Germs, Social Networks and Growth

Alessandra Fogli and Laura Veldkamp*

June 27, 2012

Abstract

Does the pattern of social connections between individuals matter for macroeconomic outcomes? If so, how does this effect operate and how big is it? Using network analysis tools, we explore how different social structures affect technology diffusion and thereby a country's rate of technological progress. The network model also explains why societies with a high prevalence of contagious disease might evolve toward growth-inhibiting social institutions and how small initial differences can produce large divergence in incomes. Empirical work uses differences in the prevalence of diseases spread by human contact and the prevalence of other diseases as an instrument to identify an effect of social structure on technology diffusion.

How does the pattern of social connections between individuals affect a country's income? Macroeconomists 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 structures and allows us to quantify the potential size of these effects. Motivated by the model, we use differences in the prevalence of diseases spread by human contact and the prevalence of other diseases as an instrument to identify an effect of social structure on technology diffusion.

There is a long history 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

*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 and seminar participants at USC, Stanford, Chicago, Western Ontario, Minnesota, Penn State, George Washington, and NYU, and participants in the Munich conference on Cultural Change and Economic Growth, the SED, SITE, NBER Macro across Time and Space, and NBER growth meetings for their comments and suggestions. We thank Corey Fincher and Damian Murray for help with the pathogen data, Pascaline Dupas, Diego Comin and Marti Mestieri, for useful comments, and Isaac Baley and David Low for invaluable research assistance. 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]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.

disseminate new technologies more efficiently than others and append a production economy where the average technology level is related to output and income. There are two problems with this explanation. First, social contacts are presumably endogenous. If so, why would a social structure that inhibits growth persist? Second, this explanation is difficult to quantify or test. How might we determine if its effects are trivial or not? While researchers have mapped social networks in schools or on-line communities (Jackson, 2008), mapping the exact social network structure for an entire economy is not feasible.

Our theory for why some societies have growth-inhibiting social structures revolves around the idea that communicable diseases and technologies spread in similar ways - through human contact. We explore an evolutionary model, where some people favor local "collectivist" social networks and others do not. People who form collectives are friends with each others' friends. The collective has fewer links with the rest of the community. This limited connectivity reduces the risk of an infection entering the collective, allowing the participants to live longer. But it also restricts the group's exposure to new technologies, making them, on average, poorer. In contrast, having an individualist social structure with fewer mutual friendships brings the benefit of faster technology diffusion. Faster diffusion increases one's expected productivity, which encourages new entrants to adopt this trait. In countries where communicable diseases are inherently more prevalent, the longer lifetimes of collectivist types will allow them to flourish and drive out the individualist types. A collectivist social structure that inhibits the spread of disease and technology will emerge. In countries where communicable diseases are less prevalent, the individualist types will be more economically successful, new entrants will adopt their more successful trait and individualist types will drive out collectivist types.

The idea that disease prevalence and social structure are related can help to isolate and quantify the effect of social structure on technology diffusion. Isolating this effect is a challenging task because technology diffusion and social structure both affect each other: Technology diffusion is a key determinant of income, which may well affect a country's social structure. To circumvent this problem, we instrument for social structure using disease prevalence data. By itself, disease prevalence would be a poor instrument because it is not likely to be exogenous: higher income levels would likely translate into better health and lower disease levels. Therefore, our instrument uses differences in the prevalence of two types of disease. The first type is diseases that are spread directly from person-to-person. These diseases might plausibly affect social structure because changing one's relationships with others can prevent transmission. The second type of diseases are those transmitted only through animals. Since direct human contact does not affect one's probability of infection, the prevalence of such diseases should not affect social structure. Thus, the main contribution of the paper is to use the difference in prevalence of communicable disease and animal-

transmitted disease as an instrument to measure the effect of social structure on income.

Our model explains why communicable disease might be correlated with social structure and how social structure can influence a country's technology diffusion and average productivity. We isolate one particular aspect of social structure, its degree of individualism versus collectivism, while holding all other aspects of the network fixed. Of course, many of these other aspects of networks may also differ across countries. We isolate collectivism because we have cross-country data measuring it. But measuring other aspects of social networks and understanding their effects on economic growth would be useful topics for further research.

Section 1 begins by considering two exogenous networks, a collectivist and an individualist one. It describes the effect of collectives on disease and technology diffusion. Then, it considers networks that evolve and explores the reverse effects: how technology and disease affect the survival of individualist and collectivist types in the network. Numerical simulations in section 2 illustrate how these forces interact. It shows that higher disease prevalence creates the conditions for collectivist networks to emerge. Collectivist networks slow technology diffusion, which over time, can explain large income differences between collectivist and individualist societies.

Section 3 describes the historical pathogen prevalence data we collected from atlases of infectious disease, the measures of a society's individualism from Hofstede (2001), and the technology diffusion measure from Comin and Mestieri (2012). Section 4 uses this data to test the model's predictions for the relationship between disease prevalence and social structure. This establishes that disease prevalence is a powerful instrument for social structure. The section then goes on to estimate the effect of social structure on technology diffusion, using the difference in communicable and non-communicable diseases as an instrument. The paper's main finding is that a 1-standard-deviation increase in individualism increases output per worker by an amount equal to 23% of the output of a US worker.

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. Spolaore and Wacziarg (2009) also take a macro perspective but proxy social distance by genetic distance. Our network theory and findings offer a specific mechanism to explain why social distance might be related to the diffusion of new ideas.

Thus in its scope, the paper is much more related to a second literature, that on technology

diffusion. Recent work by Lucas and Moll (2011) and Perla and Tonetti (2011) uses a search model framework where every agent who searches is equally likely to encounter any other agent and acquire their technology. Greenwood, Seshadri, and Yorukoglu (2005) models innovations that are known to all but are adopted when the user's income becomes sufficiently high. What sets this paper apart is its assumption that agents only encounter those in their network. Our insights about why societies adopt networks that do not facilitate the exchange of ideas and our links to empirical measures of social structure arise because of this focus on the network topology.

The third literature, on culture and its effects on national income is similarly macro in scope. Gorodnichenko and Roland (2011) focus on the psychological or preference aspects of collectivism. They use collectivism to proxy for individuals' innovation preferences and consider the effects of these preferences on income. In contrast, we view collectivism as a measure of human relationships and assess the effect of those relationships on income. Durlauf and Brock (2006) review work on social influence in macroeconomics, but bemoan the fact that no attempt is made to incorporate social interactions in models of growth. Contemporaneous work by Tabellini (2010) and Algan and Cahuc (2007) does examine the relationship between cultural characteristics and economic outcomes. Bisin and Verdier (2001), Bisin and Verdier (2000) and Fernández, Fogli, and Olivetti (2004) examine the transmission of culture. Cole, Mailath, and Postlewaite (1992) investigate how social norms affect savings choices, and in turn growth. But this literature also typically regards culture as an aspect of preferences. Greif (1994) argues that preferences and social structure are intertwined because culture is an important determinant of a society's social structure. While this may be true, we examine a different determinant of social structure that is easily measurable for an entire country, pathogen prevalence. Our evolutionary-sociological approach lends itself better to quantifying the aggregate effects of social structure on economic outcomes.

Finally, our empirical methodology is similar to work on the role of political institutions by Acemoglu, Johnson, and Robinson (2002) and Acemoglu and Johnson (2005) and the role of social infrastructure by Hall and Jones (1999). But instead of examining institutions or infrastructure, we study an equally important but distinct type of social organization, the social network structure.

1 A Network Diffusion Model

Our model serves three 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 degree of collectivism in a 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 do want to be explicit about what we intend to measure.

Second, the model motivates our choice of disease as an instrument for social structure. Specifically, it explains why disease that is spread from human-to-human might influence a society's social network in a persistent way. The disease-based instrumental variable we use is a valid instrument, regardless of the veracity of this theory. The model simply offers one possible explanation for why disease and social structure might have the robust relationship we see in the data.

The third 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. Can differences in social structure plausibly explain such large income disparities? To answer this kind of question requires a model. Section 2 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, at least since Foster and Rosenzweig (1995), a significant subbranch of the growth literature has focused on the role of personal contact in technology diffusion; see Conley and Udry (2010) or Young (2009) for a review. In his 1969 AEA presidential address, Kenneth Arrow remarked,

> "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 n agents, indexed by their location $j \epsilon \{1, 2, \ldots, n\}$ on a circle. Each agent produces output with a technology $A_j(t)$:

$$y_j(t) = A_j(t).$$

Social networks Each person i is socially connected to γ other people. If two people have a social network connection, we call them "friends." Let $\eta_{jk} = 1$ if person j and person k are friends and $= 0$ otherwise. To capture the idea that a person cannot infect themselves in the following period, we set all diagonal elements (η_{jj}) to zero. Let the network of all connections be denoted N.

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, $\ln(A_j(t+1)) = \ln(A_j(t)) + \delta$. In other words, a new invention results in a $(\delta \cdot 100)\%$ increase in productivity.

People can also learn from others in their network. If person j is friends with person k and $A_k(t) > A_j(t)$, then the next period, j can produce with k's technology, with probability ϕ. If there are multiple levels of technology used by j's social contacts, j can produce with the best of these technologies: $A_j(t+1) = \max_k \eta_{jk} A_k(t)$.

Spread of disease Each infected person transmits the disease to each of their friends with probability π. The transmission to each friend is an independent event. Thus, if m friends are diseased at time $t-1$, the probability of being healthy at time t is $(1-\pi)^m$. If no friends have a disease at time $t-1$, then the probability of contracting the disease at time t is zero.

An agent who catches a disease at time t loses their ability to produce for that period ($A_j(t) = 0$). Let $\psi_j(t) = 1$ if the person in location j is sick in period t and $= 0$ otherwise. An agent who is sick in period t dies at the end of period t. At the start of period $t+1$, they are replaced by a new person in the same location j. That new agent inherits the same social network connections as the parent node. When we discuss network evolution, we will relax this assumption. At the start of period t, the new agent begins with zero productivity and learns the technology of each of his friends with probability ϕ, just like older agents do.

1.2 Two Illustrative Networks

The previous subsection described the economic environment for a given network. Before we add a process of network evolution, it is useful to understand the properties of two networks. The evolutionary process will guide the economy to one of these two networks. They are the unique steady states of the stochastic network process. So, understanding how disease and technologies propagate in these two networks is very informative about the long-run behavior of our economy.

The two steady-state networks are extremes along a particular dimension, their degree of collectivism. This is an aspect of a social structure that has been extensively studied by sociologists. The collectivist network is one with many collectives, mutual friendships or instances of interdependence that are the hallmark of collectivist societies. To measure this interdependence, we can ask: If i is friends with j and with k, how often are j and k also friends? We refer to a structure where i, j and k are all connected to each other as a *collective*. Therefore, a measure of the extent of shared friendships, and thus the degree of collectivism, is the number of such collectives.

To count the number of collectives, we look at all the instances in a given network where one node i is connected to two other nodes j, k. Count that as a triple if j and k are connected. This

collectives measure is related to a common measure of network clustering: Divide the number of collectives by the number of possible collectives in the network to get the *overall clustering* measure (Jackson 2008).

To make our examples concrete, we will fix the number of connections γ to be 4. We explore the possibility of varying the number of connections below.

Network 1 *In the collectivist social network, each individual j is friends with the 4 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 1 is extreme in its degree of collectivism. The next result shows that there are as many collectives as there are members of the network (n).

Result 1 *In the collectivist network there are n unique collectives.*

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

At the other end of the spectrum, we examine a second network that is identical in every respect, except that it has the lowest possible degree of collectivism. We call that the individualistic network.

Network 2 *In the individualistic social network, each person is friends with the person next to them and the person m positions away from them, on either side. In other words, for any integer $m \in \{3, \ldots, n/2 - 3\}$, the network matrix has entries $\eta_{jk} = 1$ for $k = \{j-m, j-1, j+1, j+m\}$ and $\eta_{jk} = 0$ for all other k.*

Result 2 *In the individualistic network, there are zero collectives.*

These two network structures are particularly informative because of their starkly different numbers of collectives. This stark difference facilitates matching social institution data with one or the other type of network. As we will see, networks with numbers of collectives between 0 and n, are also possible along the transition path. But knowledge of the properties of these two extreme cases provides intuition about the properties of such intermediate cases as well.

Other dimensions along which networks could differ. There is a very large set of possible networks for an economy, too large to analyze completely. Therefore, we restrict attention to one dimension. We choose the prevalence of collectives because it represents the essence of collectivism, which is the sociological feature we have data on. But other dimensions of networks might also be closely related to collectivism. In particular, one might represent individualist societies as having more social linkages or capture the idea of market interactions with a time-varying, random

network. We have investigated both of these aspects of networks and found that both more linkages and random networks facilitate the spread of technology and germs. Thus, we could instead base our analysis on one of these other features and we would still come to the same conclusions: Having an individualist network exposes one to a greater risk of disease and a more productive set of technologies. In fact, preliminary analysis suggests that the quantitative effects of adding more linkages or random networks are even greater than for reducing the number of collectives.

1.3 Theoretical Results: Speed of Diffusion in Each Network

Disease spreads slowly in the collectivist network. The reason is that each contiguous group of friends is connected to at most 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 following results formalize these ideas.

Diffusion speed in each network 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 him on the circle, agent $n/2+1$. If each person has 2 friends on either side of them, then agent 1 will be friends with agent 3, who will be friends with agent 5, and this person will be friends with agent 7, etc., until we reach $n/2$. Thus, if the network size n is 6, $n/2+1$ is the farthest node. It could be reached in 2 steps: Agent 1 and agent 3 are directly connected and 3 is connected to 4. If $n/2+1$ is 6 ($n=10$), node 6 could be reached in 3 steps: from 1 to 3, 3 to 5 and 5 to 6. In general, the number of steps in this chain will be $(n-1)/4$, 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*.

Diameter is one measure of diffusion 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/γ 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.[2] Appendix A computes the diameter (as well as the average path length) of

[2]Our network is symmetric. So, the length of the path to the farthest node is the same, no matter which node

our two networks. The diameter of network 1 (collectivist), with n nodes is $(n-1)/4$, if that is an integer, or otherwise the next highest integer. Suppose, for example, that network 2 (individualist), has at least 4 nodes ($n > 4$) and $m = 4$ so that each node i is connected to $i-4$, $i-1$, $i+1$, and $i+4$. The diameter of this individualist network is round$(n/8)+1$. For large n, the diameter for the collectivist network is close to $n/4$, while the diameter of the individualist network is close to $n/8$. Therefore, as long as the network is sufficiently large, which for a country, it undoubtedly is, the individualist network will have a smaller diameter.

Figure 1 illustrates the smaller diameter and faster diffusion process in individualist networks, in the simple case where the probability of transmission is 1 and $m = 9$.[3] In both cases, a new technology arrives at one node in period 0. The "infected" person transmits that technology to all the individuals she is connected to. In period 1, 4 new people use the new technology, in both networks. But by period 2, there are 9 people using the technology in the collectivist network and 14 using it in the individualist network. In each case, an adopter of the technology transmits the technology to 4 others each period. But in the collectivist network, many of those 4 people already have the technology. The technology transmission is redundant. After 5 iterations, the new technology reaches every node in the individualist network. Thus, the diameter of the individualist network is 5. In contrast, it takes 9 iterations to reach all the nodes in the collectivist network. (The diameter is 9.) In sum, these properties tell us that, on average, ideas and diseases will diffuse more slowly through a collectivist network and an individualist one.

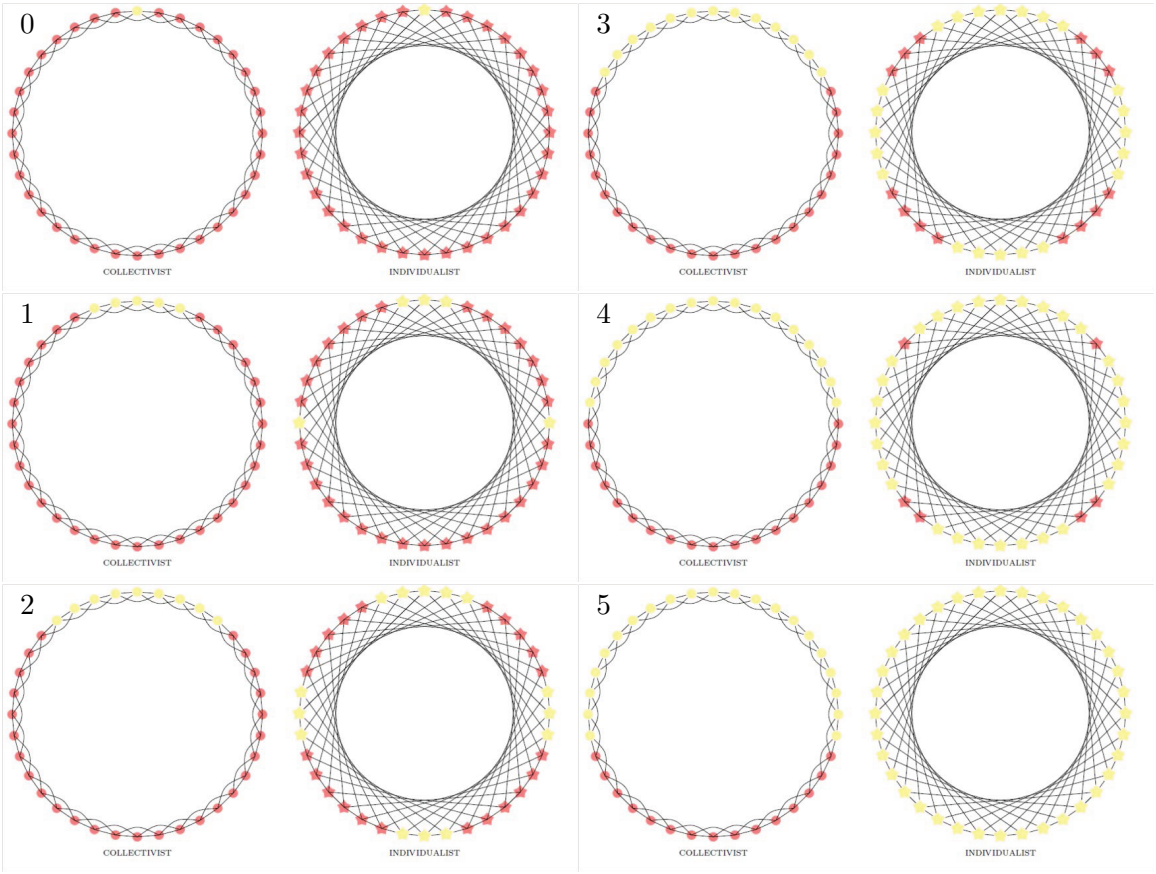
Diffusion speed and the technological frontier The diameter and average path length in a network are important determinants of the speed at which germs and ideas diffuse. In the individualistic network, because the path length between individuals is shorter, diseases and ideas disseminate more quickly. The next result uses the calculations above to characterize the mean and maximum infection times and the mean and maximum discovery time for a new technological innovation. Let $\Psi_j(t)$ be the next period in which the person living in location j at time t gets sick and dies. In other words, $\Psi_j(t) = min\{s : s \geq t, \psi_j(s) = 1\}$. Thus, $\psi(0)$ is number of periods that the person living in location j at time 0 will live. Analogously, let $\alpha_j(0)$ be the number of periods it takes for a new idea, introduced in period 0 to reach person j.

Result 3 *Consider two networks, an individualistic network (N2) and a collectivist network (N1). They have equal size $n > 8$, where $n/8$ is an integer, and equal degree $\gamma = 4$. If $\pi = 1$ and $\sum_j \psi_j(0) = 1$, then the average lifetime $E_j[\Psi_j(0)]$ and the maximum lifetime $\max_j[\Psi_j(0)]$ are*

one starts at. But in general, the diameter is the maximum path length, over all starting nodes.

[3] A larger m makes the network connections easier to identify visually. But it also accentuates the difference in diffusion speed.

Figure 1: Slower diffusion in the collectivist network (left) than the individualist network (right).



If a technology is introduced at the top of the network at time 0 and is transmitted with probability one, the light-colored nodes denote the nodes that would adopt the technology in periods 1-5. The individualist network has $m = 9$. The diameter of the individualist network is 5 because it takes 5 iterations to reach the farthest node in the network. It would take another 4 iterations to reach all the nodes in the collectivist network. Therefore the diameter of the collectivist network is 9.

longer in the collectivist network (1) than in the individualist network (2).

If $\phi = 1$*, then the average discovery time* $E_j[\alpha_j(0)]$ *and the maximum discovery time* $\max_j[\alpha_j(0)]$ *are slower in the collectivist network (1) than in the individualist network (2).*

With a collectivist network, technology invented in one location was transmitted only 2 people further each period. In the individualist network, ideas advance 4 places at a time. But faster diffusion is not the same as faster technological innovation. The reason that diffusion accelerates technology growth is that when idea diffusion is faster, redundant innovations are less frequent. Thus, more of the innovations end up advancing the technological frontier. The following result clarifies the mechanism by which the individualist network achieves a higher rate of growth.

Result 4 *Suppose that at t, a collectivist network (N1) and an individualist network (N2) have the same $A_j(t)$ $\forall j$. Then the probability that the next new idea arrival will increase the technological frontier is larger in N2 than N1.*

Together, these results explain why ideas and germs spread more quickly in the individualistic network than in the collectivist network, why diffusion might imply a higher level of technology adoption or GDP, and what evolutionary advantages each type of network might offer its adopters.

Could Collectivism Facilitate Technology Diffusion? Perhaps Arrow was not correct and technology diffusion is not a process "formally akin to the spread of infectious disease." Instead, a technology is adopted only when a person comes in contact with multiple other people who have also adopted it. It is theoretically possible that having many mutual friendships makes it more likely that groups of people adopt a technology together. But such a theory does not help to explain the empirical findings, which will show that collectivism is associated with slower technology diffusion. Furtheremore, this opposite effect would have to be strong enough to compensate for the effect our model identifies. Ultimately, this model is simply a framework for helping us think about what we find in the data. While other formulations that lead to opposite conclusions are possible, they don't help us to understand the facts at hand.

1.4 Network Evolution Model

So far, we have simply described diffusion properties of two networks. This leaves open the question of why some societies might adopt one type of network or the other. To provide one possible explanation, we next consider an evolutionary model where the network structure evolves as agents die and new ones are born in their place.

Preferences, production, endowments and the diffusion processes for technology and disease are the same as in the fixed-network model. In addition, at each date t, each person j can be one of two types: They are either a collectivist $\tau_j(t) = co$ or an individualist $\tau_j(t) = in$. All agents are linked to the two people adjacent to them. In addition, they are linked to at least one other person. Which other people depends on their type and the type of their neighbors. Individualists form links with those adjacent to them and someone four spaces to their right. For example, if the person is in location j, they are linked to $j-1$, $j+1$ and $j+4$. Collectivists form links with those adjacent to them and someone two spaces to their right. For example, if the person is in location j, they are linked to $j-1$, $j+1$ and $j+2$. In addition, a person of either type might be linked to nodes $j-2$ and/or $j-4$, depending on whether the agents in those locations are individualist or collectivist. In other words, a person's own type governs their links to the right (with indices higher than yours, except near n); others' types govern links to the left.

A person's type is fixed throughout their lifetime. The network structure only changes when someone dies. There are two reasons an individual can die. First, they can acquire the disease. Someone who acquires the disease at time t has zero output in period t. At the end of period t, they die. We also add a second potential cause of death to the model: Agents can die stochastically, for non-disease related reasons (accident, old age, etc.). With probability ξ, each person has an accident and dies at the end of each period. This probability is independent across time and individuals. When someone at node j dies in period t, then at the start of period $t+1$, a new person inhabits that node. The reason we introduce this second cause of death is to allow the network to evolve, even after the disease has died out.

A newborn person inherits the best technology from the set of people that the parent was socially connected to. He also inherits the type of the person with that best technology. In other words, if the person at node j is socially connected to nodes $\{k : \eta_{jk}(t) = 1\}$ and dies at time t, the new person at node j at time $t+1$ will start with technology $\max_{\{k:\eta_{jk}(t)=1\}} A_{kt}$. Let k^* be the argument that maximizes this expression (i.e. the friend with the highest time-t technology), then the time-$(t+1)$ type of the person is the same at the time-t type of person k^*: $\tau_j(t+1) = \tau_{k^*}(t)$.

The idea behind this process is that evolutionary models often have the feature that more "successful" types are passed on more frequently. At the same time, we want to retain the network-based idea that one's preferences are shaped by one's community. Therefore, in the model, the process by which one inherits the collectivist or individualist trait is shaped by one's community, the social network, and by the relative success (relative income) of the people in that network.

1.5 Theoretical Results: Network Evolution

The question we want this model to answer is: Why do some societies end up with a collectivist network even though it inhibits growth? What features might influence the long-run network equilibrium? These results describe the long-run properties of networks and disease. Understanding the stochastic process that governs disease and network type provides intuition for the numerical results in the next section, which will show that a higher initial prevalence of disease makes it more likely that a society will end up with a collectivist network, like that in network 1.

The first set of results show that eventually, the economy always converges to either the fully collectivist network (1) or the fully individualist one (2).

Result 5 *With probability 1, the network becomes homogeneous:* $\exists T$ *s.t.* $\tau_j(t) = \tau_k(t)$ $\forall k$ *and* $\forall t > T$.

In other words, after some date T, everyone will have the same type forever after. They might all be individualist or all be collectivist. But everyone will be the same. The reason for this is that

since traits are inherited from neighbors, when a trait dies out, it never returns. The state where all individuals have the same trait is an absorbing state. Since there are a finite number of states, and whenever there exists a j, k such that $\tau_j(t) \neq \tau_k(t)$, every state can be reached with positive probability in a finite number of steps, then with probability one, at some finite time, an absorbing state is reached and the economy stays there forever after.

Similarly, having zero infected people is an absorbing state. Since that state is always reachable from any other state, with positive probability, it is the unique steady state.

Result 6 *With probability 1, the disease dies out:* $\exists T$ *s.t.* $\psi_j(t) = 0 \quad \forall j$ *and* $\forall t > T$.

What these results tell us is that which network type will prevail is largely dependent on which dies out first, the individualist trait, or the disease. When there is a positive probability of infection, people with individualist networks have shorter lifetimes, on average. If disease is very prevalent, it kills all the individualists and the society is left with a collectivist network forever after. If disease is not very prevalent, its transmission rate is low, or by good luck, it just dies out quickly, individualists will survive. Since they are more economically successful, they are more likely to pass on their individualist trait. So, the economy is more likely to converge to an individualist network. This is not a certain outcome because of exogenous random death. It is always possible that all individualists die, even if the disease itself is no longer present.

2 Numerical Results

We use a calibrated model simulation to accomplish three objectives. First, we use the simulations to illustrate and clarify the model's mechanics. Second, we check whether differences in networks can potentially explain the magnitude of the large differences in incomes across countries. Third, we establish that societies with higher initial disease prevalence are more likely to become collectivist. The model is not rich enough to produce predicted growth rates or disease rates that are remotely accurate. Rather, the objective here is simply to confirm the direction of the model's predictions and gauge whether the predicted effects are trivial or not.

2.1 Parameter Choice

To evaluate magnitudes, we need to choose some realistic parameter values for our model. The key parameters 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 1.

Table 1: Parameters and their empirical counterparts

	Parameter	Value	Target
Initial disease prevalence	$\text{Prob}(\psi_j(0) = 1)$	0.5%	TB death rate in China
Disease transmission probability	π	32%	Disease disappears in 150 years (indiv country avg)
Innovation productivity increase	δ	30%	2.6% growth rate in individualist country
Technology transfer probability	ϕ	50%	Half-diffusion in 20 years (Comin et. al. '06)
Technology arrival rate	λ	0.25%	1 arrival every 2 years (Comin et. al. '06)
Exogenous death rate	ξ	1/70	average lifespan

For the initial pathogen prevalence rate, we use the annual tuberculosis death rate in China, a country where the disease was endemic. Tuberculosis is the most common cause of death in our sample. Note that this is a mortality rate, not an infection rate. Since individuals who get sick in the model die, this is the relevant comparison. Also, it is a conservative calibration because it uses only one disease and it would be easier to get large effects out of a higher disease prevalence rate. One would like to choose the probability of disease transmission to target a steady state rate of infection. But, as we've shown, the only steady state infection rate is zero. Thus, we set the transmission rate so that, on average, the disease disappears in 150 years. This average masks large heterogeneity. In many economies, the disease will disappear after 2 periods. In others, it will persist for hundreds of years. Thus, the economy starts with a given fraction of the population being sick and each sick person represents an independent 32% risk (π) of passing the disease on to everyone that person is friends with.

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. The rate of arrival of new technologies is calibrated so that a new technology arrives in the economy every 2 years, on average. This corresponds to the average rate of adoption of the technologies in the (Comin, Hobijn, and Rovito, 2006) data set. The magnitude of the increase in productivity from adopting a new technology is calibrated so that the individualistic network economy (more likely to be the developed economy in the data) grows at a rate of 2.2% per year. The probability of transmitting a new technology to each friend (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 20 years (Comin, Hobijn, and Rovito, 2006).

Finally, in the evolutionary model, there is a probability of exogenous death. We choose this

probability to match an average lifespan in a low-disease economy of approximately 70 years.

The economy consists of 200 people, each with 4 friends. We average the results from 200 independent runs.

2.2 An Illustrative Numerical Example

To illustrate the mechanics of technology and disease diffusion, we first describe a small-scale illustrative example. Here, we hand-pick some of the parameter values (in particular, the rate of technology arrival) to make it easier to visualize diffusion taking place. Figure 2 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).

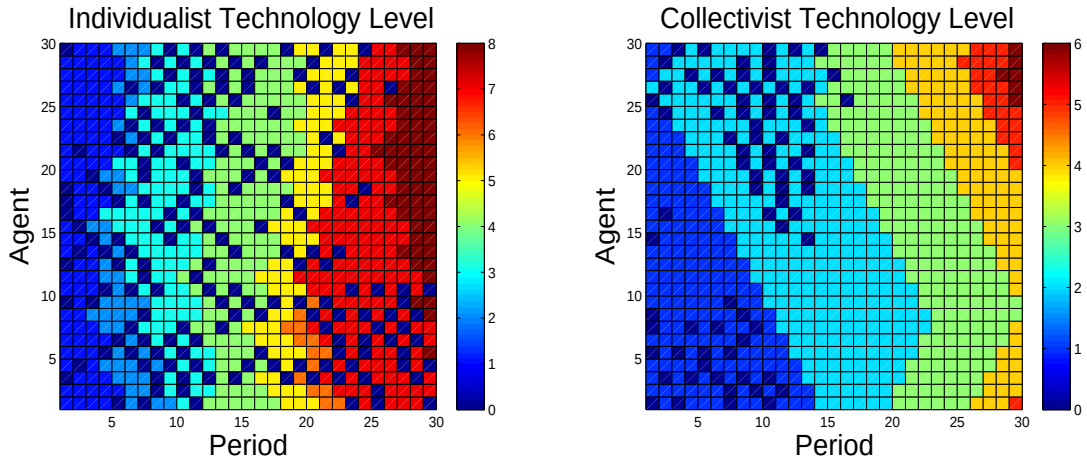


Figure 2: How disease and technology spread through networks.

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.

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 collectivist network (left figure), some agents who are 1-2 places away from agents that were sick in period 1 are now sick. In the individualistic network (right figure), some agents who are 1 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 collectivist network, some individuals are still using the initial technology level in period 8. In the individualistic network, all the healthy agents have adopted the second technology level after period 5.

After 30 periods, the most technologically advanced agents in the collectivist network only

realize 7 steps in the quality ladder. In the individualistic network, some agents operate at 9 steps. If each innovation represents a 5% productivity increase, being two steps further represents a 10% higher degree of productivity.

This example is meant to illustrate how an individualistic network spreads ideas more efficiently, and how it also spreads germs more efficiently. Of course, this is just an 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, the following simulation uses the calibrated parameters and averages the results over many agents and many simulations.

2.3 How Much Effect Might Networks Have on Output?

A potential concern about using this model to explain income differences across countries is the worry that its predicted effect is trivial, compared to the vast differences in incomes across countries. What our calibration exercise shows is that changing a society's social network structure has a small effect on the annual diffusion rate. But over time, small effects cumulate. The result is large differences in productivity levels in the long run. Thus, changes in network choice produce differences in technology diffusion rates which could explain a significant part of the disparity in countries' incomes. While idea transmission facilitates reaching higher levels of productivity, disease prevalence diminishes productivity. To see the net effect of these two forces, we simulate the model many times and examine the average outcomes.

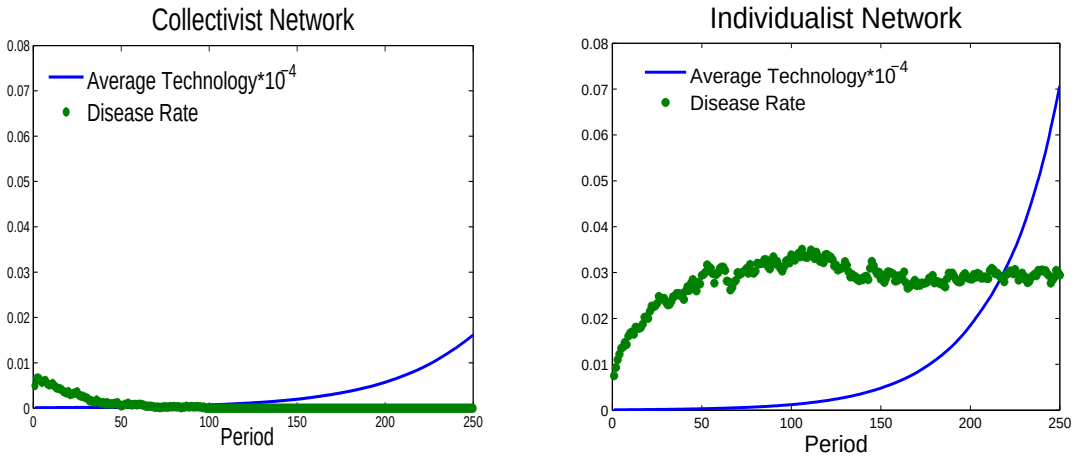


Figure 3: Average disease prevalence and productivity

Figure 3 plots the average disease prevalence and the average technology level for the whole population over 250 years. The fraction of the population infected with disease is significantly higher in the individualistic network society. In fact, the collectivist networks inhibit the spread of

disease so much that it quickly becomes extinct in most simulation runs.

However, having a individualistic network results in technology that grows at 2.6% per year. This is true by construction because it was one of the calibration targets. But the economy with the collectivist network grows at only 2.0% per year. While the difference in growth rates is small, in time, it produces large level differences. After 250 years, the average level of technology is 476% higher in the individualistic network than in the collectivist network. This simple example makes the point that a difference in network structure can create a small friction in technology diffusion. When cumulated over a longer time horizons, this small friction has the potential to explain larger differences in countries' incomes.

Of course, this also tells us that social structure is not likely to explain the nearly 100-fold difference between incomes in the poorest and richest countries. We know that corruption, war and distorted incentives explain the worst growth disasters. At the same time, 476% of national income is a large difference between seemingly similar countries. It is an extreme result in the sense that we compared a purely individualist network to a purely collectivist one. Most societies will lie somewhere in between. But it gives us an idea of the potential size of the effect. The actual effect is an empirical matter that we take up in the following sections.

2.4 Network Evolution

What we ultimately want to know from the evolutionary model is: Are high-disease societies more likely to evolve toward collectivist networks? One might wonder whether societies that start out as high-disease and adopt collectivist social structures might end up with lower disease rates in the long-run. That turns out not to be the case.

We would like to calculate the probability of arriving at each steady state (where all agents have the same type) analytically and see how that probability changes in response to changes in disease prevalence. However, to characterize the probability of a single stochastic process crossing one boundary before another is a difficult problem. Here, there are two interacting stochastic processes, one for disease and one for network types. Both have absorbing states. That added complexity makes characterizing the crossing probabilities an intractable problem.

What we can do is examine the probability of each network steady-state in the context of our numerical example. We use the same parameters as before. (See Table 1.) We set the initial fraction of individualists to 10% and simulate the economy for 250 periods 200 times. To see how the initial disease prevalence rate affects the network steady state, we consider two initial disease prevalence conditions: One is 5% (the calibration target of the original model) and the other is twice that level. Thus, there is a low disease economy, with 5% of agents infected, and a high disease economy that starts with 10% of agents infected.

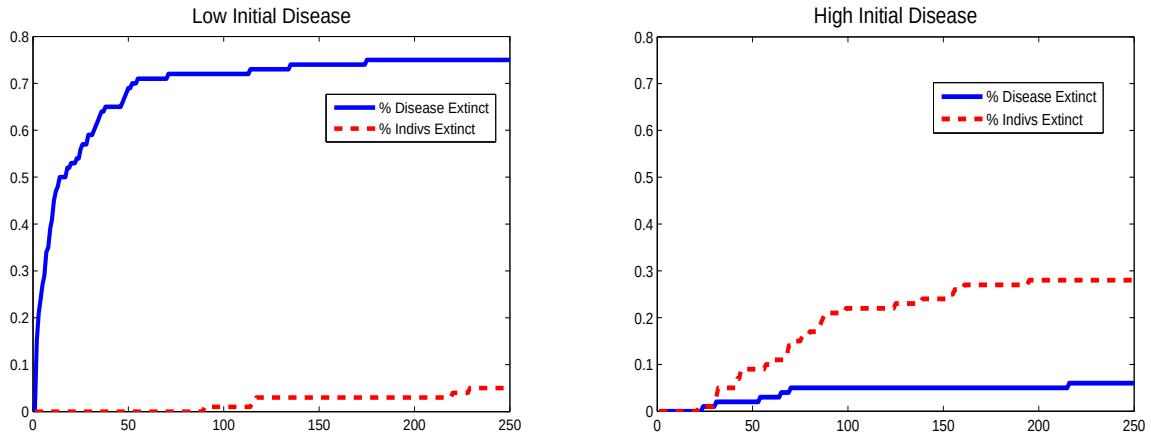


Figure 4: Simulation of the evolutionary model.

The probability that the economy has converged to a zero-disease or purely collectivist steady-state by each date. The economy on the right differs only because it began with a (2×) higher rate of disease prevalence. Because of this difference, the individualist agents typically die out and the network becomes collectivist.

Figure 4 shows the fraction of economies that have converged to a zero-disease steady-state or a purely collectivist steady-state by each date. This can also be interpreted as the probability that a given economy will converge to that steady state by that date. In the low-disease economy, much of the time, the disease dies out within a few periods. Only in a few runs does the disease persist and infect a large fraction of the population. In the high-disease economy, the disease rarely dies out within 250 periods. Conversely, individualists flourish in the low-disease economy. In the high-disease economy, after 100 periods, there is a 25% chance that all individualists have died and that economy will forever remain collectivist.

Thus, the prediction of the model is clear: *Low-disease societies are more likely to be individualist and high-disease societies are more likely to be collectivist.* The secondary effect whereby collectivism reduces disease is always dominated by the primary effect that disease kills a disproportionate number of individualists, leaving a permanently collectivist society. To see why, consider the contrary: If high-disease societies were more individualist, the disease would systematically kill the individualist types and transform the society to a collectivist one. It is simply not a stable outcome.

3 Data

Our theory is about the relationship between pathogen prevalence, social structure, and technology diffusion. This section describes how these three variables are measured.

3.1 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. By necessity, a contemporary source was used to estimate the prevalence of tuberculosis, but the prevalence of the remaining eight pathogens was estimated on the basis of old atlases of infectious diseases and other historical epidemiological information provided in Rodenwaldt and Bader's (1952-1961) World-Atlas of Epidemic Diseases and in Simmons, Whayne, Anderson, and Horack's (1944) Global Epidemiology.[4] They used a 4-point coding scheme: 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.[5] The prevalence of tuberculosis was based on a map contained in the National Geographic Society's (2005) Atlas of the World, which provided incidence information in each region for every 100,000 people. These scores, calculated for a total of 230 geo-political regions of the world, were summed up to in an overall index that summarizes a country's overall historical pathogen prevalence.[6] The countries with the highest pathogen prevalence are Brazil, India, China, Nigeria and Ghana. Countries with the lowest prevalence include Canada, Switzerland, Luxembourg, Hungary and Sweden. Figure 7 shows the historical world-wide distribution of pathogens according to the overall index. This is the data we use for most of our analysis, including our IV estimations.

For comparison, we used the same method to create an alternative measure of pathogen prevalence based explicitly on contemporary information. It delivers very similar results. See appendix C for details.

Disease reservoirs To identify the effect of disease on social structure, we will use the difference in the prevalence of various types of diseases. The "reservoir" of a diseases is a term from epidemiology that refers to the long-term host of a pathogen. When an animal hosts the disease, it may or may not become ill from it. We first follow Smith, Sax, Gaines, Guernier, and Gugan (2007) and Thornhill, Fincher, Murray, and Schaller (2010) by distinguishing diseases according to their reservoirs:

[4]Detailed information about the occurrence of various diseases across the world was collected by the Medical Intelligence Division of the United States Army.

[5]In the rare cases in which these two epidemiological sources provided contradictory information, priority was placed on data provided by the older source (Simmons et al., 1944). In cases in which the relevant maps were unavailable (this was especially true for leprosy) or insufficiently detailed (this was especially true for many of the Pacific island nations), prevalence ratings were informed also by verbal summaries found in Simmons et al.

[6]The majority of these regions are nations (e.g.,Albania, Zimbabwe); others are territories or protectorates (e.g., Falkland Islands, New Caledonia) or culturally distinct regions within a nation (e.g., Hawaii, Hong Kong).

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. In our data set, the only pathogen that resides exclusively in humans is filariae (FIL).

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. Our zoonotic pathogens are schistosomiasis (SCH) and typhus (TYP).

Multi-host Some infectious agents can use both human and non-human hosts to complete their lifecycle. We call these "multi-host" pathogens. The multi-host pathogens in our data are leishmaniasis (LEI), leprosy (LEP), trypanosomes (TRY), malaria (MAL), dengue (DEN) and tuberculosis (TB).

Since multi-host and human-specific pathogens can reside in humans, they have the potential to affect the relative benefits of a social network. Zoonotic pathogens are not carried by people, only by other animals. Therefore, their prevalence is less likely to affect the benefits of any particular social structure. Therefore, for the purposes of our analysis, we will group human-specific and multi-host diseases together and define $human\_res \equiv FIL + LEI + LEP + TRY + MAL + DEN + TB$. We compare the effects of these human- and multi-reservoir diseases to those of zoonotic diseases: $zoonotic \equiv SCH + TYP$. Both variables are measured using the historical prevalence data. We then construct their difference to use as an instrument:

$$\text{diff\_res} \equiv \text{human\_res} - \text{zoonotic} \tag{1}$$

The following standardized difference also turns out to be a useful instrumental variable:

$$\text{diff\_res\_std} \equiv \text{human\_res} - \text{zoonotic}\frac{std(\text{human\_res})}{std(\text{zoonotic})} \tag{2}$$

Disease vectors Comparing the prevalence of diseases with different reservoirs will be one instrumental variable we will use. To illustrate the robustness of the instrument to the exact classification scheme, we will also consider diseases classified by vector. Whether or not a disease affects the benefits of one social network versus another depends on how it is spread. The vector of a disease is the animal or substance that spreads the disease from one person to another. In many cases, the vector is also the reservoir. But sometimes not. For example, some diseases are hosted by people, but are spread when an insect bites the person and then bites another preson shortly after. The

insect is not the long-term host of the disease, but it is the vector that spreads it.

Many diseases can be spread in multiple ways. Therefore, in this comparison, we use only a subset of the diseases for which we have data so that we can make a clean comparison of diseases whose spread is clearly facilitated by social contact and ones that although contagious, do not require any human social contact to propagate.

Human vector The two human diseases that have only human vectors in our data set are tuberculosis and leprosy. Tuberculosis is spread when a person coughs or sneezes and another person inhales airborne droplets containing the disease. Leprosy is transmitted by prolonged close contact or by nasal droplet.

Filaria may be transmitted directly to another person, but is most commonly spread by mosquito. What makes Filaria different from our animal-vector diseases is that the animals cannot spread it between themselves. If a mosquito bites one person, that same mosquito needs to bite another person to spread the disease. Since being in close social contact with an infected person raises the probability of acquiring the disease, we include it with human-spread disease. Therefore, our variable that captures human-vector disease is: $hum\_vec = LEP + TB + FIL$.

Aquatic or animal vector One disease that is clearly not spread by social contact is schistosomiasis. This is a parasitic infection by a type of flatworm called a schistosome. It is spread when a person urinates or defecates and the larva are consumed by a snail, which is an intermediate host to the disease. Only after the parasite emerges from the snail can it infect a new human host. The parasite lives in water and directly penetrates the skin of a person that comes in contact with it. Since the substance that conveys the disease to humans is water, water is the vector. We call such a disease "aquatic." Typhus is a disease with an animal vector. Typhus is spread when an infected person is bitten by an insect, such as a louse or tick. The disease needs to incubate in the insect for multiple days before it can be spread to another insect or another person. Both can travel long distances in moving water or in migrating insect populations. Neither ever spreads from person-to-person and close social contact is not considered a risk factor for acquiring the disease. Therefore, these two diseases should not affect one's choice of social contacts. We define the following variable that captures non-human-vector disease: non human$= SCH + TYP$.

We then construct the following differences to use as instruments:

$$\text{diff\_vec} \equiv \text{human\_vec} - \text{non human}. \tag{3}$$

$$\text{diff\_vec\_std} \equiv \text{human\_vec} - \text{non human}\frac{std(\text{human\_vec})}{std(\text{non human})}. \quad (4)$$

3.2 Measuring Collectivism

In our model, collectivism is defined as a social pattern of closely linked or interdependent individuals. What distinguishes collectives from sets of people with random ties to each other is that in collectives, it is common that two friends have a third friend in common. This is the sense in which they are interdependent.

The ideal data to measure collectivism would be each country's complete social network. We would look for a high prevalence of social collectives. There are a handful of studies that map out partial social networks, but only for small geographic areas, across eight countries. (See Fischer and Shavit (1995) for a review.) Therefore, we use data from Hofstede (2001) that is available for a broad cross-section of countries. He surveyed IBM employees worldwide to find national differences in cultural values. Hofstede performed a factor analysis of the survey responses, and found two factors that together can explain 46% of the variance in survey responses. He labels one factor "Collectivism vs Individualism", and uses it to construct an index of individualism that ranges from between 0 (strongly collectivist) to 100 (strongly individualist). Hofstede describes collectivist and individualist societies as follows: "on the individualist side we find societies in which the ties between individuals are loose... On the collectivist side, we find societies in which people from birth onwards are integrated into strong, cohesive in-groups, often extended families..." This description reflects two views of a collectivist society: one where ties are strong, and one where ties are shared.

In a widely cited paper, Granovetter (1973) provides the bridge between shared ties and strong ones; he argues, "the stronger the tie between A and B, the larger the proportion of individuals [that either of them knows] to whom they will both be tied." Granovetter goes on to give three theoretical reasons to believe this is true: (1) Time. If A and B have strong ties, they will spend a lot of time together. If A and C also have strong ties, they will also spend a lot of time together. If these events are independent or positively correlated, this necessarily implies B and C will spend a lot of time together, giving them a chance to form a strong tie. (2) The tendency of an individual to interact with others like himself. If A and B have strong ties, chances are good that they are similar; the same holds for A and C. Transitivity implies B and C will be similar, and will therefore get along. (3) The theory of cognitive balance. If A is good friends with B and C, then B will want to develop a good relationship with C, in order to maintain his relationship with A. Thus, Granovetter's theory explains why Hofstede's survey questions, many of which are about the strength of social ties, are informative about the prevalence of collectives, as defined in the model.

Other questions in Hofstede's survey assess the strength of cooperation, social influence and individuals' weight on social objectives. Coleman (1988) explains why these behaviors are also linked to the presence of network collectives. He shows that effective norms depend on the presence of collectives because people enforce 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, cooperation and conforming behavior is easier to sustain than if punishments must be implemented in an uncoordinated way.

A third category of questions in Hofstede's survey are about mobility, specifically one's willingness to move or change jobs. The essence of strong social ties is that the people involved are averse to breaking those ties. Thus an unwillingness to change one's social environment is indicative of strong social network ties. In the survey, the individualism index loads positively on one's willingness to move, which is consistent with the interpretations of individualism as a society with fewer collective and thus weak ties.

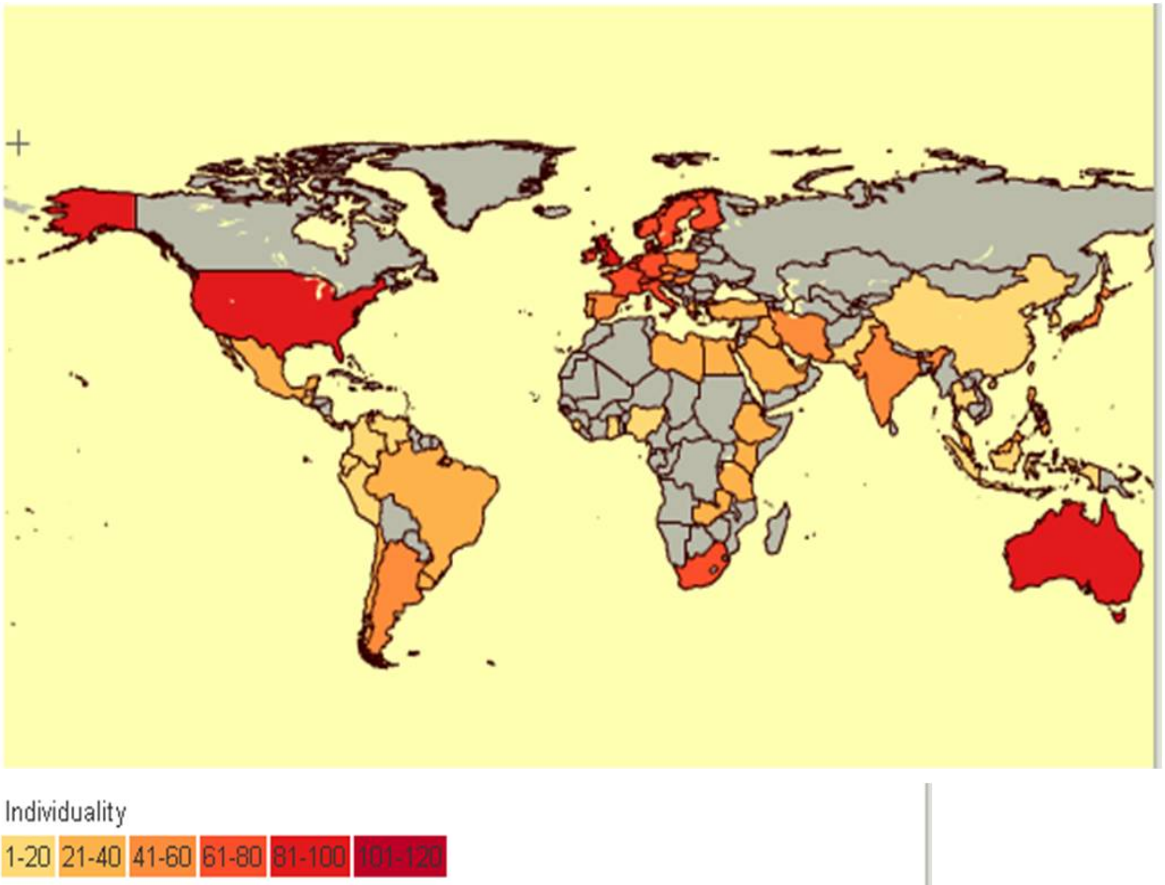
Thus, while Hofstede's survey asks questions that are not directly about the pattern of social relationships, there is a body of sociological theory and evidence that supports the connection between the behaviors that Hofstede asks about and the pattern of network collectives as described in our model. One example of such a question is " How important is it to you to work with people who cooperate well with each other?" Other questions and more details about the survey are in appendix C. This connection is bolstered by the findings of the studies that do explicitly map out social networks. Table 6 shows that highly individualist countries have lower network interdependence than more collectivist ones. Finally, other variants of the model that capture other aspects of collectivism, such as strong ties or fixed versus random networks, deliver the same effects. Networks with many weak ties, with random link formation, or with mobility, all have that ability to disseminate information or diseases more efficiently than their collectivist network counterparts (see Jackson (2008)). Appendix C contains more details about these alternative models and about the survey questions and other correlated social survey measures that shed light on the interpretation of Hofstede's index.

Figure 3.2 summarizes the findings of Hofstede's survey in a color-coded map. The most individualist countries are Canada, Netherlands, United Kingdom, Australia and United States. The most collectivist countries are Guatemala, Ecuador, Panama, Venezuela, Colombia and Pakistan.

3.3 Measuring the Rate of Technology Diffusion

We use a technology diffusion measure that is derived from the cross-country historical adoption of technology data set developed by Comin, Hobijn, and Rovito (2006). The data covers the diffusion

Figure 5: Map of Hofstede's individualism index.



of about 115 technologies in over 150 countries during the last 200 years. At a country level, there are two margins of technology adoption: the "extensive" margin (whether or not a technology is adopted at all) and the "intensive" margin (how quickly a technology is adopted, given that it is adopted.) A country can be behind in a technology even though it is adopting it quickly, if the technology was introduced to the country late.

Since our model speaks only to the intensive margin, we need to filter the extensive margin from the data. We do this with the results from Comin and Mestieri (2012), where attention is restricted to 15 technologies. Technical details are in that paper, but the idea is the following: For a given country, plotting the normalized level of a given technology (e.g. log telephone usage minus log country income) over time yields an increasing curve. For a given technology, these curves look similar across countries, except for horizontal and vertical shifts. The horizontal shifts correspond to the extensive margin of technology adoption; if country A adopts telephones in exactly the same way as country B, only twenty years later, its curve will be identical to that of B except shifted

twenty years to the right. However, if country A adopts telephones less vigorously but at the same time, its curve will be below that of B's. This is what we are interested in, so we focus only on the intensive margin of technology diffusion. Specifically, Comin and Mestieri (2012) estimate the slope of a diffusion curve. A higher slope parameter m_{ij} indicates a faster rate of adoption of technology j by country i.

Our ideal measure of the technological level of a country would be its average intensive margin from all 15 technologies. A complication is that the data set is unbalanced; if data for a country is only available for slowly-spreading technologies, it might artificially appear technologically backward. To control for this problem, we estimate $m_{ij} = \alpha_j + e_{ij}$, where α_j is a technology-specific fixed effect. Our measure of technology diffusion for a given country is the average residual $\text{diffusion}_i = \sum_j e_{ij}$.

4 Empirical Results: How Much Do Networks Affect Technology?

Our objective is to better understand how social structure affects technology diffusion and how large that effect is on economic development. The difficulty is that economic development also can potentially change the social structure. The challenge is to isolate each of these two effects. To do this, we consider the following structural model:

$$A = \beta_1 + \beta_2 S + \epsilon \tag{5}$$

where A is the speed of technology diffusion, S is social structure (individualism), as measured by the Hofstede index, the β's are unknown coefficients and ϵ is a mean-zero residual orthogonal to S. Social structure is

$$S = \gamma_1 + \gamma_2 A + \gamma_3 x + \eta, \tag{6}$$

where the γ's are unknown coefficients, x is a collection of other variables that determine social structure, and η is a mean-zero residual orthogonal to A and x. The coefficient of interest is β_2, which measures the effect of social structure S on technology diffusion A.

This model recognizes the endogeneity problem inherent in estimating the relationship between A and S. It incorporates our main hypothesis, that social structure S matters for technology A, but it also reflects the idea that perhaps technology (and income) can cause social structure to change as well. Because A depends on S and S depends on A, an OLS estimate would be biased.

Our key identifying assumption that will allow us to isolate the effect of social structure on technology diffusion is that the x variables affect technology diffusion only through social structure, and not directly. In other words, we assume that $E[x\epsilon] = 0$. The x variables we use as instruments

are measures of spoken language and the difference between the prevalence of two types of diseases: those that are spread socially and those that are not. Let x_1 be socially communicable diseases and x_2 be other diseases. We believe that technology affects the prevalence of both types of disease in a similar way:

$$x_1 = \nu A + \xi_1 \tag{7}$$

$$x_2 = \nu A + \xi_2. \tag{8}$$

Since each disease is affected equally by technological innovation, the difference in the two diseases $x_1 - x_2 = \xi_1 - \xi_2$ is uncorrelated with A or with ϵ. Thus, this difference is a valid instrument.

Similarly, it is likely that not only does technology affect disease, but also that disease direcly affects technology. In other words, $E[\xi_i \epsilon] \neq 0$. This is a second reason that x_1 or x_2 by itself would be an invalid instrument. But if we assume that diseases impede technology diffusion similarly, no matter how they are spread, that implies that $E[\xi_1 \epsilon] = E[\xi_2 \epsilon]$. Since the difference in disease prevalence is $x_1 - x_2 = \xi_1 - \xi_2$, this assumption implies that our difference instrument is orthogoal to the innovation in technology: $E[(x_1 - x_2)\epsilon] = 0$. Of course, we will also test these assumptions in the data.

Note that we do not need to know all the determinants of social structure x. Rather, any subset of the x variables can serve as valid instruments for S. Similarly, we do not need to observe S exactly. A proxy variable with random measurement noise is sufficient for an unbiased instrumental variables estimate of the coefficient β_2.

4.1 First-Stage Regressions: Disease and Social Institutions

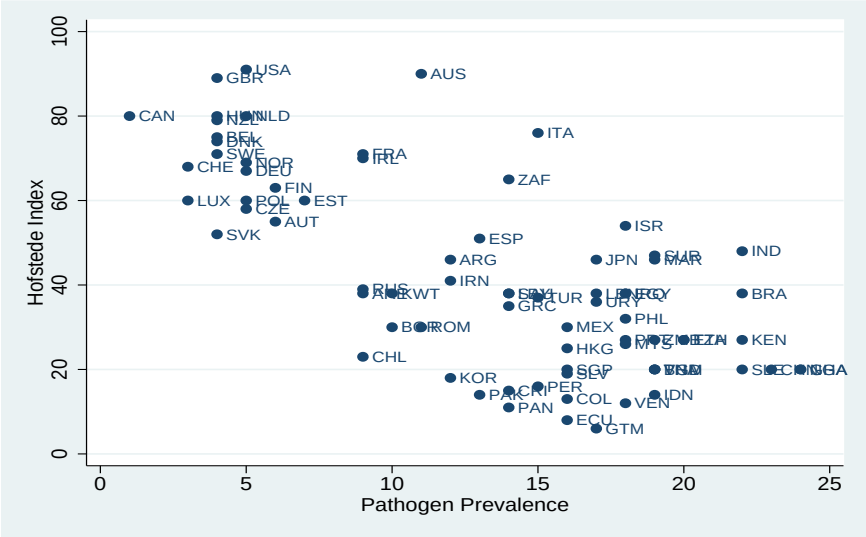
We begin by investigating the relationship between our instruments and our measure of social structure. There are two key findings: First, the instruments are powerful predictors of social structure. Second, disease is negatively correlated with individualism. Although this effect is not identified, the correlation is consistent with one key prediction of the evolutionary network model.

To illustrate the robustness of these results, we explore a handful of instrumental variable specifications. Most of the specifications have multiple instruments because that allows us to evaluate the validity of our instruments by testing the orthogonality of each instrument with the residual in equation (5). Following Hall and Jones (1999), we use two language-based variables as additional instruments to test the validity of our own disease-based instruments. The variable pronoun is a dummy variable that is 1 if it is conventional to omit first- and second-person pronouns in a country's dominant spoken language (Kashima and Kashima, 1998). For example, English and German typically do not omit pronouns, while Spanish does. In addition, we use a variable for

the fraction of the population speaking English as a first language.[7] Including English specifically contributes additional explanatory power. Because these variables are language-based, they are a product of the country's distant past and possibly its colonial heritage. As Hall and Jones (1999) argue, they are unlikely to be affected by current income or technology.

We begin by exploring the data on individualism and disease prevalence. Figure 6 illustrates the negative relationship between individualism and the sum of the prevalence of all nine pathogenic diseases in our data set. The negative relationship is consistent with our theory because the more

Figure 6: Hofstede's individualism index plotted against total pathogen prevalence.



Total pathogen prevalence is $human\_res + zoonotic$. This is a sum of the prevalence of all nine diseases described in section 3.

collectivist society, with its greater propensity for network collectives, would be a more effective structure for inhibiting the spread of disease.

Table 2 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. The explanatory power of pathogens is large; the R^2 of the regression is over 50%. The economic magnitudes are also large. A one-unit increase in our pathogen measure corresponds to one disease being endemic (always widespread), instead of epidemic (occasionally widespread). Having one more socially transmittable disease consistently prevalent corresponds to an individualism index that is 3 points lower (14% of a standard deviation).

[7]The English variable is available from the Penn World Tables, Mark 5.6.

Table 2: **First-stage regressions of pathogen prevalence variables on individualism index**

Dependent variable	Individualism (S)					
	(1)	(2)	(3)	(4)	(5)	(6)
Total pathogens	−2.73					
	(0.31)					
Human - zoonotic		-3.46	-2.15			
pathogens (diff_res)		(0.44)	(0.45)			
Human - zoonotic				-5.26		
pathogens (diff_res_std)				(2.04)		
Pathogens diff by					-2.94	
vector (diff_vec)					(0.98)	
Pathogens diff by						-4.39
vector (diff_vec_std)						(2.16)
English			25.33	28.48	25.25	26.55
			(7.50)	(8.58)	(8.24)	(8.66)
Pronoun			-19.17	-28.33	-27.04	-30.17
			(4.83)	(4.70)	(4.70)	(4.65)
Constant	77.10	67.53	69.71	59.86	65.08	61.16
R^2	0.52	0.47	0.71	0.64	0.66	0.63
Observations	72	72	62	62	62	62

The table reports OLS estimates of the γ coefficients in $S = \gamma_1 + \gamma_3 x + \eta$, where the x variables are listed in the first column of the table. The variables diff_res, diff_res_std, diff_vec and diff_vec_std are defined in equations (1), (2), (3) and (4). The instruments are pronoun drop and whether is English spoken (see appendix C). Standard errors are in parentheses. All coefficients are significant at 5% level.

Differences in disease reservoirs Our identifying assumption is that while technology diffusion and GDP may affect disease prevalence, even 40 years prior, it affects many diseases similarly. Thus, the difference in the prevalence of one type of disease or another is exogenous with respect to GDP. The first difference we consider is based on the disease reservoir: The difference between diseases that reside in humans (human-specific plus multi-host) and diseases that reside exclusively in non-human animals (zoonotic diseases).

Diseases that reside in humans could be spread through social networks and thereby affect the evolution of the network. Diseases that reside only in other animals are not spread through social networks and have no effect on network evolution. Thus, the difference in the prevalence of these two diseases should, according to the model, predict the probability of collectivism versus individualism prevailing.

In the first-stage regression, both our language variables and differences in disease prevalence are highly-significant predictors of social structure (table 2, column 3). Disease difference is a powerful instrument because the average correlation of individualism with each disease carried by humans is much larger in magnitude than the average correlation with each of the zoonotic diseases

(-0.53 vs. -0.29). The fact that the correlations are negative tells us that higher disease prevalence is associated with more collectivist societies. This is consistent with our theory, which tells us that societies should evolve toward collectivist networks when disease is more prevalent because they are better suited to prevent contagion.

Differences in disease vectors The second type of difference in disease prevalence we consider is the prevalence of diseases spread by humans, minus the prevalence of disease spread over long distances by migrating insects or by water (non-human vector).

Just as with the previous difference, the model gives the economic reason why the vector-based difference in disease prevalence is likely to be a good predictor of social structure. Diseases that are spread from human-to-human are spread through social networks. Diseases that are passed on only after they are hosted by another animal and then transmitted by contaminated water are not spread through social networks and have no effect on the probability of converging to one network or another. Thus, the vector-based difference in disease should also predict social structure.

In the first-stage regression, the vector-based difference in disease prevalence is a highly-significant predictor of social structure (table 2, column 4). The disease difference is a powerful instrument because the average correlation of individualism with each disease conveyed by humans is larger in magnitude than the correlation with the non-human diseases (-0.55 vs. -0.41).

These results are important for the next stage, identifying an effect of institutions on technology diffusion. But they are also interesting on their own because they are consistent with one reason why countries may have adopted different social institutions. Perhaps social structures have evolved, in part, as a defense against the spread of disease. But further statistical work would need to be done to say conclusively that disease prevalence is part of the reason why some societies have adopted social structures that inhibit technological diffusion and growth.

4.2 Concerns about instrument exogeneity

Even with the difference in diseases, one might be concerned about endogeneity of the instrument.

Unequal variance One concern with this instrument might be that the difference between disease prevalence rates might not be orthogonal to the sum. For example, if zoonotic disease had (hypothetically) been eradicated in every country in our sample, then *diff_res=human-zoonotic=human*. Since the prevalence of disease is likely to be correlated with income and technology diffusion, this situation would render *diff_res* an invalid instrument. For two variables x and y, $(x + y)$ is uncorrelated with $(x - y)$ when x and y have equal variances. Our human and zoonotic disease variables do not have exactly the same variance. To ameliorate this concern, we also use *diff_res_std* and

diff_vec_std as instruments in table 2 and find that they produce estimates of the importance of social structure that are even larger than the initial estimates.

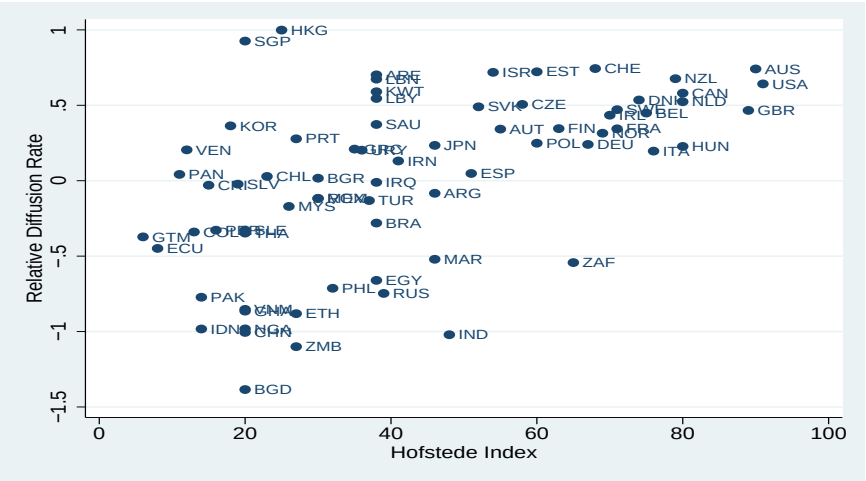
Uneven effects of technology. Greater levels of development spur public health initiatives and better health care lowers mortality rates from both types of diseases. But perhaps clean water initiatives are one of the first public health measures a country adopts when its productivity (income) rises. If this were the case, then there would be a negative correlation between aquatic illness and technology diffusion and therefore a positive correlation between (human - non-human) diseases x and shocks to technology diffusion ϵ. If $E[\epsilon x] > 0$, it would violate our identifying assumption. How would this bias the results? A positive shock to income (high ϵ) would increase the difference in disease (x), which would decrease individualism S (since we estimate $\gamma_3 < 0$). This would induce negative correlation between A and S. That negative correlation lowers the estimate of the coefficient β_2. So β_2 would be downward biased. The estimated β_2 would be a lower bound on the size of the true effect. Thus, if the instrument is invalid because economic development primarily reduces water-borne illnesses, then the true size of social structure's effect on technology diffusion is even larger than what we estimate.

Social structure affects disease. The other hypothetical cause for concern might be that faster technology diffusion and the accompanying higher income cause the social structure to change. In particular, a richer, more modern society is more likely to be market-based and individualist. The change in social social structure could affect the difference in disease prevalence by facilitating the transmission of diseases spread from human-to-human. Notice that this logic does not imply that differences in disease x are correlated with the estimation error ϵ in (5). This story suggests that social structure S depends on A, something already represented in our specification (equation 6), and it suggests that there should be an additional equation representing the idea that the instrument x depends on social structure: $x = \psi_1 + \psi_2 S + \nu$. In this structure, as long as $e[\epsilon\nu] = 0$, x is still a valid instrument for S. In other words, as long as technology diffusion affects the difference in disease through social structure, rather than directly, this form of reverse causality *does not invalidate the use of disease differences as instruments.* It only implies that γ_3 is perhaps not an unbiased estimator of the effect of disease on social institutions. Our estimates suggest that more disease is associated with less individualism. If individualism spreads disease, then this estimate is downwards-biased. In other words, the true effect of disease on social institutions would be larger than the one we estimate.

4.3 Main Results: Social Institutions and Technology Diffusion

Our main result is to quantify the effect of social structure on technology diffusion. Figure 7 illustrates the relationship between social structure and the speed of technology diffusion in a scatter plot. It reveals that more individualist societies tend to also be societies where technologies diffuse quickly. In interpreting this correlation, reverse causality is obviously 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. Therefore, we use the differences in pathogen prevalence as an instrument for social structure.

Figure 7: Technology and individualism



Comin and Mestieri (2012)'s technology diffusion measure (vertical axis) plotted against Hofstede's individualism index (horizonal axis).

The first two columns of table 3 show that the degree of individualism in a country's social structure has a large effect on a country's level of technology. A 1-standard deviation in the Hofstede index is 28.5. When we use *diff_res_std* as an instrument, a 1-standard deviation increase in individualism results in $28.5 \cdot 1.31 = 37.3\%$ increase in the speed of technology diffusion. The mean of the diffusion variable is near zero so this is not easily interpretable relative to its mean. But its standard deviation is 63.4%. Thus, a degree of individualism that is 1 standard deviation above the average is associated with technology diffusion that is 59% of a standard deviation higher than average. Across specifications, the estimates of the effect of social structure are remarkably stable. Individualism consistently explains 27-28% of the variation in technology diffusion rates.

The Sargan test statistics (in the row labeled over-ID) are chi-square statistics for the test of the null hypothesis that the instruments are uncorrelated with the regression residual ϵ. For every

Table 3: **Social Structure and Technology Diffusion (main result)**

Dependent variable:	Technology Diffusion Rate				
Instruments:	diff_res	diff_res_std	diff_vec	diff_vec_std	none
		pronoun and english			(OLS)
Individualism	1.63	1.31	1.51	1.36	1.40
	(0.33)	(0.34)	(0.34)	(0.35)	(0.28)
Over-ID p-val	0.12	0.77	0.21	0.77	
	Accept	Accept	Accept	Accept	
R^2	0.27	0.28	0.28	0.28	0.27
N	62	62	62	62	72

The first row reports $100 * \beta_2$ coefficient from an IV estimation of $A = \beta_1 + \beta_2 S + \epsilon$. Technology diffusion rate (A) comes from the Comin and Mestieri (2012) measure of the intensive technology adoption in a country. Individualism S is the Hofstede index. The variables *diff_res*, *diff_res_std* and *diff_vec* are defined in equations (1), (2) and (3). *Pronoun* and *english* are as in table 1. The over-ID test is a Sargan test statistic. The null hypothesis is that the instruments are uncorrelated with ϵ. Accept means that null hypothesis cannot be rejected at the 5% or even the 10% confidence level. All coefficients are significant at the 5% level.

IV specification, we cannot reject this null hypothesis at the 5% or even the 10% level. However, we could reject the null hypothesis at a 15% confidence level in the estimation in column (1). This suggests that the *diff_res* variable is unlikely to be a valid instrument. Note that when we use the standardized *diff_res_vec* variable as an instrument, the p-value rises to 77%, suggesting that the instrument is likely to be uncorrelated with the regression residual. We also computed Basmann statistics. They were quite close in value to the Sargan statistics in every instance.

Controlling for other possible explanatory variables. A natural question is whether social structure is simply a proxy for some other economic variable. To assess this, we choose a variety of other variables thought to explain technology adoption or income and control for their effects too. In doing so, we recognize that these control variables may themselves be endogenous. Inferring causality from these results would therefore be problematic. However, we continue to use *diff_res_std*, *pronoun* and *english* as instruments and add the following variables, one-by-one, to the first- and second-stage estimations:[8] Controlling for life expectancy at birth, social infrastructure (constructed by (Hall and Jones, 1999) to measure quality of institutions) reduces the size of the coefficient on individualism by a factor of roughly 1/2. Controlling for ethnic-linguistic fractionalization (a probability that two people belong to different ethnic or linguistic groups), latitude, disease-adjusted life expectancy, a country's degree of capitalism or socialism or population density all leave the estimate of the effect of individualism largely unchanged. Appendix C reports the

[8]Our procedure and our choice of variables here largely follow (Hall and Jones, 1999).

complete set of results for each of these estimations. In sum, there is a statistical relationship between social structure and technology diffusion that is above and beyond that which comes from other commonly-used determinants of income.

Effect of social structure on productivity and income. Our baseline results use technology diffusion as a dependent variable because it is most closely related to our theory. To interpret these results economically, it is helpful to re-run the estimation with productivity measures that are more familiar to macroeconomists: the Solow residual and output per worker.[9] We again estimate the effect of individualism on Solow residuals, instrumenting individualism with the two differences in diseases. The coefficients tell us that a 1-standard-deviation increase in the Hofstede index corresponds to a 23 (for diff_res_std, 17 for diff_vec) increase in productivity. Since the Solow residual is measured as a fraction of its value for the US, this tells us that a 1-standard-deviation increase in individualism increases productivity by 23% (25%) of the US value of productivity. For output per worker, the effects are even larger. A 1-standard deviation increase in individualism increases output per worker by 48 or 50, which represents an increase of 48% or 50% of US output per worker, depending on the set of instruments we use.

Table 4: **Social Structure, Productivity and Income**

Dependent variable:	Solow Residual		Output per capita	
Instruments:	diff_res_std pronoun, eng	diff_vec_std pronoun, eng	diff_res_std pronoun, eng	diff_vec_std pronoun, eng
Individualism	0.99	1.09	2.10	2.18
	(0.40)	(0.40)	(0.45)	(0.45)
Over-ID p-val	0.78	0.50	0.87	0.56
	Accept	Accept	Accept	Accept
R^2	0.20	0.20	0.42	0.43
N	58	58	59	59

1 std dev increase in individualism (23.0) results in 23 or 25 higher Solow residual (23% of US level) and 48 higher output per capita (48% of US level). Solow residual and output per capita come from the Penn World Tables mark 5.6. Other variables are described in table 3. All estimates are significant at 5% level.

4.4 Could Social Structure Really Change in Response to Disease?

The idea that people might choose their social circles based on disease avoidance might sound far-fetched. But researchers in animal behavior have long known that other species choose their mates with health considerations in mind (Hamilton and Zuk, 1982). Furthermore, primate research has

[9]These data come from Hall and Jones (1999).

shown that the animals most similar to human beings behave similarly to the agents in our model. Their mating strategies, group sizes, social avoidance and barriers between groups are all influenced by the presence of socially transmissible pathogens (Loehle, 1995).

One might also question whether historical societies knew enough about contagion to make informed choices about social networks. Yet, historical documents reveal a reasonable understanding of epidemiology. For example, in the sixteenth century, when smallpox reached the Americas and became a global phenomenon, people understood that the skin lesions and scabs that accompany smallpox could transmit the disease. They knew that survivors of smallpox and other infections were immune to re-infection. The practice of inoculation, whereby people were intentionally exposed to disease was practiced hundreds of years ago in China, Africa and India. Similarly, the plague was recognized to be contagious. Therefore, control measures focused primarily on quarantine and disposal of dead bodies. Even two thousand years ago, in biblical times, leprosy was understood to be contagious. Lepers, or suspected lepers, were forced to carry a bell to warn others that they were coming. Thus, the idea that one should avoid contact with others who carry particular contagious diseases is not just a modern idea.

5 Conclusions

Measuring the effect of social network structure on the economic development of countries is a challenging task. Social structure is difficult to measure and susceptible to problems with reverse causality. We use a theory of social network evolution to identify properties of social networks that can be matched with data and to select promising instrumental variables that can predict network structure. The theory predicts that societies with higher disease prevalence are more likely to become collectivist: Their social networks will have dense connections within a group, but few connections to non-group members. Such networks inhibit disease transmission, but they also inhibit idea transmission. This model guides us to choose sociological measures of individualism and collectivism to measure the prevalence of collectives in social networks. It also suggests that disease prevalence might be a useful instrument for a social network because it is one important concern that societies incorporate when they choose their network.

Of course, pathogen prevalence is not exogenous. Societies with higher incomes have better public health programs that prevent the spread of disease. Therefore, we use a difference in disease prevalence as an instrument. Since diseases that are spread from person-to-person can be avoided by choosing one's social contacts carefully, these types of diseases should affect social network formation. But diseases that are spread by contaminated water or infected animals should not affect social networks because the social network structure has no effect on the probability of

contracting the disease. Thus, the difference in the prevalence of these two diseases should predict social network formation. But this difference should be exogenous with respect to technology and income. As incomes increase, public health programs prevent both the spread of human-to-human disease and develop clean water and hygiene programs that inhibit the spread of other diseases. Thus, technology diffusion and income should not affect the difference in rates of disease prevalence. Using historical disease and difference in disease rates as instruments, we find that social structure has a significant effect on technology diffusion and an economically meaningful effect on incomes.

References

Acemoglu, D., and S. Johnson (2005): "Unbundling Institutions," *Journal of Political Economy*, 113, 949–995.

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.

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.

Comin, D., and M. Mestieri (2012): "An Intensive Exploration of Technology Diffusion," HBS Working Paper.

Conley, T., and C. Udry (2010): "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 100(1), 35–69.

Durlauf, S., and W. Brock (2006): "Social Interactions and Macroeconomics," in *Post-Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, ed. by D. Colander. New York: Cambridge University Press.

Fernández, R., A. Fogli, and C. Olivetti (2004): "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *Quarterly Journal of Economics*, 119(4), 1249–1299.

Fischer, C., and Y. Shavit (1995): "National Differences in Network Density: Israel and the United States," *Social Networks*, 17(2), 129145.

Foster, A., and M. Rosenzweig (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103(6), 1176–1209.

Gorodnichenko, Y., and G. Roland (2011): "Culture, institutions and the wealth of nations," University of California at Berkeley Working Paper.

Granovetter, M. (1973): "The Strength of Weak Ties," *American Journal of Sociology*, 78, 1360–1380.

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

Greenwood, J., A. Seshadri, and M. Yorukoglu (2005): "Engines of Liberation," *Review of Economic Studies*, 72(1), 109–133.

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.

Grinstead, C. M., and J. L. Snell (1997): *Introduction to Probability*. Russell Sage, second edn.

Gudykunst, W., G. Gao, K. Schmidt, T. Nishida, M. Bond, K. Leung, and G. W. and (1992): "The Influence of Individualism Collectivism, Self-Monitoring, and Predicted-Outcome Value on Communication in Ingroup and Outgroup Relationships," *Journal of Cross-Cultural Psychology*, 23(2), 196–213.

Hall, R., and C. Jones (1999): "Why Do Some Countries Produce So Much More Output per Worker than Others?," *Quarterly Journal of Economics*, 114, 83–116.

Hamilton, W., and M. Zuk (1982): "Heritable True Fitness and Bright Birds: A Role for Parasites?," *Science*, 218, 384–387.

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

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

Kashima, E., and Y. Kashima (1998): "Culture and Language: The Case of Cultural Dimensions and Personal Pronoun Use," *Journal of Cross-Cultural Psychology*, 29(3), 461–486.

Loehle, C. (1995): "Social Barriers to Pathogen Transmission in Wild Animal Populations," *Ecology*, 76(2), 326.

Lucas, R., and B. Moll (2011): "Knowledge Growth and the Allocation of Time," NBER Working Paper 17495.

Perla, J., and C. Tonetti (2011): "Endogenous Risk and Growth," NYU working paper.

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.

Spolaore, E., and R. Wacziarg (2009): "The Diffusion of Development," *Quarterly Journal of Economics*, 124(2), 469–529.

Tabellini, G. (2010): "Culture and Institutions: Economic Development in the Regions of Europe," *Journal of the European Economic Association*, 8(4), 677–716.

Taylor, C., and M. Hudson (1972): *World Handbook of Political andSocial Indicators*. Yale University Press (New Haven), first edn.

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.

Young, P. (2009): "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning," *The American Economic Review*, 99(5), 1899–1924.

A Proofs of Propositions

Proof of result 1 *In a collectivist network, where $\gamma = 4$, there are n unique collectives.*

Claim 1: Any three adjacent nodes are a collective.

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 collective.

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

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 collectivist network with $\gamma = 4$, 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 collective.

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 collective and there are no other collectives, there are n collectives in the network. □

Proof of result 2 *In an individualistic network, where where each person i is connected to $i-\psi$, $i-1$, $i+1$, and $i+\psi$, where $\psi > 2$, there are zero collectives.*

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 collectives among any connections of any arbitrary node i. □

Diameter of network 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 γ closest people, where γ is even, then they are connected to $\gamma/2$ people on either side. Therefore, the shortest path will be the one that advances $\gamma/2$ places around the circle, at each step in the path, until it is within $\gamma/2$ nodes of its end point. For example, agent n reach $\gamma/2$ in one step, γ in two steps and $n/2$ in $(n/2)/(\gamma/2) = n/\gamma$ steps, if n/γ is an integer. If dividing n by γ 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 $ceil(n/\gamma)$, 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 $\gamma/2$ places around the circle, and reaches the furthest node in $ceil((n-1)/2)/(\gamma/2) = ceil((n-1)/\gamma)$ steps.

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

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

Diameter of network 2. *The diameter of an individualistic 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$(n/8)+1$.□

Average path length of network 2. *In the example individualistic 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 collectivist network with $\gamma=4$, 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 3 *For a large network ($n > 8$) where $n/8$ is an integer, the individualistic network has a smaller diameter and a shorter average path length than a collectivist network with equal size n and equal degree $\gamma = 4$.*

Suppose $\psi_k(0) = 1$ for some k and $\psi_j(0) = 0 \, \forall \, j \neq k$. For a person living in location j, the sick person lives s_{jk} steps away. Since the probability of contagion is equal to 1, person j will be sick in s_{jk} periods and then die, i.e. $\Psi_j(0) = s_{jk}$. Averaging over all locations j, we have that the average lifetime is equal to the average path length from k to all other nodes: $E_j[\Psi_j(0)] = E_j[s_{jk}]$. For the maximum lifetime we have that $\max_j[\Psi_j(0)] = \max_j[s_{jk}] = diam(N)$; this is, the person whose location is furthest from k (diameter) will live the longest. Since $n > 8$ and $n/8$ is an integer, both the average path length and the diameter are longer for $N1$.

Analogously, suppose that a new idea is introduced by person k in period 0. Since the idea is transmitted with probability 1, the number of periods it takes to reach person j is given by $\alpha_j(0) = s_{jk}$. Thus the average discovery time is equal to the average path length from k to other nodes, $E_j[\alpha_j(0)] = E_j[s_{jk}]$, and the maximum discovery time $\max_j[\alpha_j(0)] = \max_j[s_{jk}] = diam(N)$. Thus the discovery process is slower in $(N1)$.

Proof of Result 4 A new technology shock advances the technological frontier if it arrives to an agent that has a technology level that is as high as any other agent in the network. Suppose that at t, the technology of each agent is the same in both types of networks and agent j (and only him[10]) is at the technological frontier. In the next period, with probability $1 - (1 - p)^4$, agent j transmits his technology to at least one of his connections and the expected number of people that have the latest technology in $t + 1$ is $1 + 4p$. That probability is the same in both networks. Each agent has an identical probability λ of inventing a new technology. Thus, the probability that a technology shock hits an agent who has the highest technology level at $t + 1$ and advances the frontier is $(1 + 4p)\lambda$, in either network.

Now consider time $t + 2$. In expectation, $1 + 12p$ people have the latest technology in N2 but only $1 + 8p$ in N1. Thus the probability of moving the frontier is $\lambda(1 + 12p)$ in N2. That probability is larger than the same probability in N1, which is given by $\lambda(1+8p)$. Continue in this fashion until every agent in the network has acquired such level of technology. At that point, all agents have the same level of technology and the probability of advancing the frontier is again equal in both networks. In every period, we find that the probability of advancing the technological frontier is weakly higher in N2 than in N1, with strict inequality in at least one period. Therefore, we conclude that the probability of a technology shock moving the frontier in N2 is than the probability of moving the frontier in N1.

Proof of Result 5 Observe that the state where all agents have the same type is absorbing. We will show that such state can be reached from any state with positive probability and therefore the process will be absorbed with probability 1 (by Lemma 1).

Lemma 1 *In an finite Markov chain that is absorbing (it has at least one absorbing state and from every state it is possible to go to an absorbing state), the probability that the process will be absorbed is 1. For proof see Grinstead and Snell (1997).*

[10]The reasoning is analogous if more than one agent receives the original shock at the same time.

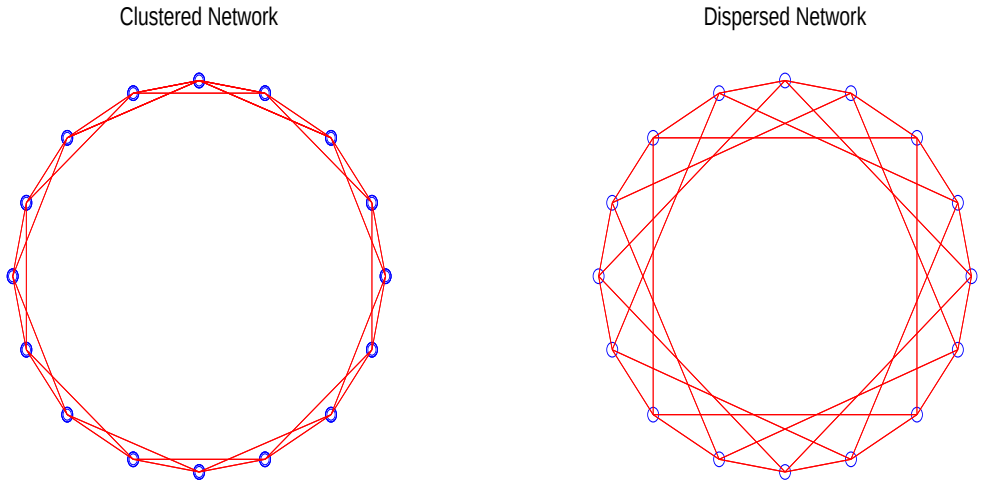
Suppose agent j is the only one whose type is different to the rest of the network. The number of j-types increases in the next period if: (i) agent j survives, (ii) all the nodes directly connected to agent j die (first tier nodes) and (iii) all the nodes connected to the nodes connected directly to agent j also die (second tier nodes). To see this, index the first tier connections with i and let $k^*(i) = argmax_{\{k:\eta_{ik}(t)=1\}} A_k(t)$. By assumption, if i dies at t, we have $\tau_i(t+1) = \tau_{k^*(i)}(t)$. Then if the three situations described happen, we have that $k^*(i) = argmax_{\{k:\eta_{ik}(t)=1\}} A_k(t) = argmax\{A_j(t), 0\} = j \ \forall i$. Therefore $\forall i$ we have $\tau_i(t+1) = \tau(k^*(i)) = \tau_j(t)$.

Now we compute a lower bound for the probability of (i)-(iii) happening at any time. First, assume $\tau_j(t) = co$. Recall that j's own type governs the links to the right and others' types govern links to the left, so in this case the first tier connections for which $\eta_{jk} = 1$ are $k = \{j-4, j-1, j+1, j+2\}$. The second tier connections (nodes connected to j's connections that are not directly connected to j) are the following: $\{j-8, j-5, j-3, j-2, j+3, j+5, j+6\}$. Therefore, with probability of at least $(1-\xi)\xi^{11}$ node j survives and all his first and second tier connections have an accident and die, reaching the absorbing state.[11]. Second, if we assume that $\tau_j(t) = in$, then his direct connections are $\eta_{jk} = 1$ for $k = \{j-2, j-1, j+1, j+4\}$ and the second tier connections are $\{j-3, j+2, j+3, j+5, j+6\}$. Therefore, with probability of at least $(1-\xi)\xi^9$ node j survives and all his first and second tier connections have an accident and die reaching the absorbing state.

In summary, we have shown that if there is one agent left with different type to the rest, with positive probability we can reach the absorbing state. If there are two or more agents whose type is different than the rest of the network, we can apply an analogous reasoning to reach the absorbing state in some finite number of steps. Since we can reach an absorbing state from any state with positive probability, the result follows from Lemma 1.

Proof of Result 6 Observe that the state with zero infected people is an absorbing state. At any given time t, for any number of sick people $m \in \{1, ..., n\}$, with probability $(1-\pi)^m > 0$ the disease is not spread and it dies out, reaching the absorbing state. Since we can reach the absorbing state from any other state with positive probability, and the number of states is finite, by Lemma 1 the probability that the process will be absorbed is 1.

B Visual Representations of our collectivist and Individualistic Networks



[11] Clearly the probability of this event is higher because of the infection process.

The connection matrix N of the collectivist 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 individualistic 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 Appendix

Summary statistics for each of the variables we use are described in table 5.

C.1 Disease Data

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

Table 5: **Summary statistics**

Variable	Obsv	Mean	Std Dev	Min	Max
Technology	75	47.84	21.33	4	95
Solow Residual	64	81.94	6.47	62.8	90.2
GDP per capita	65	92.87	8.97	70.2	104.8
Individualism	75	42.27	22.98	6	91
Pronoun	65	0.68	0.47	0	1
English	70	0 .077	0.24	0	0 .974
hum_res	75	1.28	1.16	0	3
zoo_res	75	2.87	1.49	0	6
hum_multi_res	75	10.25	5.33	1	19
diff_res	75	7.39	4.76	-1	16
diff_res_std	75	0.0011	0.995	-2.04	2.63
Life Exp	73	62.44	9.69	35.95	74.65
Soc Infra	67	0.549	0.262	0.113	1
EFL	60	36.92	29.76	0	93
daly2004	74	19,162	12,513	8,013	66,278
pathcontemp	73	32.33	6.50	23	47

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. Figure 9 uses a color-coded map to summarize the data.

Classification of Infection Diseases Epidemiologists usually classify infectious diseases according to two important epidemiologic characteristics - their means of transmission and the reservoir of the organism. According to their means of transmission, diseases can be classified into five distinct categories: contact, floor- or water-borne, airborne, vector-borne (carried by some animal that comes in contact with a person) or perinatal (contact infection during pregnancy or delivery).

The second means for epidemiological classification of infectious diseases is according to their major reservoirs in nature. The reservoir is any person, animal, plant, soil or substance in which an infectious agent normally lives and multiplies. The reservoir typically harbors the infectious agent without injury to itself and serves as a source from which other individuals can be infected. The infectious agent primarily depends on the reservoir for its survival. It is from the reservoir that the infectious substance is transmitted to a human or another susceptible host. Animals often serve as reservoirs for diseases that infect humans. The major reservoir for Yersinia pestis, the bacteria that causes plague, is wild rodents. There are also nonliving reservoirs. Soil is the reservoir for many pathogenic fungi as well as some pathogenic bacteria such as Clostridium tetani, which causes tetanus. Knowing the reservoir of the agent in addition to the means of transmission is usually needed to develop a strategy to prevent transmission. When organisms are classified according to their reservoirs in nature, four general categories are often considered: 1) Human 2) Animal (often called zoonoses) 3) Soil 4) Water.

Using information from GIDEON (Global Infectious Diseases and Epidemiology Network), we can classify the infectious diseases in the historical dataset according to agent, reservoir, vector and vehicle of transmission. GIDEON was founded in 1992 to provide health professionals with a diagnosis and reference tool for Infectious Diseases, Microbiology and Occupational Toxicology. The data in GIDEON are accessed and collated through a system of computer macros and dedicated source lists developed over the past 15 years. The sources for data included in GIDEON currently include all relevant health ministry publications (electronic and print), peer review journal publications and standard texts. A partial listing is available at *http://www.gideononline.com/resources.htm.* The quality and frequency of data input vary widely from source to source. The entire GIDEON database is updated weekly, and occasionally on a daily basis when major events occur (i.e., new case of avian influenza or an Ebola

outbreak). According to these four categories we distinguish two diseases that are transmitted from human to human, through direct contact, and have a human reservoir. These are leprosy and tuberculosis. Then we have four diseases that reproduce in humans as well as animals and are transmitted through a zoonotic vector (mosquitoes or flies): malaria, trypanosomes, leishmaniasis and dengue. One disease that has a human reservoir but a zoonotic vector (filariasis). And lastly two diseases that are transmitted through water or food (schistosomiasis and typhoid fever). Figure 8 summarizes the classification of the pathogens that we collected data on.

Figure 8: Classification of our nine infections diseases. Source: GIDEON database.

Diseases	**Agent**	**Reservoir**	**Vector**	**Vehicle**
Leprosy	BACTERIUM.	Human	None	Patient secretions
Tuberculosis	BACTERIUM	Human, Cattle	None	Air, Dairy products
Malaria	PARASITE - Protozoa	Human, Primate (Plasmodium knowlesi)	Mosquito (Anopheles)	Blood
Trypanosoma	PARASITE - Protozoa	Human, Deer, Wild carnivore, Cattle	Fly (Glossina or tsetse fly	None
Leishmaniasis	PARASITE - Protozoa	Human, Rodent, Dog, Fox	Fly	Blood
Dengue	VIRUS - RNA	Human, Mosquito, Monkey	Mosquito (Stegomyia aegypti, S. albopictus)	Blood (rare)
Filariasis	PARASITE - Nematoda	Human	Mosquito (Anopheles, Aedes, Culex)	None
Schistosomiasis	PARASITE - Platyhelminthes, Trematoda	Snail (Bulinus, Planorbarius) Rarely baboon or monkey	None	Water
Typhoid fever	BACTERIUM	Human	None	Water, Food

Contemporaneous disease data. To assess the accuracy of our historical disease prevalence series, we compare them to contemporaneous data that is presumably better-measured. Data were obtained from the Global Infectious Diseases and Epidemiology Online Network (GIDEON, http://www.gideononline.com). The database is updated weekly; the data used were obtained during the period April–June 2007. GIDEON reports current distributions of infectious diseases in each country of the world. The data cover seven classes of pathogens (leishmanias, trypanosomes, malaria, schistosomes, filariae, spirochetes and leprosy) and codes the relative prevalence of each spe-

cific pathogenic disease within each class. A total of 22 specific pathogenic diseases are coded, each on the same three-point prevalence scale. These values are summed within each country to create a composite index estimating the contemporary prevalence of pathogens.

Our two pathogen prevalence indices appear to be accurate because they are highly correlated (0.77) with the corresponding indices, based on contemporaneous data. They are also highly correlated with a similar index created by Gangestad & Buss (1993) to assess pathogen prevalence within a smaller sample of 29 regions. Correlations are 0.89 with our index from 1930's data and 0.83 with our index of 2000's data.

Even though the contemporaneous data probably has less measurement error, the historical pathogen data has a stronger statistical relationship with social structure. The coefficient on the historical nine-pathogen index in table **??** is -2.64. The analogous coefficient on the contemporaneous index is -2.24. The contemporaneous data could be reflecting more reverse causality since a less individualistic social structure in the 1970's could be responsible for lower disease prevalence today. That would explain why the coefficient is less negative.

C.2 Measuring Individualism

Hofstede (2001) 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 Hofstede individualism index values are based on the results of a factor analysis of work goals across countries. The index was constructed from data collected during an employee attitude survey program conducted by a large multinational organization (IBM) within its subsidiaries in 72 countries. The survey took place in two waves, in 1969 and 1972 and included questions about demographics, satisfaction and work goals. The answers to the 14 questions about "work goals" form the basis for the construction of the individualism index. The individual answers were aggregated at the country level after matching respondents by occupation, age and gender. The countries mean scores for the 14 "work goals" were then analyzed using factor analysis that resulted in the identification of two factors of equal strength that together explained 46% of the variance. The individualism factor is mapped onto a scale from 1 to 100 to create the individualism index (hereafter IDV) for each country. The highest IDV values are for the United States (91), Australia (90), and Great Britain (89); the lowest are for Guatemala (6), Ecuador (8) and Panama (11). Subsequent studies involving commercial airline pilots and students (23 countries), civil service managers (14 counties) and consumers (15 countries) have validated Hofstede's results.

IBM survey text (a subset). The original Hofstede survey is too lengthy to include in its entirety. Below, we list a subset of the questions asked. We categorize questions according to which aspect of collectivism they measure, as described in section 3.2. That grouping is not in the original survey. The survey instructions read as follows:

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 to:

Category 1: Questions about the importance of personal freedom and individual benefits from the organization

1. Have considerable freedom to adopt your own approach to the job (I)
2. Have a job which leaves you sufficient time for your personal or family life (I)
3. Have challenging work to do (I)

   In contrast, the last example question emphasizes the opposite, how the organization benefits from the individual's skills:

4. Fully use your skills and abilities on the job (C)

Category 2: Value of cooperation

1. Work with people who cooperate well with each other (C)
2. Have training opportunities (C)

Category 3: Willingness to change job or location

1. Live in an area desirable to you and your family (I)

We have followed the question with (I) when high importance (a low numerical score) indicates more individualism. When the higher importance indicates less individualism (more collectivist) we denote that with (C). We report these particular questions because all have factor loadings of 0.35 or more in absolute value.

Cross-Country Network Analysis There is a small literature that analyzes and compares social network structures across countries. It is summarized and extended by Fischer and Shavit (1995). Surveys typically ask respondents to name people with whom they confided, were friends, asked for help, ect. The survey takers would then interview the named friends to find out their networks and interview the friends they named as well. By repeating this process many times, the researchers could map out fairly complete social networks in specific geographic locations. For our purposes, the key finding from these studies is that the frequency of network collectives varies greatly across countries. These studies do not typically report the number of collectives. They report a related measure, network density. Density is the fraction of possible links between individuals that are present. Importantly, a network that is fully dense also has the maximum possible number of collectives. Because this research design involves lengthy interviews of many respondents, it has been done only on a handful of countries. But it is useful to see how the prevalence of network collectives correlates with Hofstede's individualism index.

Table 6: **Measures of network interdependence and individualism**

Region	Country	Network interdependence	Individualism (for country)
Haifa	Israel	0.57	54
N. California	U.S.	0.44	91
all	U.S.	0.40	91
E.York, Toronto	Canada	0.33	80
London	U.K.	0.34	89
Taijin	China	0.58	20
West Africa		0.45-0.77	20

The theory predicts a negative relationship between network interdependence (closely related to collectivism) and the individualism index. Interdependence is measured as the fraction of all possible links in a social network that are present. It is also referred to as "network density." West Africa here includes Ghana, Nigeria and Sierra Leone.

Correlation of individualism with other measures of culture. To better understand what Hofstede's individualism index (IDV) measures, we examine related cultural measures that are highly correlated with the index.

Family structure. In a collectivistic society, people grow up with members of an extended family and sometimes also neighbors, housemates, other villagers, lords and servants. Collectivists have strong ties and frequent contact with family members. In individualistic societies, people grow up in nuclear families. Their family ties are weaker. Extended family live elsewhere and visit infrequently.

Group identity. In collectivist societies, people learn to think about themselves as part of collective, with a group identity. That identity is determined by birth. Similarly, friendships come from existing group ties. Members of the collective are distinct from non-members. In the individualistic society, people learn to think about themselves as an individual, not a member of a group. There is no distinction between group members and and non-members. Gudykunst, Gao, Schmidt, Nishida, Bond, Leung, and and (1992) surveyed 200 students in each of 4 countries: Australia and US (high IDV) and Hong Kong and Japan (lower IDV). Half of the respondents were asked to imagine a group member; the others were asked to imagine a non-member. They were then asked to report if they would: talk about themselves with the person, ask about the other, expect *shared attitudes and networks*, and have confidence in the other. The differences between how respondents viewed group members and non-members correlated exactly (negatively) with their country's IDV scores.

Other ways of modeling individualism and collectivism in networks. **Weak vs. strong ties** Granovetter (1973) introduced the idea of strong ties and weak ties in networks. Strong ties are close friends, while weak ties are acquaintances. Granovetter argues that more novel information comes from weak ties than from strong ties. The reasoning is very similar to that in our model. Because people who are very closely socially related have similar information sets, they are more likely to convey redundant information and are less likely to have novel information. Weak ties are more likely to be connected to people that we do not know and therefore are possible conduits for new information. Granovetter argues that people with few weak ties are at an informational disadvantage because they have difficulty accessing information in other parts of the social network. Thus, a society comprised of agents with mostly strong ties and few weak ties will not transmit information (or disease) as easily. Thus another way to formulate our model that would lead to the same conclusions would be to characterize collectivist societies as ones with strong ties and individualist societies as one with weak ties.

Random vs. fixed networks Another characteristic of individualist economies is that more commerce is mediated by a market, rather than being based on personal relationships. One could think of a market as being like a random search model. Buyers encounter suppliers with various prices and decide to do business or not. A random search model looks almost identical to a random network, where agents are connected to others in the network with some probability. In contrast, the collectivist economy is one where transactions take place only between people who are connected and those connections do not change over time. This captures the essence of market vs. relational transactions. For most network structures, the random network will achieve faster diffusion of technology and diseases than the fixed network (see Jackson (2008)). Thus, modeling individualist and collectivist societies as fixed or random networks would also not change the basic message of the paper.

C.3 Measuring Technology

C.4 Other Control Variables

An inevitable question arises: "What if you also control for X?" We would like to know if individualism is highly correlated with and thus proxying for some other economic phenomenon. The problem with answering this question is that what we would like to control for is likely an endogenous variable. We could treat it as such and instrument for it. But in most cases, our instruments are not strong predictors. Or, we could just, suspend disbelief, assume that these are exogenous variables, abandon any pretense of saying anything about causality, and just see what statistical relationship they have with the other variables in the estimation. We take the second approach. Each row of table 7 reports the coefficients of a second stage regression of technology diffusion on hofstede, one other control variable, and a constant. Since we have assumed that the control variable is exogenous, we use it as an instrument in the first stage, in addition to a constant and our standard instruments: pronoun, english and the standardized difference in pathogens variable, diff_res_std.

The control variables are social infrastructure, a measure of the efficient functioning of political and social institutions, constructed by Hall and Jones (1999); ethno-linguistic fractionalization, a measure of the probability that two randomly-chosen people in the country will belong to different ethnic or linguistic groups, constructed by Taylor and Hudson (1972); latitude, which is the absolute value of the country's latitude, divided by 90; disability-adjusted life expectancy, which is the expected length of time an individual lives free of disability, is measured by the World Health Organization in 2004 (*http://www.who.int/healthinfo/global_burden_disease/estimates_country/en/index.html*); capitalist, which is the "economic organization" variable constructed by Freedom House, scores more capitalist countries higher and more socialist countries lower; and population density is the 1970 population per square mile, as reported by the World Bank.

Table 7: **Controlling for other economic variables**

Dependent variable	Technology Diffusion						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individualism (S)	0.59	0.69	1.23*	1.35*	1.02*	1.24*	1.46*
	(0.34)	(0.39)	(0.30)	(0.36)	(0.27)	(0.36)	(0.31)
Life expectancy	4.29*						
at birth (LEB)	(0.78)						
Social		112.2*					
Infrastructure (SocI)		(30.17)					
Ethno-linguistic			−1.08*				
fractionalization (EFL)			(0.21)				
Latitude				0.21			
				(0.26)			
Disease-adj					−0.0030*		
life expectancy (DALY)					(0.0006)		
Capitalist						5.89	
(EcOrg)						(4.44)	
Population							0.040*
Density							(0.010)
Constant	-300.7	-98.51	-15.40	-67.68	7.05	-76.77	-72.64
R^2	0.58	0.47	0.52	0.33	0.63	0.34	0.43
Observations	62	60	55	61	61	61	62

2SLS estimates of $100 * \gamma$ coefficients in Diffusion $= \gamma_1 + \gamma_2 \mathrm{S} + \gamma_3 x + \eta$, where the x variables are listed in the first column of the table. The first stage regression is $S = h_1 + h_2 x + h_3 \mathrm{diff\_res\_std} + h_4 pronoun + h_5 english + e$. Standard errors in parentheses. $*$ denotes significance at 5% level.

Figure 9: A world map of historical pathogen prevalence.