Germs, Social Networks, and Growth

Alessandra Fogli
Federal Reserve Bank of Minneapolis

Laura Veldkamp
Columbia Graduate School of Business

Staff Report 572
November 2018

DOI: https://doi.org/10.21034/sr.572
Keywords: Growth; Development; Technology diffusion; Economic networks; Social networks; Pathogens; Disease
JEL classification: E02, O1, O33, I1

Federal Reserve Bank of Minneapolis • 90 Hennepin Avenue • Minneapolis, MN 55480-0291
https://www.minneapolisfed.org/research/

Germs, Social Networks and Growth

Alessandra Fogli and Laura Veldkamp*

November 9, 2018

Abstract

Does the pattern of social connections between individuals matter for macroeconomic outcomes? If so, where do these differences come from and how large are their effects? Using network analysis tools, we explore how different social network structures affect technology diffusion and thereby a country's rate of growth. The correlation between high-diffusion networks and income is strongly positive. But when we use a model to isolate the effect of a change in social networks, the effect can be positive, negative, or zero. The reason is that networks diffuse ideas and disease. Low-diffusion networks have evolved in countries where disease is prevalent because limited connectivity protects residents from epidemics. But a low-diffusion network in a low-disease environment needlessly compromises the diffusion of good ideas. In general, social networks have evolved to fit their economic and epidemiological environment. Trying to change networks in one country to mimic those in a higher-income country may well be counterproductive.

How does a country's culture affect its income? Many papers in macroeconomics have tackled this question by modeling various aspects of culture and measuring its economic consequences.[1] This paper explores one aspect of a country's culture: its pattern of social connections. Using tools from network analysis, we explore how and to what extent different social network structures might affect a country's rate of technological progress. Our network model explains why societies might adopt growth-inhibiting structures and allows us to quantify the effect of networks on income.

Measuring the speed of information or technology diffusion within various kinds of networks has a long history(Granovetter, 2005). To measure the macroeconomic effect of Granovetter's theory, we

*Corresponding author: afogli00@gmail.com, Federal Reserve Bank of Minneapolis, 90 Hennepin Ave., Minneapolis, MN 55405. lv2405@columbia.edu, Columbia Graduate School of Business, 3022 Broadway, New York, NY 10027. We thank participants at the Minnesota Workshop in Macroeconomic Theory, NBER EF&G meetings, SED meetings, the Conference on the Economics of Interactions and Culture and Einaudi Institute, the Munich conference on Cultural Change and Economic Growth, SITE, NBER Macro across Time and Space, and NBER growth meetings and seminar participants at Bocconi, Brown, USC, Stanford, Chicago, Western Ontario, Minnesota, Penn State, George Washington, and NYU for their comments and suggestions. We thank Corey Fincher and Damian Murray for help with the pathogen data, Diego Comin, Pascaline Dupas, Chad Jones, and Marti Mestieri for useful comments, and Isaac Baley, Callum Jones, Hyunju Lee, David Low, Amanda Michaud, and Arnav Sood for invaluable research assistance. Laura Veldkamp thanks the Hoover Institution for its 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]See, for example, Bisin and Verdier (2001), Tertilt (2005), Doepke and Tertilt (2009), Greenwood and Guner (2010), or Doepke and Zilibotti (2014) for a literature review.

could take a simple production economy, embed within it fast-diffusion and slow-diffusion networks, calibrate it, and measure the difference in output growth. This simple measurement exercise faces two key challenges. The first challenge is determining what network to use for a whole country. Although researchers have mapped complete social networks in schools or online communities, mapping the exact social network structure for a large cross section of countries is not feasible. Instead, we propose three aggregate features of networks that regulate the network's diffusion speed and have measurable counterparts in cross-country data. We then measure a country's social network by rating the network along those three dimensions. The second challenge is that there are multiple channels of endogeneity: Income affects social network structure; networks have long been known to affect disease transmission, which in turn affects productivity and income. Finally, technology diffusion reduces disease prevalence, and lower death rates alter how social networks evolve. All of these channels compromise the direct measurement of social networks' effects. In response, we build a model with each of these various sources of endogeneity in the model. Then we calibrate the model, assess its performance with overidentifying moments, and use the model to measure the effect of the network alone.

Our theory for how networks evolve endogenously revolves around the idea that communicable diseases and technologies spread in similar ways – through human contact. We explore an evolutionary model in which networks that are stable, local, and have fewer connections reduce the risk of infection, allowing the participants to live longer. But such low-diffusion networks also restrict the group's exposure to new technologies. In countries where communicable diseases are inherently more prevalent, the high risk of infection makes nodes with many, unstable, or distant linkages more likely to die out. A network that inhibits the spread of disease and technology will emerge. In countries where communicable diseases are less prevalent, nodes with few, stable, and local connections will be less economically and reproductively successful and will die off in the long run.

Section 1 begins by considering an evolutionary model of social networks where the networks govern disease and technology diffusion, but diseases and technology also govern network evolution. Section 2 examines the effect of each network feature on technology and disease diffusion and explores the reverse effects: how technology and disease affect the types of networks that emerge. Specifically, disease prevalence creates the conditions that are conducive to growth-inhibiting networks. Section 3 describes our measures of pathogen prevalence, social networks, and technology diffusion and how we use them to calibrate the model. Section 4 then uses the calibrated model to investigate how much an exogenous change in network structure would affect technology and output, and explores why this answer depends on the country. Finally, we estimate the effect of social networks on technology diffusion in the data, using the difference in communicable and

noncommunicable diseases as an instrument. Since diseases spread by humans are systematically more likely to infect people who are more central to their social network, human-to-human diseases are widely thought to affect the evolution of social networks, more than other diseases do. The model and data agree that changes in the social network can raise a country's productivity and output growth by 50-100%. But when applied to an economy with high disease prevalence, the same change can undermine output by propagating disease.

Related Literature The paper contributes to four growing literatures. Primarily, the paper is about a technology diffusion process. It is perhaps most closely related to economic history work by de la Croix et al. (2017) on the role that guilds and extended families played in technology diffusion in preindustrial economies. Recent work by Lucas and Moll (2014) and Perla and Tonetti (2011) uses a search model framework in which every agent who searches is equally likely to encounter any other agent and acquire the agent's technology. Greenwood et al. (2005) model innovations that are known to all but are adopted when the user's income becomes sufficiently high. In Comin et al. (2013), innovations diffuse spatially. What sets our paper apart is its assumption that agents encounter only those in their own network. Our main results all arise from this focus on the network topology. Many recent papers use networks to represent the input/output structure of the economy instead of social connections.[2] Our focus on social networks creates new measurement challenges and leads us to examine different forms of networks. For example, Oberfield (2013) models firms that optimally choose a single firm to connect to, which is appropriate for his question but precludes thinking about the network features we examine.

The paper also contributes to the literature on culture and its macroeconomic effects (e.g., Bisin and Verdier (2001), Greenwood and Guner (2010), or Doepke and Tertilt (2009)). Gorodnichenko and Roland (2017) focus on the psychological or preference aspects of collectivism, one of the four network measures we use as well. They use collectivism to proxy for individuals' innovation preferences and consider the effects of these preferences on income. In contrast, we use collectivism as one of many dimensions of a social network and assess the effect of those relationships on the speed of technology *diffusion*. Similarly, most work on culture and macroeconomics regards culture as an aspect of preferences.[3] Greif (1994) argues that preferences and social networks are intertwined because culture is an important determinant of a society's network structure. Although this may be true, we examine a different determinant of networks – pathogen prevalence – that is easily measurable for an entire country. Our evolutionary-sociological approach lends itself to quantifying the aggregate effects of social networks on economic outcomes.

[2]See, for example, Chaney (2014) or Kelley et al. (2013).

[3]See, for example, Tabellini (2010) and Algan and Cahuc (2007). Brock and Durlauf (2006) review work on social influence in macroeconomics but bemoan the lack of work that incorporates social network interactions.

Our empirical methodology clearly draws much of its inspiration from work on the role of political institutions by Acemoglu et al. (2002), Acemoglu and Johnson (2005) and the role of social infrastructure by Hall and Jones (1999). But instead of examining institutions or infrastructure, which are not about the pattern of social connections between individuals, we study an equally important but distinct type of social organization, the social network structure.

There is a micro literature that considers the effects of social networks on economic outcomes (e.g., see Granovetter, 2005; 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 networks affect technology diffusion economy-wide. Ashraf and Galor (2012) and Spolaore and Wacziarg (2009) also take a macro perspective but measure social distance with genetic distance. Our network theory and findings complement this work by offering an endogenous mechanism to explain the origins of social distance and why it might be related to the diffusion of new ideas.

Finally, many papers argue that disease prevalence is growth inhibiting.[4] To this discussion, our paper adds a new, social-networks-based channel through which health affects economic outcomes.

1 A Network Diffusion Model

The model is a framework for measurement, which allows us to quantify the effect of networks on growth. While our model is surely simpler than reality, our quantitative exercises give us a sense of the order of magnitude of the macroeconomic effect. The concept of social network structure is a fungible one. We pick particular aspects of networks on which to anchor our analysis. In doing this, we do not exclude the possibility that other aspects of social or cultural institutions are important for technology diffusion and income. We focus on dimensions we can measure. The model teaches us about how three different aspects of social networks facilitate technology diffusion. It also explains why disease that spreads from human to human might influence a society's social network in a persistent way.

A key feature of our model linking social networks to technological progress is that technologies are spread by human contact. This feature is not obvious, since new ideas could be spread by print or electronic media. However, Ryan and Gross (1943) first established the importance of social contacts in the spread of technology. Since then, a large literature in sociology and consumer research, starting with Rogers (1976), equates a diffusion model with innovations that spread

[4]Bloom et al. (2004) summarize this literature, which typically uses aggregate economic data and aggregate health data, as well as standard controls, and estimates that a one-year increase in life expectancy raises output by 4%. In contrast, other authors such as Weil (2007) use micro-evidence of health effects on individual output and then simulate the implied aggregate effect on GDP.

through a social system.[5] In his 1969 American Economic Association 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. (Arrow, 1969, p. 33)

While the idea that certain network properties facilitate information diffusion is well known, the origin of such social networks is less-explored territory. Since a key problem with determining the economic effect of social networks in the data is the endogeneity of networks, it is important for our model to have endogenous social networks as well. Therefore, a second key feature of our model is an evolutionary process for social networks, in which economically successful agents pass on their pattern of social connections. This evolutionary model also explains why growth-inhibiting social networks emerge in response to disease. The idea that social circles might evolve based on disease avoidance is based on economic and biological evidence (Birchenall, 2014).[6] Motivated by this evidence, 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 J agents, indexed by their location $j \epsilon \{1, 2, \ldots, J\}$ on a circle. Agents have social connections that we describe with a network.

Social Networks Each person i is socially connected to others. If two people have a social network connection, we call them "friends." Friends could also be family, coworkers or any pair of people with close and repeated social contact. Let $\eta_{jk} = \eta_{kj} = 1$ 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 symmetric network of all connections be denoted N.

[5] In economics, Foster and Rosenzweig (1995) spawned a branch of the growth and development literatures that focuses on the role of personal contact in technology diffusion; see Conley and Udry (2010) or Young (2009) for a review.

[6] Animal behavior researchers have long known that many species form social connections that depend on the group's health status (Hamilton and Zuk, 1982). For primates in particular, mating strategies, group sizes, social avoidance, and barriers between groups are all influenced by the presence of socially transmissible pathogens (Loehle, 1995). Thornhill et al. (2010) document this effect in human societies. An alternative to the evolutionary approach would be to work with a network choice model. But equilibria in such models often do not exist, and when they do, they are typically not unique.

In a large economy, it is not possible to measure the complete pattern of linkages between every agent. Instead, we categorize agents, or nodes, into types whose linkages follow particular patterns. This approach allows us to characterize a large, aggregate network by its fraction of nodes of each type. This characterization enables comparisons with survey data that will enable us to measure the properties of social networks across countries.

At each date t, each person j has a type $\tau_j(t)$. That type is a three-dimensional state. The three dimensions to node type are: a binary collectivist type θ_j, a degree type n_{fj}, and a mobility type m_j, where n_{fj} and m_j come from a discrete, finite set of possible types. Note that node type does not include technology.

When we think about node types governing social connections, we are faced with an inevitable dilemma: What happens if i's type dictates that he should be friends with j, but j should not be friends with i? To resolve this impasse, we assume that each agent's type dictates her links in one direction but not the other. This dilemma is a product of the fact that social connections are inherently bidirectional. To break this impasse, we define node types by the patterns of links that nodes form to one side (their left, if we order nodes clockwise). These definitions do not rule out other links. In fact, every node will have additional links on the other side. But the nature of those (right-side) links is governed by other nodes' types.

We explore three aspects of social networks because they are important determinants of diffusion speed and we have cross-country data measuring them. Of course, this means that we hold fixed many other aspects of networks that may also differ across countries. Measuring these other aspects of social networks and understanding their effects on economic growth would be useful topics for further research.

Network Feature 1: Degree. The degree of a node is the number of connections that node has to other nodes. In the context of a social network, degree is the number of friends a person has. In our model, the degree type of a node regulates the number of links it has on one side.

Definition 1 *If agent j has degree n_{fj} (even), he is connected to $n_{fj}/2$ other nodes to his left.*

The number of links one creates on one side is given by the degree divided by two because that ensures that if every node has degree-type n_f, every node is connected to n_f others.

Network Feature 2: Collectivism / Individualism. Collectivism and its opposite, individualism, define a feature of social networks that has been studied extensively by sociologists. Mutual friendships and interdependence are hallmarks of collectivist societies. Individualists have friendships, but those friends are less likely to know one another. To measure collectivism, 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 in which i, j, and k are all connected to each other as a *collective*.

A measure of the extent of shared friendships, and thus the degree of collectivism, is the number of such collectives. The number of collectives 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). A fully collectivist network is illustrated in Figure 1.

Definition 2 *If agent j is collectivist (which we denote by $\theta_j = 1$), then he is connected to the $n_{fj}/2$ closest nodes to his left: $\eta_{jk} = 1$ for $k = \{j+1, \ldots, j+n_{fj}/2\}$, modulus the size of the network.*

All nodes are connected to their immediate neighbors. This is what constitutes a neighbor in this model. The modulus phrase in the definition simply means that, for example, if $j + n_f/2 = N + 1$, then since there is no $N + 1$st node, the link simply wraps around the other side of the circle and connects with node 1. Therefore, when a node j is connected to $j + 1$ and $j + 2$, that will always form a collective, since those neighbors are always themselves connected. Thus, collectivist nodes are ones whose ties complete a collective. The opposite of collectivist is individualist: A node that is individualist is connected to some nodes that are not themselves connected.

Definition 3 *If agent j is an individualist (not collectivist: $\theta_j = 0$) with degree n_f, then she is connected to the closest node $j + 1$ and $n_{fj}/2 - 1$ nodes that are all at least $\Delta_j > 2$ spaces away: $\eta_{jk} = 1$ for $k = \{j + 1, j + \Delta_j, \ldots, (j + \Delta_j + n_f/2 - 2)\}$, modulus the size of the network.*

Let $\bar{\theta} \equiv \int \theta_i di$ denote the fraction of collectivists in a network.

Network Feature 3: Mobility The third network feature we introduce is shortcut links that span the network. These long links represent the long-distance social ties that arise when some agents in a society are mobile. Their frequent travels bring them in close social contact with others who are not in their social neighborhood. We model this as a small probability of a long link to a randomly chosen node in the network.

Definition 4 *If an agent j has mobility m_j, then with probability m_j, one of j's existing links is broken and reassigned to any other node which with j is unconnected, with equal probability. This long link endures until the next social change shock arrives.*

This type of ring network, with random long links, is a small-world network (Watts and Strogatz, 1998). Sociologists frequently use small-world networks as an approximation to large social networks because of their high degree of collectivism and small average path length, both pervasive features of real social networks.

These are the three network features that we will map to data to assess the economic importance of social networks. A node can be any combination of the three features. However, some

combinations do create tension. When we construct the network, we first construct collectivist or individualist links with the right degree. Then, we randomly draw mobility shocks and rewire some social linkages. A mobile node may end up connecting with a low-degree or collectivist node that now has additional long links. These random connections do not change the type of the node connected to. They do highlight that our categorizations of social relationships are only aggregate and approximate characteristics of the network.

Network Evolution A node has an opportunity to change type when a social change shock arrives. This shock, $\xi_{rj}(t) \in \{0, 1\}$, is drawn independently across individuals and time and takes on the value 1 with probability p_ξ and 0 otherwise. When $\xi_{rj}(t) = 1$, the agent at node j adopts the type of the most successful node that j is connected to. In other words, if the person at node j is socially connected to nodes $\{k : \eta_{jk}(t) = 1\}$ and gets hit with a social change shock at time t, then at time $t + 1$ he adopts the type of $k^* = argmax_{\{k:\eta_{jk}(t)=1\}} A_k(t)$ (i.e., the friend with the highest time-t technology). Then the time-$(t + 1)$ type of person j is the same as the time-t type of person k^*: $\tau_j(t + 1) = \tau_{k^*}(t)$.[7] Whether or not one's type changes, when social change arrives ($\xi_{rj}(t) = 1$), agent j takes a new, independent draw of social mobility. Whether they have a long link and which node that link connects them to can change.

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

Technology and Output Each agent produces output with a technology $A_j(t)$:

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

Technological progress occurs when someone improves on an existing technology. To make this improvement, the person needs to know about the existing technology. Thus, if a person is producing with technology $A_j(t)$, she will invent the next technology with an i.i.d. probability λ_t each period. If she invents 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

[7] Another logical specification for the social change shock is to link it to disease, so that someone who gets sick dies and then is reborn with, potentially, a different type. The problem with this formulation of the model is that epidemics prompt rapid social change. Since this is counterfactual, our model makes social change independent of disease.

$A_k(t) > A_j(t)$, then with an i.i.d. probability ϕ, j can produce with k's technology in the following period: $A_j(t+1) = A_k(t)$. With multiple friends, each offers an independent chance of technology transmission. If multiple new technologies are transmitted to the same node, the best technology is adopted.

A social change shock does not alter the technology diffusion process. It only changes the network, which affects technology diffusion indirectly.

Disease, Death, and Renewal Each infected person transmits the disease to each of his friends with probability π_t. The transmission to each friend is an independent event. Thus, if $\dot{n}_d$ friends are diseased at time $t-1$, the probability of being healthy at time t is $(1-\pi_t)^{\dot{n}_d}$. If no friends have a disease at time $t-1$, then the probability of contracting the disease at time t is zero.

Let $d_j(t) = 1$ if the person in location j acquires a transmittable disease (is sick) in period t and $= 0$ otherwise. An agent j who acquires a disease is sick and loses the ability to produce for the remainder of her life span ($A_j(s) = 0,\ \forall s \epsilon [t, t_r]$). The date at which life ends and a new node appears t_r is governed by the social change shock $\xi_{rj}(t)$. When social change shock hits ($\xi_{rj}(t) = 1$), the agent j is replaced by a new, healthy person in the same location j. Just like when a healthy node is renewed, the new agent j inherits the technology and network type of the most productive node that the parent was connected to.

We link the rebirth and social change process because social change only propagates when it affects productive nodes that can pass on their technology and type to others. Therefore, when we allow a node to adjust its type, we also cure it of disease if it is sick. It is as if the node dies and is reborn, healthy, with a type dictated by its community. This assumption facilitates the process of social change.

Feedbacks between Disease and Technology The challenge in measuring the effect of networks on income is endogeneity. Diseases affect not only networks but also technological innovation. Similarly, technology, which obviously raises income, also helps to eradicate disease, which in turn, influences social networks. Therefore, we build these effects into our model, calibrate them, and then account for them when we do model experiments that gauge the effect of a change in a social network.

Feedback 1 (Innovation): The rate of technological innovation depends on disease prevalence. The time-varying probability of innovation is

$$\lambda_t = 2\lambda_0(1 - \Phi(\bar{d}(t)/\kappa_T)), \tag{1}$$

where Φ is a standard normal cumulative density and $\bar{d}(t) \equiv 1/n \sum_j d_j(t)$ (recall that $d_j(t) = 1$ if

agent j is infected at time t and 0 otherwise). This captures the idea that having a large number of infected people in a society is not conducive to innovation.

Feedback 2 (Infection): The probability of disease transmission depends on the aggregate technology level, which is defined as $A(t) \equiv 1/n \sum_j A_j(t)$. The endogenous disease transmission probability is

$$\pi_t = 2\pi_0(1 - \Phi(A(t)/\kappa_G)), \tag{2}$$

where Π_A is a parameter that governs the strength of the technology effect on disease transmission. This captures the idea that as technology improves, public health innovations, vaccines, and medical advances make new infections less likely.

2 Theoretical Results

Before using the model to measure network effects, we briefly explain the model's basic properties. We begin by exploring how each network characteristic affects a network summary statistic known as average path length, and how that path length is related to both the expected time to infection and the expected rate of technological progress. These results formally break down the logic for why some types of social networks can increase aggregate income. We then explore the long-run convergence of networks. The proofs of all results are in Appendix A.

2.1 Diffusion Speed, Infections, and Innovations

We begin with a definition of average path length. Average path length is the average number of steps along the shortest paths for all possible pairs of network nodes. Let p_{ij} represent the shortest path length between nodes i and j and $\mathcal{N} = \{1, \ldots, J\}$ represent the set of J nodes. Then,

$$\text{Average path length} = \frac{1}{J} \sum_{i \in \mathcal{N}} \frac{\sum_{j \in \mathcal{N}/i} p_{ij}}{J - 1}. \tag{3}$$

Average path length is important because it governs the mean infection time from a disease and the mean discovery time for a new technological innovation. Let $\bar{L}$ be the average healthy lifetime of an agent in network N (specifically, it is the number of consecutive healthy periods of a node j). Similarly, let $\bar{\alpha}$ be the average number of periods it takes for a new idea to reach a given agent in network N. We call $\bar{\alpha}$ the average adoption lag.

Result 1 *If $\pi = 1$ and $\sum_j d_j(0) = 1$, then the average healthy lifetime $\bar{L}(N)$ is monotonically increasing in the average path length of the network.*

If $\phi = 1$, then the average technology adoption lag $\bar{\alpha}(N)$ is monotonically decreasing in the average path length of a network.

Of course, diffusion is not the same as innovation. Diffusion accelerates technology growth because when idea diffusion is faster, redundant innovations are less frequent. Thus, more of the innovations end up advancing the technological frontier.

Next, we explore how the three characteristics of nodes in our network regulate the average path length of the network. The next result shows that higher degree, individualism, and greater mobility all have the effect of reducing path lengths, on average.

Result 2 *The following network features reduce the average path length of the network:*

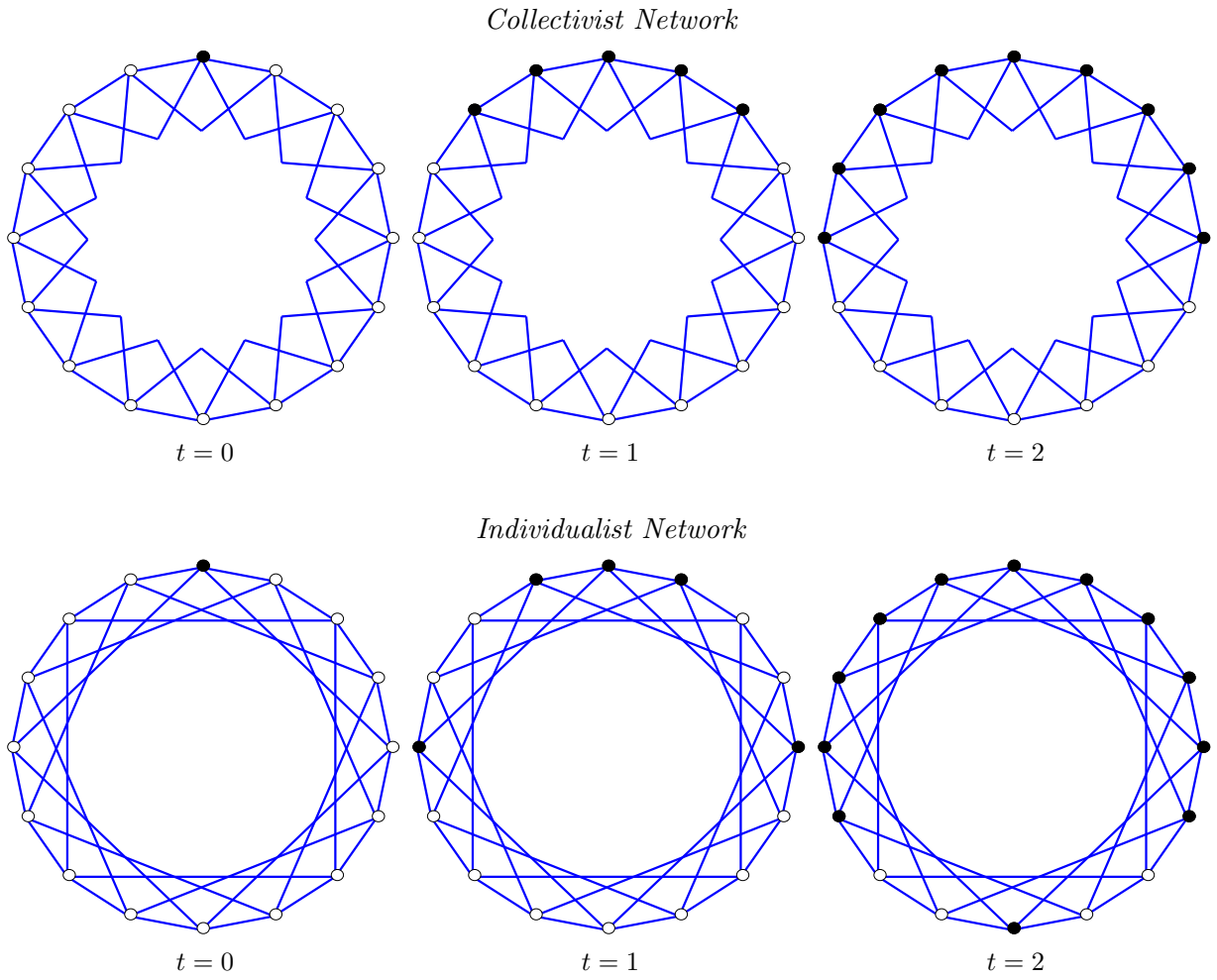
1. *Higher degree: The average path length in a ring network is a decreasing function of n_f.*
2. *Individualism: $\theta_j = 1, \forall j$ is a network with a longer path length than $\theta_j = 0, \forall j$, for a given degree, no mobility, and a network of size $J > \bar{J}$.*
3. *Greater mobility: The expected average path length of the network is a decreasing function of the mobility probability $m_j \; \forall j$.*

A social network with a higher degree has a lower average path length. With more connections, it requires fewer steps to reach other nodes.

For individualism, the logic is slightly different. Disease and technology spread more slowly in the collectivist network because each contiguous group of friends is connected to, at most, four nongroup 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 faraway location. In the meantime, someone else may have reinvented 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.

Figure 1 illustrates the smaller path length and faster diffusion process in a purely individualist network, compared with a purely collectivist one, where both networks have degree 4 and no mobility. In this simple case, in which the probability of transmission is 1, each frame shows the transmission of an idea or a disease introduced to one node at time 0. The "infected" person transmits that technology to all the individuals to whom she is connected. In period 1, 4 new people use the new technology, in both networks. But by period 2, 9 people are using the technology in the collectivist network and 12 are using it in the individualist network. In each case, an adopter

Figure 1: Slower diffusion in the collectivist network ($\theta_j = 0 \forall j$, top) than in the individualist network ($\theta_j = 1 \forall j$, bottom). Black dots represent nodes that have acquired a new technology that was discovered by the top node.



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, so diffusion is slowed.[8]

For mobility, a greater probability of forming long links (higher mobility) is similar to the

[8]What if technology diffusion is not a process "formally akin to the spread of an infectious disease"(Arrow, 1969). Instead, a technology is adopted only when a person comes in contact with multiple other adopters. This idea is called "complex contagion." While Centola and Macy (2007) demonstrate that collectives can theoretically facilitate technology adoption, they admit, "We know of no empirical studies that have directly tested the need for wide bridges [collectives] in the spread of complex contagions." In other words, the theoretical possibility lacks empirical support. In contrast, the idea that technology is adopted when information about the success of the technology arrives from a single social contact is a well-documented phenomenon (see, e.g., Foster and Rosenzweig, 1995; Munshi, 2004; Conley and Udry, 2010).

individualist's long links. It decreases the average path length between nodes that would have otherwise been far apart. All the results in this section are about symmetric networks simply because there are too many possible asymmetric networks to consider them all. However, the results in Table 3 show how each of these network features speeds diffusion, even in a network that is asymmetric.

Taken together, these results explain why ideas and germs spread more quickly in low-path-length networks, why fast diffusion might imply faster technological progress and output growth, and what evolutionary advantages each type of network might offer its adopters.

2.2 Understanding Network Evolution

Why do some societies end up with networks that inhibit growth? If disease prevalence can permanently alter social structure, then diseases that were prevalent long ago may have created networks that are no longer ideal. This rationalizes differences in social structures that persist even after diseases have been eliminated. It also motivates our use of historical disease data, in the next section. The first result shows that eventually, the economy always converges to a uniform-type network.

Result 3 Networks converge to a uniform type*: 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 finite 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. Traits are inherited from neighbors, so when a trait dies out, it never returns. The state in which all individuals have the same trait is an absorbing state. The result relies on there being a finite network and a finite type space.

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.[9]

Result 4 Disease dies out*: With probability 1,* $\exists T$ *s.t.* $d_j(t) = 0 \quad \forall j$ *and* $\forall t > T$.

These results show us that which network type will prevail is largely dependent on which dies out first: the disease-prone trait (individualism, high degree, mobility) or the disease. If disease is very prevalent, it kills off all disease-prone types. The society is left with a disease-resistant but

[9]Epidemiological models with an infinite number of agents often have a second steady state with a positive infection rate. For example, in Goenka and Liu (2015), any positive fraction of infected agents is still an infinite number of infections. Since the probability that none of the infinite infections is passed on to another node is zero, the infinite-agent model never reaches extinction. Our model predicts extinction because we have a finite number of agents.

diffusion-inhibiting network forever after. If disease is not very prevalent, its transmission rate is low, or by good luck it just dies out quickly, then the disease-prone will survive. Since they are more economically successful, the economy is more likely to converge to a disease-prone, high-diffusion network. The key consequence is that networks can persist long after the conditions to which they were adapted have changed.

3 Data, Measurement, and Calibration

After providing a qualitative analysis of the forces at work in the model, we bring the model to the data and use it to quantify the macroeconomic effects of social networks. Quantifying the model involves assembling an assortment of data, not typically used in economics. Our theory is about the relationship between pathogen prevalence, social networks, and technology diffusion. We assembled a data set that contains each of these variables for 71 countries. This section describes how each variable is measured and used to calibrate model parameters. We explore alternative parameterizations in Appendix B.

3.1 Measuring Social Networks

Our model features three dimensions of social network connections: individualism, degree, and mobility. To measure individualism, we use the Hofstede individualism index. Hofstede (2003) analyzed the systematic variation across countries in individual values and behaviors and identified four dimensions along which cultural values could be analyzed and compared across countries. One of these dimensions is the individualism-collectivism dimension, which captures the degree of interdependence among members of a society. Hofstede developed an index to measure this dimension across countries, and we adopt his index in our empirical analysis to capture the degree of clustering in a network, which is our theoretical measure of individualism.[10]

Measures of individual mobility and number of friends are not available at the country level from a unique source that allows cross-country comparison. For this reason, we use US data on immigrants from different countries to measure systematic variation in these variables across countries of origin. The underlying assumption is that people who move to the United States from countries with higher degree/mobility maintain higher degree/mobility and pass these social network preferences on to their children. This approach of using data for US residents follows other studies and helps to control for institutional differences that might otherwise explain different

[10]The other cultural dimensions identified by Hofstede are: uncertainty avoidance, power distance (strength of social hierarchy), and masculinity-femininity (task orientation versus person orientation).

behavior across countries.[11] At the same time, immigrants are a selected sample who likely have a higher propensity for mobility and have been partly assimilated into American culture. As such, our measure likely underestimates the differences in social network structure across countries.

Measuring Individualism In our model, collectivism is defined as a social pattern of closely linked or interdependent individuals. Individualism is its opposite. What distinguishes collectives from sets of people with random ties to each other is that in collectives, two friends often have a third friend in common. This is the sense in which they are interdependent. Hofstede developed his theory using factor analysis to examine the results of a worldwide survey of employee values conducted by IBM between 1967 and 1973.

He defines collectivism as

> 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. (Hofstede, 2003, pp. 9-10)

As a proxy for individualism, we use country i's Hofstede individualism index. This 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 goals. The individual answers were aggregated at the country level after matching respondents by occupation, age, and gender. The countries' mean scores for 14 questions were then analyzed using factor analysis, which resulted in the identification of two factors of equal strength that together explained 46% of the variance.[12] The individualism factor is mapped onto a scale from 1 to 100 to create an index for each country.

Measuring Mobility Most people form their strongest social ties through repeated social contact with neighbors. Long-distance ties are likely to be with one's former neighbors. Thus, mobility

[11] See, for example, Fernández and Fogli (2009) for an application of this approach to identify cross-country differences in labor force participation that are driven by cultural factors.

[12] Of course, IBM employees are not representative of country residents as a whole. Subsequent studies involving commercial airline pilots and students (23 countries), civil service managers (14 countries), and consumers (15 countries) have validated Hofstede's results. Appendix C describes these additional studies, as well as some of Hofstede's questions. The appendix also summarizes sociological theories that link these questions to network structure. Finally, it documents studies that map out partial social networks for small geographic areas. Taken together, these network mapping studies reveal that highly collectivist countries, according to Hofstede, have a higher average prevalence of network collectives.

governs the quantity of social ties in far-off locations. For US data, the census tells us the annual probability of a US resident moving across state lines. For other countries, we do not have a large cross-country panel of mobility data. But we do have extensive data on mobility for US residents, including immigrants. The data come from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS), from 1994 to 2013. We select individuals between ages 18 to 35 and use the variable BPL to assign country of origin. To measure mobility, we use the variable MIGRATE1, which indicates whether the respondent had changed residence in the past year. Movers are those who reported moving outside their state borders. Our measure of mobility is the fraction of first-generation US immigrants from each country that move in a given year.

Measuring Network Degree A network's degree is the average number of connections of a node in the network, n_f in the model. Our empirical proxy for network degree in each country is the average number of close friends reported by US residents that report having ancestors coming from that country. Our data come from the General Social Survey (GSS).[13] The variable *numgiven* asks, "From time to time, most people discuss important matters with other people. Looking back over the last six months - who are the people with whom you discussed matters important to you? Just tell me their first names or initials." Based on this variable, we select respondents that report having ancestors coming from another country and average their responses to construct an index for network degree for each country in our sample.

In the GSS data set, only 28 countries are found. For the remaining countries, we impute degree by estimating α, β_1, and β_2 in : $degree_i = \alpha + \beta_1 hofstede_i + \beta_2 mobility_i + \epsilon_i$. Then, we use the predicted values from that regression to fill in our missing observations.

Mapping Network Data to Model Variables We calibrate the model parameters to replicate features of the US economy. The United States is characterized by low disease prevalence and fast diffusion networks. Network degree (n_f) and mobility (m_j) allow the model to match the modal number of friends and the probability of moving. To keep the number of node types limited, we allow a person to be either high or low degree ($n_f = \{4, 6\}$) and either high or low mobility (probability of forming new links $m_j = \{0, 0.1\}$). We choose these high and low values because they straddle the variation in our data. Each node is linked to its immediate neighbors by definition. Because n_f is an even number, the minimum number of degree that differentiates collectivists and individualists is 4. On the other hand, the maximum number of degree in the GSS data set for the United States is 6.

[13] We use the variable *ethnic*, which asks the country of origin of ancestors, to associate individuals with different countries.

The spirit of the exercise is to simulate an economy that starts with a small fraction of high-diffusion network types. In different epidemiological environments, we calibrate the parameters to match a low-disease economy and generate time series for networks, output, and technology over 500 periods. We then change the initial disease prevalence, and compare the economic outcomes with the benchmark economy of low disease. We use the United States as our benchmark economy and set the initial fraction of individualist (long-link type) to 30%.[14] We use the same initial fraction for the high mobility type. Since in the data the annual interstate migration rate in the United States in 2000 is 3%, we set the high-mobility type to have a probability of moving equal to 10%.

The length Δ_j of the individualists' long link governs the average path length in an individualist network. Seminal research done by Travers and Milgram (1969) on average US social network path length found that a letter dropped in the middle of the United States typically found its recipient after being passed on five to six times. Thus, we set Δ_j so that, in the US-calibrated highest diffusion social network, characterized by only high-diffusion types and without any germs, the average length of the path between any two nodes is about 5.1.[15]

Mapping Data to Technology Parameters Every node/person starts with a technology level of 0. Each period, any given person may discover a new technology that raises his productivity with probability λ. The rate of arrival (λ) is calibrated so that the average time between advances in the technology frontier is 21 years.[16]. According to Comin et al. (2006), this corresponds to the average time between invention and first adoption (beginning of diffusion) of new technologies in the United States.

The magnitude of the increase in productivity from adopting a new technology (δ) is calibrated to match the US GDP growth rate of 2.6% per year. The probability of transmitting a new technology to each friend (ϕ) is chosen to match the fact that for the average household technology, the time between invention and diffusion to half of the population is 40.18 years (Greenwood et al., 2005).[17]

For the initial disease prevalence rate ($d_h(0)$), we set the low germs pervalence to 0.05% since in the United States, infectious diseases have been almost entirely eradicated. In our experiment, we set the high disease prevalence to 18%, which matches the malaria prevalence in Ghana in 2006.[18]

[14] This fraction should be small enough to leave room for endogenous diffusion of individualists. But at the same time, it needs to be large enough so that the individualists survive initial disease prevalence.

[15] We start the economy with all individualists, all high-degree, and all high-mobility. There are no germs. The formula for the average path length is: Average path length $= 1/J \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}/i} p_{ij}/(J-1)$.

[16] From the US adoption lag in Comin and Hobijn (2010)

[17] We calculate this half-diffusion from the Greenwood et al. (2005) data set by averaging the number of years from introduction until a 50% adoption rate for 13 of their 15 technologies. We exclude the vacuum and washer because their adoption rates were more than 30% in the year they first appeared in the data.

[18] See Appendix C for the data source.

Table 1: Correlations

	Degree (1)	Individualism (2)	Mobility (3)	Index loadings
Individualism	1.000			0.575
Mobility	0.364	1.000		0.540
Degree	0.545	0.471	1.000	0.615

The table reports correlations of the three measures of social network structure described in Section 3.1 and the loadings of each measure on the network index.

To calibrate the probability of disease transmission (π), a natural target is the steady state rate of infection. But, as we have shown, the only steady state infection rate is zero. Therefore, we set the transmission rate so that, on average, the disease disappears in 150 years in the high germs economy (Ghana).

Finally, our social change shock p_ξ regulates the rate of network evolution. When we calibrate p_ξ to 0.2, it means that on average, a social change shock arrives once every 5 years. Although the node can change type every 5 years, most of the time, the highest-productivity node among one's friends has the same type, so no change of type occurs. This p_ξ value is set such that each generation (25 years), 5% of the population changes type.[19]

Network Diffusion Index Our three network features – collectivism/individualism, degree, and mobility – all accelerate the diffusion of new technologies. Summarizing their effects jointly facilitates graphical representation and later avoids econometric identification problems. Therefore, we combine our network measures into a single network index. The index we construct is the first principal component of the three measures. The last column of Table 1 lists the linear weights.

$$\text{Network Index}(\tilde{N}) = 0.58 * \widetilde{Individualism} + 0.61 * \widetilde{Degree} + 0.54 * \widetilde{Mobility}, \tag{4}$$

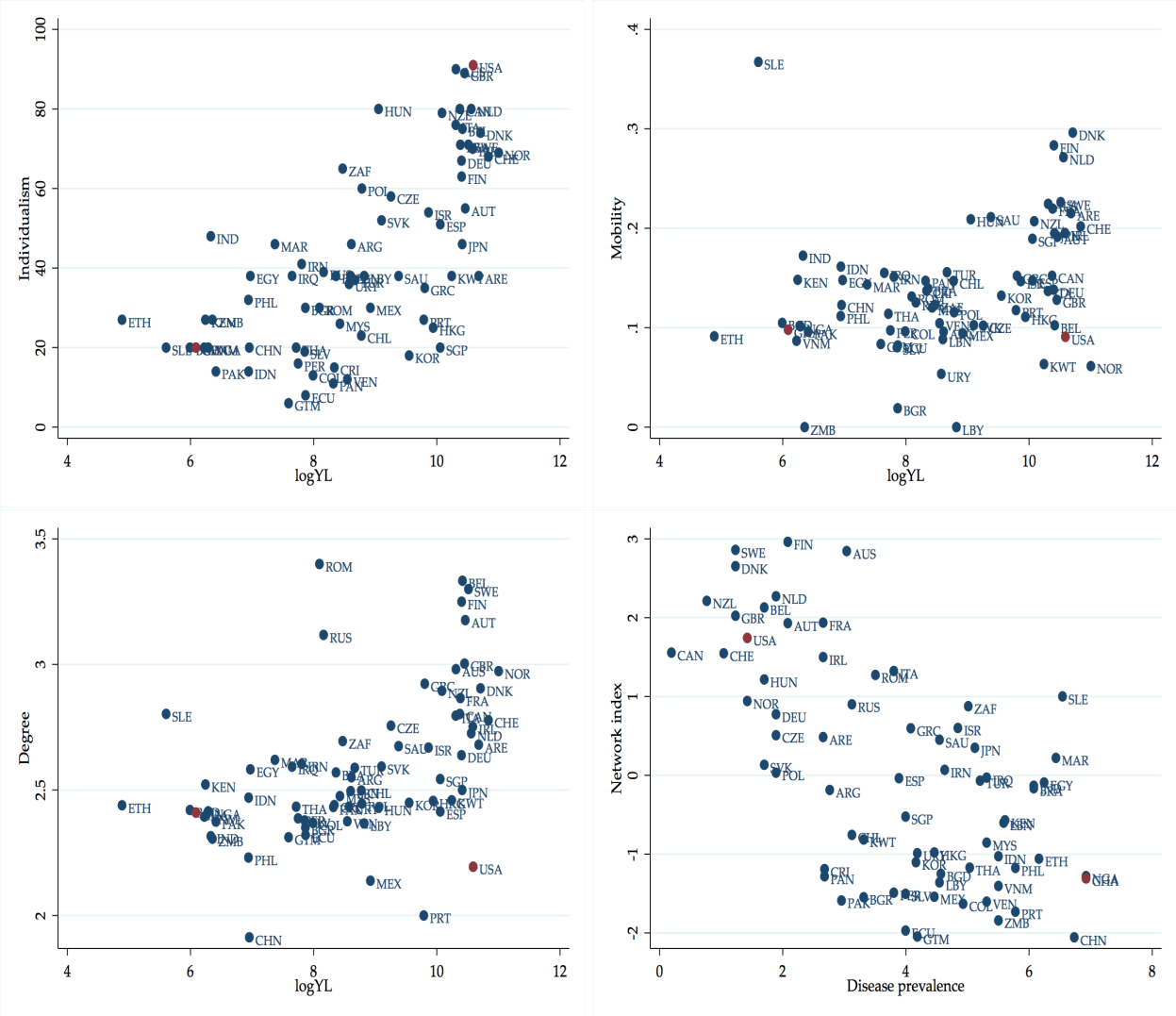
where $\tilde{x} \equiv (x - \text{mean}(x))/\text{std}(x)$. The dimension of greatest network variation in the data turns out to be roughly the same dimension that ranks networks by their diffusion speed. This index accounts for 64% of the variation in the three measures.

One might be concerned that all three measures capture the same variation. Table 1 also describes the cross-correlation of our three measures of social networks. While the measures are not uncorrelated, there is also some independent variation between them.

Comparing correlations in the model and data. Before we continue describing how we measure the effect of networks, it is useful to get some sense of what these network and disease

[19] We conduct a robustness exercise in Appendix B.1.

Figure 2: **Social network features, income and disease (data).** The first three panels are each of our three network measures (degree, collectivism/individualism, and mobility), ordered by log of income per capita. The last panel illustrates the relationship between disease (pathogen) prevalence and the individualism (Hofstede) index.



data look like. Figure 2 illustrates the network features of each country in our data set. The last panel reports the disease prevalence measure, as well as the combination of all the network features into a single network index.

3.2 Measuring Technology Diffusion

We use technology measures derived from the cross-country historical adoption of the technology data set developed by Comin et al. (2006). The data cover the diffusion of about 115 technologies

in over 150 countries during the last 200 years. 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 diffuses, given that it is adopted). If the technology was introduced to the country late, a country can be behind in a technology even though it is adopting it quickly.

To calibrate our model, we use the extensive measure, called the adoption lag. Comin et al. (2006) define country x's adoption lag to be the number of years in between invention and the date when the first adopter in country x adopts the technology. We average over the various technologies to arrive at the country's average adoption lag.

Disease-technology feedbacks. Section 1.1 introduced two feedback effects: an endogenous innovation rate that falls as the infection rate rises and an endogenous disease infection probability that falls as technology improves. Each of these feedbacks introduces one new parameter to calibrate. For endogenous innovation, we choose the parameter $\kappa_T = 5.56$ (from (1)) to match the average difference in adoption lags between the United States and Ghana, which is 32 years. The new parameter that governs endogenous infection rates $\kappa_G = 400$ (from (2)) targets a complete extinction of germs in all US simulations. Notice that the innovation probability evolves as a function of disease prevalence, and when the germs are completely eradicated, it is equal to the US innovation probability (λ_0). Given the disease prevalence in Ghana, this implies an initial value of innovation probability of 0.03%, which evolves over time as the germs disappear.On the other hand, the germs transmission probability (π) starts with 5% and decreases faster in the low-disease economy, the United States, as the output level increases more rapidly compared with Ghana. At the end of the simulation period, the germs transmission probability approaches close to 0.

These parameters are summarized in Table 2.

4 Main Results: How Do Networks Affect Income?

Our objective is to better understand how social networks affect technology diffusion and economic development. The difficulty is that economic development also changes social network structures, both directly and through disease prevalence. Using the quantified model, we can separate these two effects by exogenously changing the network structure and observing the predicted effect on technology diffusion and output. To do this, we proceed in three stages. First, we look at the model-implied correlations between networks and income/productivity measures. This tells us how much variation in cross-country income our model, with all its mechanisms, can generate. Second, we use the model to isolate the network effect. Holding all other features fixed, we compare outcomes of an economy in a high-diffusion social network with those from a low-diffusion social network. The main finding is that, while high-diffusion social networks are a strong statistical predictor of output,

Table 2: Calibrated Model Parameters

Description	Parameter	Value	Target (Data)	Target (Model)
Degree, low	$n_f(L)$	4	General Social Survey (GSS)	
Degree, high	$n_f(H)$	6		
Mobility, low	$m_j(L)$	0		
Mobility, high	$m_j(H)$	0.1	Interstate migration rates	
Disease transmission probability	π_0	5%	Disease disappears in 150 years	142
Innovation productivity increase	δ	80	2.6% growth rate in low germ country	3.0%
Technology transfer probability	ϕ	12%	Half-diffusion in 40.18 years (Greenwood et al. 2005) in low germ	42.35
Number of nodes to furthest friend of an individualist	Δ	7	Average path length of 5 (Travers and Milgram 1969) for low germ, high degree, low mobility, and individualistic society	5.1
Technology arrival rate	λ_0	0.08%	US technology adoption lag of 21 years (Comin and Hobijn, 2010), low germ	21.5
Exogenous death rate	p_ξ	20%	5% population type change every 25 years, low germ	8.7%
Tech feedback	κ_T	5.56	difference of adoption lag 32 years	34
Germs feedback	κ_G	400	Disease disappears in all low-germs simulations (0 survival of germs)	0
Initial conditions for endogenous variables:				
Fraction of individualist	f_i	30%		
Fraction of high degree	f_d	20%		
Fraction of high mobility	f_m	70%		

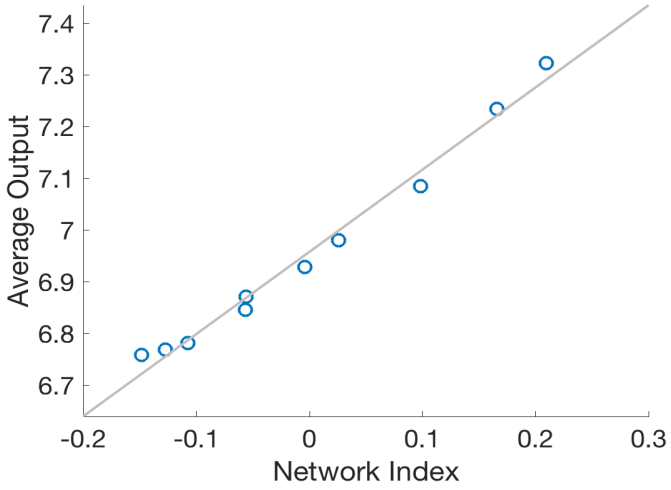
this does not imply that changing a country's social structure will boost incomes. If a country's social network is low-diffusion, replacing it with a higher-diffusion pattern of social connection can impoverish the country. Third, we explore why the effect of network changes depends on disease prevalence. If a country has a low-diffusion network, that network evolved in order to protect the country from disease epidemics. Replacing the low-diffusion network with a high-diffusion one can be disastrous for the economy and the residents because the new network is maladapted to their environment.

4.1 The Covariance between Networks and Income

How much of the variation in income can this mechanism explain? Here, we show that the forces of the model can jointly explain large differences in income across countries. Of course, just as in the data, the model's networks are endogenous. This result does not imply that changing networks alone can produce this variation in cross-country income. The causal effect comes in the next subsection.

To explore the relationship between networks and income, the model needs some variation in networks. But networks are endogenous outcomes of different disease environments (and random chance). Therefore, we construct a set of economies with different initial disease prevalence, simulate them, and measure the network and incomes at the end of the simulation. Each initial disease level is meant to represent a different country that evolved to have a different social network. Specifically, we consider 10 equally spaced levels of initial disease prevalence between 0.05% and 18%. These end points correspond to the highest and lowest levels of prevalence of the diseases in our data. For each initial level of disease, we run 200 simulations, where each simulation is computed on a 400-person network, over 500 periods. At the end of the 500 periods, we use the fraction of individualists, high-degree, and high-mobility nodes to construct the network index (defined in 4). Figure 3 shows that a 0.1-standard-deviation change in the network index (0.15 points) correlates with a 0.25 log point (25%) higher end-of-simulation income.

Figure 3: **Networks and Income: Covariance.** Both graphs report the end-of-simulation income (left) against the end-of-simulation network index. The points on the left of the x-axis come from a high-disease economy that starts with a maximum disease prevalence of 18% and evolves to have low-diffusion networks (low network index). The points on the right come from low-disease economies (minimum is 0.05% prevalence), where high-diffusion networks emerged (high network index).



In large part, networks and income are correlated simply because both are correlated with

disease. At the same time, this correlation is a useful starting point because the real world also has joint and reverse causality. It makes clear that, later, the difference in the causal results arises not simply because the model differs from the real world. The causal results and correlations differ greatly even inside the model. The model is valuable, in part, because it explains why the true causal effects might deviate from the correlation that the data alone would suggest.

4.2 How Exogenous Network Changes Affect Technology and Income

To measure the causal effect of social networks on technology and output in the model, we change the network and observe the result. Surprisingly, we find that the effect of high-diffusion networks can be positive, negative, or zero. In other words, while the correlation between networks and income is strongly positive, the causal effect may not be. The result depends on the environment the network is in. The reason is that high-diffusion networks can expose disease-prone countries to income-reducing disease epidemics.

We begin by looking at social network effects in a low-disease environment. We start with our network calibrated to the United States and then vary each network feature one at a time. For each value of individualism, degree, and mobility, we double their values, one at a time. Specifically, we simulate networks with nodes that have either $n_f = \{4, 6\}$. High-degree networks have a higher fraction of degree 6 nodes. For mobility, we consider annual moving probabilities $p = \{0, 0.1\}$. High-mobility networks have more of the nodes that have a 1/10th chance of moving each year. Finally, nodes are either individualist or collectivist. Individualist networks have a higher fraction of individualist nodes. When we simulate a new economy with a different degree or mobility parameter, we change the probability that every node in the network has a particular characteristic, in a symmetric way.

The top of Table 3 describes the effect of higher-diffusion networks in a low-disease economy. This is the result one would expect: network features that facilitate diffusion speed the diffusion of new technologies and boost income. The greatest gains come from mobility and degree, where a doubling of more mobile or connected agents raises growth by 0.23-0.24%. Cumulated over 200 years, this network change raises incomes by 56-58%. This is a more modest effect than the correlations suggest, but it is directionally consistent.

The bottom of Table 3 paints a different picture. In most of these experiments, altering the social network to facilitate faster diffusion lowers income. The effect is most drastic for degree. For high-disease economies, doubling degree causes output to fall by 90%. Likewise, individualism impairs growth for these two economies. The reason is that both perpetuate disease, which compromises growth. Only mobility facilitates economic development for these two economies. Of course, mobile agents can also introduce disease to new communities. But ideas are more prevalent than disease.

Table 3: **Main Result: Effect of social network features on output per capita.**

		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	3.35 %	3.44 %	3.59 %	3.58 %
(0.05% initial prevalence)	log(income)	6.70	6.88	7.18	7.16
	tech diffusion	1/25.2	1/24.1	1/22.5	1/23.2
	disease extinction	59.4	58.5	96.3	51.2
High-disease economy	growth	1.33 %	1.29 %	0.12 %	1.46 %
(0.18% initial prevalence)	log(income)	2.66	2.58	0.24	2.92
	tech diffusion	1/73.4	1/75.8	1/177.8	1/69.1
	disease extinction	196.3	196.9	201.0	196.0

Entries report average annual growth, ending level of output per capita, $ln(A_{t+1}/A_t)$, the number of periods it takes for half-diffusion of a new technology (inverted), and the average number of periods until disease extinction. All are averaged across 200 simulations, each of which runs 200 periods. Social networks are held constant. The network is either the US social network or a network that is a US benchmark where one of the following is doubled: the proportion of nodes that are individualist (from 44.5% to 89%), the proportion of high-degree $n_f = 6$ nodes (from 60% to 100%), or the proportion of high-mobility $m_j = 0.1$ nodes (from 31.2% to 62.5%). All other model features and parameters are constant and reported in Table 2.

Mobile agents spread ideas more than germs. Because of the feedback from technology to the infection rate, the faster technology diffusion from mobile agents speeds up disease extinction and promotes economic growth.

The bottom line is that the way in which networks affect economic growth depends on the disease environment. High-diffusion networks in high-disease economies spread more disease. Because that is the dominant effect, such networks – especially high-degree networks – can impoverish high-disease countries. These same networks also transmit more disease in low-disease countries. We can see that effect in a longer time to extinction. The difference is that when new ideas are prevalent and disease is rare, the net effect of spreading ideas and disease is positive. Thus, the same networks that impoverish poor countries can facilitate growth in rich ones where epidemics are rare. To thrive, each country needs a social network that is well adapted to its environment.

These causal effects can be substantially different from what the correlations suggest. Of course, depending on when we measure, we could get small or arbitrarily large differences in income levels. However, reporting income after 200 periods facilitates comparison with the previous results. For high-disease economies, the effect of high-diffusion networks is the opposite of what the correlations would predict. For low-disease economies, the results look more similar. After 200 periods, the log incomes in the low-disease economy range from 6.7 to about 7.2. This translates into income levels (not in logs) ranging from 812 to 1339, a roughly 50% difference in income per capita. This is similar to the effects predicted by the correlation results in Figure 3.

Parameter sensitivity. The qualitative nature of the underlying data on social structure makes this model a tricky one to calibrate, which makes exploring alternative parameter values particularly important. We explored the robustness of our results to changes in model parameters. In particular, we explore doubling and halving the values of the probability of social change p_ξ, the technology transmission probability ϕ, the probability of disease diffusion π, and initial disease prevalence $d(0)$. The effect of networks on technology and output is remarkably stable. It varies by less than 20% across all of these different specifications. Table 11 in Appendix B reports results for each exercise individually.

Instead of doubling, we could have instead halved the prevalence of each of our network features. Table 12 in Appendix B also shows that most of these results are the same as well, in the sense that, if doubling was detrimental to growth, halving is beneficial.

The Role of Disease-Technology Feedbacks. The disease-technology feedbacks help the model to generate more nuanced results. In particular, when disease transmission rates fall as technology improves, that force helps to prevent the extinction of high-diffusion types, who are more susceptible to disease.

Turning off the feedbacks primarily affects high-disease economies. Turning off feedback 1, so that the innovation rate does not depend on disease, makes high-disease economies grow faster. It also makes high-diffusion networks more economically beneficial: If innovation does not depend on disease, then when a high-diffusion network spreads disease, this has a less negative effect on output. When we turn off feedback 2, so that the infection rate does not depend on technology, growth in high-disease economies falls because these economies cannot innovate their way out of epidemics. This is particularly costly for economies with high-diffusion networks because the benefits of their technology diffusion are smaller, while the disease-promotive costs remain.

Details of these results and the calibration are in Table 13 of Appendix B. The message from this exercise is that the feedbacks strengthen our main conclusions but are not the sole drivers of the interaction between networks, technology, and disease.

4.3 How Social Networks Adapt to Their Environment

Our main question is about the effect, not the origin, of networks. But to understand where and why changing the structure of social networks might be harmful, it is useful to understand how a country's social network emerges.

In the model, what creates variation in networks is the initial level of disease and random chance. Therefore, we simulate two economies that vary in their initial infection rates. In Figure 4, the results labeled "low disease" have an initial disease prevalence of 0.05%. The results labeled "high disease" have an initial disease prevalence of 18%. All other parameters are the same across

the two sets of simulations. While disease rates evolve endogenously through the model's contagion process, the initial differences create persistent differences in social network structure and thus in output.

Figure 4: **Initial Disease, Network Evolution, Extinction, and Output.** These figures illustrate the persistent effects of initial disease on social networks and output. The top two graphs report the fraction of the model's economies of individualists, high-degree agents, and high-mobility agents at each date t. The top left panel is a low-disease economy that starts with a disease prevalence of 0.05%. The top right panel is a high-disease economy that starts with 18% prevalence. The bottom two graphs plot the rate of disease extinction and output for both the low-disease (low germs) and high-disease (high germs) economies.

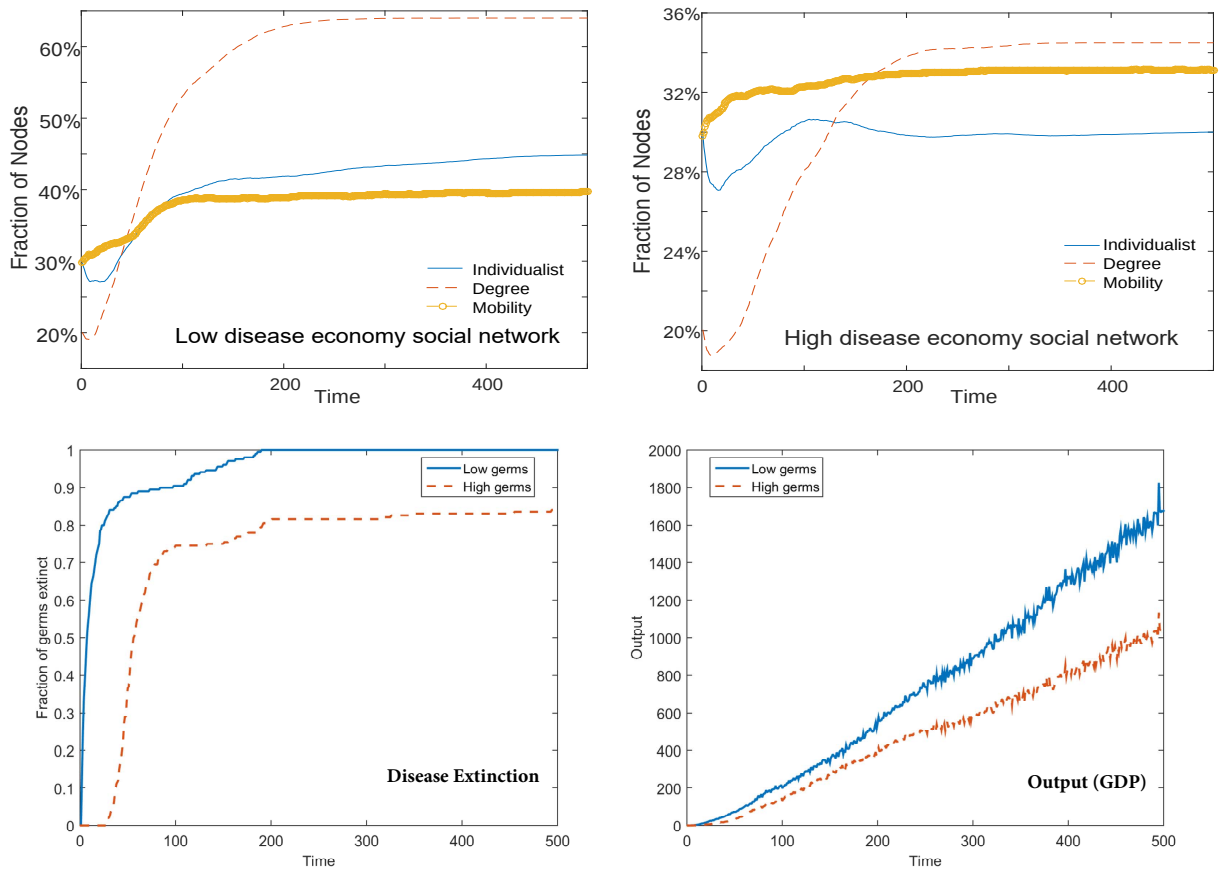


Figure 4 illustrates how social networks adapt to the disease environment. A low-disease environment has an increasing number of individualist nodes with high degree and high mobility. Such nodes thrive because they get new ideas faster than low-diffusion types. New ideas boost their output. When other nodes change type, they adopt the type of the most economically successful nodes, which are the high-diffusion nodes. Thus, in low-disease environments, high-diffusion network characteristics thrive.

But in high-disease environments (top right panel), high-diffusion nodes get sick quickly. They may also get new ideas. But if a node is sick, it is unproductive, regardless of its technology.

One has to be alive and well to be productive. Therefore, when nodes change type, the most economically successful types are the low-diffusion types because those agents are less likely to be sick. In the high-disease environment, most nodes retain low-diffusion characteristics.

Since networks adapt to the disease environment, changing the network without changing the disease environment can be disastrous. A high-diffusion network, in a place where disease is prevalent, is a recipe for epidemics and humanitarian crisis. At the same time, we also see that even after disease is eradicated, the network persists. Table 3 reports that on average, diseases are eradicated in 50-200 years. Yet, we see social network features persisting because they are passed down, generation by generation. So, networks are adapted to their envirnoment. But, when the environment changes, especially if the change is not life-threatening, social networks may be slow to respond.

5 An Alternative Approach: Estimation with Instruments

So far, we learned that although the correlation between social networks and income is strong, the causal evidence is more mixed: some economies benefit from high-diffusion networks, and others do not. Before ending, we bring one additional perspective on the question of how social networks matter for aggregate outcomes. This last approach uses instrumental variables to try to identify a causal relationship in a nonstructural way. Both approaches are imperfect. In an environment with equilibrium effects everywhere, no instrument will be exogenous to everything. Every model is also surely wrong in some ways. We briefly pursue this second approach to measuring the importance of networks in order to compare the two answers. We find that the model-based approach and the identification-based approach deliver broadly consistent answers to the following question: For the median country that has mostly eradicated communicable disease, how much do social networks matter for the macroeconomy?

Before we continue describing how we identify the effect of networks, it is useful to get some sense of the covariances. By comparing the model estimation with the data estimation of the same relationship, we can also get a sense for how quantitatively plausible the model is. Using the initial disease levels and endogenous network and output outcomes, we run the same linear regression in the model as in the cross-country data, to compare coefficients. We estimate the following linear models:

$$\text{output} = \beta_1 + \beta_2 \text{Network Index} + \epsilon, \tag{5}$$

$$\text{output} = \beta_3 + \beta_4 \text{sumgerm} + \tilde{\epsilon}, \tag{6}$$

where output per capita is $A(t)$, Network Index is defined in (4), *sumgerm* is total disease prevalence, the β's are unknown coefficients, and ϵ, $\tilde{\epsilon}$ are mean-zero residuals.

Table 4: **Comparing Endogenous Network Outcomes to Data.** Entries are OLS estimated coefficients of equations (5) or (6). Output in the data is real output per capita $\log(Y/L)$. In the model, output is the average node's technology $A(t)$. Disease prevalence in the data is the disease prevalence index, defined in Section 3.1. In the model, disease prevalence is the initial fraction of sick agents, which lies on a grid of 10 equally spaced points between 0.05% and 18%. Standard errors are in parentheses.

Dependent variable: GDP per capita				
	Data	Model	Data	Model
network index	0.713***	1.591***		
	(0.103)	(0.067)		
germs sum			-0.640***	-3.129***
			(0.0758)	(0.352)
Observations	71	10	71	10
R^2	0.409	0.986	0.508	0.908

The strong positive correlation between network diffusion and income comports with what we see in the data. This finding is reassuring, since the model was not calibrated to match this correlation or any closely related moments. In calibration parlance, these are "overidentifying moments." The similar orders of magnitude in Table 3 suggest that the model is a reasonable framework for quantifying network effects on income.

How large is this effect economically? Suppose every person in a given country had one more friend. How much additional income would these estimates suggest? Since the standard deviation of degree is 0.30, the adjusted $\widetilde{degree}$ measure changes by 3.3. According to Table 4, this increases the *Network Index* by $0.56 \times 3.3 = 1.85$. Multiplying this by the model's estimated coefficient on *Network Index* delivers an increase of average output per capita of $1.85 \times 1.59 = 2.94$. How large is that increase? Since the United States has average income ($\log Y/L$) of 10.5, this represents a 28% increase in US GDP per capita.

Of course, that was a three-standard deviation change in a network feature, a rather large change. Suppose that we change the fraction of individualists by 0.24, an amount that corresponds to one standard deviation of individualism in the data. This increases the *Network Index* by 0.61. When multiplied by the coefficient on *Network Index* estimated from model output, the predicted effect on average output is $0.61 \times 1.59 = 0.97$. That represents a more modest 9.2% of US GDP. A similar calculation implies that a 6% change in the probability of moving (also one standard deviation in the data) would predict a change in GDP equal to 8.5% of US income. The bottom line is that the estimation coefficients of the model and data reveal that, statistically, networks can

predict lots of variation in cross-country income.

Disease Difference Instrument. Our theory suggests that an instrument with power to predict social network structure $\tilde{N}$ is total disease prevalence. But this instrument is not likely to be valid, both because technology affects disease (vaccines are a technology, for example) and because poor health reduces productivity and diminishes one's capacity for invention.

However, according to sociobiologists, the effect of disease on social networks depends on how the disease is transmitted.[20] Sociobiologists often classify infectious diseases by reservoir. The reservoir is any person, animal, plant, soil, or substance in which an infectious agent normally lives and multiplies. From the reservoir, the disease is transmitted to humans. Pathogens that use humans as their reservoir (perhaps in addition to other reservoirs) have the potential to affect social networks because they are passed on to more connected individuals more frequently. Zoonotic pathogens are those not carried by people, only by other animals. Their prevalence is less likely to affect social networks in any systematic way.

Therefore, we construct two disease prevalence variables: the prevalence of all human-reservoir diseases in a country, d_h, and the prevalence of all zoonotic diseases in the country, d_z. To distinguish between the direct effect of disease on output and the effect of disease on networks, we construct the raw difference in disease prevalence as $\tilde{\Delta}\text{germ} \equiv d_h - d_z$. Our results are cleaner if this difference is uncorrelated with the sum of disease prevalence. Constructing the standardized difference ensures this independence.[21] Thus,

$$\Delta\text{germ} \equiv \frac{1}{std(\tilde{\Delta}\text{germ})}\left(d_h - d_z - \text{mean}(\tilde{\Delta}\text{germ})\right). \tag{7}$$

Thus, our identifying assumption is

$$E[\epsilon(d_h - d_z)] = 0. \tag{8}$$

We restricted the coefficients on d_h and d_z to be the same, meaning that human disease prevalence and zoonotic disease prevalence have the same effect on technology. Hence, the total effect on technology is determined by the sum $d_h + d_z$. This is orthogonal to the composition of the effect between the two types of disease, $d_h - d_z$, which has no direct effect on A. Therefore, since the diseases have different effects on networks $\tilde{N}$ and similar effects on the speed of technology diffusion

[20] See, for example, Smith et al. (2007) or Thornhill et al. (2010). Also, Birchenall (2014) and Murray and Schaller (2010) argue that human-to-human transmitted diseases have a disproportionate effect on the pattern of social relationships.

[21] When $var(d_h) = var(d_z)$, the difference $(d_h - d_z)$ is orthogonal to the sum $(d_h + d_z)$.

A, the instrument $(d_h - d_z)$ can be a powerful and valid instrument. [22]

To measure the prevalence of both human and animal-transmitted diseases, we use historical data. In recent times, disease prevalence dropped drastically with the medical advances of the 20th century. Yet, our model predicts that disease patterns generations ago might still affect social networks today. The oldest comprehensive cross-country disease data come from the 1930s and includes the following nine life-threatening diseases: leishmanias, leprosy, trypanosomes, malaria, schistosomes, filariae, dengue, typhus, and tuberculosis. Most of the prevalence data come from Murray and Schaller (2010),[23] who use a four-point coding scheme: 0 = completely absent or never reported, 1 = rarely reported, 2 = sporadically or moderately reported, and 3 = present at severe levels or epidemic levels at least once. We have prevalence of all nine diseases in 160 countries. For each country, we add up the prevalence score from each disease to get that country's disease prevalence index. Appendix C contains more information about the diseases and their characteristics.

What Do Linear Estimates Teach Us? The full-sample coefficients in Table 5 tell us that a one-standard-deviation increase in the network index (1.45 units) increases log output per worker by $1.45 * 0.684 = 0.99$, which represents a 99% increase in real GDP per capita. Like the correlations from both the data and the model, that is a large effect. These large estimates suggest that social network structures might be relevant for macroeconomists and that policy makers might want to craft policies to alter these networks to promote growth.

However, Table 6 suggests a more subtle message that echoes the results of the model. The positive effect of social networks only appears for the low-disease prevalence countries. For countries with high disease prevalence, the estimates of the effect of higher diffusion networks on income are negative and insignificant. When we include other controls, such as English and total disease prevalence, the pattern of significant effects only for low-disease countries is robust. These results echo the finding of the model that changing social networks can be highly beneficial in some countries. However, installing a high-diffusion social network in countries where such social structures are maladapted, at best shows no benefit and at worst could be destructive.

[22] We do not need to know all the determinants of social structure. Rather, any subset of the determining variables can serve as a valid instrument for $\tilde{N}$. Similarly, we do not need to observe $\tilde{N}$ exactly. A proxy variable with random measurement noise is sufficient for an unbiased instrumental variables estimate of the coefficient β_2. To ensure that this approach is also model consistent, Appendix B.3 verifies that in the model, this disease difference is a valid instrument. Appendix C.7 discusses some additional endogeneity concerns and the ways in which we address them.

[23] The Murray and Schaller (2010) data are based primarily on epidemiological maps provided in Rodenwaldt and Jusatz (1961) and Simmons et al. (1945), and originally collected by the Medical Intelligence Division of the United States Army. The one exception is tuberculosis, which comes from the National Geographic Society's (2005) *Atlas of the World*. For each region, they coded the prevalence of tuberculosis on a three-point scale: $1 = 3 - 49$, $2 = 50 - 99$, $3 = 100$ or more per 100,000 people.

Table 5: **Social networks and GDP per capita (data)**

Columns report the β_2 coefficient from an IV estimation of $\log(Y/L) = \beta_1 + \beta_2\tilde{N} + \epsilon$. $\tilde{N}$ is estimated from a first-stage regression $Network\ Index = \beta_3 + \beta_4\Delta germs + \epsilon_2$. The Network Index is defined in equation (4). The instrument $\Delta germ$ is defined in equation (7). English, which is 1 if the country is English-speaking, allows for a Sargan test for validity of the $\Delta germ$ instrument. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Dependent variable:	First stage: Network index		Second stage: GDP per capita	
Instruments:	Δgerm	Δgerm & English	Δgerm	Δgerm & English
Δgerm	-0.422***	-0.448***		
	(0.123)	(0.118)		
germs_sum	-0.503***	-0.424***	-0.297*	-0.361***
	(0.0706)	(0.0736)	(0.171)	(0.138)
english		1.446***		
		(0.537)		
network index			0.684**	0.556**
			(0.307)	(0.236)
Constant	1.973***	1.561***	9.849***	10.10***
	(0.299)	(0.325)	(0.672)	(0.549)
Observations	71	71	71	71
F-stat	31.12	25.07	38.17	41.09
R^2	0.478	0.529	0.517	0.551
Sargan-p				0.520

Table 6: **Social networks and GDP per capita: Split sample**

Columns report the β_2 coefficient from an IV estimation of $\log(Y/L) = \beta_1 + \beta_2\tilde{N} + \epsilon$. Variables and first-stage regression are described in Table 5. High (low) disease are countries with a disease score above 14 (below 12).

Dependent variable: GDP per capita	High disease	Low disease
Network Index	-0.371	0.882***
	(1.682)	(0.327)
Constant	7.455***	7.001**
	(1.438)	(2.881)
Observations	34	26

Is disease difference a valid instrument? One reason the difference in the prevalence of human and zoonotic diseases might be invalid would be if human diseases were much more virulent, so that the difference predicted severe disability and thus reduced output. Epidemiologists have concluded that there is no evidence of a link between the mode of transmission and the severity of the disease.[24] Furthermore, if the problem is that the instrument makes the predicted network variable correlated with disease virulence, which makes the network a stronger predictor of GDP, then controlling for a measure of disease-free productive working years should drive out the network variable. Table 18 shows it does not. Similarly, if only human disease were virulent, then subtracting or adding zoonotic disease would make little difference. Yet, when we control for total disease prevalence, this does not drive out, nor is it driven out by, the disease difference.

Robustness to Inclusion of Other Variables. A natural question is whether social networks are simply a proxy for some other economic variable. To assess this question, we choose a variety of other variables thought to explain technology adoption or income and control for their effects too. However, we continue to use $\Delta germ$ as an instrument and our network index as an explanatory variable. We add the following variables, one by one, to the first- and second-stage estimations.[25] To control for a direct effect of disease on technology diffusion, we control for disease-adjusted *life expectancy* at birth. Because urban and rural social networks may be quite different, and we might worry that perhaps network features are proxying for urbanization, we control for *population density*. To distinguish our results from the preference-based theory of technology diffusion (Gorodnichenko and Roland, 2017), we control for *blood distance*, an instrument they use to capture genetic difference in a population. To distinguish social networks or social institutions from the effects of political institutions, we control for *executive constraint*, *inflation*, and *trade openness*.

Tables 17 and 18 in Appendix C report the complete set of results for each of these estimations. The network index survives the inclusion of every one of these variables, and the size of the estimated effect is remarkably stable. The effect of degree remains statistically significant at the 5% level every time, and the difference in disease prevalence is consistently a significant predictor of the network index. In sum, there is a statistical relationship between disease, social network structure,

[24]The Dieckmann et al. (2005) epidemiology textbook shows a bar chart with five categories of diseases (p. 19). In two categories, the directly transmitted (human-to-human) diseases have a higher rate of death per infection, and in the other three categories, the vector-borne (zoonotic) diseases are slightly more lethal. In discussing this mixed evidence, they write, "It is possible, for example, that vectorborne or waterborne transmission is associated with a greater within-host genetic variation than directly transmitted pathogens, and this genetic variation could favor increased virulence. There is no empirical evidence for this association" (p. 18).

[25]Our procedure here follows Hall and Jones (1999). If we add all the controls in together, almost nothing is significant, neither our variables nor almost any others. This reflects the fact that there is only so much variation in the dependent variable. When many imperfectly correlated series are all included as explanatory variables, any one of them explains little. In doing this, we also recognize that these control variables may themselves be endogenous. Inferring causality from these results would therefore be problematic.

and technology diffusion that is above and beyond that which comes from other commonly used cross-country determinants of income.

6 Conclusions

Measuring the effect of social network structure on the economic development of countries is a challenging task. Networks are 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 adopt low-diffusion social networks. Such networks inhibit disease transmission, but they also inhibit idea transmission. This model reveals which social features should speed or slow diffusion. It also suggests that disease prevalence might be a useful instrument for a social network because it affects how social networks evolve.

Quantifying the model reveals that small initial differences in the epidemiological environment can give rise to large differences in network structure that persist. Over time, these persistent network differences can generate substantial divergence in technology diffusion and output. We find evidence of this social network effect in the data. Exploiting the differential mode of transmission of germs, we are able to identify a significant effect of social network structure on technology diffusion and income. Specifically, we find that a one-standard-deviation change in social network structure can increase the growth of output per worker by 1/2% per year.

More broadly, the paper's contribution is to offer a theory of the origins of social institutions, propose one way in which these institutions might interact with the macroeconomy, and show how to quantify and test this relationship.

References

Acemoglu, D. and S. Johnson (2005). Unbundling institutions. *Journal of Political Economy 113*(5), 949–995.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–1401.

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics 117*(4), 1231–1294.

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

Arrow, K. J. (1969). Classificatory notes on the production and transmission of technological knowledge. *American Economic Review Papers and Proceedings 59*(2), 29–35.

Ashraf, Q. and O. Galor (2012). The 'out of africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review 103*(1), 1–46.

Birchenall, J. (2014). Disease and diversity in long-term economic development. Working Paper, University of California at Santa Barbara.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory 97(2)*, 298–319.

Bloom, D. E., D. Canning, and J. Sevilla (2004). The effect of health on economic growth: A production function approach. *World development 32*(1), 1–13.

Brock, W. A. and S. N. Durlauf (2006). Social interactions and macroeconomics. In D. Colander (Ed.), *Post-Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, pp. 97–115. New York: Cambridge University Press.

Centola, D. and M. Macy (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology 113(3)*, 702–734.

Chaney, T. (2014, November). The network structure of international trade. *American Economic Review 104*(11), 3600–3634.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology 94*(Supplement), S95–S120.

Comin, D., M. Dmitriev, and E. Rossi-Hansberg (2013). The spatial diffusion of technology. Working paper, Princeton University.

Comin, D. and B. Hobijn (2010, December). An exploration of technology diffusion. *American Economic Review 100*(5), 2031–59.

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. Working paper, Harvard Business School.

Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in ghana. *American Economic Review 100(1)*, 35–69.

de la Croix, D., M. Doepke, and J. Mokyr (2017). Clans, guilds, and markets: Apprenticeship institutions and growth in the pre-industrial economy. Northwestern University Working Paper.

Dieckmann, U., J. A. Metz, and M. W. Sabelis (2005). *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*, Volume 2. Cambridge University Press.

Doepke, M. and M. Tertilt (2009). Women's liberation: What's in it for men? *Quarterly Journal of Economics 124(4)*, 1541–1591.

Doepke, M. and F. Zilibotti (2014). Culture, entrepreneurship, and growth. *in Handbook of Economic Growth 2 ed. by P. Aghion and S.N. Durlauf, pp. 1-48. Amsterdam: Elsevier.*

Fernández, R. and A. Fogli (2009, January). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics 1*(1), 146–77.

Fischer, C. S. and Y. Shavit (1995). National differences in network density: Israel and the united states. *Social Networks 17(2)*, 129–145.

Foster, A. D. and M. R. Rosenzweig (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy 103(6)*, 1176–1209.

Gangestad, S. and D. Buss (1993). Pathogen prevalence and human mate preferences. *Ethology and Sociobiology 14(2)*, 89–96.

Ghana Health Service (2008). *2008 GHS Annual Report*.

Ghana Statistical Service (2008). *Ghana in Figures*.

Goenka, A. and L. Liu (2015). Infectious diseases, human capital and economic growth. Working paper, University of Rochester.

Gorodnichenko, Y. and G. Roland (2017). Culture, institutions, and the wealth of nations. *Review of Economics and Statistics 99*(3), 402–416.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–1380.

Granovetter, M. S. (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives 19(1)*, 33–50.

Greenwood, J. and N. Guner (2010). Social change. *International Economic Review 51(4)*, 893–923.

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*(5), 912–950.

Grinstead, C. M. and J. L. Snell (1997). *Introduction to Probability* (2nd ed.). New York: Russell Sage.

Gudykunst, W. B., G. Gao, K. L. Schmidt, T. Nishida, M. H. Bond, K. Leung, G. Wang, and R. A. Barraclough (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. E. and C. I. Jones (1999). Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics 114*(1), 83–116.

Hamilton, W. D. and M. Zuk (1982). Heritable true fitness and bright birds: A role for parasites? *Science 218*(4570), 384–387.

Hofstede, G. (2003). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations* (2nd ed.). Thousand Oaks, CA: Sage publications.

Jackson, M. (2008). *Social and Economic Networks* (1st ed.). Princeton, NJ: Princeton University Press.

Kelley, B., H. Lustig, and S. Van Nieuwerburgh (2013). Firm volatility in granular networks. Working paper, New York University.

Loehle, C. (1995). Social barriers to pathogen transmission in wild animal populations. *Ecology 76(2)*, 326–335.

Lucas, R. and B. Moll (2014). Knowledge Growth and the Allocation of Time. *Journal of Political Economy 122*(1), 1–51.

Mali, S., S. Steele, L. Slutsker, and P. M. Arguin (2008). Malaria surveillance - united states, 2006. *Surveillance Summaries*, 24–39.

Munshi, K. (2004). Social learning in a heterogeneous population: Technology diffusion in the indian green revolution. *Journal of Development Economics 73*(1), 185–213.

Murray, D. R. and M. Schaller (2010). Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture. *Journal of Cross-Cultural Psychology 41(1)*, 99–108.

National Geographic Society (2005). *Atlas of the World* (8th ed.). Washington, DC: National Geographic Society.

Oberfield, E. (2013). Business networks, production chains, and productivity: A theory of input-output architecture. Working paper, Federal Reserve Bank of Chicago.

Perla, J. and C. Tonetti (2011). Endogenous risk and growth. Working paper, New York University.

Rauch, J. E. and A. Casella (2001). *Networks and Markets* (1st ed.). New York: Russell Sage Foundation.

Rodenwaldt, E. and H. Jusatz (1961). *World Atlas of Epidemic Diseases, 1952-1961*. Hamburg: Falk Verlag.

Rogers, E. M. (1976). New product adoption and diffusion. *Journal of Consumer Research 2(4)*, 290–301.

Ryan, B. and N. Gross (1943). The diffusion of hybrid seed corn in two iowa communities. *Rural Sociology 8*, 15–24.

Simmons, J. S., T. F. Whayne, G. W. Anderson, and H. M. Horack (1945). *Global Epidemiology: A Geography of Disease and Sanitation* (1st ed.). Philadelphia: Lippincott.

Smith, K. F., D. F. Sax, S. D. Gaines, V. Guernier, and J.-F. Guegan (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.

Tertilt, M. (2005). Polygyny, fertility, and savings. *Journal of Political Economy 113(6)*, 1341–1371.

Thornhill, R., C. L. Fincher, D. R. 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.

Travers, J. and S. Milgram (1969). An experimental study of the small world problem. *Sociometry 32(4)*, 425–443.

U.S. Census Bureau (2006). Current population survey, annual social and economic supplement.

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of "small-world" networks. *Nature 393*(6684), 440–442.

Weil, D. N. (2007). Accounting for the effect of health on economic growth. *The Quarterly Journal of Economics 122*(3), 1265–1306.

Young, H. P. (2009). Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review 99(5)*, 1899–1924.

Online Appendix

All of this material is for a separate, online appendix. It is not intended to be printed with the paper.

Contents

A Appendix: Proofs

Proof of Result 1 *Step 1: Mapping model primitives into average lifetimes and discovery times*: $L_j(t)$ represent the last day of a healthy life for a person located at node j at time t. It is the first period in which the person living in location j at t gets sick, minus one: $L_j(t) = min\{s : s \geq t, d_j(s) = 1\}$. Thus, $L_j(0)$ is the number of periods that the person living in location j at time 0 will be healthy and productive. Then the average agent's healthy lifetime is $E_j[L_j(0)]$. Analogously, let $\alpha_j(0)$ be the number of periods it takes for a new idea, introduced in period 0, to reach person j. Then the average discovery time is $\bar{\alpha} = E_j[\alpha_j(0)]$.

Step 2: Compute average healthy lifetime. Suppose that at time 0, there is one sick person k and everyone else in the network is healthy. Then, $d_k(0) = 1$ for some k and $d_j(0) = 0 \, \forall \, j \neq k$. 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, that is, the lifetime is $L_j(0) = s_{jk}$. Averaging over all locations j, we have that the average lifetime is equal to the average path length from k to all other nodes: $\bar{L} = E_j[L_j(0)] = E_j[s_{jk}]$.

What if the probability of transmission is less than one? Note that if the probability of disease transmission is less than one, then there is a positive probability that the disease dies out before it is spread to anyone. Since there is no other source of death, this implies that lifetime is infinite. With a positive probability of infinite lifetime, $E_j[L_j(0)] = \infty, \forall \, \pi < 1$.

Step 3: Compute average discovery time. 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, $E_j[\alpha_j(0)] = E_j[s_{jk}]$.

Proof of Result 2 We start by deriving the path length in each network and then compare.

Average path length of pure collectivist network. Consider the distance from the last node, J, of a symmetric, fully collectivist network with degree n_f. The node J can be connected to nodes 1 though $n_f/2$ and $J-1$ through $J - n_f/2$ in one step. More generally, it can be connected to nodes $(s-1)n_f/2 + 1$ through $sn_f/2$ and $n - (s-1)n_f/2 - 1$ through $n - sn_f/2$, in s steps. For each s, there are n_f nodes for which the shortest path length to J is s steps. We know from result 1 that when n_f is even and n/n_f is an integer, the longest path length (the diameter) is n/n_f. Thus, the average length of the path from J to any other node is $1/n \sum_{s=1}^{J/n_f} n_f s$. By symmetry, this is the same average distance from any node to others. Using the summation formula, this is $(n_f/J)(J/n_f)(J/n_f + 1)/2 = 1/2 + n/(2n_f)$.

Average path length of zero-collectives (individualist) network. Consider the path length from node J to any other node in the network between 1 and $J/2$. By symmetry, the path length starting from any other node is the same, and the average path length to the nodes between $J/2$ and J is the same as for the nodes in the first half. Consider taking steps of length Δ_j until one reaches or passes the node $J/2 - \Delta_j/2$. The number of steps in this path is $\tilde{s} \equiv \text{round}(J/(2\Delta_j))$, where round is the nearest integer value. All points not on this path (interior nodes) can be reached in, at most, $\tilde{s} + \Delta_j/2$ steps.

This implies a sequence of path lengths of the following form:

$$\{1, \ldots, \frac{\Delta_j}{2}\}$$

$$1 + 2\{1, \ldots, \frac{\Delta_j}{2}\}$$

$$\vdots$$

$$\tilde{s} + 2\{1, \ldots, \frac{\Delta_j}{2}\}$$

This is an upper bound on the total path lengths of the network because $\tilde{s}$ may be greater than $J/2 - \Delta_j/2$. The average path length is the sum of all path lengths, divided by the number of nodes. In this case, that is

$$PL \leq \frac{1}{J/2} \left[\sum_{i=1}^{\tilde{s}} i + (2\tilde{s} + 1) \sum_{i=1}^{\Delta_j/2} i \right].$$

We can use the summation formula to replace the sums:

$$PL \leq \frac{2}{J} \left[\frac{\tilde{s}(\tilde{s}+1)}{2} + (2\tilde{s}+1) \frac{\Delta_j/2(\Delta_j/2+1)}{2} \right].$$

Note also that any number is rounded down by, at most, 1/2. Therefore, an upper bound on $round(x)$ is $x+1/2$. Similarly, we know that $\tilde{s} \leq J/(2\Delta_j) + 1/2$. Since the path length expression is increasing in $\tilde{s}$,

$$PL \leq \frac{1}{4J}\left[\frac{(J+\Delta_j)(n+3\Delta_j)}{\Delta_j^2} + \frac{J+2\Delta_j}{\Delta_j}\Delta_j(\Delta_j+2)\right].$$

Comparing path lengths. A sufficient condition for the individualist path length to be smaller is

$$\frac{1}{4J}\left[\frac{(J+\Delta_j)(J+3\Delta_j)}{\Delta_j^2} + \frac{J+2\Delta_j}{\Delta_j}\Delta_j(\Delta_j+2)\right] < \frac{J}{8}.$$

Rearranging, this implies that

$$\left(\frac{1}{2} - \frac{1}{\Delta_j^2}\right)n^2 - \left(\frac{4}{\Delta_j} + \Delta_j + 2\right)J - 2\Delta_j(\Delta_j+2) - 3 > 0.$$

Since we assumed that $\Delta_j > 2$, the coefficient on the J^2 term is positive. Therefore, there is a sufficiently large J such that the inequality holds.

Higher Degree Speeds Diffusion Take a network and its matrix of shortest path lengths $\{p_{ij}\}_{i,j=1}^{J}$. For one node i, decrease its degree n_f by 2, by breaking the two farther links, the links to nodes $j \pm n_f/2$. Then the shortest path length between nodes i and $j \pm n_f/2$ increases by one. Furthermore, breaking these two links can only increase the shortest path length(s) for any other node $j \neq i$. Therefore, for the new matrix of shortest path lengths $\{\tilde{p}_{ij}\}_{i,j=1}^{N}$, $p_{ij} \leq \tilde{p}_{ij}$ for all i, j. The average path length (3) increases.

Mobility Speeds Diffusion Increasing mobility means replacing short links with random, long links to the network with a higher probability. A network with long links is an individualist network with a large and random Δ. Thus, following the logic of the proof above, as the probability of long links increases, the probability of a shorter path length increases and the expected average path length declines.

Proof of Result 3 (Network Becomes Homogeneous) 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.

Lemma 1 *In a 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).*

Suppose agent j is the only one whose type is different from the rest of the network. One scenario in which the number of j-types increases, for certain, in the next period is if (i) agent j stays healthy, (ii) some node j' directly connected to agent j gets hit by the social change shock, and (iii) all the nodes connected to j' get sick. Of course, this is not the only way to increase the prevalence of j's type. But it works because in this senario node j is the only non-sick node, and thus the only productive node that some other node is connected to when it is type changes. Since j is the most productive, its type is adopted by another node.

Now we compute a lower bound for the probability of (i)-(iii) happening at any time. First, even if all n_{fj} nodes connected to j are sick, there is a probability $(1-\pi_t)^{n_{fj}}$ that none of these nodes transmits disease. Second, for a given node j', the probability of social change is p_ξ. Third, if the disease has not died out, there is a positive probability that any node is sick. Thus, there is a positive probability that j' is sick. If so, there is a probability $\pi_t^{n_{fj'}}$ that all the contacts of j' get sick in the next period. Since each of these events is independent and each has positive probability, the join event of all three happening has positive probability. Thus, with at least one node of a given type, there is a positive probability that more nodes of that type will be present in the following period.

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

Proof of Result 4 (Disease Dies Out) 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 Model Simulation and Numerical Results

Algorithm to Construct Network At a given date, with a fixed set of node types, this is how one constructs the social network.

1. Construct a ring network where all nodes are connected to their immediate neighbors. Let N be an $n \times n$ matrix where n is the number of nodes. $N_{ij} = 1$ means that i and j are socially connected. Then, the binary matrix that represents the ring network is a matrix of zeros with 1's on the diagonals just above and below the main diagonal: $N_{ij} = 1$ if $j = mod(i \pm 1, n)$, and $= 0$ otherwise, where we define $mod(n, n) = mod(0, n) = n$.
2. Add individualists and collectivists. If node i is a collectivist, set $N_{ij} = N_{ji} = 1$ for $j = mod(i+2, n)$. If node i is an individualist, set $N_{ij} = N_{ji} = 1$ for $j = mod(i + \Delta_j, n)$.
3. Adjust the degree of the network. If node i has degree $n_f > 4$, then add more local links. Set $N_{ij} = N_{ji} = 1$ for $j = \{mod(i+3, n), \ldots, mod(i + n_f/2, n)$. Note that n_f must be even.
4. Adjust links for mobility. For every node j, draw a random, binary variable that is 1 with probability m_j. If the realization is 1, choose one non-adjacent link to break; that is, choose an i from the set $\{i \neq \{j-1, j+1\} : N_{ij} = N_{ji} = 1\}$, each with equal probability. Then break that link; set $N_{ij} = N_{ji} = 0$. Then, choose any nodes k, unlinked to j, with uniform probability, and form a new link between those nodes: Set $N_{kj} = N_{jk} = 1$.

Laws of Motion for State Variables Each node j in date t has a health state $d_j(t)$, a productivity $A_j(t)$, and a type $\tau_j(t)$. The shocks to these types come from 1) disease transmission, which is $N\xi(t)$, where $\xi(t)$ is an i.i.d., $n \times 1$ vector of binary variables that are one with probability π; 2) technology transmission $N\xi_\tau(t)$, where $\xi_\tau(t)$ is an i.i.d $n \times 1$ vector of binary variables that are 1 with probability ϕ; 3) innovation $\iota_j(t)$, which increases technology by δ units with probability λ; and 4) social change shocks $\xi_{rj}(t)$, a vector of binary variables that are 1 with probability p_ξ. Then the laws of motion for disease, technology, and type are

$$d_j(t+1) = min(1, d_j(t) + N\xi(t))(1 - \xi_{rj}(t))) \tag{9}$$

$$A_j(t+1) = (1 - d_j(t))\left(\delta\iota_j(t) + max(A_j(t), NA(t)\xi_\tau(t))\right) \tag{10}$$

$$\tau_j(t+1) = (1 - \xi_{rj}(t)))\tau_j(t) + \xi_{rj}(t)\hat{\tau_j}(t), \tag{11}$$

where $\hat{\tau_j}(t) = \tau_k : \ k = argmax_{k':N_{jk'}=1} A_{k'}(t)$. Of all of j's friends (nodes s.t. $N_{jk'} = 1$), k is the one with the highest productivity. Therefore, j adopts k's social network type, τ_k.

For the exogenous network model, the first two laws of motion are the same and the types τ are fixed: $\tau_j(t+1) = \tau_j(t)$.

Complete Simulation Results for Calibrated Model The results in Tables 7, 8, 9, and 10 are computed from the full model, with the endogenous network, and two-way feedbacks between technology and disease. We ran 200 simulations, for 500 periods each and reported simulation averages. Results market low disease differ from those marked high disease only in the initial number of infected nodes. For low disease simulations, 0.05% of nodes start out infected. In high disease simulations, 18% of nodes start the simulation infected.

Table 7: Fractions of Nodes of Each Type: Average and Ending. Average results are the fraction of nodes of that type, averaged over the 500 periods in all 200 simulations. Ending results are the fraction in period 500, averaged over the 200 simulations. Two-way technology-disease feedbacks active, number of people = 400

	Individualists average	High Degree average	Low Mobility average	Infected average	Individualists ending	High Degree ending	Low Mobility ending	Infected ending
low disease	0.4103	0.5725	0.3825	0.0013	0.4484	0.6400	0.3975	0.0000
high disease	0.2975	0.3146	0.3273	0.0843	0.3000	0.3450	0.3313	0.0668

Table 8: Disease extinction, Output and Technology Diffusion. germs.surv is the number of simulations where at least one agent is still infected in period 500. extin.yrs is the average number of periods before disease extinction (max is 500). Growth is the average growth rate of technology over the periods and simulations, in percentage. Lag is the average adoption lag of technology over periods and simulations, in number of periods. Half.diff is the number of periods that one idea takes to reach half the population. Type.chn is the number of changes of node type per 25 years. Avpath is the average path length in the network, averaged over 200 simulations and 500 periods. Two-way technology-disease feedbacks active, number of people = 400

	germs.surv	extin.yrs	growth	log(Y/L)	lag	half.diff	type.chn	avpath
low disease	0	24.7600	3.0276	7.3224	21.4775	42.3500	0.0867	5.1200
high disease	32	142.0650	1.7392	6.7681	56.3792	97.4250	0.0498	5.1200

Table 9: Network Convergence Probabilities. For each of the three node types, entries list the fraction of simulations where at period 500, either all of the nodes were of that type or none of the nodes had that type. 200 simulations.Two-way technology-disease feedbacks active, number of people = 400.

	Individualism		High Degree		Low Mobility	
	all	none	all	none	all	none
low disease	0.3207	0.4868	0.4891	0.3427	0.2790	0.5037
high disease	0.2641	0.6504	0.2763	0.6258	0.2925	0.6183

Table 10: Technology-Disease Feedback Parameters: $\pi(t)$ is the per-period probability of disease transmission, conditional on being connected to a sick node, at time t (defined in eq. 2). $\lambda(t)$ is the probability of a given agent innovating in period t (eq. 1) 200 simulations, 500 periods. feedback for trans. = 1, feedback for tech. = 1, number of people = 400

	$\pi(150)$	$\pi(500)$	$\lambda(2)$	$\lambda(150)$	$\lambda(500)$
low disease	0.0188	0.0009	0.0008	0.0008	0.0008
high disease	0.0318	0.0090	0.0003	0.0006	0.0007

B.1 Robustness to Lower and Higher Parameter Values

Next, we explore the effect of changing each of the model's parameters on our main results. For consistency, we varied each parameter by exploring values half as large and twice as large. Table 11 reports the effect of a change in each of the three network characteristics in output per capita, for these alternative parameter values.

Table 11: Robustness

100 additional nodes		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	3.43%	3.58%	3.81%	3.67%
(0.05% initial prevalence)	log(income)	6.53	6.57	6.64	6.6
	tech diffusion	1/24.7	1/23.8	1/21.5	1/22.5
	average germs	0.0037	0.0040	0.0210	0.0039
	extinction years	47.5	47.0	93.4	43.9
High-disease economy	growth	1.47%	1.38%	0.15%	1.59%
(0.18% initial prevalence)	log(income)	5.68	5.62	3.43	5.76
	tech diffusion	1/71.5	1/73.2	1/187.9	1/64.1
	average germs	0.2209	0.2355	0.4719	0.2212
	extinction years	197.1	198.3	201.0	196.5

double technology arrival rate		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	4.53%	4.67%	5.08%	4.77%
(0.05% initial prevalence)	log(income)	6.81	6.84	6.92	6.86
	tech diffusion	1/19.6	1/19.1	1/17.1	1/18.8
	average germs	0.0037	0.0044	0.0161	0.0041
	extinction years	39.3	43.4	72.4	43.5
High-disease economy	growth	2.4%	2.48%	0.25%	2.7%
(0.18% initial prevalence)	log(income)	6.17	6.20	3.92	6.29
	tech diffusion	1/54.5	1/53.3	1/160.6	1/47.5
	average germs	0.1818	0.1906	0.4611	0.1739
	extinction years	182.2	181.6	201.0	176.4

When we change parameters, the results change very little because these results are statements about what happens in the long run. The effect of these parameter changes mostly shows up in the speed of the transition to the long-run levels. What determines the long-run levels and growth rates is the network. But this exercise takes the network as exogenous because we are trying to assess the robustness of the exogenous network results at the end of the paper. If we instead changed parameters in the endogenous network, the effect would be much larger. But that is not a statement about the fragility of our predictions. It simply reflects the importance of network structure for long-run outcomes.

Lastly, the main results double the prevalence of each of the high-diffusion network features. It raises the question, what if we halved them instead? The model's results are not linear. For most features where doubling was harmful, such as degree in the high-disease economy, halving was beneficial. But along other dimensions, such as individualism in low-disease economies, where doubling was beneficial, halving had a negligible effect.

B.2 The Role of Disease-Technology Feedbacks

Table 13 reveals how the two disease-technology feedbacks affect our results.

double technology transfer probability		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	4.62%	4.79%	5.02%	5.00%
(0.05% initial prevalence)	log(income)	6.83	6.86	6.91	6.91
	tech diffusion	1/17.9	1/17.3	1/16.0	1/16.3
	average germs	0.0039	0.0037	0.0139	0.0031
	extinction years	43.1	40.9	68.4	36.2
High-disease economy	growth	2.32%	2.24%	0.15%	2.64%
(0.18% initial prevalence)	log(income)	6.14	6.11	3.37	6.27
	tech diffusion	1/53.4	1/55.5	1/165.1	1/51.6
	average germs	0.1900	0.2027	0.4712	0.1795
	extinction years	180.1	182.5	201.0	173.2

double length of individual link		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	3.59%	3.88%	3.67%	3.79
(0.05% initial prevalence)	log(income)	6.58	6.65	6.6	6.63
	tech diffusion	1/23.2	1/21.4	1/20.7	1/21.5
	average germs	0.0070	0.0071	0.0436	0.0065
	extinction years	55.3	49.8	97.1	49.6
High-disease economy	growth	1.33%	1.19%	0.12%	1.48
(0.18% initial prevalence)	log(income)	5.58	5.47	3.2	5.69
	tech diffusion	1/70.6	1/73.3	1/182.1	1/68.1
	average germs	0.2437	0.2658	0.4766	0.2396
	extinction years	196.8	198.4	201.0	195.4

double type change rate		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	2.18%	2.26%	2.53%	2.31%
(0.05% initial prevalence)	log(income)	6.08	6.11	6.23	6.14
	tech diffusion	1/34.0	1/32.9	1/29.4	1/31.6
	average germs	0.0001	0.0001	0.0002	0.0001
	extinction years	5.44	5.64	7.77	6.02
High-disease economy	growth	2.15%	2.22%	2.47%	2.27%
(0.18% initial prevalence)	log(income)	6.06	6.09	6.20	6.12
	tech diffusion	1/38.3	1/36.6	1/34.3	1/35.0
	average germs	0.0054	0.0053	0.0078	0.0053
	extinction years	25.2	24.6	33.7	25.3

We observe the following changes in our results. First, the high-disease economy benefits when there is no technology feedback. Its growth rate and income are higher because the high rate of disease does not slow innovation. In such a setting, increasing the network diffusion helps both economies to achieve higher growth and income because the downside of diffusion is disease and disease has a diminished impact on growth.

When there is no effect of technology on disease transmission, the high-disease economy has a significantly lower growth rate and income than the low-disease economy. This is because the high-disease economy cannot innovate its way out of its disease epidemics and stays poor longer. In this setting, higher diffusion networks increase growth

Table 12: Halving diffusion exercise

		benchmark	halve individualists	halve high-degree	halve high-mobility
Low-disease economy	growth	3.35 %	3.27 %	2.79 %	3.15 %
	tech diffusion	1/25.2	1/25.8	1/30.1	1/26.9
High-disease economy	growth	1.33 %	1.36 %	2.43 %	1.13 %
	tech diffusion	1/73.4	1/75.6	1/46.3	1/80.9

Entries report growth of output per capita, $ln(A_{t+1}/A_t)$, or the inverse of the number of periods it takes for half-diffusion of a new technology. Both are averaged across 200 simulations and 200 periods. Social network is held constant thoughout the 200 periods. The network is either the US social network or a network that is a US benchmark where one of the following is halved: the proportion of nodes that are individualist (from 44% to 22%), the proportion of high-degree $n_f = 6$ nodes (from 64% to 32%), or the proportion of high-mobility $m_j = 0.1$ nodes (from 40% to 20%). All other model features and parameters are constant and reported in Table 2.

Table 13: Main results without feedback

feedback 1 off		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	3.37%	3.49%	3.86%	3.6%
(0.05% initial prevalence)	log(income)	6.51	6.55	6.65	6.58
	tech diffusion	1/25.0	1/24.0	1/21.8	1/22.9
	average germs	0.0057	0.005	0.0201	0.0047
	extinction years	50.5	49.3	81.3	43.9
High-disease economy	growth	3.07%	3.21%	3.25%	3.3%
(0.18% initial prevalence)	log(income)	6.42	6.47	6.48	6.49
	tech diffusion	1/39.5	1/38.3	1/71.5	1/35.7
	average germs	0.125	0.1277	0.257	0.1188
	extinction years	146.9	146.0	175.9	142.3

.

feedback 2 off		benchmark	double individualists	double high-degree	double high-mobility
Low-disease economy	growth	2.64%	2.75%	1.90%	2.76%
(0.05% initial prevalence)	log(income)	6.27	6.31	5.94	6.31
	tech diffusion	1/25.5	1/24.4	1/22.9	1/23.5
	average germs	0.0583	0.0616	0.2003	0.069
	extinction years	97.1	89.4	127.2	87.4
High-disease economy	growth	0.50%	0.48%	0.10%	0.51%
(0.18% initial prevalence)	log(income)	4.60	4.56	3.00	4.63
	tech diffusion	1/78.9	1/80.3	1/198.2	1/74.6
	average germs	0.3251	0.3344	0.4862	0.3305
	extinction years	201.0	201.0	201.0	201.0

and income only for the low-disease economy. For high-disease economies, the downside of more disease diffusion is more costly than the benefit of technology diffusion. Eliminating this feedback reduces the benefits of technology

innovation.

B.3 Is the Difference between Diseases a Valid Instrument in the Model?

Since the paper is about connecting model and data, it is useful to show that the model is consistent with our use of the difference in disease prevalence as an instrument for social networks. Of course, this does not prove that the instrument is valid.

The difference in the prevalence of socially transmittable and other diseases is a valid instrument if equation (8) holds. To show that this condition holds in our model, we hold the network fixed and vary the initial prevalence of both types of disease. The set of simulation nodes used for initial prevalence of both types of disease is $[.004 : .004 : .02]$. The resulting average prevalence of transmissible disease ranges is the range $[0, 0.106]$. But since socially transmittable disease spreads and typically becomes more prevalent over time, but the other diseases do not spread, comparing the rates of initial prevalence is not a valid comparison. Therefore, our statistical analysis considers the relationship between average prevalence of the disease in the first 100 periods and productivity growth. The other parameters used in the simulation are those in Table 2.

Both the transmissible and zoonotic diseases reduce productivity significantly. A 10% increase in prevalence reduces average GDP growth by 0.2-0.6%. This is not surprising since by assumption, a sick agent has zero productivity. What is important here is that the difference between transmissible and zoonotic disease prevalence is not a significant predictor of productivity growth. This coefficient is not significant at the 5% or even 10% significance levels, even though we generated 10,000 independent simulations of the model under different starting conditions on which to run these regressions. What we learn from this finding is that, if the network connections are held fixed, there should be no significant direct effect of the difference in diseases on productivity growth. Both diseases affect productivity in the same way: by making people sick and thus unproductive. Since the two diseases have similar effects, the difference in prevalence has no effect. Thus, the model motivates our choice of disease difference as an instrument to capture network effects without affecting technology directly. Of course, there may be forces outside the model that invalidate our instrument. We address those forces in the next section.

Table 14: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
germs diff	0.067	0.995	-1.952	2.678	71
germs sum	3.868	1.74	0.191	6.926	71
mobility	0.139	0.064	0	0.367	71
hofstede	41.915	23.36	6	91	71
degree	2.59	0.299	1.913	3.4	71
network index	0	1.401	-2.056	2.963	71
GDP per capita	8.701	1.563	4.887	11.004	71
tech_diffusion	-0.025	0.633	-2.39	0.999	71
english	0.076	0.239	0	0.974	71

C Data Appendix

The 71 countries in our sample are Argentina, Australia, Austria, Bangladesh, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Czech Republic, Denmark, Ecuador, Arab Republic of Egypt, El Salvador, Ethiopia, Finland, France, Germany, Ghana, Greece, Guatemala, Hong Kong (China), Hungary, India, Indonesia, Islamic Rep. of Iran, Iraq, Ireland, Israel, Italy, Japan, Kenya, Rep. of Korea, Kuwait, Lebanon, Libya, Malaysia, Mexico, Morocco, Netherlands, New Zealand, Nigeria, Norway, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Romania, Russian Federation, Saudi Arabia, Sierra Leone, Singapore, Slovak Republic, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, United Arab Emirates, United Kingdom, United States, Uruguay, Venezuela, Vietnam, and Zambia.

Summary statistics for each of the variables we use are described in Table 14.

- Initial disease prevalence
  - High germs: Ghana 2006 malaria prevalence 18%
    - 2006 Ghana population
      (Ghana Statistical Service, 2008): 21.9 mil
    - 2006 Malaria cases
      (Ghana Health Service, 2008): 3.87 million
    - Prevalence as a percentage of population: 3.87/21.9 = 17.7%
  - Low germs: US 2006 malaria prevalence 0.0005%
    - 2006 US population
      (U.S. Census Bureau, 2006): 293.834 mil
    - 2006 malaria cases
      (Mali et al., 2008): 1564
    - Prevalence as a percentage of population: 1.564/293,834 = 0.0005%
  - Note that in every simulation, the code makes sure that there is at least one sick person.
- Note on feedback: Where π_0 is the initial transmission probability, κ_G is the parameter for the germs feedback elasticity and $\Phi(\cdot)$ is the Normal distribution cdf. Note that $\Phi(0) = 1/2$
- Note: When germs = 0, then $\lambda = \lambda_0$. The same applies to the technology feedback, where $\pi = \pi_0$ when avg tech = 0. Notice that the innovation probability λ has λ_0 as the "ceiling," which is reached for the low germs. On the other hand, germs transmission probability, π, reaches 0 when the technology goes up.

C.1 Disease Data

Disease Reservoirs Animals often serve as reservoirs for diseases. There are also nonliving reservoirs, such as soil, which is a reservoir for fungi and tetanus.

Human-specific d_{hs} Many diseases have only human reservoirs, even though they historically may have arisen in other species, such as measles, which originated in cattle. Such diseases may be spread with the help of an animal (called a vector), such as a mosquito that injects one person's blood into another person. But it is in the human, not in the mosquito, where the disease flourishes. Human-specific diseases in our data set include diptheria, filaria, measles, and smallpox. The variable d_{hs} is defined as $d_{hs} \equiv \sum_{l \epsilon \mathcal{HS}} \text{prevalence}_l$, where l is a disease and $\mathcal{HS}$ is the set of all human-specific diseases.

Zoonotic p_ξ Other diseases, although they infect and kill humans, develop, mature, and reproduce entirely in nonhuman hosts. These are zoonotic diseases. Humans are a dead-end host for infectious agents in this group. Our zoonotic diseases include anthrax, rabies, schistosomiasis, tetanus, and typhus. The variable d_z is defined as $d_z \equiv \sum_{l \epsilon \mathcal{Z}} \text{prevalence}_l$, where l is a disease and $\mathcal{Z}$ is the set of all zoonotic diseases.

Multihost d_m Some infectious agents can use both human and nonhuman hosts to complete their life cycle. We call these multihost pathogens. Our multihost diseases include leishmaniasis, leprosy, trypanosomes, malaria, dengue, and tuberculosis. The variable d_m is defined as $d_m \equiv \sum_{l \epsilon \mathcal{M}} \text{prevalence}_l$, where l is a disease and $\mathcal{M}$ is the set of all multihost diseases.

Since multihost 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. Their prevalence is less likely to affect the benefits of any particular social network. Therefore, for the purposes of our analysis, we will group human-specific and multihost diseases together. We define the variable $\bar{d} \equiv d_{hs} + d_m$. It is the sum of 22 human and multihost diseases, whereas p_ξ is the sum of 12 diseases.

Contemporaneous Disease Data We redid our analysis with more recent data on a broader array of diseases from the Global Infectious Disease and Epidemiology Network in 2011-12 and report primarily 2011 prevalence rates. The sources for data included in GIDEON currently include health ministry publications (electronic and print) and peer review journal publications. A partial listing is available at `http://www.gideononline.com/resources.htm`. The quality and frequency of data input vary by source. A total of 34 specific pathogenic diseases are coded, each on a 1-3 prevalence scale. GIDEON classifies some diseases on a six-point scale, according to the per capita reported infection rate. The cutoff rates for each level vary by disease; for example, a 4 for rabies means an infection rate between .01 and .02 per 100,000 people, whereas the same range delimits a 3 for tetanus. We convert from the 1-6 scale to a 1-3 scale as follows: a 1 remains a 1, a 2 or a 3 is coded as a 2, and any number above 3 is coded as a 3. The total pathogen prevalence variable is the sum of the values for each disease within each country.

Our two pathogen prevalence indices appear to be accurate because they are highly correlated (0.77). They are also highly correlated with a similar index created by Gangestad and Buss (1993) to assess pathogen prevalence within a smaller sample of 29 regions. Correlations are 0.89 with our index from 1930s data and 0.83 with our index from 2011 data. This high correlation explains why the results with contemporaneous data are nearly identical.

C.2 Measuring Individualism

Hofstede (2003) 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 countries), and consumers (15 countries) have validated Hofstede's results.

Hofstede's 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 (5). How important is it 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.

Theories linking Hofstede questions to network structure These questions reflect two views of a collectivist society: one in which ties are strong and one in which ties are shared. In a widely cited paper, Granovetter (1973) provides the bridge between shared ties and strong ones; he argues that "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 that 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 is 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. Many of the questions about work goals provide evidence of these preference weights. One example of such a question is, "How important is it to you to work with people who cooperate well with each other?" Coleman (1988) explains why cooperative behavior is 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 are easier to sustain than if punishments must be implemented in an uncoordinated way.

A third category of questions in Hofstede's survey is 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.

Cross-Country Network Analysis A small literature analyzes and compares social network structures across countries. It is summarized and extended by Fischer and Shavit (1995) and reported in Table 15. Surveys typically ask respondents to name people with whom they confided, were friends, asked for help, and so on. 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 typically do not 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 15: 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 with 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 members live elsewhere and visit infrequently.

Group Identity. In collectivist societies, people learn to think about themselves as part of a 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 nonmembers. In the individualistic society, people learn to think about themselves as an individual, not as a member of a group. There is no distinction between group members and nonmembers. Gudykunst et al. (1992) surveyed 200 students in each of four countries: Australia and the United States (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 nonmember. 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 nonmembers correlated exactly (negatively) with their country's IDV scores.

C.3 Measuring Network Degree

The variable degree uses the combination of two survey questions from the General Social Survey (GSS). The variable *numfriend* asks the respondent: "How many good friends do you have?" while the variable *numgiven* asks, "From time to time, most people discuss important matters with other people. Looking back over the last six months - who are the people with whom you discussed matters important to you? Just tell me their first names or initials." The resulting variable lists the number of people mentioned in response to this question.

Table 16: Social networks and technology diffusion

	First stage: Network index		Second stage: Technology diffusion	
VARIABLES	germs	germs & english	germs	germs & english
germs_diff	-0.422***	-0.448***		
	(0.123)	(0.118)		
germs_sum	-0.503***	-0.424***	-0.0570	-0.0852
	(0.0706)	(0.0736)	(0.0884)	(0.0708)
english		1.446***		
		(0.537)		
network index			0.325**	0.269**
			(0.159)	(0.121)
Constant	1.973***	1.561***	0.196	0.305
	(0.299)	(0.325)	(0.348)	(0.281)
Observations	71	71	71	71
F-stat	31.12	25.07	17.77	19.72
R^2	0.478	0.529	0.210	0.283
Sargan-p				0.382

Columns report β_2 coefficient from an IV estimation of $A = \beta_1 + \beta_2 \tilde{N} + \epsilon$. The technology diffusion rate (A) comes from the Comin and Mestieri (2012) measure of the intensive technology adoption in a country. $\tilde{N}$ is estimated from a first-stage regression $Network\ Index = \beta_3 + \beta_4 \Delta germs + \epsilon_2$. The Network Index is defined in (4). The instrument $\Delta germs$ is defined in equation (7). First stage estimates are in the appendix. The Sargan test statistic uses both Δgerm and *English* as instruments. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

All the respondents are US residents. To assign respondents to different countries, we use the variable *ethnic*, which asks, "From what countries or part of the world did your ancestors come?" Ethnicities are supposed to be listed in order of importance. Thus, in cases in which multiple ethnicities are reported, we use only the first one. Sometimes respondents report regions rather than countries as an ethnicity. We map regions into countries as described in the following section. We use the response of the US residents who declare a country as their ethnic origin to proxy for the number of social connections for an average resident of that country.

C.4 Measuring Mobility

Using the hypothesis that people break social network ties when they move from one community to another, we construct the following proxy for social mobility. The data on mobility come from the GSS. The variable we use is *MIGRATE1*, which indicates whether the respondent had changed residence in the past year. Those who were living in the same house as one year ago were considered nonmovers. Movers were asked about the city, county, and state and/or the US territory or foreign country where they resided one year ago. We considered "stayers" those that did not move or moved inside county borders.

C.5 Complete IV Estimation Results (First and Second Stage)

The first stage results in Table 16 are interesting because they are consistent with one reason why countries may have adopted different social institutions. Our endogenous network theory predicts that since nodes with low individualism, low degree, and low mobility are less likely to be sick, their types are more likely to be adopted in high human disease environments. This correlation is borne out in the data. We make no claim to identify any causal link here. But the data are consistent with a world where social networks have evolved, in part, as a defense against the spread of directly communicable diseases.

C.6 Other Control Variables and Complete Results for Output per Worker

An inevitable question arises: "What if you also control for X?" We would like to know if our social network measure 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. Alternatively, we can 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 in Tables 17 and 18 reports the coefficients of a first- or second-stage regression of income per capita on the network index, 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, total disease prevalence (sumgerm), and our standard instrument, the difference in pathogens variable, $\Delta germ$.

Tables 17 and 18 reveal that the difference in human versus zoonotic disease prevalence is consistently a strong predictor of the network index and the network index has a significant and quantitatively similar effect on output or technology, despite controlling for a variety of other development-related variables. The dependent variable in the first stage is the Network Index, as defined in Table 4. The dependent variable in the second stage of Table 18 is real output per capita in 1999 from the World Bank.

The control variables are life expectancy at birth (LEB), disability-adjusted life expectancy (daly), which is the expected length of time an individual lives free of disability, as measured by the World Health Organization in 2004, social infrastructure (SocInf) and executive power (Executive), measures of the efficient functioning of social and political institutions, constructed by Hall and Jones (1999); the genetic distance between inhabitants of the country, as measured by variation in blood type (blood) from Gorodnichenko and Roland (2017); the average annual inflation rate from 1970 to 1998 (inflation) of the country; and trade openness (openness), which is imports plus exports as a fraction of GDP.

We run the regressions with the additional instrument variable English, which is the fraction of the population in the country that is English-speaking.

Table 17: First stage, with additional control variables

The dependent variable is Network Index, defined in (4). Disease variables are defined in (7), with details in Appendix C1. Control variables are life expectancy at birth (LEB), disability-adjusted life expectancy (daly), which is the expected length of time an individual lives free of disability, as measured by the World Health Organization in 2004, social infrastructure (SocInf) and executive power (Executive), measures of the efficient functioning of social and political institutions, constructed by Hall and Jones (1999); the genetic distance between inhabitants of the country, as measured by variation in blood type (blood) from Gorodnichenko and Roland (2017); the average annual inflation rate from 1970 to 1998 (inflation) of the country; and trade openness (openness), which is imports plus exports as a fraction of GDP.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
germs_diff	-0.410***	-0.433***	-0.427***	-0.452***	-0.442***	-0.463***	-0.367***	-0.402***	-0.433***	-0.456***	-0.437***	-0.460***	-0.478***	-0.496***
	(0.127)	(0.122)	(0.125)	(0.120)	(0.124)	(0.120)	(0.125)	(0.121)	(0.128)	(0.123)	(0.128)	(0.123)	(0.126)	(0.122)
germs_sum	-0.452***	-0.372***	-0.542***	-0.461***	-0.390***	-0.337***	-0.448***	-0.390***	-0.469***	-0.396***	-0.476***	-0.406***	-0.458***	-0.391***
	(0.105)	(0.105)	(0.0864)	(0.0885)	(0.0869)	(0.0878)	(0.0747)	(0.0763)	(0.0877)	(0.0889)	(0.0761)	(0.0791)	(0.0819)	(0.0841)
english		1.364**		1.424**		1.174**		1.284**		1.405**		1.325**		1.271**
		(0.540)		(0.543)		(0.537)		(0.544)		(0.554)		(0.559)		(0.547)
LEB	0.0190	0.0192												
	(0.0186)	(0.0179)												
daly			1.02e-05	9.09e-06										
			(1.23e-05)	(1.18e-05)										
SocInf					1.412**	1.245**								
					(0.594)	(0.581)								
blood							-6.942*	-5.436						
							(3.537)	(3.480)						
executive									0.0611	0.0502				
									(0.0804)	(0.0773)				
inflation											-0.168	-0.133		
											(0.109)	(0.106)		
openness													0.576	0.503
													(0.410)	(0.397)
Constant	0.594	0.172	1.940***	1.537***	0.846	0.636	2.328***	1.885***	1.532**	1.201*	2.362***	1.890***	1.649***	1.318***
	(1.485)	(1.437)	(0.304)	(0.330)	(0.604)	(0.594)	(0.345)	(0.383)	(0.691)	(0.676)	(0.386)	(0.422)	(0.443)	(0.450)
Observations	69	69	70	70	64	64	70	70	68	68	65	65	64	64
F	5.24e-10	1.41e-10	1.78e-09	3.72e-10	2.16e-10	1.25e-10	3.97e-10	1.55e-10	3.77e-09	9.52e-10	2.94e-09	1.08e-09	1.19e-09	5.00e-10
R2	0.506	0.551	0.481	0.531	0.548	0.582	0.505	0.544	0.480	0.528	0.500	0.543	0.521	0.561

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18: Income and Networks, with Controls.

Coefficient Estimates with GDP as the dependent variable. GDP data from World Bank. Other variables defined in Table 17.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
network index	0.568**	0.476**	0.598***	0.537***	0.450**	0.352**	0.656*	0.493*	0.687**	0.561**	0.577**	0.490**	0.615***	0.516***
	(0.245)	(0.188)	(0.223)	(0.177)	(0.213)	(0.175)	(0.348)	(0.263)	(0.295)	(0.230)	(0.292)	(0.238)	(0.219)	(0.180)
germs_sum	0.0590	0.0200	-0.0855	-0.118	-0.159	-0.195**	-0.291*	-0.361***	-0.230	-0.289**	-0.315**	-0.356***	-0.167	-0.212**
	(0.134)	(0.113)	(0.135)	(0.114)	(0.102)	(0.0899)	(0.168)	(0.135)	(0.165)	(0.138)	(0.158)	(0.135)	(0.121)	(0.105)
LEB	0.101***	0.103***												
	(0.0159)	(0.0148)												
daly			-6.52e-05***	-6.47e-05***										
			(9.47e-06)	(9.25e-06)										
SocInf					3.001***	3.165***								
					(0.571)	(0.519)								
blood_distUSA							-3.402	-4.894						
							(4.773)	(4.162)						
xcon1990sj									0.122	0.128*				
									(0.0814)	(0.0779)				
logavgcpinflat7098											-0.126	-0.142		
											(0.122)	(0.116)		
YrsOpen													1.780***	1.836***
													(0.362)	(0.341)
Constant	2.173*	2.180*	10.23***	10.35***	7.585***	7.640***	10.07***	10.46***	8.881***	9.080***	10.23***	10.44***	8.514***	8.672***
	(1.172)	(1.120)	(0.487)	(0.408)	(0.471)	(0.447)	(0.912)	(0.720)	(0.832)	(0.756)	(0.791)	(0.678)	(0.507)	(0.453)
Observations	69	69	70	70	64	64	70	70	68	68	65	65	64	64
Sargan-p		0.419		0.219		0.900		0.642		0.580		0.297		0.870

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

C.7 Addressing Econometric Concerns

Our identifying assumption is that, although technology diffusion and GDP may affect disease prevalence, it affects many diseases similarly. Likewise, the direct effect on GDP of different types of disease is also similar. Thus, the difference in the prevalence of two types of disease is exogenous with respect to GDP. The difference we consider is the difference between diseases that reside in humans (human-specific plus multihost) and diseases that reside exclusively in nonhuman animals (zoonotic diseases).

The results reflect covariances with GDP. Surely income explains variation in both social networks and technology. One might worry that this drives our results. This concern is somewhat alleviated by the fact that the technology measure is designed to remove the income effect on technology adoption. Comin and Mestieri (2012) use a growth model to design an estimator that captures the slope of a diffusion curve of a new technology, after that technology has been introduced to the country, and net of the effect that we would expect higher income to have on the demand for the new technology. Given that the diffusion measure is constructed to eliminate the direct effect of income, it does not make sense to also control for GDP in the regression. Essentially, the technology diffusion measure is already a residual from an estimation procedure that removed GDP effects.

Of course, the difference in disease instrument is also constructed with the idea that it is orthogonal to income. However, it is impossible to prove that independence for sure.

Technology explains differences in disease prevalence. Our empirical strategy is based on the assumptions that human-transmitted disease $\bar{d}$ and zoonotic disease p_ξ have the same relationship with technology A but different relationships with the social network $\tilde{N}$. One may think that this relationship does not necessarily hold. It sounds reasonable, but one cannot be sure.

The most salient example of this problem is that clean water initiatives might be one of the first public health measures adopted when income rises. If this were the case, then there would be a negative correlation between zoonotic illness and technology diffusion, and therefore a positive correlation between (human - zoonotic) diseases ($\Delta germ$) and shocks to technology diffusion ϵ. If $E[\epsilon x] > 0$, how would this bias the results? A positive shock to income (high ϵ) would increase the difference in disease (x), which would decrease individualism $\tilde{N}$ (since we estimate $\beta_5 < 0$). This would induce a negative correlation between A and $\tilde{N}$, which would lower the estimated coefficient β_2 in equation (5). So β_2 would be downwardly biased. Thus, if the instrument is invalid because economic development primarily reduces waterborne illnesses, then the true size of the network's effect on technology diffusion is even larger than what we estimate.

Of course, it is possible that there is some other force that causes technology to have a greater effect on human diseases that works in the opposite direction to bias our estimates upward.

The difference in disease has a direct effect on diffusion. If human-to-human diseases are more deadly than zoonotic onces, then perhaps the difference in human and zoonotic diseases might decrease average income and decrease the speed of technology diffusion. Of course, the technology diffusion measure is supposed to be the part of diffusion not explained by differences in income. So in principle, it should net out this effect. But perhaps it does not. It could also be that debilitating diseases cost more in forgone output than death does. In this case, our results would underestimate the true effect of networks on diffusion.

We address this concern in three ways. One set of results that speak to this problem are estimates of the effect of social networks using the colonial settler mortality rates from Acemoglu et al. (2001). Because these are the mortality rates of Europeans arriving in newly discovered lands hundreds of years ago, it is very unlikely that these disease mortality rates directly affect income or the process of technology diffusion today. Appendix B reports the results, which show that the coefficients on the social network measures with Acemoglu, Johnson and Robinson's instrument are very similar to our original IV estimates. Another set of results that address this concern are those that control for disease-adjusted life expectancy. Table 18 demonstrates that our network index survives statistically and has economically similar effects, even after controlling for disease-adjusted life expectancy (daly variable).

Social networks affect disease The other hypothetical cause for concern might be that faster technology diffusion and the accompanying higher income cause the social structure to change, which could, in turn, 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 network structure $\tilde{N}$ depends on A, something already represented in our specification, and it suggests that there should be an additional equation representing the idea that the instrument x depends on the

network: $x = \psi_1 + \psi_2 S + \nu$. In this structure, as long as $e[\epsilon\nu] = 0$, x is still a valid instrument for $\tilde{N}$. In other words, as long as technology diffusion affects the difference in disease through networks, rather than directly, this form of reverse causality *does not invalidate the use of disease differences as instruments*. It only implies that β_5 does not identify the effect of disease on social institutions.