

GERMS, SOCIAL NETWORKS AND GROWTH

Alessandra Fogli and Laura Veldkamp*

July 12, 2016

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 model also explains how different social networks may emerge endogenously in response to the prevalence of infectious disease. Initial differences in disease prevalence can produce different network structures, leading to divergent levels of income. We compare calibrated model predictions with data. The model and data agree that a one-standard-deviation increase in our index of network diffusion speed results in output growth that is 1/2% higher per year.

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 potential size of these effects and compare the model's predictions with data.

Measuring the speed of information or technology diffusion within various kinds of networks has a long history (Granovetter, 2005). Given these findings, a simple way to explain the effect of social structure on GDP is to show that some types of social networks disseminate new technologies more

*Corresponding author: afogli00@gmail.com, Federal Reserve Bank of Minneapolis, 90 Hennepin Ave., Minneapolis, MN 55405. lveldkam@stern.nyu.edu, NYU Stern School of Business, 44 West Fourth St., Rm. 7-77, New York, NY 10012. 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, and Amanda Michaud 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.

¹See, 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.

efficiently than others and append a production economy in which the average technology level is related to output and income. This explanation is problematic in two ways. First, social contacts are presumably endogenous. If that is true, why would a social network structure that inhibits growth emerge and persist? Second, this explanation is difficult to quantify or test. How might we determine whether its effects are trivial or not? Although researchers have mapped social networks in schools or online communities, mapping the exact social network structure for an entire economy is not feasible.

Our theory for why some societies have growth-inhibiting social structures revolves around the idea that communicable diseases and technologies spread in similar ways – through human contact. We explore an evolutionary model in which some people favor small, stable, local social networks and others do not. Stable, local, and fractionalized networks are more insular. They have fewer links with the rest of the community. This limited connectivity reduces the risk of an infection entering the collective, allowing the participants to live longer. But it also restricts the group’s exposure to new technologies. In countries where communicable diseases are inherently more prevalent, the high risk of infection makes nodes with distant linkages more likely to die out. A stable, local, and fractionalized social network that inhibits the spread of disease and technology will emerge. In countries where communicable diseases are less prevalent, nodes with only local connections will be less economically and reproductively successful. The greater reproductive success of nodes that diffuse ideas and germs quickly leads them to dominate social networks in the long run.

The idea that disease prevalence and social networks are related can help to isolate and quantify the effect of social networks on technology diffusion. Isolating this effect is a challenging task because technology diffusion and social networks both affect each other: technology diffusion is a key determinant of income, which may well affect a country’s social network structure. To circumvent this problem, we build a model where networks are endogenous. By calibrating the model to data, we can use the model to perform quantitative experiments where we isolate the effect of changes in networks.

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.

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 compares the model's predictions for the relationship between disease prevalence and social network structure and the relationship between network structure and technology diffusion with the data. After showing that the model has quantitatively realistic predictions, Section 5 then uses the model to calibrate how much an exogenous change in network structure would affect technology and output. 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. The model and data agree that a one-standard-deviation change in the network index results in productivity and therefore output growth that is 1/2% higher per year. This difference quickly cumulates to create large gaps in income.

Related Literature The paper contributes to four growing literatures. Primarily, the paper is about a technology diffusion process. Recent work by Lucas and Moll (2011) 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 (2011) 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

²See, for example, Chaney (2013) or Kelley *et al.* (2013).

³See, 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.

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.

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.

1 A Network Diffusion Model

Our model helps us determine the effect of social networks on macroeconomic outcomes in three ways. First, the model fixes ideas. The concept of social network structure is a fungible one. We want to 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. But we do want to be explicit about what we intend to measure.

Second, the model guides the choice of variables that we explore empirically. The model teaches us about how three different aspects of social networks facilitate technology diffusion. It also motivates our choice of disease as an instrument for social network structure. Specifically, it explains why disease that spreads from human to human might influence a society’s social network in a persistent way.

The final role of the model is 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 predicted growth effect.

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 through a social system. 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. 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)

With this description of the process of technological diffusion in mind, we propose the following model.

1.1 Economic Environment

Time, denoted by $t = \{1, \dots, T\}$, is discrete and finite. At any given time t , there are n agents, indexed by their location $j \in \{1, 2, \dots, n\}$ on a circle. Each agent produces output with a technology $A_j(t)$:

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

Social Networks Each person i is socially connected to n other people. If two people have a social network connection, we call them “friends.” Our network friends could also be family, co-workers or any pair of people with close and repeated social contact. Let $\eta_{jk} = 1$ if person j and person k are friends and $= 0$ otherwise. To capture the idea that a person cannot infect themselves in the following period, we set all diagonal elements (η_{jj}) to zero. Let the network of all connections be denoted N .

Spread of Technology Technological progress occurs when someone improves on an existing technology. To make this improvement, 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 a Poisson probability λ 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 $A_k(t) > A_j(t)$, then with probability ϕ , j can produce with k 's technology in the following period: $A_j(t+1) = A_k(t)$.

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

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 ($A_j(s) = 0, \forall s \in [t, t_r]$). The date at which life ends and a new node appears t_r is governed by a random renewal process $\xi_{rj}(t)$. The variable $\xi_{rj}(t)$ is an i.i.d. binary random variable with $Pr(\xi_{rj}(t) = 1) = p_r$. When the node renewal process hits ($\xi_{rj}(t) = 1$), the agent j is replaced by a new, healthy person in the same location j . That new agent j inherits the same technology that the healthy parent node j had before infection.

1.2 Node Types and Aggregate Network Properties

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

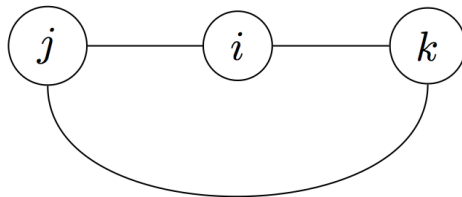
When we think about node types governing social connections, we are faced with an inevitable dilemma: What happens if i 's type dictates that he should be friends with j , but j should not be friends with i ? To resolve this impasse, we assume that each agent's type dictates her links in one direction 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 node forms 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.

Network Feature 1: Collectivism Collectivism is an aspect of a social network structure that has been extensively studied by sociologists. Mutual friendships and interdependence are hallmarks of collectivist societies. To measure this interdependence, we can ask: If i is friends with j and with k , how often are j and k also friends? We refer to a structure in which i , j , and k are all connected to each other as a *collective*. Figure 1 illustrates this collective structure.

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

In all the networks we consider, all nodes are always connected to their immediate neighbors.

Figure 1: A collective



This is what constitutes a neighbor in this model. The modulus phrase in the definition simply means that if $j + \Delta_j > N$, then instead of that link going off the network, it simply wraps around the other side of the ring.⁴ Therefore, when a node is connected to its two closest neighbors, that will always form a collective, since those neighbors are always themselves connected. Thus, collectivist nodes are ones whose ties complete a collective. In contrast, a node that is not collectivist is connected to two nodes that are not themselves connected.

Definition 2 *If agent j is not collectivist ($\theta_j = 0$), then she is connected to the closest nodes to her left and the node $\Delta_j > 2$ spaces away: $\eta_{jk} = 1$ for $k = \{j + 1, j + \Delta_j\}$, modulus the size of the network.*

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 2.

Network Feature 2: 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 3 *If an agent has degree n_f , he is connected to $n_f/2$ of the closest nodes to his left: $\eta_{jk} = 1$ for $k = \{j + 1, \dots, j + n_f/2\}$, modulus the size of the network.*

The number of links one creates on one side is given by the degree divided by two because that ensures that when every node has degree-type n_f , every node is connected to n_f other nodes and thus the network as a whole has degree n_f .

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

⁴If we consider a ring of size 100, the last node on the ring is node 100. This node cannot be connected to $100 + \Delta_j$ because there is no node with this index. Instead, node 100 would be connected to node $\text{mod}(100 + \Delta_j, 100) = \Delta_j$. So, if $\Delta_j = 4$, then node 100 is connected to node 4 and node 99 to node 3, and so forth.

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 each of the links to her left has a probability m_j of being broken and reassigned to any other node which with j is unconnected, with equal probability.*

This type of ring network 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. But first we consider the question of where such networks come from.

1.3 Network Evolution

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, we endogenize networks with an evolutionary model in which the network changes as agents die and new ones are born in their place. This evolutionary model also helps to explain why growth-inhibiting social networks might persist long after most diseases have died out.⁵ The idea that social circles might evolve based on disease avoidance is based on biological evidence. Animal behavior researchers have long known that many species choose mates with health considerations in mind (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). Motivated by this evidence, we propose the following model.

At each date t , each person j has a type $\tau_j(t)$. That type is a three-dimensional state comprising a binary collectivist type θ_j , a degree type n_f and a mobility type m_j , where n_f and m come from a discrete, finite set of possible types. Note that type does not include technology. A social change shock does not alter the technology diffusion process described in Section 1.1. It only changes the network, which affects technology diffusion indirectly.

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

⁵Another 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.

probability p_r 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 $k^* = \operatorname{argmax}_{\{k:\eta_{jk}(t)=1\}} A_k(t)$ (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^* : $\tau_j(t + 1) = \tau_{k^*}(t)$.⁶

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 helps to speed up the process of social change.

The idea behind this process is that evolutionary models often have the feature that more successful types are passed on more frequently. At the same time, we want to retain the network-based idea that one's traits are shaped by one's community. Therefore, in the 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.

1.4 Feedbacks between Disease and Technology

One of the problems that we will face in our measurement exercise is the possibility of reverse causality and endogeneity problems that arise from the direct effects of disease prevalence on technological innovation and the effect of new technologies on the spread of disease. To understand how these issues arise, it is useful to first understand what effect such feedbacks have in our simple model. Therefore, we consider two extensions of our baseline model. We calibrate the model without these two feedbacks and then turn them on to see how they affect our predictions.

Feedback 1 (Innovation): When the first feedback is turned on, the rate of technological innovation depends on disease prevalence. Instead of a fixed probability λ of a new invention each period, we introduce a time-varying probability

$$\lambda(\bar{d}) = (1 - \Phi(\bar{d}(t)/\Lambda_d)), \tag{1}$$

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.

⁶Another 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.

Feedback 2 (Infection): When the second feedback is turned on, the probability of disease transmission depends on the aggregate technology level. Instead of a fixed probability π that one infected person infects a social contact, we introduce a probability that depends on average productivity $A(t) \equiv 1/n \sum_j A_j(t)$. The endogenous disease transmission probability is

$$\pi(A(t)) = (1 - \Phi(A(t)/\Pi_A)). \quad (2)$$

This captures the idea that as technology improves, public health innovations, vaccines, and medical advances make new infections less likely.

1.5 Exogenous Network Model

Understanding where social networks come from is important for a couple of reasons: It answers the potential criticism that if some networks are bad for growth, societies would not adopt them. It also informs our approach to identifying network effects in the data. At the same time, we also want to use the model to measure the effect of a change in the social network. If the network is endogenous, we can only observe the effect of the network changing, in conjunction with the conditions that prompted the change. That does not isolate the effect of social networks on income. To identify the effect of social networks in our model, we need to be able to alter the social network exogenously. Here, we describe exactly how we do that.

Essentially, the model is the same, except that there is no change in node types. The technological and disease diffusion processes are identical. In this model, we shut off the two feedback mechanisms (fixed λ and π). The one thing that is slightly different is the interpretation of the node renewal process $\xi_{rj}(t)$. In the endogenous network model, renewal means that nodes have the opportunity to change type. Since node types are fixed in the exogenous network model, that is not relevant. But node renewal also marked the date at which sick nodes are replaced with healthy ones. In the exogenous network model, the $\xi_{rj}(t)$ shock still plays that second role. It governs the rate of recovery – or rebirth – from disease. We mention this because it explains why we calibrate the probability of renewal p_r in such different ways for the two models. It simply plays two different roles.

2 Theoretical Results

We begin by exploring the long-run properties of the network evolution model. Once we know what network types the system converges to, we then explore how these networks spread technologies and diseases. We show that a network summary statistic known as average path length is related

to both the expected time to infection and the expected rate of technological progress. Then, we relate each of the three network features to average path length. To do this second part of the analysis, we use our exogenous network model. This simpler setting is useful because it allows us to focus on a few key economic forces and to isolate the effect of the network on aggregate economic outcomes.

2.1 Network Evolution and Steady State Convergence

Why do some societies end up with a collectivist network even though it inhibits growth? These results describe the long-run properties of networks and disease. They explain how disease prevalence can permanently alter social structure. This idea is important because it rationalizes differences in social structures that persist even after diseases have been eliminated. The first result shows that eventually, the economy always converges to a uniform-type network. For example, if the type space was only individualist or collectivist, then every network would converge to either the fully collectivist network (1) or the fully individualist one (2).

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

In other words, after some date T , everyone will have the same type forever after. They might all be individualist or all be collectivist. But everyone will be the same. 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. Since there are a finite number of states, and whenever there exists a j, k such that $\tau_j(t) \neq \tau_k(t)$, every state can be reached with positive probability in a finite number of steps, then with probability one, at some finite time, an absorbing state is reached and the economy stays there forever after. 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.⁷

Result 2 *With probability 1, the disease dies out: $\exists T$ s.t. $d_j(t) = 0 \forall j$ and $\forall t > T$.*

These results teach us that which network type will prevail is largely dependent on which dies out first: the individualist trait or the disease. When there is a positive probability of infection,

⁷Epidemiological 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.

people with individualist networks have shorter productive lifetimes on average. If disease is very prevalent, it drives all the individualists out, because they are more likely to be infected. The society is left with a collectivist network forever after. If disease is not very prevalent, its transmission rate is low, or by good luck it just dies out quickly, then individualists will survive. Since they are more economically successful, they are more likely to pass on their individualist trait. So the economy is more likely to converge to an individualist network. This outcome is not certain because network type change is random. It is always possible that individualists die out, even if the disease itself is no longer present. These results hold, no matter if τ represents collectivist types, high-degree, or mobile types. The main takeaway here is that networks can persist long after the conditions to which they were adapted have changed.

2.2 Network Features and Diffusion Speed

Next, we consider three specific characteristics that networks, or nodes, may have and how they regulate the average path length of the network and thus the speed of diffusion. The 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 $\mathcal{N} = \{1, \dots, n\}$ represent the set of n nodes. Then,

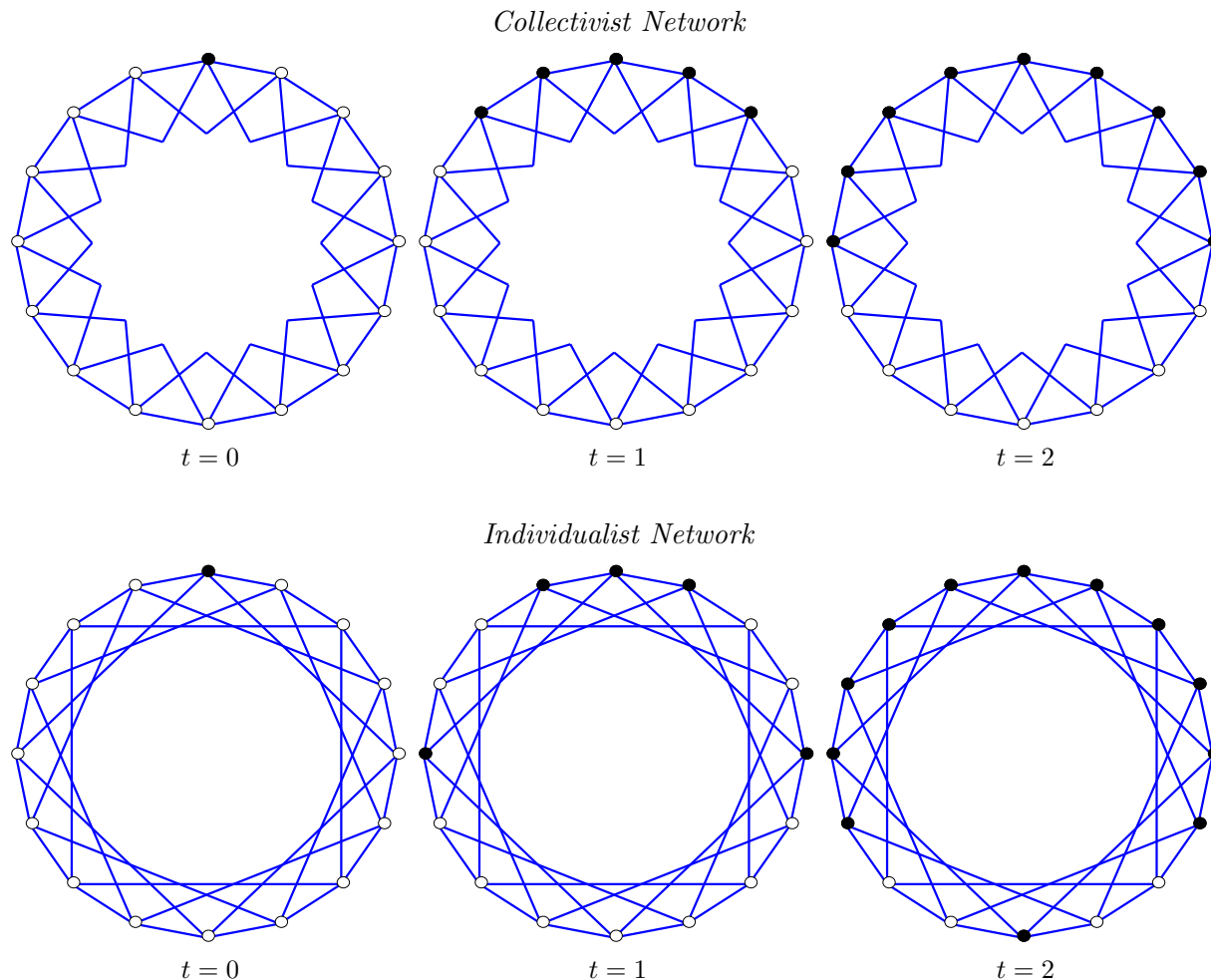
$$\text{Average path length} = \frac{1}{n} \sum_{i \in \mathcal{N}} \frac{\sum_{j \in \mathcal{N}/i} p_{ij}}{n-1}. \quad (3)$$

If the average path length between individuals is shorter, diseases and ideas disseminate more quickly because they require fewer transmissions to reach most nodes. Therefore, when we talk about a faster diffusion speed, it refers to a shorter average path length.

Network Feature 1: Collectivism 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.

Result 3 (*Collectivism slows diffusion*) For any $m \in (2, n/4)$, there is a network size \bar{n} such that,

Figure 2: 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.



for any $n > \bar{n}$, the average path length in the purely collectivist network ($\theta_j = 1, \forall j$) is longer than the average path length in the no-collectivist network ($\theta_j = 0, \forall j$).

Figure 2 illustrates the smaller path length and faster diffusion process in a purely individualist network, compared with a purely collectivist one. In a 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 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. This example illustrates why, on average, ideas and diseases will diffuse more slowly through a collectivist network than through an individualist one.

Could Collectivism Facilitate Technology Diffusion? Perhaps Arrow (1969) was not correct and technology diffusion is not a process “formally akin to the spread of an infectious disease.” Instead, a technology is adopted only when a person comes in contact with multiple other people who have also adopted it. This idea is called complex contagion. Centola and Macy (2007) demonstrate the theoretical possibility that having many mutual friendships makes it more likely that groups of people adopt a technology together. However, these same authors admit, “We know of no empirical studies that have directly tested the need for wide bridges 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).

Network Feature 2: Degree A ring lattice social network with a higher degree has a lower average path length. With more connections, it requires fewer steps to reach other nodes. This network will speed the diffusion of germs and technologies.

Consider a ring lattice in which every node has degree n_f and where n_f is even. Each individual j is friends with the n_f closest nodes. In other words, $\eta_{jk} = 1$ for $k = \{j - n_f/2, \dots, j - 1, j + 1, \dots, j + n_f/2\}$ and $\eta_{jk} = 0$ for all other k . In such a network:

Result 4 (*Higher degree speeds diffusion*) *The average path length in a ring network is a decreasing function of degree n_f .*

Network Feature 3: Mobility The idea here is that most links are local, but people occasionally move, change jobs, or randomly encounter new friends that were not in their original social circle. These new links are the randomly chosen long links.

The procedure to add symmetric mobility m_j to network N is to take each node j in N and add a new link with probability m_j . The new link connects j to any other node in the network with uniform probability.

Result 5 (*Mobility speeds diffusion*) *The expected average path length of the network is a decreasing function of the mobility index m_j .*

A lower probability of forming shortcuts (lower mobility) increases the average path length between nodes. As such, it decreases the speed of diffusion.

All the results in this section are about symmetric networks simply because there are too many possible asymmetric networks to consider them all. However, these same relationships hold in our calibrated endogenous network, which is asymmetric.

2.3 Diffusion Speed, Infections, and Innovations

The last set of results connect average path length to 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 6 *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.

The proofs on this and all future results are in Appendix A.

But faster diffusion is not the same as faster technological 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. The following result clarifies the mechanism by which the networks with lower average path length achieve a higher rate of growth.

Result 7 *Suppose that at t , two networks have the same $A_j(t) \forall j$. Then the probability that the next new idea arrival will increase the technological frontier is larger in the network with the smaller average path length.*

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. Next, we describe what observable features of a network cause its average path length to be long or short.

3 Connecting Model and Data

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 69 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 Pathogen Prevalence

Using old epidemiological atlases, we compiled the historical prevalence of nine different pathogens in the 1930s. The historical nature of the data alleviates some of the concerns one might have about direct effects of disease on income. We study nine pathogens: leishmaniasis, leprosy, trypanosomes, malaria, schistosomes, filariae, dengue, typhus, and tuberculosis. We choose these diseases because we have good worldwide data on their incidence, and they are serious, potentially life-threatening diseases that people would go to great length to avoid.

The historical pathogen prevalence data are from Murray and Schaller (2010), who build on existing data sets and employ old epidemiological atlases to rate the prevalence of nine infectious diseases in each of 230 geopolitical regions in the world. For all except tuberculosis, the prevalence estimate is based primarily on epidemiological maps provided in Rodenwaldt and Jusatz (1961) and Simmons *et al.* (1945). Much of their data were, in turn, collected by the Medical Intelligence Division of the United States Army. A four-point coding scheme was employed: 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. The prevalence of tuberculosis 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. We have prevalence of all nine diseases in 160 political regions.

To identify the effect of disease on social network structure in the data, it is useful to distinguish between diseases that are transmitted in different ways. Epidemiologists often classify infectious diseases by reservoir.⁸ 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 . Next, we construct the raw difference in disease prevalence as $\tilde{\Delta}_{\text{germ}} \equiv d_h - d_z$. Then from that, we construct the standardized difference

⁸See, for example, Smith *et al.* (2007) or Thornhill *et al.* (2010).

to use as an instrumental variable:

$$\Delta\text{germ} \equiv \frac{1}{\text{std}(\tilde{\Delta}\text{germ})} \left(d_h - d_z - \text{mean}(\tilde{\Delta}\text{germ}) \right) \quad (4)$$

Finally, to control for the total disease prevalence of all diseases in each country, we use

$$\text{sumgerm} \equiv d_h + d_z. \quad (5)$$

We provide additional information about the characteristics of the diseases and redo our analysis with contemporaneous data, which has many more diseases, in Appendix C.

3.2 Measuring Social Networks

While collectivism data measure features of foreign social relations directly, degree and link stability measure the social relationships and moving patterns of U.S. residents who are immigrants or children of immigrants from other countries. These latter two measures help to mitigate some of the concerns that arise when one does cross-country empirical analysis.

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

Hofstede (2001) 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, 2001, p. 9-10)

As a proxy for collectivism, we use $\text{collectivism}_i = 1 - 0.01\text{hofstede}_i$, where hofstede_i is 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 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 an index for each country. The least collectivist (highest Hofstede index) countries are the United States (91), Australia (90), and Great Britain (89); the most collectivist are 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.

Of course, IBM employees are not representative of country residents as a whole. Appendix C describes additional studies on other population subgroups that corroborate Hofstede’s findings. The appendix also describes some of Hofstede’s questions and then 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.

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 U.S. residents that report having ancestors coming from that country. Our data come from the General Social Survey (GSS).⁹ 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 : $degree_i = \alpha + \beta_1 hofstede_i + \beta_2 ELF_i + \epsilon_i$, where *ELF* is an ethno-linguistic fractionalization measure, which reflects “the probability that two randomly selected individuals from a population belonged to different groups” (Alesina *et al.*, 2003). Then, we use the predicted values from that regression to fill in for our missing observations.

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 governs the quantity of social ties in far-off locations. We do not have a large cross-country panel of

⁹GSS, <http://www3.norc.org/gss+website/>.

Table 1: Correlations

	Individualism (1)	Mobility (2)	Degree (3)	Index loadings
Individualism	1.00			0.61
Mobility	0.553	1.00		0.56
Degree	0.856	0.662	1.00	0.56

The table reports correlations of the three measures of social network structure described in section 3.2 and the loadings of each measure on the network index.

mobility data. But we do have extensive data on mobility for U.S. residents, including immigrants. The data come from the Current Population Survey (CPS). The variable we use is MIGRATE1, which indicates whether the respondent had changed residence in the past year. Movers are those who reported moving outside their county borders. Our measure of mobility is the fraction of first-generation U.S. immigrants from each country that move across a county border in a given year.

There are 66 mobility observations. For three countries (Estonia, United States, and Libya), the values are imputed by estimating a , b_1 , and b_2 in $mobility_i = a + b_1 hofstede_i + b_2 ELF_i + \xi_i$, and using the predicted values.

The underlying assumption is that people who move to the United States from countries with higher mobility/degree maintain higher degree/mobility and pass these social network preferences on to their children. This approach of using data for U.S. residents follows many other studies and helps to control for many institutional differences that might otherwise explain different behavior across countries. 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.

Network Diffusion Index. Our three network features—collectivism, degree, and mobility—all accelerate the diffusion of new technologies. Table 1 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. For some of our model and data comparisons, it will be useful to combine our network measures into a single network index to facilitate valid inference. The index we construct is the first principal component of the three measures. The last column of Table 1 lists the linear weights.

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 70% of the variation in the three measures.

3.3 Measuring the Rate of 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. At a country level, there are two margins of technology adoption: the extensive margin (whether or not a technology is adopted at all) and the intensive margin (how quickly a technology 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.

We do not calibrate to match the diffusion rate of a technology, that is, its intensive margin of adoption. Instead we use diffusion rate data to evaluate the model's fit. The technology diffusion rate is the following: For a given country, plotting the normalized level of a given technology (e.g., log telephone usage minus log country income) over time yields an increasing curve. For a given technology, these curves look similar across countries, except for horizontal and vertical shifts. The horizontal shifts correspond to the extensive margin of technology adoption; if country A adopts telephones in exactly the same way as country B does, only 20 years later, its curve will be identical to that of country B's except it will be shifted 20 years to the right. The measured diffusion rates will be identical. However, if country A adopts telephones, starting at the same time as country B, but less vigorously, its curve will be below that of B's. Measured diffusion will be slower. Specifically, Comin and Mestieri (2012) estimate the slope of a nonlinear diffusion curve. A higher slope parameter \dot{q}_{ij} indicates a faster diffusion rate of technology j in country i . In addition, they use an equilibrium model of technology adoption to control for the effect of aggregate demand on technology adoption. So, this diffusion measure should not be subject to the criticism that it is GDP differences that create differences in technology diffusion speeds.

A complication is that the diffusion data set is unbalanced; if data for a country are available only for slowly spreading technologies, the country might artificially appear technologically backward. To control for this problem, we estimate $\dot{q}_{ij} = \alpha_j + e_{ij}$, where α_j is a technology-specific fixed effect. Our measure of the technology diffusion speed for a given country is the average residual:

$$\text{diffusion}_i = \sum_j e_{ij}. \tag{6}$$

3.4 Parameter Choice

The key parameters are summarized in Table 2. We calibrate the models with exogenous networks and endogenous network evolution separately so that they each give us a reasonable match to the data. Our strategy is to match moments of our least collectivist network to U.S. data. Then the model predicts endogenous outcomes for other countries. For calibration, we left the feedbacks inactive since our main results are without feedback.

Table 2: Calibrated Model Parameters

Description		Exog Model	Endog Model	Data Target
Innovation size	δ	70	80	2.6% GDP growth in U.S.
Technology arrival rate	λ	0.1%	0.08%	U.S. tech adoption lag, 21 yrs (Comin and Hobijn, 2010)
Technology transfer prob	ϕ	8.5%	35%	Half-diffusion in 40 years (Greenwood <i>et al.</i> , 2005)
Disease transmission prob	π	0.5%	12%	Disease extinct in 150 yrs
Individualist's longer link	Δ_j	15	7	Average U.S. path length of 5
Node type change rate	p_r	0.014	0.20	1 change per generation
Colltv fraction	f_C	14%	endog.	(1-Hofstede) for U.S.
Mobility	m_j	20%	n/a	Moving rate in the U.S.
Degree	n_f	4	4	Average number of friends

Technology parameters. Everyone starts with a technology level of 1. 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. 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 so that the least collectivist network economy grows by 2.6% per year. The probability of transmitting a new technology to each friend (ϕ) is chosen to explain the fact that for the average household technology, the time between invention and when adoption hits 50% is 40.18 years (Greenwood *et al.*, 2005).¹⁰

Disease parameters. 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 least collectivist network with the calibrated mobility and degree. While we use the same target in both exogenous and endogenous network economies, the resulting estimates differ

¹⁰We calculate this half-diffusion from the Greenwood *et al.* (2005) data set by averaging the number of years from introduction until 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.

because when the endogenous network evolves, diffusion speeds evolve with it.

For the initial disease prevalence rate ($d_h(0)$), we explore high and low values between 0.05% and 39%. Since we are calibrating the model to the United States, we use as benchmark the lower initial disease prevalence of 0.05% for calibration. For comparison, one of the most prevalent diseases in any country was tuberculosis in China, where the disease was endemic. Tuberculosis is just one of the nine diseases we consider but is the most common cause of death in our sample. The death rate in China from tuberculosis was 5%.¹¹ In the evolutionary network model, we start the model off at 0.05% and explore the effect of higher and lower starting values in Appendix B.1.

Network parameters. Network degree (n_f) and mobility (m_j) allow the model to match the modal number of friends and the probability of a U.S. resident moving across county borders in the GSS data.

The node replacement process p_r plays a different role in the exogenous and endogenous network models. In the endogenous (evolutionary) model, p_r regulates the rate of network evolution, or social change. Therefore, we calibrate the p_r process to match a reasonable rate of social change. In the evolutionary network, p_r is set to 0.2. 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_r value is set such that each generation (25 years), 5% of the population changes type.

In the exogenous network model, the network cannot evolve. So calibrating to a rate of social evolution makes no sense. Instead p_r is just the probability that a node is reborn and replaced with a new healthy node that adopts the technology of its more productive neighbor. Since this shock looks more like a birth process, we calibrate the exogenous network p_r parameter so that the average life span (time between renewal) of a node is about 70 years. If we were instead to have the nodes renewed more frequently, as in the endogenous network model, our numbers would change a bit, but all the predictions would be qualitatively the same.

The fraction of individualists f_I is fixed in the exogenous network model. We set it so that the fraction of individualists equals the Hofstede individualism index (divided by 100) for an average individualist country. In the endogenous network model, f_I evolves. We start it at 30% and then let the model tell us where the average fraction of individualists should be.

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 U.S. 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 the average length of the

¹¹Note that this is a mortality rate, not an infection rate. This calibration is conservative because it uses only one disease and it would be easier to get large effects out of a higher disease prevalence rate.

path between any two nodes is about 5.5.

3.5 Simulation and Constructing Model Metrics

The results that follow come from a 400-person network simulated for 500 periods. Because the exogenous network model is quick to compute, we simulate each economy 500 times. For the slower endogenous simulations, we repeat each one 100 times.

The evolutionary social network model allows for node types to vary along three dimensions. But to explore the relationship between network characteristics and technology, it is useful to vary node types along just one dimension at a time. For the evolutionary results only, we hold the network degree to 4 and the mobility probability to zero. This allows us to see clearly how one network feature, collectivism, evolves in response to changes in disease and technology. When we turn to isolating network effects in Section 5, we allow all three network features to vary again.

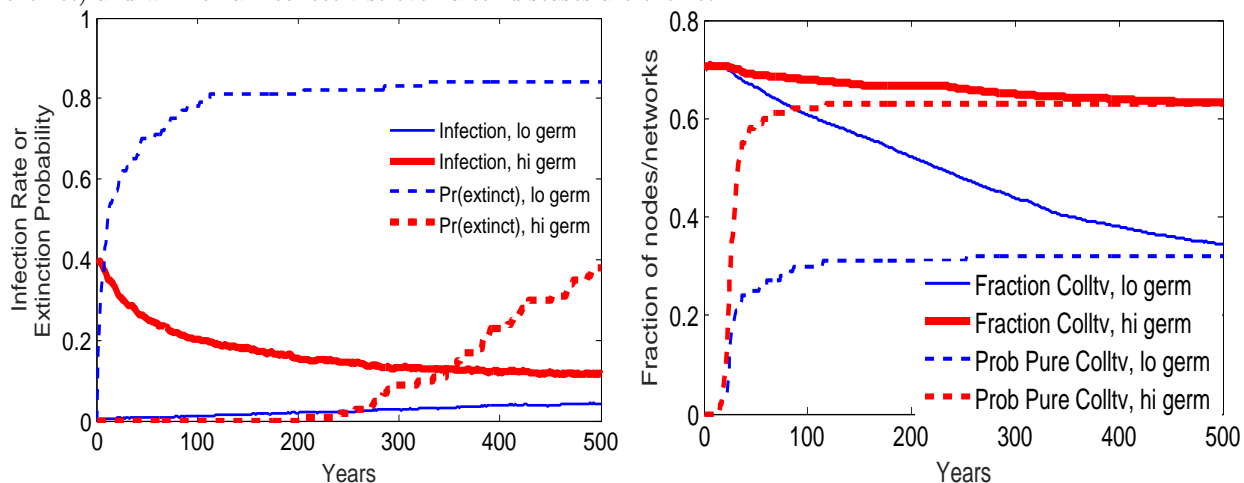
The dependent variables that we compare in model and data are log real GDP per capita ($\text{Log}(Y/L)$) and technology diffusion TechDiff . In the model, the GDP and diffusion rates depend sensitively on when we measure them. In most of the countries in our data sample, the diseases we measure in the 1930s are no longer common and are in many cases eradicated. In our model, the relationship between networks and productivity depends on the level of disease prevalence. To facilitate comparison with our data, we run the model until disease extinction and then use GDP and diffusion measures from the post-extinction periods.¹²

4 Comparing Moments in the Model and Data

So far, we have described a model where disease affects network evolution, networks regulate technology diffusion, and networks that are unconducive to rapid growth can persist long after disease has disappeared. The next question is whether the data support this mechanism and how big the effect is. The challenge is endogeneity. Income and technology obviously affect both networks and disease. We take two approaches to tackling this problem. In this section, we use our model, which has many of these endogenous relationships built in. We vary disease prevalence to observe its effect on network structure and the resulting effects on technology diffusion and output. Then we estimate the relationships (correlations) between disease and networks and network and output, in both the model and the data. These results compare endogenous model outcomes with

¹²Specifically, we map the model-generated average time- t technology level $A(t)$ into GDP per capita and a diffusion speed by constructing $\text{Log}(Y/L) = \log\left(\frac{\text{sum}((\text{germs}_t=0).*A(t))}{\text{sum}(\text{germs}_t=0)}\right)$ and $\text{TechDiff} = \log\left(\frac{\text{sum}((\text{germs}_t=0).*(A(t)-A(t-1)))}{\text{sum}(\text{germs}_t=0)}\right)$.

Figure 3: **Evolution of Infection Rates and Networks: Initial Disease Matters.** The Left panel plots the infection rate in economies with high (red, thick lines) and low (blue, thin lines) initial disease infection rates. The dashed lines report the fraction of economies where disease dies out. The right panel plots the fraction of collectivist nodes (dashed) and the fraction of economies that have converged to pure collectivism (individualists are extinct) and will remain collectivist even after diseases are extinct.



endogenous data outcomes. In the next section, we consider the effect of exogenous changes in a network on macro outcomes and attempt to isolate some of these exogenous changes in the data.

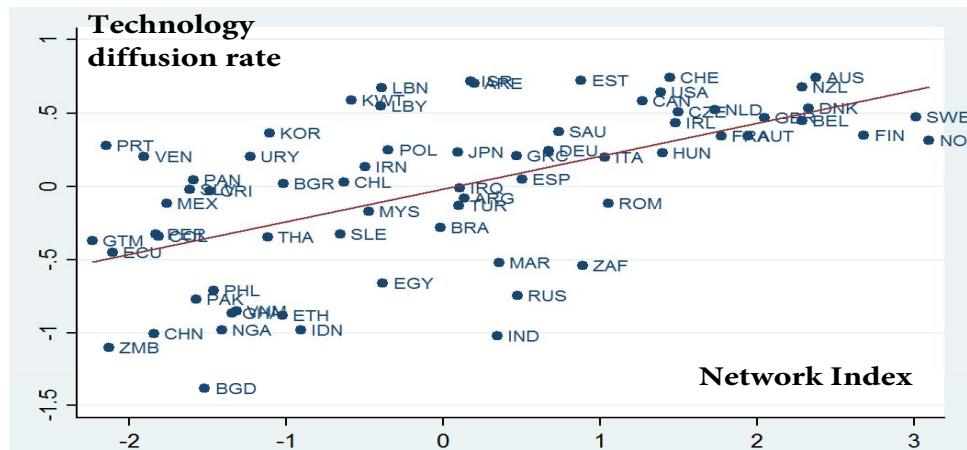
4.1 Technology and Networks

Model. In the model, what creates variation in both networks and technology is the initial level of disease and random chance. Therefore, we simulate two economies that vary in their initial infection rates. In Figure 3, the results labeled “lo germ” have an initial disease prevalence of 0.05%. The results labeled “hi germ” have an initial disease prevalence of 39%. 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.

The two panels of Figure 3 illustrate the model’s average dynamics. In the left panel, the high initial disease (red lines) quickly declines while low initial disease (blue) is stable on average. But these averages mask enormous heterogeneity across simulations. The lines with dashes show that in many low initial disease economies, the disease becomes extinct after only a few periods, while in others, it spreads and becomes endemic. Many high initial disease economies also extinguish the disease, but only after many periods.

In every simulation, the number of collectivist nodes starts at 70% and immediately changes as nodes are infected. Infected nodes have zero productivity, so their types are not passed on to others. Since collectivists are less likely to be infected, the evolutionary process systematically favors them

Figure 4: **Network index and technology diffusion (data)**. Comin and Mestieri (2012)'s technology diffusion measure (vertical axis) plotted against our social network index (horizontal axis).



when disease is prevalent. This shows up as a higher average rate of collectivism in the high initial disease (blue line) economy. As the infection rate declines and diseases are eradicated, evolution disadvantages collectivist nodes. Their shorter links make them less likely to have adopted the latest technology and therefore less likely to transmit their type as the highest-productivity node in their neighborhood. This shows up in the extinction rate (dashes). Initially, when disease is prevalent, many societies become exclusively collectivist. Since this is an absorbing state, they stay purely collectivist forever after. As the disease eradication rate rises, the number of purely collectivist societies stabilizes. In societies where individualists remain, these individualists thrive after disease eradication. Those societies rarely become collectivist because without disease, the technology adoption advantage of individualist nodes is the dominant evolutionary force.

The dynamics of networks and disease both affect output. Low-disease economies produce more for two reasons. First, being sick makes workers unproductive. Low-disease economies tend to see disease extinction more frequently and earlier. Since extinction is an absorbing state, these economies have persistently less disease and higher productivity. Second, low-disease economies are less likely to become collectivist. The benefit of being a collectivist node is less exposure to disease. When disease is scarce, individualist nodes with their better access to new technologies thrive. The faster diffusion of new ideas also boosts growth. Since networks that converge to purely collectivist or purely individualist stay that way forever, output growth in the low-disease societies is permanently higher on average.

Data. Figure 4 illustrates the raw data on the relationship between social network structure and the speed of technology diffusion in a scatter plot. It reveals the basic pattern that less collectivist, higher degree, and more mobile societies also tend to be societies in which technologies diffuse

quickly. As we show next, more individualist, higher degree, and more mobile societies also tend to be societies in which human diseases are less prevalent relative to zoonotic ones.

Comparison. To compare the model and data quantitatively, we simulate the endogenous network model with many different starting levels of disease and then run the same linear regression in the model as in the data and compare coefficients. Different initial disease levels – 32 equally spaced points between 0.05% and 40% – create variation in networks. Since the network index is constantly changing during each simulation, we choose the final network index (period 500). Likewise, output and technology are period-500 values. This fits with the fact that the network, technology, and output data we have were all measured in the last few decades. We do this comparison first in the model without feedbacks and then with feedbacks. We estimate the following linear model:

$$A = \beta_1 + \beta_2 \text{Network Index} + \epsilon, \quad (7)$$

where A is the speed of technology diffusion, Network Index is defined in Table 3 and denoted \tilde{N} , the β 's are unknown coefficients, and ϵ is a mean-zero residual orthogonal to \tilde{N} .

Table 3: Comparing Endogenous Network Outcomes to Data. Note: All entries are OLS estimated coefficients in a univariate regression with a constant. In the data, Network Index = $0.61 * \widetilde{Collectivism} + 0.56 * \widetilde{Degree} + 0.56 * \widetilde{Mobility}$, where $\tilde{x} \equiv (x - \text{mean}(x))/\text{std}(x)$. Δ germs is defined in (4). In the model, output per capita is A ; network index is one minus the fraction of collectivist nodes at the end of the simulation ($t = 500$); and Δ germ is the initial prevalence of the (communicable) disease. In the model, Δ germ lies on a grid of 32 equally spaced points between 0.05% and 40%.

	Data	Model	Data	Model
Dependent variable: Output per capita				
Network index	0.743*** (0.096)	0.603*** (0.084)		
Δ germ			-0.633*** (0.075)	-0.322*** (0.070)
Dependent variable: Network index				
Δ germ	-0.479*** (0.124)	-0.588*** (0.054)		

Table 3 reveals that the model and data predict similar relationships between network output per capita. The model explains about 80% of the magnitude of the coefficient estimated in the data. In contrast, the model predicts that networks vary less with disease prevalence than the data support. Since disease eventually dies out, and because our disease data are from the last century, Δ germ in the model is based on initial disease prevalence (period 1). In both model and data, output and networks are endogenous. These results do not suggest any direction of causality. However, the fact that the direction and magnitude of the coefficients resemble the data, even

though they were in no way targeted in the calibration, tells us that this model is quantitatively plausible.

4.2 The Effect of Disease-Technology Feedbacks

Section 1.4 introduced two extensions of the model: 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 two extensions introduces one new parameter to calibrate. For endogenous innovation, we choose the parameter $\lambda_d = 0.15$ (from eq.(1)) to match the average difference in adoption lags between the collectivist countries (those with individualism indices below the median) and noncollectivists. The new parameter that governs endogenous infection rates $\Pi_A = 200$ (from eq.(2)) targets disease extinction in 100-150 years on average.

These feedbacks help the model to generate more powerful and more nuanced results. When disease transmission rates fall as technology improves, that force helps to prevent the extinction of individualist agents, who are more susceptible to disease. In the benchmark model, the initial disease rate needs to be quite low to prevent individualist extinction in many simulations. This feedback allows the model forces to compete in societies with high and low initial disease, without the collectivist agents consistently taking over.

When we add the endogenous innovation rate as well, it amplifies the effect of disease and network differences on technology and growth. It creates a second channel besides the original diffusion and network evolution channel through which networks, and in turn disease, can influence technology creation. In the benchmark model, the society that starts with lower disease and becomes collectivist grows 0.7% faster in the long run. In the model with both feedbacks operative, that difference in growth rates doubles.

Details of these results and the calibration are in Appendix B.2. The message from this exercise is that the model can accommodate a more nuanced state process, and doing so strengthens our main conclusions.

5 Isolating Network Effects on 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. In the model, we can separate these two effects theoretically by exogenously changing the network structure and observing the predicted effect on technology diffusion and output. In the data, we attempt to isolate the network effect using an instrumental variables approach. We can again compare the magnitude of the model and data predictions to see whether our explanation for how

networks matter is consistent with the data. We find that the model-based approach and the identification-based approach deliver broadly consistent answers to the question: How much do social networks matter for the macroeconomy?

5.1 Effect of Exogenous Network Changes on Model Technology

To measure the effect of social networks on technology and output in the model, we need to exogenously change the network. To do this systematically, we start with a ring network and then vary each network feature one at a time. For each value of individualism, degree, and mobility, we choose a grid of values to examine. For degree, we simulate networks with $n_f = \{2, 4\}$. For mobility, we consider moving probabilities $p = [0 : 0.05 : 0.30]$. For individualism, we vary the fraction of individualist nodes in increments of 1/10th between $[0.06 : 0.1 : 0.86]$. When we simulate a new economy with a different degree or mobility parameter, we change the type of every node in the network in a symmetric way. We use the model's realized growth rates and construct the network index for each economy. The collection of growth rates and network indices for all runs of each economy (each network configuration) is our data set. We estimate the same linear regression of technology and output on the network index, first in the model and then in the data.

Although changes in individualism, degree, and mobility have small effects on growth, a 10% rise in the probability of a long-distance link increases annual growth by 0.06 percentage point. Similarly, giving each agent in the economy one additional social connection would increase growth by 0.16 percentage point. This simple exercise makes the point that a difference in network structure can create a small but permanent friction in technology diffusion. When cumulated over a long time horizon, this small friction has the potential to explain large differences in countries' incomes. When combined, these network effects are large. To compare the model and data, we compare the effects of changes in the network index. But first we examine the data.

5.2 Isolating Network Effects in the Data

So far, we have proposed a model of disease and technology diffusion, compared its predictions with the data, and then used the model to answer our question of interest: How much do networks matter for macro outcomes? We could stop here. But our theory suggests one additional way to compare the model's predictions with data using an instrumental variables approach.

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. We capture the correlation between disease prevalence and

technology from both directions of causality in the following relationship, which says that, after controlling for networks, there is a residual ϵ , which is correlated with technology and disease:

$$\epsilon = \delta_1 + \delta_2(d_h + d_z) + \xi. \quad (8)$$

If $E[\epsilon(d_h + d_z)] \neq 0$, in other words, if $\delta_2 \neq 0$, then disease prevalence is an invalid instrument.

To resolve this problem, we use the difference in human disease prevalence and zoonotic disease prevalence ($d_h - d_z$) as our instrument. When $var(d_h) = var(d_z)$, the difference ($d_h - d_z$) is orthogonal to the sum ($d_h + d_z$). Therefore, in our final exercise, we scale d_z to give it the same variance as d_h to ensure that the orthogonality holds. Thus, our identifying assumption is

$$E[\epsilon(d_h - d_z)] = 0. \quad (9)$$

Since in equation (8) we restrict the coefficients on d_h and d_z to be the same, we assume 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 A , the instrument ($d_h - d_z$) can be a powerful and valid instrument.¹³

We proceed by using the network index, defined in Section 3.2. It combines the three network features into one measure that we can construct from our data and investigate empirically.

The first stage results are also interesting on their own 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 (high d_h) 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.

Is disease difference a valid instrument? The instrumental variables estimates in Table 4 are only valid if the difference in the prevalence of human and zoonotic diseases is uncorrelated with income and technology. One reason this difference might be invalid would be if human diseases

¹³We do not need to know all the determinants of social structure. Rather, any subset of the determining variables can serve as valid instruments 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 . Just to be sure that this approach is also model consistent, Appendix B.2 verifies that in the model, this disease difference is a valid instrument. Appendix C.6 discusses some additional endogeneity concerns and the ways in which we address them.

Table 4: Social networks affect technology diffusion and output (data)

	Tech diffusion (A)			Output ($\log(Y/L)$)		
	(1)	(2)	(3)	(4)	(5)	(6)
First Stage						
Sum germs		-0.545*** (0.0698)	-0.487*** (0.0745)		-0.545*** (0.0698)	-0.487*** (0.0745)
Δ germs	-0.458*** (0.170)	-0.479*** (0.124)	-0.497*** (0.121)	-0.458*** (0.170)	-0.479*** (0.124)	-0.497*** (0.121)
English			1.060* (0.541)			1.060* (0.541)
Constant	0.0227	2.132***	1.826***	0.0227	2.132***	1.826***
Second Stage						
Network Index	0.357** (0.147)	0.360** (0.141)	0.334*** (0.123)	0.708** (0.303)	0.729** (0.266)	0.676*** (0.234)
Sum germs		-0.0326 (0.085)	-0.0465 (0.076)		-0.255 (0.160)	-0.284** (0.145)
Constant	-0.050	0.076	0.130	8.658***	9.645***	9.757***
Observations	69	69	69	69	69	69
Prob > F	0.20	0.000	0.000	0.024	0.000	0.000
Sargan			0.671			0.644
R^2	0.203	0.207	0.249	0.472	0.555	0.570

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 the Network Index defined in Table 3. The instrument Δ germs is defined in equation (4). The Sargan test statistic uses both Δ germ and *English* as instruments. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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. 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).

Furthermore, if the problem is that the instrument makes the predicted network variable correlated with disease virulence, and this 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 9 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.

Table 5: Comparing Regression Coefficients in Model and Data. Data entries are the coefficients on Network Index in columns (3) and (6) of Table 4. Model entries are from the same second-stage regression run on simulated model data where we exogenously vary the network. Y/L is log (RGDP) per worker and \tilde{N} are the measures of social network structure. Y/L data come from the Penn World Tables Mark 5.6. The instrument is $\Delta germ$, as in (4). *** Standard errors are in parentheses. $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	Data		Model	
	Output Log(Y/L)	Tech Diffusion Log(gY)	Output Log(Y/L)	Tech Diffusion Log(gY)
Network Index	0.676*** (0.164)	0.334** (0.234)	0.373*** (0.028)	0.346*** (0.026)

5.3 Comparing the Results from Model and Data

How do the model’s predictions compare with the second-stage estimates of the data? Table 5 reveals that the model accounts for about half of the empirical relationship between social networks and output per capita, and explains the observed relationship between networks and technology diffusion quite well. The model’s coefficient estimate is only about 5% different from the empirical estimate.

To interpret these results economically, it is easier to consider the effect of social network structure on log real output per worker (denoted $\text{Log}(Y/L)$ in Table 5). The model coefficients (on the right in Table 5) tell us that a one-standard-deviation increase in the network index (1.45 units) increases log output per worker by $1.45 \times 0.373 = 0.54$, which represents a 54% increase in real GDP per capita. The technology diffusion coefficient is similarly large. A one-standard-deviation increase in the network index raises technology growth by 0.5% in the model and 0.48% in the data. The data coefficient suggests an even larger increase. These large estimates suggest that social network structures might be relevant for macroeconomists and policy makers.

5.4 Robustness Checks

Model Robustness The qualitative nature of the underlying data on social structure makes this a tricky model to calibrate. That 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_z , the technology transmission probability ϕ , the probability of disease diffusion π , and initial disease prevalence \bar{d}_0 . For each of these alternatives, we reestimate the effect of social networks on technology diffusion and output, using the two-stage estimation on simulated model quantities.

The effect of networks on technology and output is remarkably stable. It varies by less than 20% across all of these different specifications. Table 6 in Appendix B reports results for each exercise

individually.

Robustness of Empirical Results 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. In doing so, we recognize that these control variables may themselves be endogenous. Inferring causality from these results would therefore be problematic. However, we continue to use $\Delta germ$ as an instrument and our network index as explanatory variables. We add the following variables, one by one, to the first- and second-stage estimations.¹⁴ 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, 2011), 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*. Finally, since regional differences might well be correlated with social network characteristics, we control for each of the world’s *continents*.

Tables 11 and 12 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. In the first-stage regressions, the difference in disease prevalence is consistently a significant predictor of the network index. Finally, when we reestimate with continent fixed effects, the network index remains as large as or larger than the original estimates while remaining statistically significant (Table 13, Appendix C). In sum, there is a statistical relationship between disease, social network structure, 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 struc-

¹⁴Our procedure here follows Hall and Jones (1999).

ture. 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 JOHNSON, S. (2005). Unbundling institutions. *Journal of Political Economy*, **113** (5), 949–995.
- , — and ROBINSON, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, **91** (5), 1369–1401.
- , — and — (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics*, **117** (4), 1231–1294.
- ALESINA, A., DEVLEESCHAUWER, A., EASTERLY, W., KURLAT, S. and WACZIARG, R. (2003). Fractionalization. *Journal of Economic Growth*, **8**(2), 155–194.
- ALGAN, Y. and CAHUC, P. (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 GALOR, O. (2012). The ‘out of africa’ hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, **103** (1), 1–46.
- BISIN, A. and VERDIER, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, **97**(2), 298–319.
- BROCK, W. A. and DURLAUF, S. N. (2006). Social interactions and macroeconomics. In D. Colander (ed.), *Post-Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, New York: Cambridge University Press, pp. 97–115.

- CENTOLA, D. and MACY, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, **113**(3), 702–734.
- CHANEY, T. (2013). The network structure of international trade, working paper, Toulouse School of Economics.
- COLEMAN, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, **94** (Supplement), S95–S120.
- COMIN, D., DMITRIEV, M. and ROSSI-HANSBERG, E. (2013). The spatial diffusion of technology, working paper, Princeton University.
- and HOBIJN, B. (2010). An exploration of technology diffusion. *American Economic Review*, **100** (5), 2031–59.
- , — and ROVITO, E. (2006). Five facts you need to know about technology diffusion, nBER Working Paper 11928.
- and MESTIERI, M. (2012). *An Intensive Exploration of Technology Diffusion*. Working paper, Harvard Business School.
- CONLEY, T. G. and UDRY, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review*, **100**(1), 35–69.
- DIECKMANN, U., METZ, J. A. and SABELIS, M. W. (2005). *Adaptive dynamics of infectious diseases: in pursuit of virulence management*, vol. 2. Cambridge University Press.
- DOEPKE, M. and TERTILT, M. (2009). Women’s liberation: What’s in it for men? *Quarterly Journal of Economics*, **124**(4), 1541–1591.
- and ZILIBOTTI, F. (2014). Culture, entrepreneurship, and growth. in *Handbook of Economic Growth*, vol. 2, ed. by P. Aghion and S.N. Durlauf, pp. 1–48. Amsterdam: Elsevier.
- FISCHER, C. S. and SHAVIT, Y. (1995). National differences in network density: Israel and the United States. *Social Networks*, **17**(2), 129–145.
- FOSTER, A. D. and ROSENZWEIG, M. R. (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 BUSS, D. (1993). Pathogen prevalence and human mate preferences. *Ethology and Sociobiology*, **14**(2), 89–96.
- GOENKA, A. and LIU, L. (2015). *Infectious Diseases, Human Capital and Economic Growth*. Working paper, University of Rochester.
- GORODNICHENKO, Y. and ROLAND, G. (2011). *Culture, Institutions and the Wealth of Nations*. Working paper, University of California at Berkeley.
- GRANOVETTER, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, **78** (6), 1360–1380.
- (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, **19**(1), 33–50.
- GREENWOOD, J. and GUNER, N. (2010). Social change. *International Economic Review*, **51**(4), 893–923.
- , SESHADRI, A. and YORUKOGLU, M. (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 SNELL, J. L. (1997). *Introduction to Probability*. New York: Russell Sage, 2nd edn.
- GUDYKUNST, W. B., GAO, G., SCHMIDT, K. L., NISHIDA, T., BOND, M. H., LEUNG, K., WANG, G. and BARRACLOUGH, R. A. (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 JONES, C. I. (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 ZUK, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science*, **218** (4570), 384–387.
- HOFSTEDE, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Thousand Oaks, CA: Sage Publications, 2nd edn.
- JACKSON, M. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press, 1st edn.
- KELLEY, B., LUSTIG, H. and VAN NIEUWERBURGH, S. (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 MOLL, B. (2011). *Knowledge Growth and the Allocation of Time*. NBER Working Paper 17495., National Bureau of Economic Research.
- 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 SCHALLER, M. (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*. Washington, DC: National Geographic Society, 8th edn.
- 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 TONETTI, C. (2011). *Endogenous Risk and Growth*. Working paper, New York University.
- RAUCH, J. E. and CASELLA, A. (2001). *Networks and Markets*. New York: Russell Sage Foundation, 1st edn.
- RODENWALDT, E. and JUSATZ, H. (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 GROSS, N. (1943). The diffusion of hybrid seed corn in two iowa communities. *Rural Sociology*, **8**, 15–24.

- SIMMONS, J. S., WHAYNE, T. F., ANDERSON, G. W. and HORACK, H. M. (1945). *Global Epidemiology: A Geography of Disease and Sanitation*. Philadelphia: Lippincott, 1st edn.
- SMITH, K. F., SAX, D. F., GAINES, S. D., GUERNIER, V. and GUEGAN, J.-F. (2007). Globalization of human infectious disease. *Ecology*, **88(8)**, 1903–1910.
- SPOLAORE, E. and WACZIARG, R. (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., FINCHER, C. L., MURRAY, D. R. and SCHALLER, M. (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 MILGRAM, S. (1969). An experimental study of the small world problem. *Sociometry*, **32(4)**, 425–443.
- WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of small-world networks. *Nature*, **393** (6684), 440–442.
- YOUNG, H. P. (2009). Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, **99(5)**, 1899–1924.

A Proofs of Propositions

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 A new technology shock advances the technological frontier if it arrives to an agent that has a technology level that is as high as any other agent in the network. Suppose that at t , the technology of each agent is the same in both types of networks and agent j (and only him¹⁵) is at the technological frontier. In the next period, with probability $1 - (1 - p)^4$, agent j transmits his technology to at least one of his connections and

¹⁵The reasoning is analogous if more than one agent receives the original shock at the same time.

the expected number of people that have the latest technology in $t + 1$ is $1 + 4p$. That probability is the same in both networks. Each agent has an identical probability λ of inventing a new technology. Thus, the probability that a technology shock hits an agent who has the highest technology level at $t + 1$ and advances the frontier is $(1 + 4p)\lambda$ in either network.

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

Proof of Result 3 We start by deriving the path length in each network and then compare the two.

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

Average Path Length of zero-collectives network. Consider the path length from node n to any other node in the network between 1 and $n/2$. By symmetry, the path length starting from any other node is the same, and the average path length to the nodes between $n/2$ and n 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 $n/2 - m/2$. The number of steps in this path is $\tilde{m} \equiv \text{round}(n/(2\Delta_j))$, where round is the nearest integer value. All points not on this path (interior nodes) can be reached in, at most, $s + \Delta_j/2$ steps.

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

$$\begin{aligned} & \{1, \dots, \frac{\Delta_j}{2}\} \\ & 1 + 2\{1, \dots, \frac{\Delta_j}{2}\} \\ & \vdots \\ & \tilde{m} + 2\{1, \dots, \frac{\Delta_j}{2}\} \end{aligned}$$

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

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

We can use the summation formula to replace the sums:

$$PL \leq \frac{2}{n} \left[\frac{\tilde{m}(\tilde{m} + 1)}{2} + (2\tilde{m} + 1) \frac{\Delta_j/2(\Delta_j/2 + 1)}{2} \right]$$

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

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

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

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

Rearranging, this implies that

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

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

Proof of Result 4 (Higher Degree Speeds Diffusion) Take a network and its matrix of shortest path lengths $\{p_{ij}\}_{i,j=1}^N$. 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.

Proof of Result 5 (Mobility speeds diffusion) Increasing mobility means adding random, long links to the network with a higher probability. A network with additional links always has a weakly shorter path length. Thus, as the probability of long links increases, the probability of shorter path length increases and the expected average path length declines.

Proof of Result 6 (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 Grimstead and Snell (1997).*

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

Now we compute a lower bound for the probability of (i)-(iii) happening at any time. First, assume $\tau_j(t) = co$. Recall that j 's own type governs the links to the right and others' types govern links to the left, so in this case the first-tier connections for which $\eta_{jk} = 1$ are $k = \{j-4, j-1, j+1, j+2\}$. The second-tier connections (nodes connected to j 's connections that are not directly connected to j) are the following: $\{j-8, j-5, j-3, j-2, j+3, j+5, j+6\}$. Therefore, with probability of at least $(1-p_r)p_r^{11}$, node j survives (is not hit by the renewal shock $\xi_{rj}(t)$) and all his first- and second-tier connections are renewed, reaching the absorbing state. Second, if we assume that $\tau_j(t) = in$, then his direct connections are $\eta_{jk} = 1$ for $k = \{j-2, j-1, j+1, j+4\}$ and the second-tier connections are $\{j-3, j+2, j+3, j+5, j+6\}$. Therefore, with probability of at least $(1-p_r)p_r^9$, node j survives without renewal and all his first- and second-tier connections are renewed, reaching the absorbing state.

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 the absorbing state. 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 in some finite number of steps. Since we can reach an absorbing state from any state with positive probability, the result follows from Lemma 1.

Proof of Result 7 (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 $m \in \{1, \dots, n\}$, 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 $j = \text{mod}(i \pm 1, n)$, and = 0 otherwise, where we define $\text{mod}(n, n) = \text{mod}(0, n) = n$.
2. Add individualists and collectivists. If node i is a collectivist, set $N_{ij} = N_{ji} = 1$ for $j = \text{mod}(i + 2, n)$. If node i is an individualist, set $N_{ij} = N_{ji} = 1$ for $j = \text{mod}(i + \Delta_j, n)$.
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 $j = \{\text{mod}(i + 3, n), \dots, \text{mod}(i + n_f/2, n)\}$. Note that n_f must be even.
4. Adjust links for mobility. For every unique link between two nodes $N_{ij} = N_{ji} = 1$, draw a random, binary variable that is 1 with probability m_j . If the realization is 1, break the link: Set $N_{ij} = N_{ji} = 0$. Then, choose any pair of unlinked nodes k, l , with uniform probability, and form a new link between those nodes: Set $N_{kl} = N_{lk} = 1$.

Laws of Motion for State Variables For the endogenous network model, 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 one 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_r . Then the laws of motion for disease, technology, and type are

$$d_j(t+1) = \min(1, d_j(t) + N\xi(t))(1 - \xi_{rj}(t)) \quad (10)$$

$$A_j(t+1) = (1 - d_j(t))(\delta \iota_j(t) + \max(A_j(t), NA(t)\xi_\tau(t))) \quad (11)$$

$$\tau_j(t+1) = (1 - \xi_{rj}(t))\tau_j(t) + \xi_{rj}(t)\hat{\tau}_j(t) \quad (12)$$

where $\hat{\tau}_j(t) = \tau_k : k = \text{argmax}_{k': N_{jk'}=1} A_{k'}(t)$. 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: $\tau_j(t+1) = \tau_j(t)$.

B.1 Robustness Checks

Next, we explore the effect of changing each of the model's parameters in the exogenous network structure. For consistency, we varied each parameter by exploring values half as large and twice as large. Table 6 reports the estimated coefficient on the network index variable in a 2SLS estimation. The estimation performed in the model is the same as that performed on the data in columns (1) and (4) of Table 4. Entries are estimated coefficients from the same second-stage regression run on simulated model data where we exogenously vary the network. The dependent variables are $\text{Log}(Y/L)$ is log output per agent and $g(Y)$, the average rate of technological progress. The instrument is Δgerm , as in (4). This is the same output as reported in Table 5, except with altered model parameters. Standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

When we change parameters, the results change very little. The reason is that 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.

Table 6: Robustness to lower and higher parameter values.

Dependent variable: Independent variable:	Log(Y/L)	Log(gY)	Network Index	Log(Y/L)	Log(gY)
$p_r = 1/140$	0.334*** (0.026)	0.344*** (0.025)	$z = 1/35$	0.395*** (0.029)	0.352*** (0.026)
$\phi = 4.25\%$	0.387*** (0.028)	0.352*** (0.025)	$\phi = 17\%$	0.357*** (0.028)	0.337*** (0.026)
$\pi = 0.25\%$	0.387*** (0.028)	0.347*** (0.026)	$\pi = 1\%$	0.339*** (0.028)	0.345*** (0.026)
$\bar{d} = 0.015\%$	0.374*** (0.028)	0.345*** (0.026)	$\bar{d} = 0.06\%$	0.376*** (0.028)	0.346*** (0.026)

Table 7: Model with Endogenous Innovation and Infection Feedbacks.

Feedback:	None	Infection	Both	Both
Initial disease prevalence	18%	18 %	0.05%	39%
Avg Fraction Individualists	33%	37 %	47%	29%
Years to Disease Extinction	431	106	111	435
Avg. Growth Rate	2.0%	2.7%	2.6%	1.2%
Adoption Lag	24	22	23	56
Fraction of Extinct Collectivist Networks	30%	29%	32%	27%
Fraction of Pure Collectivist Networks	62%	53%	35%	67%
Life expectancy (years)	49	53	58	46

B.2 Disease-Technology Feedback Results

Moments of the calibrated model with the infection rate feedback and with both the infection and innovation rate feedbacks are reported in Table 7. The first column reports our calibrated benchmark model, with no feedbacks, but with a level of disease halfway between the high and low rates we considered in the main text.

In the original benchmark model, there was only a four-year difference between the adoption lag in a purely collectivist economy and one with no collectivists at all. In the data, countries with above-median collectivism (below-median Hofstede individualism indices) have adoption lags that are 34 years longer on average. The model with both feedbacks can match this 34-year difference. In the model, the adoption lag is the time in between the arrivals of innovations in the technology frontier. That time is 56 years in the high-disease collectivist economy, 23 in the low-disease individualist economy.

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 (9) 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 are $[\cdot 004 : \cdot 004 : \cdot 02]$. 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 disease does 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 (Table 8). 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

Table 8: The effect of socially transmittable and other disease on GDP growth in the model. The Table reports $100 \times \beta_2$ coefficient in the regression $\bar{gA} = \beta_1 + \beta_2 x + \epsilon$, where \bar{gA} is the average growth rate of output and x is the average prevalence (fraction of the population infected at a given time) of either transmissible disease \bar{d}_0 or other (zoonotic) disease p_r .

Dependent variable: Productivity growth (\bar{gA})		
Transmissible disease	\bar{d}_0	-1.95 (0.51)
Zoonotic disease	p_r	-6.33 (0.70)
Difference	$\bar{d}_0 - p_r$	-0.94 (0.55)

Table 9: Summary Statistics

	count	mean	std	min	max
Zoonotic disease	69	1.91	1.02	0	3.95
Human disease	69	1.96	0.998	0.191	3.63
Disease difference	69	0.0496	0.990	-1.95	2.68
Disease sum	69	3.87	1.76	0.191	6.93
Network index	69	-3.93e-09	1.45	-2.23	3.10
Individualism	69	42.59	23.52	6	91
Degree	69	2.58	0.302	1.91	3.4
Migration	69	0.150	0.0609	0	0.333
English	72	0.075	0.238	0.000	0.974

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.

C Data Appendix

The 69 countries in our sample are Argentina, Australia, Austria, Bangladesh, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Czech Republic, Denmark, Ecuador, Egypt, El Salvador, Estonia, Ethiopia, Finland, France, Germany, Ghana, Greece, Guatemala, Hungary, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Japan, Kenya, S. 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, 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 9.

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. Table 14 summarizes the properties and standard epidemiological

classification of all the pathogens for which we collected data.

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 in another person. But it is in the human, not in the mosquito, where the disease flourishes. Human-specific diseases in our data set include diphtheria, filaria, measles, and smallpox. The variable d_{hs} is defined as $d_{hs} \equiv \sum_{l \in \mathcal{HS}} \text{prevalence}_l$, where l is a disease and \mathcal{HS} is the set of all human-specific diseases.

Zoonotic p_r 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 $d_z \equiv \sum_{l \in \mathcal{Z}} \text{prevalence}_l$, 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 $d_m \equiv \sum_{l \in \mathcal{M}} \text{prevalence}_l$, 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_r 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 of 2011 data. This high correlation explains why the results with contemporaneous data are nearly identical.

C.2 Measuring Individualism

Hofstede (2001) defines individualism in the following way:

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

The Hofstede individualism index values are based on the results of a factor analysis of work goals across countries. The index was constructed from data collected during an employee attitude survey program conducted by a large multinational organization (IBM) within its subsidiaries in 72 countries. The survey took place in two waves, in 1969 and 1972, and included questions about demographics, satisfaction, and work goals. The answers to the 14 questions about work goals form the basis for the construction of the individualism index. The individual answers were aggregated at the country level after matching respondents by occupation, age, and gender. The countries’ mean scores for the 14 work goals were then analyzed using factor analysis that resulted in the identification of two factors of equal strength that together explained 46% of the variance. The individualism factor is mapped onto a

scale from 1 to 100 to create the individualism index (hereafter IDV) for each country. The highest IDV values are for the United States (91), Australia (90), and Great Britain (89); the lowest are for Guatemala (6), Ecuador (8), and Panama (11). Subsequent studies involving commercial airline pilots and students (23 countries), civil service managers (14 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. 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 10. 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 10: Measures of network interdependence and individualism

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

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

Correlation of Individualism with Other Measures of Culture To better understand what Hofstede’s individualism index (IDV) measures, we examine related cultural measures that are highly correlated with the index.

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

All the respondents are U.S. 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 U.S. 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 Link Stability

Using the hypothesis that people break social network ties when they move from one community to another, we construct the following proxy for social link stability. The data on link stability 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 U.S. 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 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 individualism is highly correlated with and thus proxying for some other economic phenomenon. The problem with answering this question is that what we would like to control for is likely an endogenous variable. We could treat it as such and instrument for it. But in most cases, our instruments are not strong predictors. Or we could just suspend disbelief, assume that these are exogenous variables, abandon any pretense of saying anything about causality, and just see what statistical relationship they have with the other variables in the estimation. We take the second approach. Each row in Tables 11 and 12 reports the coefficients of a first- or second-stage regression of income per capita or technology diffusion 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 11 and 12 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 3. The dependent variable in the second stage of Table 11 is real output per capita from the Penn World Tables, mark 5.6.

The control variables are life expectancy at birth (LifeExp), disability-adjusted life expectancy (Disability), which is the expected length of time an individual lives free of disability, as measured by the World Health Organization in 2004 (http://www.who.int/healthinfo/global_burden_disease/estimates_country/en/index.html); social infrastructure (SocInf) and executive power (Executive), measures of the efficient functioning of political and social 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 (2011); 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.

The variables with Sargan at the end of their name are not included as controls, but are used for a test of exogeneity of our disease instrument. For these Sargan tests, we also use the variable English, which is the fraction of the population that is English-speaking.

Finally, we reestimate the main results with continent fixed effects in Table 13. For each specification, we find that the effect of the network index on technology diffusion and output per capita is at least as strong as it is without these fixed effects.

Table 11: Estimating the effect of networks on income (Y/L), while controlling for other economic variables

CONTROL VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	LifeExp	LifeExp_Sargan	Disability	Disabl_Sargan	SocInf	SocInf_Sargan	Blood	Blood_Sargan	Executive	Executive_Sargan	Inflation	Inflation_Sargan	Openness	Openness_Sargan
First Stage Dependent Variable: Network Index														
Δ germ	-0.429*** (0.126)	-0.448*** (0.124)	-0.482*** (0.125)	-0.499*** (0.123)	-0.447*** (0.126)	-0.463*** (0.125)	-0.441*** (0.125)	-0.464*** (0.124)	-0.495*** (0.128)	-0.509*** (0.126)	-0.473*** (0.128)	-0.487*** (0.127)	-0.493*** (0.129)	-0.507*** (0.127)
sumgerm	-0.446*** (0.104)	-0.388*** (0.106)	-0.557*** (0.0856)	-0.495*** (0.0896)	-0.415*** (0.0884)	-0.378*** (0.0908)	-0.504*** (0.0746)	-0.460*** (0.0776)	-0.522*** (0.0867)	-0.469*** (0.0900)	-0.526*** (0.0753)	-0.479*** (0.0802)	-0.494*** (0.0834)	-0.445*** (0.0872)
LifeExp	0.0275 (0.0187)	0.0275 (0.0184)												
English		1.033* (0.537)		1.057* (0.545)		0.830 (0.550)		0.945* (0.548)		1.011* (0.554)		0.872 (0.564)		0.946* (0.564)
Disability			2.89e-06 (1.21e-05)	2.00e-06 (1.19e-05)										
SocInf					1.564** (0.616)	1.437** (0.616)								
Blood							-5.277 (3.532)	-4.196 (3.535)						
Executive									0.0526 (0.0798)	0.0456 (0.0784)				
Inflation											-0.201* (0.109)	-0.176 (0.108)		
Openness													0.600 (0.427)	0.541 (0.422)
Constant	0.0182 (1.494)	-0.283 (1.472)	2.122*** (0.301)	1.821*** (0.333)	0.814 (0.620)	0.673 (0.621)	2.392*** (0.341)	2.067*** (0.385)	1.768** (0.690)	1.521** (0.691)	2.630*** (0.386)	2.311*** (0.433)	1.729*** (0.452)	1.485*** (0.469)
Second Stage Dependent Variable: Output per Capita														
Network Index	0.592** (0.237)	0.565*** (0.205)	0.606*** (0.206)	0.596*** (0.184)	0.463** (0.214)	0.410** (0.192)	0.654** (0.284)	0.583** (0.250)	0.721*** (0.256)	0.669*** (0.228)	0.651** (0.270)	0.627** (0.247)	0.621*** (0.212)	0.574*** (0.190)
sumgerm	0.0527 (0.129)	0.0414 (0.118)	-0.0864 (0.129)	-0.0916 (0.119)	-0.147 (0.105)	-0.167* (0.0978)	-0.257* (0.155)	-0.291** (0.140)	-0.165 (0.160)	-0.193 (0.147)	-0.267* (0.158)	-0.279* (0.147)	-0.148 (0.123)	-0.171 (0.113)
LifeExp	0.0906*** (0.0173)	0.0916*** (0.0165)												
Disability			-5.87e-05*** (9.57e-06)	-5.87e-05*** (9.53e-06)										
SocInf					2.860*** (0.616)	2.962*** (0.581)								
Blood							-4.965 (4.105)	-5.522 (3.902)						
Executive									0.153* (0.0793)	0.155** (0.0777)				
Inflation											-0.0528 (0.125)	-0.0584 (0.121)		
Openness													1.700*** (0.372)	1.731*** (0.359)
Constant	2.830** (1.207)	2.811** (1.187)	10.11*** (0.492)	10.13*** (0.452)	7.632*** (0.473)	7.656*** (0.460)	10.05*** (0.781)	10.23*** (0.702)	8.443*** (0.834)	8.542*** (0.794)	9.808*** (0.809)	9.871*** (0.756)	8.495*** (0.503)	8.572*** (0.473)
Observations	68	68	69	69	63	63	69	69	67	67	63	63	63	63
R^2 stage 1	0.554	0.579	0.532	0.558	0.581	0.597	0.547	0.567	0.537	0.561	0.562	0.579	0.551	0.572
R^2 stage 2	0.732	0.739	0.733	0.735	0.768	0.779	0.588	0.606	0.579	0.595	0.575	0.582	0.715	0.728
Sargan-p		0.809		0.914		0.510		0.556		0.616		0.818		0.563

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

Table 12: Estimating the effect of networks on technology diffusion (g_Y), while controlling for other economic variables

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	LifeExp	LifeExp_Sargan	daly2004	daly2004_Sargan	SocInf	SocInf_Sargan	blood	blood_Sargan	Executive	Executive_Sargan	Inflation	Inflation_Sargan	Openness	Openness_Sargan
First Stage Dependent Variable: Network Index														
Δ germ	-0.429*** (0.126)	-0.448*** (0.124)	-0.482*** (0.125)	-0.499*** (0.123)	-0.447*** (0.126)	-0.463*** (0.125)	-0.441*** (0.125)	-0.464*** (0.124)	-0.495*** (0.128)	-0.509*** (0.126)	-0.473*** (0.128)	-0.487*** (0.127)	-0.493*** (0.129)	-0.507*** (0.127)
sumgerm	-0.446*** (0.104)	-0.388*** (0.106)	-0.557*** (0.0856)	-0.495*** (0.0896)	-0.415*** (0.0884)	-0.378*** (0.0908)	-0.504*** (0.0746)	-0.460*** (0.0776)	-0.522*** (0.0867)	-0.469*** (0.0900)	-0.526*** (0.0753)	-0.479*** (0.0802)	-0.494*** (0.0834)	-0.445*** (0.0872)
LifeExp	0.0275 (0.0187)	0.0275 (0.0184)												
English		1.033* (0.537)		1.057* (0.545)		0.830 (0.550)		0.945* (0.548)		1.011* (0.554)		0.872 (0.564)		0.946* (0.564)
Disability			2.89e-06 (1.21e-05)	2.00e-06 (1.19e-05)										
SocInf					1.564** (0.616)	1.437** (0.616)								
Blood							-5.277 (3.532)	-4.196 (3.535)						
Executive									0.0526 (0.0798)	0.0456 (0.0784)				
Inflation											-0.201* (0.109)	-0.176 (0.108)		
Openness													0.600 (0.427)	0.541 (0.422)
Constant	0.0182 (1.494)	-0.283 (1.472)	2.122*** (0.301)	1.821*** (0.333)	0.814 (0.620)	0.673 (0.621)	2.392*** (0.341)	2.067*** (0.385)	1.768** (0.690)	1.521** (0.691)	2.630*** (0.386)	2.311*** (0.433)	1.729*** (0.452)	1.485*** (0.469)
Second Stage Dependent Variable: Tech Diffusion														
Network Index	0.342** (0.153)	0.318** (0.130)	0.318** (0.124)	0.307*** (0.110)	0.255* (0.133)	0.254** (0.122)	0.336** (0.152)	0.304** (0.132)	0.334** (0.133)	0.311*** (0.118)	0.316** (0.137)	0.310** (0.126)	0.275** (0.120)	0.273** (0.110)
sumgerm	0.0546 (0.0832)	0.0446 (0.0752)	0.0249 (0.0777)	0.0191 (0.0711)	-0.0544 (0.0653)	-0.0550 (0.0622)	-0.0331 (0.0826)	-0.0486 (0.0738)	-0.0427 (0.0836)	-0.0548 (0.0762)	-0.0693 (0.0803)	-0.0725 (0.0751)	-0.0463 (0.0697)	-0.0476 (0.0656)
LifeExp	0.0219* (0.0112)	0.0227** (0.0105)												
Disability			-2.00e-05*** (5.78e-06)	-2.00e-05*** (5.71e-06)										
SocInf					0.398 (0.383)	0.401 (0.369)								
Blood							-1.535 (2.194)	-1.787 (2.061)						
Executive									0.0197 (0.0414)	0.0204 (0.0403)				
Inflation											0.0434 (0.0636)	0.0420 (0.0619)		
Openness													0.336 (0.211)	0.338 (0.208)
Constant	-1.613** (0.779)	-1.630** (0.755)	0.234 (0.297)	0.257 (0.270)	-0.107 (0.294)	-0.106 (0.292)	0.201 (0.417)	0.281 (0.371)	-0.00492 (0.435)	0.0378 (0.412)	0.0640 (0.412)	0.0801 (0.386)	-0.0586 (0.285)	-0.0541 (0.274)
Observations	68	68	69	69	63	63	69	69	67	67	63	63	63	63
R^2 stage 1	0.554	0.579	0.532	0.558	0.581	0.597	0.547	0.567	0.537	0.561	0.562	0.579	0.551	0.572
R^2 stage 2	0.291	0.331	0.380	0.395	0.401	0.403	0.253	0.302	0.268	0.303	0.300	0.309	0.387	0.390
Sargan-p		0.737		0.841		0.977		0.613		0.676		0.907		0.953

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 13: Social network effects with continent fixed effects

	Tech diffusion (A)		Output ($\log(Y/L)$)	
	(1)	(2)	(3)	(4)
Network Index	0.609** (0.283)	0.431** (0.157)	1.277** (0.519)	1.055** (0.308)
Sum germs	0.104 (0.158)		0.195 (0.290)	
Constant	-0.796	-0.234	7.807**	8.559**
Observations	69	69	69	69
R^2	0.063	0.316	0.505	0.587

Columns report β_2 coefficient from an IV estimation of continent fixed effects and the left-column variables on technology diffusion or output per capita. The technology diffusion rate (A) comes from the Comin and Mestieri (2012) measure of the intensive technology adoption in a country. Network Index is an index combining three measures of social network structure described in Section 3.2. The instruments are $\Delta germ$ s as defined in equation (4), sumgerm, English, and continent fixed effects. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C.6 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. One variable that is surely not driven by GDP is latitude. The three-stage least squares procedure that uses latitude as an instrument for disease, which is an instrument for technology diffusion, should allay concerns about GDP being a causal variable that explains all of the results. It supports the statistically significant and economically large effects of network features on technology diffusion.

Technology explains differences in disease prevalence. Our empirical strategy is based on the assumptions that human-transmitted disease \bar{d} and zoonotic disease p_r 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 (7). 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. Here, the results of the three-stage least squares estimation are helpful. That procedure acknowledges that the difference in disease may not be an exogenous instrument, but that latitude surely is. When we use latitude as the exogenous instrument, we find 3SLS results that are quite similar in magnitude and even more statistically significant than the original estimates.

The difference in disease has a direct effect on diffusion. If human-to-human diseases are more deadly than zoonotic ones, 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. The second row of each of the four panels in Table 11 demonstrates that individualism, fractionalization, and network stability all survive statistically and have economically similar effects, even after controlling for disease-adjusted life expectancy. Degree is similar in magnitude, but no longer statistically significant, in part, because there are only 26 observations for that estimation. Finally, the three-stage procedure where we use latitude as an instrument for disease difference, which in turn is an instrument for networks, should also avoid any endogeneity concerns.

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 (7). 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.

Table 14: Classification of diseases

Disease	Agent	Reservoir	Spread By	Type	Historical
Diphtheria	Bacteria	Man	Droplet, Contact, Dairy, Clothing	H	N
Filaria - Bancroftian	Nematoda	Man	Mosquito	H	Y
Filaria - Brugia Timori	Nematoda	Man	Mosquito	H	Y
Measles	Virus - RNA	Man	Droplet	H	N
Meningitis - Bacterial	Bacteria	Man	Air, Secretions	H	N
Meningitis - Viral	Virus - RNA	Man	Fecal-oral, Droplets	H	N
Pertussis	Bacteria	Man	Air, Secretions	H	N
Poliovmyelitis	Virus - RNA	Man	Fecal-oral, Food, Water, Flies	H	N
Smallpox	Virus - DNA	Man	Contact, Secretions, Fomite	H	N
Syphilis	Bacteria	Man	Sexual Contact, Secretions	H	N
Typhoid fever	Bacteria	Man	Fecal-oral, Food, Flies, Water	H	N
Dengue	Virus - RNA	Man, Monkey, Mosquito	Mosquito, Blood (rare)	M	Y
Filaria - Brugia Malayi	Nematoda	Man, Primate, Cat, Civet	Mosquito	M	Y
Leishmania - Cutaneous	Protozoa	Man, Rodent, Other Mammals	Fly	M	Y
Leishmania - Mucocutaneous	Protozoa	Man, Rodent, Sloth, Marsupial	Fly	M	Y
Leishmania - Visceral	Protozoa	Man, Rodent, Dog, Fox	Fly, Blood	M	Y
Leprosy	Bacteria	Man, Armadillo	Patient Secretions	M	Y
Malaria	Protozoa	Man, Mosquito	Mosquito, Blood	M	Y
Trypanosoma - African	Protozoa	Man, Deer, Cattle, Carnivores	Fly	M	Y
Trypanosoma - American	Protozoa	Man, Dog, Cat, Other Mammals	Kissing Bug, Blood, Fruit	M	Y
Tuberculosis	Bacteria	Man, Cattle	Air, Dairy Products	M	Y
Typhus - Epidemic	Bacteria	Man, Flying Squirrel	Louse	M	Y
Anthrax	Bacteria	Soil, Water, Other Mammals	Fly, Hair, Hides, Bone, Air, Meat	Z	N
Leptospirosis	Bacteria	Frog, Cattle, Other Mammals	Water, Soil, Urine, Contact	Z	N
Rabies	Virus - RNA	Dog, Fox, Other Mammals	Saliva, Bite, Transplants, Air	Z	N
Schistosomiasis - Haematobium	Flatworms	Snail, Baboon, Monkey	Water (Skin Contact)	Z	Y
Schistosomiasis - Intercalatum	Flatworms	Snail	Water (Skin Contact)	Z	Y
Schistosomiasis - Japonicum	Flatworms	Snail, Other Mammals	Water (Skin Contact)	Z	Y
Schistosomiasis - Mansoni	Flatworms	Snail, Other Mammals	Water (Skin Contact)	Z	Y
Schistosomiasis - Matthei	Flatworms	Snail, Other Mammals	Water (Skin Contact)	Z	Y
Schistosomiasis - Mekongi	Flatworms	Snail, Dog	Water	Z	Y
Tetanus	Bacteria	Animal Feces, Soil	Injury	Z	N
Typhus - Endemic	Bacteria	Rat	Flea	Z	Y
Typhus - Scrub	Bacteria	Rodent, Carnivores, Mite	Mite	Z	Y

Disease type is *H* if we classify the disease as human, *M* for multitype, and *Z* for zoonotic. The classification is based on the disease reservoir. Historical is *Y* if the disease is present in our 1930s historical data. For disease categories with multiple sub-types (i.e., Filaria - Bancroftian and Filaria - Brugia Timori), there is one combined value in the historical data. That value is the maximum of the value for each strain. It lists 4 if any strain is epidemic, 3 if any strain is endemic, and so on. All listed diseases are present in the 2011 data. Different strains are treated like different diseases. All 2011 results are robust to combining strains (results available on request). Source: GIDEON database.