting individuals to the possibility of an innovation, it seems to be personal contact that is most relevant in leading to its adoption. Thus, the diffusion of an innovation becomes a process formally akin to the spread of an infectious disease." With this description of the process of technological diffusion in mind, we propose the following model.

1.1 Economic Environment

Time, denoted by $t = \{1, \ldots, T\}$, is discrete and finite. At any given time t, there are A agents, indexed by their location $j \epsilon \{1, 2, \ldots, A\}$ on a circle. Each agent produces consumption goods with a technology $A_j(t)$ and labor input $l_j(t)$:

$$y_j(t) = A_j(t) l_j(t)^{\alpha}$$

Each healthy agent is endowed with 1 unit of labor, which they supply inelastically ($l_j(t) = 1$). Furthermore, there is no savings technology. Thus, consumption for healthy agents is $c_j(t) = y_j(t) = A_j(t)$.

An agent who catches a disease at time t loses their endowment of labor for one period ($n_j(t) = 0$ and thus $c_j(t) = 0$). At the end of this period, they die and are replaced by a new person in the following period. Let $\psi_j(t) = 1$ if the person in location j is sick in period t and $= 0$ otherwise and $\Psi_j(t) = min\{s : s \geq t, \psi_j(s) = 1\}$ be the period in which the person living in location j at time t gets sick and dies.

Then, the objective function of person j is

$$U_j = \sum_{\tau=t}^{\Psi_j - 1} \beta^{\tau - t} \frac{(c_j(\tau))^{1-\gamma}}{1-\gamma} \tag{1}$$

where β is the time discount factor. Note that we need $\gamma \leq 1$ for death to be a bad outcome. Otherwise, utility when living is always negative and death is desirable.

Social networks Each person i knows ϕ_i other people. Let $\eta_{jk} = 1$ if person j and person k know each other and $= 0$ otherwise. Let the network of all connections be denoted N. Agent i's expected utility can then be expressed indirectly as a function of N: $EU_i(N)$.

Spread of disease Each infected person in one's social network transmits the disease to any other network member with probability π. Thus, if m network members are diseased at time $t-1$, then the probability of being healthy at time t is $(1-\pi)^m$. If no one in the social network has a disease at time $t-1$, then the probability of contracting the disease at time t is zero.

Agents have rational expectations about the aggregate disease rate. (We can relax this later, but it keeps things very simple.) In other words, they correctly anticipate the prevalence of disease country-wide. But they do not know who is sick.

Spread of technology Technological progress occurs when someone improves on an existing technology. To make this improvement, they need to know about the existing technology. Thus, if a person is producing with technology $A_j(t)$, they will invent the next technology with a Poisson probability λ each period. If they invent the new technology, $\ln(A_j(t+1)) = \ln(A_j(t)) + 0.05$. In other words, a new invention results in a 5% increase in productivity.

People can also learn from others in their network. If person j is connected to person k and $A_k(t) > A_j(t)$, then the next period, j can produce with k's technology. If there are multiple levels of technology used by j's social contacts, j can produce with the best of these technologies:

$A_j(t+1) = \max_k \eta_{jk} A_k(t)$.

As with disease, agents' expectations about others' technology are rational. In other words, they correctly anticipate the fraction of people producing with each technology level. But they do not know who has which technologies.

Definition of Equilibrium We focus on symmetric equilibria where all agents have the same number of connections ($\eta_i = \eta$). An equilibrium is a network that is stable under pairwise deviations. In other words, there are no two people who both prefer to form a new connection with each other, and there is no individual who would be better off from severing some existing network connection.

1.2 Results: Speed of Diffusion in Clustered and Diffuse Networks

Ideally, we would like to have a model where differences in initial disease prevalence cause agents to choose different types of social networks from the set of all possible networks. The problem is that such network choice models frequently have multiple equilibria. Furthermore, if one had such a model, it would not be clear how the variety of possible networks should be mapped into data. To clarify the mapping between the model and the data, we choose analyze only networks. We choose networks that are extremes along a particular dimension, their degree of clustering, because that is an aspect of a social network that bears a close resemblance to features of societies measured by sociologists and anthropologists.

To make our examples concrete, we will fix the number of connections ϕ to be 4. While it would also be interesting to analyze the variation in the number of connections each individual has, we restrict attention to the degree of network clustering because it corresponds most closely to the ideas measured in our data.

Network 1 *In the clustered social network, each individual j is connected to the $\phi > 2$ people located closest to them. In other words, $\eta_{jk} = 1$ for $k = \{j-2, j-1, j+1, j+2\}$ and $\eta_{jk} = 0$ for all other k.*

Network 2 *In the dispersed social network, each person is friends with the person next to them and the person 4 positions away from them, on either side. In other words, $\eta_{jk} = 1$ for $k = \{j-4, j-1, j+1, j+4\}$ and $\eta_{jk} = 0$ for all other k.*

Diffusion speed in the Clustered Network Disease spreads slowly in the clustered network. The reason is that each contiguous group of friends is connected to only 4 non-group members. Those are the two people adjacent to the group, on either side. Since there are few links with

outsiders, the probability that a disease within the group is passed to someone outside the group is small.

Likewise, ideas disseminate slowly. Something invented in one location takes a long time to travel to a far-away location. In the meantime, someone else may have re-invented the same technology level, rather than building on existing knowledge and advancing technology to the next level. Such redundant innovations slow the rate of technological progress and lower average consumption.

The speed at which germs and ideas disseminate can be measured by the number of social connections in the shortest path between any two people. Consider an agent in position 1 and the agent farthest away from them on the circle, agent $n/2$. If η is even so that each person has $\eta/2$ friends on either side of them, then agent 1 will be friends with agent $1 + \eta/2$, who will be friends with agent $1 + \eta$, and this person will be friends with agent $1 + 3\eta/2$, ect., until we reach $n/2$. The number of friends in this chain will be n/η, if that is an integer, or otherwise the next highest integer. The distance to this farthest person in the network is called the *network diameter*.

Result 1 *The diameter of a clustered network, with n nodes, where each agent has an even number ϕ of connections is $(n-1)/\phi$, if that is an integer, or otherwise the next highest integer.*

The proof of this and all subsequent results are in appendix A.

Diameter is one measure of dispersion speed because it tells us how many periods a new idea takes to travel to every last person in the network. If each person communicates the idea to each of their friends each period, then in n/ϕ periods, the farthest person in the network will have learned the idea, along with every other agent. Since disease is spread only probabilistically, from friend to friend, the diameter gives us the smallest number of periods in which every person is infected, with positive probability.

Another related measure of the speed of diffusion is the average path length. Instead of measuring the number of nodes in the most direct path to the farthest person, this measure computes the number of nodes in the shortest path to every person and averages those lengths. If ϕ is even and $n/(2\phi)$ is an integer then $n/(2\phi)$ is the average path length in a clustered network.

Result 2 *If ϕ is even and $n/(2\phi)$ is an integer then $1/2 + n/(2\phi)$ is the average path length in a clustered network.*

Diffusion speed in the Dispersed Network In this environment, dissemination of ideas or disease is fast. Each group of friends is connected to many outsiders, making the probability that a disease within the group is passed to someone outside the group is high. Likewise, ideas disseminate quickly because they travel many positions around the circle each period.

To measure the speed of dissemination, we compute the diameter and average path length in the network. First, we define the operator round(x) to be the integer y closest to x. In other words, if y is an integer, then round$(x) = y$ iff $x\epsilon[y - 1/2, y + 1/2)$.

Result 3 *The diameter of a dispersed network, with $n > 4$ nodes where each node i is connected to $i - 4$, $i - 1$, $i + 1$, and $i + 4$, is round$(n/8) + 1$.*

Result 4 *In the example dispersed network, when $n/8$ is an integer, the average path length is $7/8 + n/16$.*

These measures tell us why ideas and germs spread more quickly in the dispersed network than in the clustered network. Whereas, with a clustered network, technology invented in one location was transmitted only $\phi/2$ people further each period, in this network, ideas advance 4 places at a time. Because redundant innovations are less frequent, the rate of technological progress is faster.

Result 5 *For a large network ($n > 8$) where $n/8$ is an integer, the dispersed network has a smaller diameter and a shorter average path length than a clustered network with equal size n and equal degree $\phi = 4$.*

2 Data and Its Relationship to the Model

The model is about the relationship between three main variables: pathogen prevalence, social structure, and the technological frontier. The section describes how these three variables are measured and how each measure corresponds to its theoretical counterpart.

2.1 Measuring Pathogen Prevalence

To measure the prevalence of disease, we use the historical prevalence of 9 pathogens: leishmanias, leprosy, trypanosomes, malaria, schistosomes, filariae, dengue, typhus and tuberculosis. We choose these diseases because we have good worldwide data on their incidence, and they are serious, potentially life-threatening diseases that people would go to great length to avoid.

Our data comes from 1930-40 atlases of infectious diseases and the Gideon health statistics database. For each disease, we have estimates coded on a 3 point scale (not endemic, sporadic, endemic), standardized across countries. The mean of standardized scores across diseases captures a country's relative pathogen prevalence.

To identify the effect of disease on social structure, we follow Smith, Sax, Gaines, Guernier, and Gugan (2007) and Thornhill, Fincher, Murray, and Schaller (2010) by distinguishing between three types of diseases:

Human specific Many infectious agents known to afflict mankind are currently entirely restricted to human reservoir hosts (i.e., contagious only between persons), even though they historically may have arisen in other species, such as measles which originated in cattle. Examples of human-specific infectious agents represented in the GIDEON database include measles, smallpox, and syphilis.

Zoonotic Infectious agents that develop, mature, and reproduce entirely in non-human hosts, but nonetheless have the potential to spill over and infect human populations, are referred to herein as zoonotic infectious agents. Humans are a dead-end host for infectious agents in this group. Examples of zoonotic infectious agents in the GIDEON database include rabies, plague, and hantavirus.

Multi-host Some infectious agents can use both human and non-human hosts to complete their lifecycle. Oftentimes these infectious agents are lumped with zoonotics, but for the purposes of this study we distinguish them with the term multi-host infectious agent ("multi" referring to both human and non-human hosts). Examples of multi-host infectious agents in the GIDEON database include the three forms of leishman iasis (cutaneous, mucocutaneous, and visceral) that can use humans, wild, and/or domestic animals as reservoir hosts.

2.2 A Sociological Measure of Clustering: Collectivism

Collectivism is defined as a social pattern of closely linked individuals; interdependent members of a collective. Collectivistic societies are ones in which individuals are integrated into communities. What distinguishes communities from sets of people with random ties to each other is that in communities, people have mutual friendships. In other words, it is common that two friends have a third friend in common. This is the sense in which they are interdependent.

Individualism is the opposite of collectivism. Individualistic societies are ones where the ties between individuals are loose. Everyone is expected to look after him/herself and immediate family members. In individualistic societies, people interact though market mechanisms. Through markets, they interact with a variety of people who are unlikely to know each other. Thus, individualistic societies are ones where social networks have fewer mutual friendships.

To measure where various societies fall on the individualism/collectivism spectrum, Hofstede (2001) performed a survey of IBM employees worldwide. He used the 33 survey questions to form an index of individualism that ranges from between 0 (strongly collectivist) to 100 (strongly individualist). Figure 1 summarizes the findings of his survey in a color-coded map.

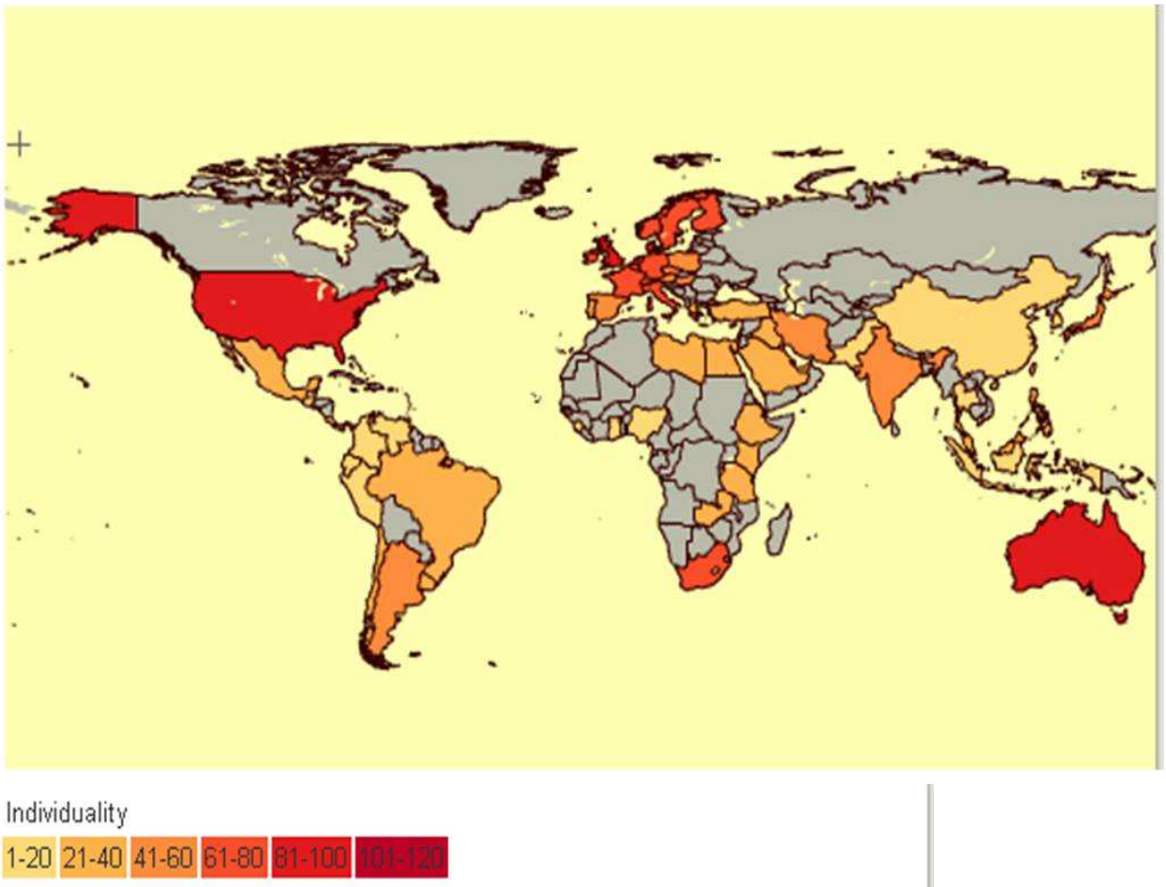


Figure 1: Map of Hofstede's individualism index.

Measuring collectivism in the model In the model, we can look for the same pattern of mutual friendships that is the hallmark of collectivist societies. In each network, we can ask: If A is friends with B and with C, how often are B and C also friends? In the networks literature, a structure where A, B and C are all connected to each other is called a *triple*. Therefore, a measure of the extent of shared friendships, and thus the degree of collectivism, is the number of network triples. We begin by examining the number of triples in each of the two types of networks we have considered.

To count the number of triples, look at all the instances where one node i is connected to two other nodes j, k. Count that as a triple if j and k are connected. This triples measure is related to a common measure of network clustering: Divide the number of triples by the number of possible triples in the network to get the *overall clustering* measure (Jackson 2008).

Result 6 *In a clustered network, where $\phi = 4$, there are n unique triples.*

Result 7 *In a dispersed network, where where each person i is connected to $i - \psi$, $i - 1$, $i + 1$, and*

$i + \psi$, where $\psi > 2$, there are zero triples.

In fact, we chose these two network structures because of their starkly different triples results. This stark difference facilitates matching social institution data with one or the other type of network. Of course, other intermediate cases, with numbers of triples between 0 and n, are also possible in reality. But knowledge of the properties of these two extreme cases sheds light on the likely outcomes of intermediate cases as well.

Collectivism as strong social norms Another way to interpret collectivism and the notion of interdependence that it entails is to relate it to the strength of social norms. Perhaps being members of an integrated collective means adopting similar behaviors and norms. In fact, Hofstede's individualism index is highly correlated with measures of social conformity in the GSS survey.

Social conformity is easier to sustain in clustered networks. Coleman (1988) shows that the presence of effective norms and thus the accumulation of social capital depend on network "closure." Closure is present when your friends are also your friends' friends. In other words, it depends on the presence of triples. Coleman explains that people enforce strong group norms through collective punishments of deviators. If j observes i deviating from a social norm, then j can directly contact other friends of i to enact some joint retribution for the misdeed. When collective punishments are implementable, conforming behavior is easier to sustain than if punishments must be implemented in an uncoordinated way. Thus, if we interpret collectivism as strong social conformity, such collectivism is more likely to emerge in networks with many triples.

2.3 Measuring the Technological Frontier

We use the cross-country historical adoption of technology (hereafter CHAT) data set developed by Comin, Hobijn, and Rovito (2006). CHAT covers the diffusion of about 115 technologies in over 150 countries during the last 200 years. We use the number of adopted technologies per country to measure how far up the technological ladder the country's most advanced agents are. This measure seems to reliably capture countries' technological ranking because there are universal leaders and universal followers. In other words, countries' ranking in terms of their speed of adoption is stable across technologies and over time.

3 Empirical Results

Our objective is to better understand how social structure affects development and how large that effect is. The difficulty is that economic development also can potentially change the social structure. The challenge is to isolate each of these two effects. We take two approaches. The

first uses differences in the prevalence of human and zoonotic diseases as an instrument for social structure. The second approach, explored in the following section, uses a calibrated model to determine how much of the relationship between social structure and GDP is due to the technology diffusion effect.

Before we look at the effect of social structure on technology diffusion, we first establish an empirical relationship between disease and social structure that justifies our use of disease prevalence as an instrumental variable.

3.1 The Relationship between Disease and Social Institutions

The first exercise is to do a basic regression of the Hofstede index of individualism on pathogen prevalence, to see if these two variables are related to each other, in any way. Figure 2 illustrates the positive statistical correlation between individualism and the prevalence of pathogenic disease.

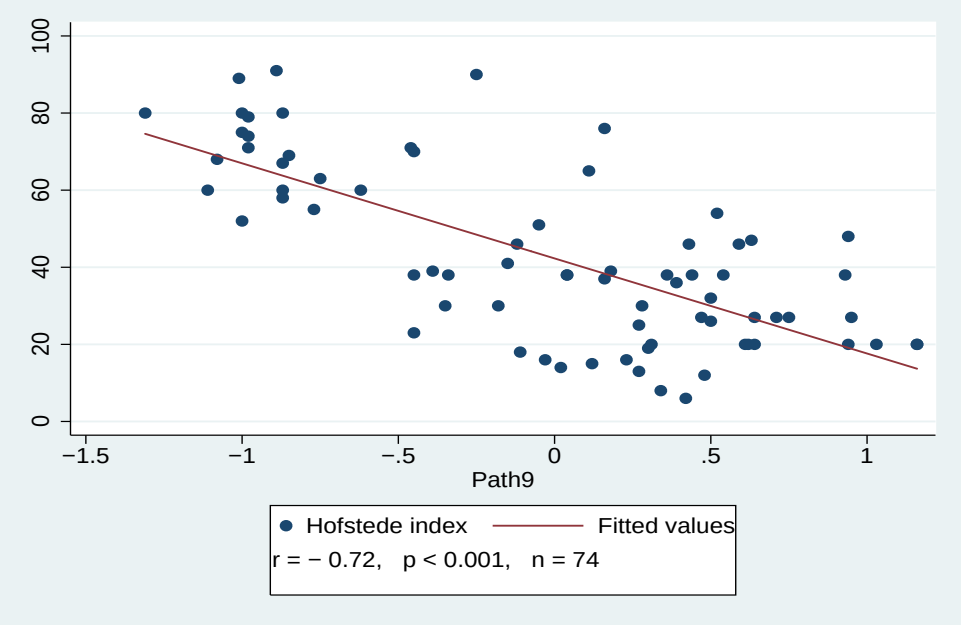


Figure 2: Hofstede's individualism index plotted against pathogen prevalence.

Table 1 quantifies this relationship. Column 1 shows that pathogen prevalence and individualism are related in a statistically significant way. The negative sign on the pathogen coefficient means that the increased presence of pathogens is associated with a less individualistic (more collectivist) society. That is consistent with our theory because the more collectivist society, with its greater propensity for network triples, would be a more effective structure for inhibiting the spread of disease. The explanatory power of pathogens is large; the R^2 of the regression is over 50%.

Of course, it is possible that both disease and social structure are governed by GDP, or that

Dependent variable	Individualism (Hofstede Index)			
	OLS	OLS	IV	OLS
Pathogens	-25.10	-21.76	-42.74	-8.35
	(2.80)	(3.98)	(7.33)	(4.42)
1970 GDP		3.80	-5.25	3.31
		(2.50)	(3.79)	(2.17)
1970 population density		-0.003	-0.002	-0.001
		(0.003)	(0.004)	(0.003)
Latitude				0.71
				(0.15)
R^2	0.52	0.55	0.37	0.65
Observations	78	73	73	73

Table 1: **Relationship Between Pathogen Prevalence and Hofstede Individualism Index**

GDP is the PPP-adjusted GDP from the Penn World Tables. The IV regression uses a 2SLS procedure, where latitude is an instrument for pathogens. Each equation includes a constant as well.

higher population density lends itself to a different social structure and more disease prevalence. To determine whether pathogens might have an effect, beyond that governed by GDP and density, we estimate a second regression (column 2) where we control for GDP and population density. We use the figures from 1970, the same time as the Hofstede survey was being collected. Controlling for GDP and population density only slightly lessens the significance of the relationship between disease and social structure. Surprisingly, when we include both pathogens and GDP in the regression, the effect of GDP on social structure gets crowded out. This suggests that GDP might affect social structure through the prevalence of pathogens, rather than the other way around.

Differences in diseases One can still not deduce from these results that pathogen prevalence is a determinant of social structure. It is possible that lower rates of disease and social structure changes might both be caused by some other common factor. We also know that social structure can affect the transmission and therefore the prevalence of disease.

Our approach to identification is three-pronged. First, we use a simple timing argument. The pathogen prevalence is historical, from the 1930's and 40's. Therefore, the timing makes it more likely that the pathogens affected the social structure 30-40 years later, than the other way around. But of course, social structure is very persistent. So, it is still possible that social structure prior to the 1930's is responsible for the historical pathogen prevalence.

The second approach is to instrument for pathogen prevalence. Our instrument is latitude, which is clearly an exogenous, immutable feature of a country. Latitude is also a good predictor of pathogen prevalence. Countries with lower latitudes (nearer the equator) are warmer and more

conducive to the growth and spread of disease. When we use latitude as an instrument, pathogens remain a highly-significant predictor of social structure (column 3).

One concern with this approach is that latitude alone is a good predictor of social structure. To be sure that pathogens have some explanatory, above and beyond that of latitude, the results in column 4 include both variables. The importance of pathogens does decline when latitude is included separately. But pathogens are still significant with a P-value of 6.3%.

The third approach, is to exploit the difference between diseases that could be spread by social contact with others and those that are either not contagious or spread by other means, such as by flies, rodents or ingestion of contaminated water. The former should rationally influence your social network, while the latter should not, except through indirect effects, such as through population density. The additional effect of the communicable diseases on network structure should cause the impact of those diseases on income to be stronger. Therefore, the next step in our analysis tests whether the coefficient on communicable disease is significantly larger than the coefficient on other diseases.

Dependent variable	Individualism (Hofstede Index)		
	OLS	OLS	OLS
Human	-6.19		
	(2.35)		
Multi-host		-3.50	
		(0.66)	
Zoonotic			-3.24
			(1.68)
1970 GDP	9.39	5.43	11.11
	(2.56)	(2.36)	(2.42)
R^2	0.41	0.54	0.38
Observations	70	70	70

Table 2: **The Relationship Between Various Types of Pathogens and Hofstede's Individualism Index** GDP is the PPP-adjusted GDP from the Penn World Tables. Human indicates pathogens that are spread directly from human to human. Zoonotic pathogens that develop in non-human hosts and are then spread to humans. Multi-host refers to pathogens that can develop in either human or non-human hosts. See section 2.1 for more details.

The key result illustrated in table 2 is that the human to human pathogens have a stronger effect on the Hofstede index than do zoonotic pathogens. The economic effect of human-to-human pathogens is nearly twice as large as the other two categories of pathogens. Furthermore, the effect on individualism is statistically significant at the 95% confidence level for human and multi-host pathogens, but not for the zoonotic pathogens.

However, by limiting ourselves to historical pathogen data, we face data availability constraints. In particular, it severely limits the number of diseases in each category. Our next step is to expand

the list of pathogens, by using more recent infectious disease data. A concern with the current analysis is that the diseases in each category are not all equally serious. With more recent data, we could then also control for the virulence, infectivity and pathogenicity of each disease.

These results are important for the next stage, identifying an effect of institutions on technology diffusion. But they are also important on their own because they point to reason why countries may have chosen different social institutions. To the extent that we have identified some causal relationship of pathogens on social structure, it suggests that social structures have evolved, in part, as a defense against the spread of disease. This seems to be at least part of reason why some societies have adopted social structures that are less well-suited to promoting technological diffusion and growth.

3.2 The Relationship between Social Institutions and Technology Diffusion

Our main result is to establish an effect of social structure on technology diffusion. Figure 3 illustrates the statistical relationship between social structure and the speed of technology diffusion. It reveals that more individualistic societies (those with more dispersed social networks) tend to also be societies where technologies diffuse quickly. In table 3, a simple regression of the CHAT measure of technology diffusion on the Hofstede index of individualism confirms that this relationship is statistically significant.

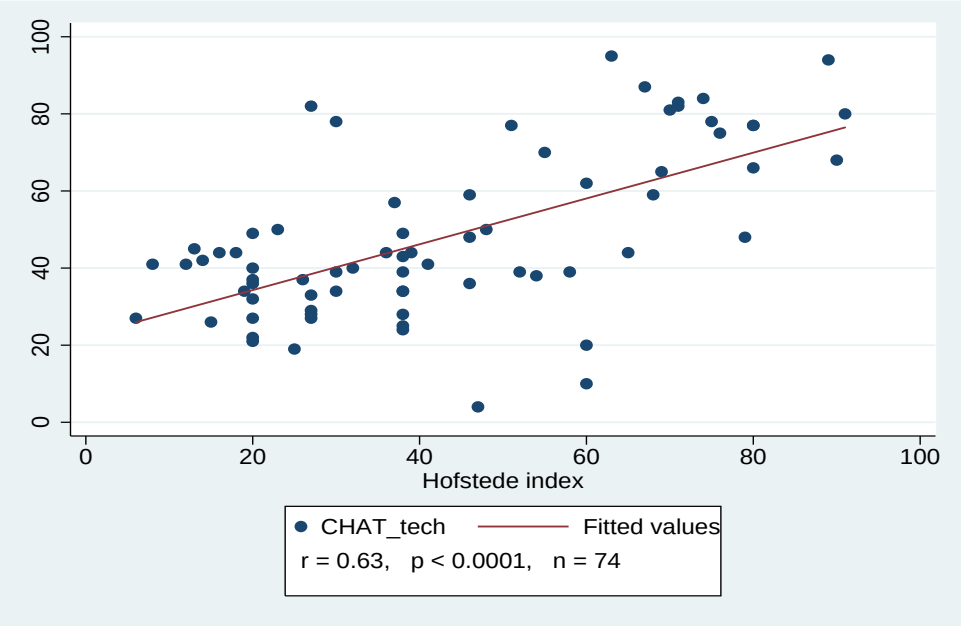


Figure 3: Comin, Hobijn, and Rovito (2006)'s technology diffusion (CHAT) measure plotted against Hofstede's individualism index.

Reverse causality is again a concern. Faster technology diffusion raises incomes, which might

well change the social structure. Likewise, the economic development that results from technology diffusion could produce a wave of urbanization, which influences social structure. The results in column 2 show that there is an effect of social structure on technology diffusion, above and beyond that which is captured by higher income and higher population density. The result that social structure predicts technology diffusion better than income or density does is a surprising one.

Dependent variable:	Technology			
	OLS	OLS	IV	IVdiff
Individualism	0.586*	0.626*	0.880*	0.742*
	(0.084)	(0.107)	(0.201)	(0.223)
1970 GDP		-0.474	-3.919	-2.05
		(2.37)	(3.32)	(3.53)
1970 density		-0.0037	-0.0023	-0.0029
		(0.0036)	(0.0038)	(0.0038)
R^2	0.40	0.46	0.41	0.45
N	75	70	70	67

Table 3: **Relationship between Social Structure and Technology Diffusion**

Technology is Comin, Hobijn, and Rovito (2006)'s measure of the number of technologies adopted in a country. Individualism is the Hofstede index. Density is the country's population density in people per square kilometer. IV uses pathogen prevalence as an instrument for individualism. IVdiff uses the difference in human/multi-host and zoonotic disease prevalence as an instrument. * indicates significance at the 5% level.

One might still be concerned that technology diffusion might affect social structure through some non-income-based channel. To alleviate concerns about this alternative type of reverse causality, we can use pathogen prevalence as an instrument for social structure. The last column of table 3 shows that instrumenting for social structure only increases the size of the effect that social structure has on technology diffusion. This effect continues to be highly significant.

The instrumental variables estimate is also important because it tells us how much of the variation in technology diffusion is due to our mechanism. It isolates the part of variation in social structure due to differences in disease prevalence and quantifies the importance of this variation for technology diffusion. We find that 46% of the variation in technology diffusion is due to social structure, income and density. When we isolate the part of social structure due to differences in pathogen prevalence, we still explain 41% of that variation.

Of course, the prevalence of disease itself is not exogenous. Even in the theory, it is a result of the social network structure. In reality, it is heavily dependent on a country's income and health infrastructure. Therefore, the final step uses the difference in human-to-human and zoonotic pathogen prevalence as an exogenous instrument to isolate the effect of social structure on technological advancement. While greater levels of development spur public health initiatives, these measures prevent the human transmission and the animal transmission of diseases. Likewise, better

health care lowers mortality rates from both types of diseases.

In the column labeled IVdiff in table 3, the coefficient on individualism is highly statistically significant and large economically. The Hofstede index ranges from 16 (Indonesia) to 148 (Norway). Its standard deviation is 28. Thus an one-standard-deviation increase in individualism results in 21 (28×0.742) additional technologies being in use in a country.

4 Quantifying the Potential Effect on Technology

An important potential concern about using this model to explain income differences across countries is the worry that its effect is trivial. This concern is especially pressing because disease prevalence rates typically differ only by a few of percentage points across countries while differences in incomes can be 100-fold.

What our calibration exercise shows is that small differences in disease prevalence can produce strikingly different technology diffusion rates. The reason is that the utility cost of catching a disease and dying from it is much greater than the utility benefit of producing with an incrementally better technology. Since utility is very sensitive to the disease state, choices react greatly to small changes in disease prevalence. These changes in network choice produce differences in technology diffusion rates, which could explain a modest part of the disparity in countries' incomes.

4.1 Calibrated Parameters

To know whether small differences in disease produce big differences in technology, we need to choose some realistic parameter values for our model and analyze the simulated model outcomes. The key parameters in the model are the probabilities of disease and technology transmission, the initial pathogen prevalence rate and the rate of arrival of new technologies. These parameters are summarized in table 4.1.

	Parameter	Value	Target
Initial disease prevalence	$\text{Prob}(n_j(0) = 0)$	0.5% (high) 0.035% (low)	TB death rate China in New Zealand
Disease transmission probability	p_ψ	31%	disease steady state in dispersed netwk
Technology arrival rate	λ	5%	2% growth rate in low-disease country
Technology transfer probability	p	50%	Half-diffusion in 19 years (Comin et. al. '06)

For the initial pathogen prevalence rate, we will use a high and a low value and compare them. These high and low values are the max and min across all countries of the deaths from tuberculosis, per 1,000 inhabitants per year. Tuberculosis is the most common cause of death in our sample. Note that these are mortality rates, not infection rates. Since individuals who get sick in the model die, this is the relevant comparison. Also, it is a conservative calibration because it would be easier to get large effects out of the higher disease prevalence rates. The probability of disease transmission is chosen to make the initial prevalence rate equal to the steady state rate of infection. Thus, the economy starts with a given fraction of the population being sick and each sick person represents an independent 31% risk (π) of passing the disease on to everyone that person is directly connected to.

Everyone starts with a technology level of 1. But each period, there is a chance that any given person may discover a new technology that raises their productivity by one percent. The rate of arrival of new technologies is calibrated so that the dispersed network economy (more likely to be the developed economy in the data) grows at a rate of 2% per year. The probability of transmitting a new technology to each person that one is connected to (λ) is chosen to explain the fact that for the average technology, the time between invention and when half the population has adopted the technology is approximately 19 years (Comin, Hobijn, and Rovito, 2006).

We simulate the high and low disease prevalence economies each with clustered and dispersed networks. In this example, the economy consists of 1000 people, each with 4 friends. The next step will be to compare the utility from the clustered and dispersed networks. That exercise requires calibrating the utility function. There we will use values that are standard in the literature: The discount rate β is 0.99 and the CRRA preference parameter γ is 0.5.

4.2 Simulation Results

First, we show the process by which technologies and diseases spread in a small-scale illustrative example. Then, we consider the calibrated simulation with many agents and many periods, averaged over many runs to get a more precise idea of the aggregate effect of a network.

Figure 4 illustrates the diffusion of technology and disease. Each box represents a person/date combination. Time is on the horizontal axis. People are lined up on the vertical axis according to their location. In the first period (first column of boxes on the left), everyone starts with the same technology level. But there are a few agents who have a disease (the darkest boxes).

By the second period, new ideas start to arrive. In the second column of boxes, there are a couple of lighter-colored boxes that indicate that these agents have reached the next technology level. In the clustered network (left figure), some agents who are adjacent to or 2 places away from agents that were sick in period 1 are now sick. In the dispersed network (right figure), some

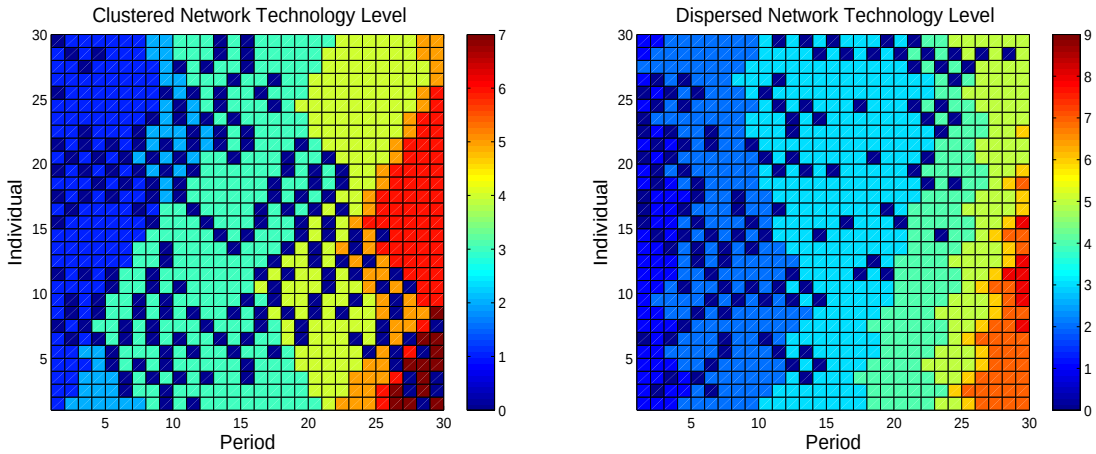


Figure 4: Spread of disease and evolution of technology in a clustered network (left) and a dispersed network (right). The darkest boxes indicate individuals who acquired the disease in period t and therefore have zero time-t productivity. Warmer colors indicate higher levels of technology.

agents who are adjacent to or 4 places away from agents that were sick in period 1 are now sick. In period 3, the new ideas that arrived in period 2 start to diffuse to nearby locations. In the clustered network, individuals are still using the initial technology level in period 8. In the dispersed network, all the healthy agents have adopted the second technology level after period 5. (In the calibrated model, this diffusion process takes longer. We sped up technology diffusion in this example to make it easier to see.)

After 30 periods, the most technologically advanced agents in the clustered network only realize 7 steps in the quality ladder. In the dispersed network, some agents operate at 9 steps. Since each innovation represents a 5% productivity increase, being two steps further represents a 10% higher degree of productivity. Of course, this is just an illustrative example. It is a comparison of the maximum level of technology from a small number of agents. To get a sense of the aggregate effect, we average the technology level over 1000 agents and 30 independent runs.

Figure 5 plots the average disease prevalence (times 10,000) and the average technology level for the whole population over 200 years. The fraction of the population infected with disease is significantly higher in the dispersed network society. In fact, the clustered networks inhibit the spread of disease so much that it becomes extinct in this calibration.

However, having a dispersed network results in technology that grows at 2.0% per year. This is true by construction because it was one of the calibration targets. But the economy with the clustered network grows at only 1.8% per year. While the difference in growth rates is small, in time, it produces large level differences. After 200 years, the average level of technology is about 60% higher in the dispersed network than in the clustered network. This simple example makes the point that a difference in network structure can create a small friction in technology diffusion.

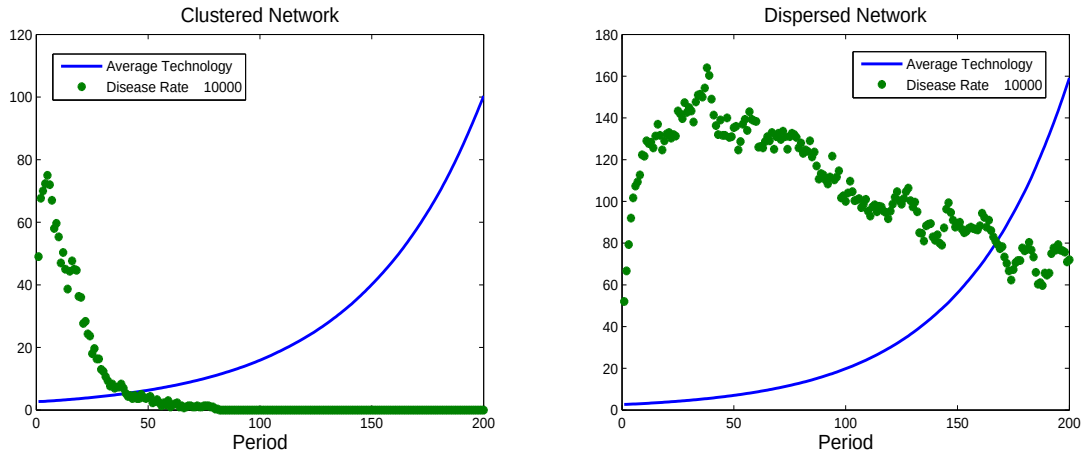


Figure 5: Prevalence of disease ($\times 10$) and average technology level in a clustered network (left) and a dispersed network (right).

When cumulated over a long time horizon, this small friction has the potential to explain large differences in national incomes.

5 Conclusions

Our results are consistent with the idea that countries with high pathogen prevalence tend to choose social structures where people have more friends in common. These social structures may be very persistent. This allows them to inhibit or facilitate technology diffusion and become an important determinant of a country's level of development.

The next steps in this project include calibrating the model to data on disease prevalence, infectiousness and virulence so that we can accurately predict the type of social network each country would optimally form. Then, using facts about technology diffusion, we could calibrate the rate of technology transmission and get some estimates for the amount of variation in national productivity levels that social structure might account for.

Another step left to be done is to use the difference between diseases that are socially transmitted from those that are not to identify the effect of social structure on technology diffusion. While both types of diseases make people sick, retard productivity growth and reduce income, only the socially transmissible diseases should rationally affect the types of social connections people choose to form. Thus, the additional effect of socially transmissible diseases on social structure and in turn, on technology diffusion and income levels could be attributed to our mechanism.

References

Acemoglu, D., and S. Johnson (2005): "Unbundling Institutions," *Journal of Political Economy*.

Acemoglu, D., S. Johnson, and J. Robinson (2002): "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distributions," *Quarterly Journal of Economics*, CXVII(4), 1231–1294.

Algan, Y., and P. Cahuc (2007): "Social attitudes and Macroeconomic performance: An epidemiological approach," Paris East and PSE Working Paper.

Binford, L. (2001): *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Ethnographic and Environmental Data Sets*. University of California Press.

Bisin, A., and T. Verdier (2000): "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, 115(3), 955–988.

Bisin, A., and T. Verdier (2001): "The Economics of Cultural Transmission and the Evolution of Preferences," *Journal of Economic Theory*, 97(2), 298–319.

Cole, H., G. Mailath, and A. Postlewaite (1992): "Social Norms, Savings Behavior, and Growth," *Journal of Political Economy*, 100(6), 1092–1125.

Coleman, J. (1988): "Social Capital in the Creation of Human Capital," *American Journal of Sociology*, 94, S95–S120.

Comin, D., B. Hobijn, and E. Rovito (2006): "Five Facts You Need to Know About Technology Diffusion," NBER Working Paper 11928.

Fernández, R., and A. Fogli (2005): "An Empirical Investigation of Beliefs, Work and Fertility," NBER Working Paper 11268.

Granovetter, M. (2005): "The Impact of Social Structure on Economic Outcomes," *The Journal of Economic Perspectives*, 19(1), 33–50.

Greif, A. (1994): "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies," *Journal of Political Economy*, 102, 912–950.

Hofstede, G. (2001): *Culture's consequences : comparing values, behaviors, institutions, and organizations across nations*. Sage Publications.

Jackson, M. (2008): *Social and Economic Networks*. Princeton University Press.

Rauch, J., and A. Casella (2001): *Networks and Markets*. Russell Sage, first edn.

Smith, K., D. Sax, S. Gaines, V. Guernier, and J.-F. Gugan (2007): "Globalization of Human Infectious Disease," *Ecology*, 88(8), 1903–1910.

Tabellini, G. (2005): "Culture and Institutions: Economic Development in the Regions of Europe," CESifo Working Paper No.1492.

Thornhill, R., C. Fincher, D. Murray, and M. Schaller (2010): "Zoonotic and Non-Zoonotic Diseases in Relation to Human Personality and Societal Values: Support for the Parasite-Stress Model," *Evolutionary Psychology*, 8(2), 151–169.

A Proofs of Propositions

Proof of result 1. Proof: Without loss of generality, consider the agent in the last position, the agent with location n on the circle.*Case 1: n even.* If n is even, then the farthest node from n is $n/2$. If each person is connected to the ϕ closest people, where ϕ is even, then they are connected to $\phi/2$ people on either side. Therefore, the shortest path will be the one that advances $\phi/2$ places around the circle, at each step in the path, until it is within $\phi/2$ nodes of its end point. For example, agent n reach $\phi/2$ in one step, ϕ in two steps and $n/2$ in $(n/2)/(\phi/2) = n/\phi$ steps, if n/ϕ is an integer. If dividing n by ϕ leaves a remainder m, then one step in the path to reach $n/2$ must be only $m < n/2$ nodes away. Thus, when n is even, the shortest path to the furthest node $n/2$ is $ceil(n/\phi)$, where $ceil(x) = x$ if x is an integer, and is otherwise, the next largest integer.

Case 2: n odd. If n is odd, then $(n-1)/2$ and $(n+1)/2$ are equally far from node n. Each is $(n-1)/2$ nodes away. Following the same logic as before, the shortest path will be the one that advances $\phi/2$ places around the circle, and reaches the furthest node in $ceil((n-1)/2)/(\phi/2) = ceil((n-1)/\phi)$ steps.

Lastly, note that when n is even, $ceil(n/\phi) = ceil((n-1)/\phi)$. Note that, since $\phi > 1$ and both ϕ and n are integers, $ceil(n/\phi)$ and $ceil((n-1)/\phi)$ will only differ if $(n-1)/\phi$ is an integer, so that adding $1/\phi$ to it will make $ceil(n/\phi)$ the next largest integer. But if ϕ is even and $(n-1)/\phi$ is an integer, then $n-1$ must be even, which makes n odd. Thus, $ceil(n/\phi) = ceil((n-1)/\phi)$.□

Proof of result 2. Proof: Without loss of generality, consider the distance from the last node, n. n can be connected to nodes 1 though $\phi/2$ and $n-1$ through $n-\phi/2$ in 1 step. More generally, it can be connected to nodes $(s-1)\phi/2+1$ through $s\phi/2$ and $n-(s-1)\phi/2-1$ through $n-s\phi/2$, in s steps. For each s, there are ϕ nodes for which the shortest path length to n is s steps. We know from result 1 that when ϕ is even and n/ϕ is an integer, the longest path length (the diameter) is n/ϕ. Thus, the average length of the path from n to any other node is $1/n\sum_{s=1}^{n/\phi}\phi s$. Using the summation formula, this is $(\phi/n)(n/\phi)(n/\phi+1)/2 = 1/2 + n/(2\phi)$. □

Proof of result 3. *The diameter of a dispersed network, with $n > 4$ nodes where each node i is connected to $i-4$, $i-1$, $i+1$, and $i+4$, is round$(n/8)+1$.*

Proof: Without loss of generality, consider distances from the agent located at node n. n can reach nodes 1, 4, $n-1$ and $n-4$ in one step. It can reach nodes 2, 3, 5, 8 and $n-2$, $n-3$, $n-5$ and $n-8$ in two steps. In any number of steps $s > 1$, agent n can reach nodes $4(s-2)+2, 4(s-1)-1, 4(s-1)+1, 4s$ (moving clockwise around the circle) as well as $n-4(s-2)-2, n-4(s-1)+1, n-4(s-1)-1, n-4s$ (moving counter-clockwise).

Let the operator $floor(x)$ be the largest integer y such that $y \leq x$. Define $\tilde{n} \equiv 4 * floor(n/8)$. Then $\tilde{r} \equiv n - 2 * \tilde{n}$ is the remainder when n is divided by 8. There are eight cases to consider, one for each possible value of $\tilde{r}$.

Case 1: $\tilde{r} = 0$. If the total number of nodes in the network n is a multiple of 8, then it takes $(1/4) * n/2$ steps to connect node n with node $n/2$, the geographically farthest node in the network. But it takes one more step to reach $n/2-1$, $n/2+1$. The nodes $n/2-2$ and $n/2+2$ can be reached in 2 steps from $n/2-4$ and $n/2+4$, each of which is one step closer to n than $n/2$ is. Thus, every node can be reached in $n/8+1$ steps, making the diameter of the network $n/8+1$.

Case 2: $\tilde{r} = 1$. In this case, $\tilde{n}$ and $\tilde{n}+1$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps. But it takes one more step to reach $\tilde{n}-1$, $\tilde{n}-2$, $\tilde{n}+2$ or $\tilde{n}+3$. Since $\tilde{n} = 4floor(n/8)$, $\tilde{n}/4 = floor(n/8)$, and thus the diameter is one step more than that, which is $floor(n/8)+1$.

Case 3: $\tilde{r} = 2$. In this case, $\tilde{n}$ and $\tilde{n}+2$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps. But it takes one more step to reach $\tilde{n}-1$, $\tilde{n}-2$, $\tilde{n}+1$, $\tilde{n}+3$ or $\tilde{n}+4$. Thus, the diameter is again $floor(n/8)+1$.

Case 4: $\tilde{r} = 3$. In this case, $\tilde{n}$ and $\tilde{n}+3$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. It is still the case that it takes one more step to reach $\tilde{n}-1$, $\tilde{n}-2$ and $\tilde{n}+1$. $\tilde{n}+2$ can be reached in one

additional step from $\tilde{n}+3$, as can $\tilde{n}+4$. And $\tilde{n}+5$ can be reached in 2 additional steps from $\tilde{n}+4$, which is one step closer to n than $\tilde{n}+3$. Thus, every node can still be reached in $floor(n/8)+1$ steps.

Case 5: $\tilde{r}=4$**.** In this case, $\tilde{n}$ and $\tilde{n}+4$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. But now, getting to $\tilde{n}+2$ requires 2 additional steps. Thus, the diameter of this network is $floor(n/8)+2$.

Case 6: $\tilde{r}=5$**.** In this case, $\tilde{n}$ and $\tilde{n}+5$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. Getting to either $\tilde{n}+2$ or $\tilde{n}+3$ requires 2 additional steps. Thus, the diameter of this network is $floor(n/8)+2$.

Case 7: $\tilde{r}=6$**.** In this case, $\tilde{n}$ and $\tilde{n}+6$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. In one additional step, one can connect from $\tilde{n}$ to $\tilde{n}+1$ or $\tilde{n}+4$ or from $\tilde{n}+6$ to $\tilde{n}+2$ or $\tilde{n}+5$. It takes two additional steps from $\tilde{n}$ to connect to $\tilde{n}+3$. Thus, the diameter of this network is $floor(n/8)+2$.

Case 8: $\tilde{r}=7$**.** In this case, $\tilde{n}$ and $\tilde{n}+7$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. In one additional step, one can connect from $\tilde{n}$ to $\tilde{n}+1$ or $\tilde{n}+4$ or from $\tilde{n}+7$ to $\tilde{n}+3$ or $\tilde{n}+6$. It takes two additional steps from either $\tilde{n}$ or $\tilde{n}+7$ to connect to $\tilde{n}+2$ or $\tilde{n}+5$. Thus, the diameter of this network is $floor(n/8)+2$.

The one condition that encapsulates all 8 of these cases is diameter=round$(n/8)+1$. To see this, recall that $\tilde{r}$ is the remainder when n is divided by 8. When this remainder is zero, then $(n/8)+1$ =round$(n/8)+1$. When this remainder is less than 4, then floor$(n/8)+1$ =round$(n/8)+1$. When this remainder is 4 or more (4-7), then round$(n/8)$ =floor$(n/8)+1$, and therefore floor$(n/8)+2$ =round$(n/8)+1$. Thus, in each case of the 8 cases, the diameter of the network is equal to round$(n/8)+1$.□

Proof of result 4. *In the example dispersed network, when $n/8$ is an integer, the average path length is $7/8+n/16$. This is less than the average path length in a clustered network with $\phi=4$, when the network is large ($n>6$).*

Proof: Without loss of generality, consider distances of each node from node n. n can reach 4 different nodes: 1, 4, $n-1$ and $n-4$ in one step. It can reach 8 different nodes 2, 3, 5, 8 and $n-2$, $n-3$, $n-5$ and $n-8$ in two steps. More generally, for a number of steps $s\geq 2$, agent n can reach 8 new nodes with each step. These nodes are: $4(s-2)+2, 4(s-1)-1, 4(s-1)+1, 4s$ (moving clockwise around the circle) as well as $n-4(s-2)-2, n-4(s-1)+1, n-4(s-1)-1, n-4s$ (moving counter-clockwise). This rule holds until the number of steps s reaches $n/8$, the number of steps to travel approximately half way around the circle. At that point, the number of additional nodes that can be reached in an additional step depends on the size of the network. There are 8 cases to consider.

Recall that $\tilde{n}\equiv 4*floor(n/8)$ and that $\tilde{r}\equiv n-2*\tilde{n}$ is the remainder when n is divided by 8. There are eight cases to consider, one for each possible value of $\tilde{r}$.

If the total number of nodes in the network n is a multiple of 8, then it takes $n/8$ steps to connect node n with node $n/2$. Using the algorithm above, it also takes $n/8$ steps to connect with nodes $n/2-6, n/2-5, n/2-3, n/2+6, n/2+5$ and $n/2+3$. But this is 7 total nodes instead of 8 total nodes because when the total number of steps being considered is $n/8$ ($s=n/8$) nodes $4s$ and $n-4s$ are both equal to node $n/2$.

It takes one more step to reach $n/2-1$, $n/2+1$. The nodes $n/2-2$ and $n/2+2$ can be reached in 2 steps from $n/2-4$ and $n/2+4$, each of which is one step closer to n than $n/2$ is. Thus, 4 additional nodes can be reached in $n/8+1$ steps.

Counting up, there is 1 node (n) reachable in zero steps, 4 nodes reachable in 1 step, 8 nodes reachable in s steps for $s\epsilon\{2,3,\ldots,n/8-1\}$, 7 nodes reachable in $n/8$ steps and 4 nodes reachable in $n/8+1$ steps. That makes the average path length $1/n$ times the sum of all the path lengths to the n nodes: $1/n[4+8\sum_{s=2}^{n/8-1}s\ +7*n/8+4*(n/8+1)]$. Applying the summation formula, $8\sum_{s=2}^{n/8-1}s=8(n/8)(n/8-1)/2-8$, where the -8 corrects for the fact that the sum begins at $s=2$, rather than at $s=1$. Substituting in this formula and collecting terms, this is $1/n[4+8(n/8)(n/8-1)/2-8+11n/8+4]=1/8n[n(n-8)/2+11n]=7/8+n/16$. □

Proof of result 5 *For a large network ($n>8$) where $n/8$ is an integer, the dispersed network has a smaller diameter and a shorter average path length than a clustered network with equal size n and equal degree $\phi=4$.*

Proof: We begin with the diameter. If $n/8$ is an integer, then n must also be a multiple of 4. Since the diameter of the clustered network with $\phi=4$ is $(n-1)/4$ or the next highest integer, that integer-valued diameter will be $n/4$. Likewise, in the dispersed network, round$(n/8)+1=n/8+1$. Thus, the diameter of the dispersed network is smaller iff $n/8+1<n/4$, which is true iff $n>8$.

Next, we turn to average path length. If $n/8$ is an integer and $\phi=4$, then result 2 tells us that the average path length of a clustered network is $1/2+n/8$. Result 4 tells us that the average path length of the dispersed network is $7/8+n/16$. Thus, the dispersed path length is smaller iff $7/8+n/16<1/2+n/8$, which is true iff $n>6$. Thus, $n\geq 8$ is a sufficient condition for the dispersed network to have a shorter average path length. □

Proof of result 6 *In a clustered network, where $\phi = 4$, there are n unique triples.*

Claim 1: Any three adjacent nodes are a triple.

Proof: Consider nodes j, $j+1$ and $j+2$. Since every node is connected to its adjacent nodes, $j+1$ is connected to j and $j+2$. And since every node is also connected to nodes 2 places away, j is connected to $j+2$. Since all 3 nodes are connected to each other, this is a triple.

Claim 2: Any sets of 3 nodes that are not 3 adjacent nodes are not a triple.

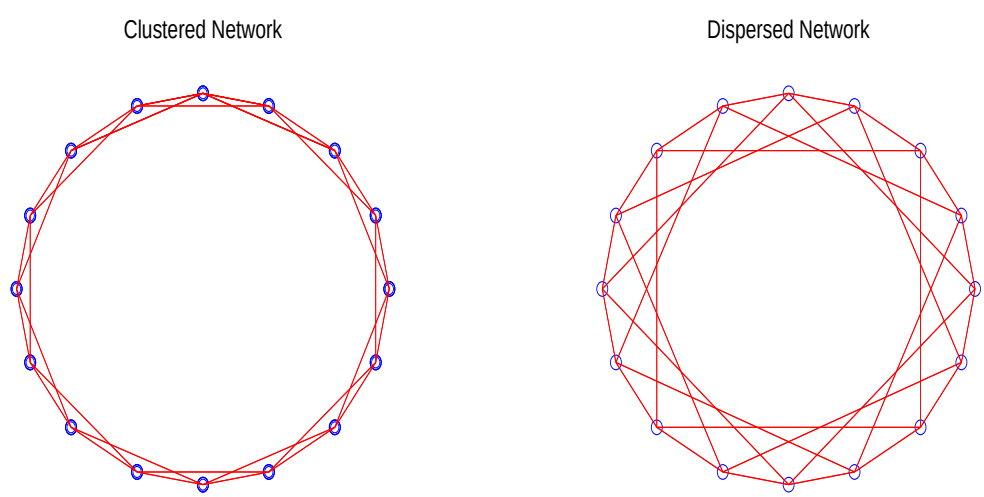
Proof: Consider a set of 3 nodes. If the nodes are not adjacent, then two of the nodes must be more than 2 places away from each other. Since in a clustered network with $\phi = 4$, nodes are only connected with other nodes that are 2 or fewer places away, these nodes must not be connected. Therefore, this is not a triple.

Thus, there are n unique sets of 3 adjacent nodes (for each j there is one set of 3 nodes centered around j: $\{j-1, j, j+1\}$). Since every set of 3 adjacent nodes is a triple and there are no other triples, there are n triples in the network. □

Proof of result 7 *In a dispersed network, where where each person i is connected to $i-\psi$, $i-1$, $i+1$, and $i+\psi$, where $\psi > 2$, there are zero triples.*

Proof: Consider each node connected to an arbitrary i, and whether it is connected to another node, which is itself connected to i. In addition to being connected to i, node $i-\psi$ is connected to $i-2\psi$, $i-\psi-1$, and $i-\psi+1$. None of these is connected to i. Node $i-1$ is also connected to $i-2$, $i-\psi-1$ and $i+\psi-1$. But none of these is connected to i. Node $i+1$ is also connected to $i+2$, $i-\psi+1$ and $i+\psi+1$. But none of these is connected to i. Finally, node $i-\psi$ is also connected to $i+\psi-1$, $i+\psi+1$ and $i+2\psi$. But none of these is connected to i. Therefore, there are no triples among any connections of any arbitrary node i. □

B Visual Representations of our Clustered and Dispersed Networks



The connection matrix N of the clustered network is

$$\begin{bmatrix} 0 & 1 & 1 & 0 & \dots & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & \dots & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & \dots \\ & & & \vdots & & & & \\ 1 & 1 & 0 & \dots & 0 & 1 & 1 & 0 \end{bmatrix}.$$

This matrix had zeros on the diagonal. (Typically, we don't consider one's relationship with oneself to be a connection.) It has two ones just to the left and right of the diagonal, indicating that each person is connected to the two people to their left and the two people to their right. The three entries in the top-right and bottom-left corners also have ones. This captures the connection between agents 1 and 2, who are located adjacent to agents $n-1$ and n on the circle. The rest of the entries are zeros, indicating that these individuals are not directly connected in the network.

The connection matrix N of our example dispersed network is

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 \\ & & & & & \vdots & & & & & & \\ 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Again, there are zeros on the diagonal. There is one 1 entry just to the left and to the right of the diagonal. This represents each agent's connection with their immediate neighbor. There is also a 1 four columns to the left and four columns to the right of the diagonal, indicating the connection between agent j and $j+4$, and between agent j and $j-4$. As before, there are a handful of 1's in the top-left and bottom-right corners, indicating the connections between agents near n and those near 1, who are one or four spots away from each other on the circle. The rest of the entries are zeros, indicating that these individuals are not directly connected in the network.

C Data Details

Pathogen prevalence The pathogen prevalence measure used in these baseline regressions is from Murray and Schaller "Historical Prevalence of Infectious Diseases within 230 geopolitical regions: A Tool for investigating the origins of culture". They extended the work of Gangestad and Buss (1993) who employed old epidemiological atlases to rate the prevalence of seven different kinds of disease-causing pathogens and combined estimates into a single measure indicating the historical prevalence of pathogens in each of 29 countries. More recently, Murray and Schaller used a similar procedure to rate the prevalence of nine infectious diseases in each of 230 geopolitical regions world. The nine diseases coded were leishmanias, schistosomes, trypanosomes, leprosy, malaria, typhus, filariae, dengue, and tuberculosis. Epidemiological atlases were used to estimate the prevalence of each of these nine diseases in each region. For eight of them (excluding tuberculosis), prevalence of each disease was based primarily on epidemiological maps provided in Rodenwaldt and Bader's (1952-1961) *World-Atlas of Epidemic Diseases* and in Simmons and others (1944) Global Epidemiology. A 4-point coding scheme was employed: 0 = completely absent or never reported, 1 = rarely reported, 2= sporadically or moderately reported, 3 = present at severe levels or epidemic levels at least once. The prevalence of tuberculosis was based on a map contained in the National Geographic Society's (2005) *Atlas of the World*, which provides incidence information in each region for every 100,000 people. Prevalence of tuberculosis was coded according to a 3-point scheme: $1 = 3-39$, $2 = 50-99$, $3 = 100$ or more. For 160 political regions, they were able to estimate the prevalence of all nine diseases. The remaining 70 regions typically lacked historical data on the prevalence of either tuberculosis or leprosy; 6 of these regions lacked data on malaria as well. Therefore, in addition to create a 9 item index of disease prevalence (computed for 160), they also created a seven item index (excluding both leprosy and tuberculosis) for 224 regions and a six item index (excluding also malaria) for 230 regions. To ensure all different disease prevalence indices were computed on a common scale of measurement, all nine disease prevalence ratings were standardized by converting them to z scores. Each overall disease prevalence index was then computed as the mean of z scores of the items included in the index. Thus, for each index the mean is approximately 0, positive scores indicate disease prevalence that is higher than the mean and negative scores indicate disease prevalence that is lower than the mean.

Hofstede Index Hofstede defines individualism in the following way: "Individualism (IDV) on the one side versus its opposite, collectivism, that is the degree to which individuals are integrated into groups. On the individualist side we find societies in which the ties between individuals are loose: everyone is expected to look after him/herself and his/her immediate family. On the collectivist side, we find societies in which people from birth onwards are integrated into strong, cohesive in-groups, often extended families (with uncles, aunts and grandparents) which continue protecting them in exchange for unquestioning loyalty."

The original questions from the 1966-1973 Hermes (IBM) attitude survey questionnaires used for the international comparison of work-related values were listed in Hofstede (1980, Appendix 1). Appendix 4 of the same book presented the first Values Survey Module for future cross-cultural studies. It contained 27 content questions and 6 demographic questions. This VSM 80 was a selection from the IBM questionnaires, with a few questions added from other sources about issues missing in the IBM list and judged by the author to be of potential importance. In the 1984 abridged paperback edition of Hofstede (1980) the original IBM questions were not included, but the VSM 80 was.

A weakness of the VSM 80 was its dependence on the more or less accidental set of questions used in the IBM surveys. Therefore in 1981 Hofstede through the newly-founded Institute for Research on Intercultural Cooperation (IRIC) issued an experimental extended version of the VSM (VSM 81). On the basis of an analysis of its first results, a new version was issued in 1982, the VSM 82. This was widely used.

The VSM 82 questionnaire is too long to include in its entirety. However, factor analysis of 14 "work goals" questions from the survey produced 2 factors that together explained 46% of the variance. The first factor was demographic characteristics. The second set pertain to work goals. Here are those key work goals questions:

In this section, we have listed a number of factors which people might want in their work. We are asking you to indicate how important each of these is to you. Possible answers: of utmost importance to me (1), very important (2), of moderate importance (3), of little importance (4), of very little or no importance. "How important is to you":

1. Have challenging work to do
2. Live in an area desirable to you and your family
3. Have an opportunity for high earnings
4. Work with people who cooperate well with each other
5. Have training opportunities
6. Have good fringe benefits
7. Get the recognition you deserve when you do a good job
8. Have good physical working conditions
9. Have considerable freedom to adopt your own approach to the job
10. Have the security that you will be able to work for your company as long as you want to
11. Have an opportunity for advancement to higher level job
12. Have a good working relationship with your manager
13. Fully use your skills and abilities on the job
14. Have a job which leaves you sufficient time for your personal or family life

The answers to these questions were used to develop the Hofstede index of individualism for each country.

D Robustness: Alternative Measures of Social Structure

The hardest variable to quantify is clearly the social structure. While Hofstede's definition of collectivism is a reasonable fit with the notion of clustered networks in the model, exploring alternative interpretations of social structure and network type would give us added confidence that we are measuring a concept similar to that in the model. To that effect, we consider alternative indicators of social structure. Below, we describe some of these alternative measures.

D.1 A Geographic Measure of Clustering: Range Size

Another way of interpreting network clustering is geographically. How physically separated is one community from another? The greater the degree of separation, the more difficult it will be for pathogens and technologies from diffuse from one society to another.

Looking at evidence from historical hunter-gatherer societies, Binford (2001) measures the geographic clustering as the total land in square kilometers occupied by the group. He calls this measure "range size." The larger the range size, the less connected societies are with one another and the greater the degree of clustering. Societies with a small range size are more likely to have connections outside of their communities, as do individuals in the dispersed network.

The 339 groups Binford examines are primarily in developing countries and their economies are based mainly on hunting, fishing and gathering. The groups' populations range from 22 to 13,000 members, with an average size of 812.

The pathogens data measures prevalence by country. These clustering measures are not country-level measures of clustering, but rather group-level measures. The groups are not representative of the country. However, we have geographic characteristics of the land occupied by each group that we can use to impute the degree of pathogen prevalence.

First, we regress pathogen prevalence at the national level on national-level geographic characteristics. Then, we use the coefficients from the regression and the same geographical characteristics measured for each group to produce a predicted level of pathogen prevalence for the group. The advantage of this procedure is that the geography measures used to impute pathogen prevalence are exogenous, immutable features of the group's surroundings. By using exogenous variables to impute pathogens, we remedy concerns about reverse causality.

Dependent variable:	Range Size (IV)
Pathogens	0.189
	(0.373)
Population	0.650
	(0.051)
Hunters	0.044
	(0.003)
Gatherers	0.0094
	(0.0040)
R^2	0.59
Observations	339

Table 4: More disease corresponds to societies with greater geographic clustering. Geographic variables are instruments for pathogen prevalence.

Table 4 shows that this geographic measure of social structure exhibits the same relationship with pathogen prevalence. When pathogens are more prevalent, societies are more geographically isolated, more clustered. This result also holds after controlling for number of moves and distanced moved, and for residuals clustered by country.

D.2 World Values Survey

The 2005 World Values Survey asked a number of questions that are indicative to the type of social structure in the respondent's country. Below, we describe each survey question we use and how it relates to our network model.

Nuclear family This question asks how often the respondent visits their nuclear family. People who visit their nuclear family often have stronger bonds with those family members, who are connected to each other. Such a society would be one with more triples.

Associations This question asks if the respondent has participated more than twice in a social organization, such as a sports, political, or charitable organization. Such organizations create communities of people who know each other and increase the presence of triples.

Jobs This question asks the respondent how he or she found their current job. Did they find out about the job from friends/relatives, or from an agency, school or advertisement? If the job was discovered through a personal connection, this indicates less of a market-based society and one where connections between people who know each other is more important. If you need to be introduced by a friend or family member to someone in order to have an economic relationship, this signifies the importance of trust and the presence of many relationship triples.