Connectivity:
A Key Factor in International Development

John Gerring

Department of Political Science
Boston University
jgerring@bu.edu

Draft: Please do not cite without permission
At the turn of the twenty-first century, the quality of life enjoyed by persons around the world varies enormously. While there may be a slow convergence in human development (Gerring et al. 2008; Goesling, Firebaugh 2000; Neumayer 2003), it remains the case that one’s chances of surviving past infancy, maintaining good health, getting educated, finding a job, and escaping poverty are largely contingent upon what part of the planet one happens to be born in.

The markedly spatial pattern of development, both within and across countries, suggests three broad lines of explanation. It could be that areas with high and low levels of development are intrinsically different -- ecologically, economically, culturally, sociologically, politically, or historically -- in ways that have enhanced/diminished their development potential. London (England) is more developed than Bangui (Central African Republic) because London is different from Bangui. There are many versions of this causal story, which may rest on the intrinsic properties of a land and its climate (Hibbs, Olsson, Diamond 2004; Landes 1999), population growth in tandem with technological innovation (Boserup 1965; Galor, Weil 1999; Goodfriend, McDermott 1995; Jones 2001; Kremer 1993), public policies and political institutions (Acemoglu, Johnson, Robinson 2004; Glaeser, Lopez-de-Silanes, Shleifer 2004), cultural properties (Iyigun 2007; Landes 2006), and historical contingencies (Pomeranz 2000).

A second possibility is that development is an essentially zero-sum distributional game in which regions of the world compete. Developed regions are therefore those possessing superior military and geo-political capacities, allowing them to monopolize technology, physical capital, human capital, and other factors that contribute to development. London is more developed than Bangui because London (along with London’s allies) is militarily and organizationally dominant. The strongest and most systematic version of this story is associated with the dependency and world-systems school of thought (Benton 1996; Bergeson 1980; Prebisch 1950; Snyder, Kick 1979; Van Rossem 1996; Wallerstein 1974).

A final possibility is that development is the product of a region’s “connectivity.” London is more developed than Bangui because London has greater access to the world. This, in turn, is a product of its favorable location, transport infrastructure, and the location of adjacent populations. Connectivity is thus understood as a structural feature of the landscape -- a product of geographic, infrastructural, and demographic features that make it easy or difficult for individuals living in a region to reach those outside the boundaries of that region. (A region may be any size.) We expect that relative ease of movement will facilitate contact and that contact will, in turn, hasten development. To motivate the analysis, let us consider briefly several reasons why this might be so.

First, connectivity facilitates the diffusion of technology, ideas, and norms. Second, connectivity fosters the standardization of technology, ideas, norms, formal regulations, and language across regions and this by itself (regardless of the quality of these factors) enhances efficiencies and presumably also speeds the ongoing process of diffusion, since ideas travel more quickly where cultural barriers are less severe. Third, connectivity reduces transaction costs, lowering the price and increasing the variety and availability of consumer goods and the potential markets for producer goods. Fourth, connectivity enhances the mobility of labor and capital, thus reducing inequalities -- locally, nationally, and internationally (wherever connectivities exist) -- according to Ricardian theory. Fifth, connectivity probably strengthens the capacity of governments. It is easier for states to perform basic functions like keeping order, collecting taxes, and surveynance, as well as more complex policy interventions, when the population under its purview is well-connected. (We assume that state capacity generally has a positive relationship to social and economic development.) Sixth, connectivity enhances the access of non-market, non-governmental actors whose involvement in the project of development is essentially voluntaristic (or at least nominally so). Insofar as missionaries, health workers, educators, and other actors rooted in civil society have played a
positive role in development, this too may be related to the relative connectivity of different regions. At the same time, connectivity reduces the power of governments, religions, businesses, and other organizations to restrict the flow of ideas and technologies. In this respect, connectivity is inherently anti-monopolistic, fostering a spirit of open competition across jurisdictions that are in contact with one another. Finally, connectivity may nurture a more universalistic set of values and norms, which may in turn play a positive role in the complex and intertwined processes of development. Regions with high connectivity breed cosmopolitan citizens of the world, while regions with low connectivity encourage a parochial outlook on life, love, and duty.

Several important caveats must be attached to these optimistic prognostications. First, it should be understood that these are not invariant laws. We readily acknowledge situations in which high connectivity has served as a detriment to short-term -- and perhaps even long-term -- development. Access to coastal areas in Africa facilitated the slave trade (Nunn, Puga 2007); access to European explorers facilitated the spread of epidemics in the New World (Diamond 1997); and access to the outside world today facilitates the spread of HIV/AIDS. These are important exceptions to our argument.

Second, to the extent that our argument holds, it should not be interpreted as a monocausal view of historical development. Connectivity matters, all else equal. But all else is not equal in the real world, and many additional factors presumably impinge upon long-term patterns of development. Our ambition is therefore limited to explaining some portion of the variance in this complex set of outcomes. It should be noted that we do not test the impact of connectivity relative to many other factors such as population, culture, institutions, policies, and technology, which are viewed as endogenous to the theory.

Third, it must be appreciated that many of the benefits that arising from connectivity do not manifest themselves straightaway. Indeed, the immediate effect of enhanced connectivity is often to exacerbate latent social conflicts. (One thinks of the effect of China’s recently completed rail connection to Tibet, which seems to have served as a stimulus for protest.) Over time, however, we assume that the effects of connectivity will generally be positive – at least, for those who survive the often-violent nature of initial contact.

Finally, it must be stressed that our use of the concept of development is limited to the material dimensions of human life -- e.g., health, mortality, education, housing, employment, income, and assets -- and does not encompass the spiritual or emotional quality of life, happiness, liberty, empowerment, cultural integrity, or the quality of the environment.


The theory departs from extant work in three principal respects. First, we adopt a broad conception of development, extending well beyond the usual economic outcomes (usually measured
by GDP per capita or trade flows). Indeed, it seems probable that the causal effect of connectivity may be greater on human development than on economic development, at least in the contemporary era. Second, the theory presents a broad theoretical framework, encompassing elements from many traditions, including those listed above. It thereby articulates a more complete -- and arguably more compelling -- account of the role of cross-district contact in creating and sustaining international development. Finally, we hope to present a more elaborate and robust set of empirical tests featuring a wide variety of connectivity measures (focused on local, national, and global connectivities and on diverse transport infrastructures), both short- and long-term causal effects, individual and country-level outcomes, and multi-level modeling techniques.

Measuring Connectivity

The initial unit of analysis envisioned in this study is the district (to be defined, but perhaps about twenty-five square miles). For each district on earth, we intend to calculate various connectivity measures by looking at the time that would be required to travel from one district to various destinations outside that district, using modes of transport usual to that time and place. Travel-time to various destinations thus comprises the basic metric by which connectivity will be measured. Destinations of concern are divided into five categories: transport routes, urbanization measures, country measures, sub-continental measures, and global measures.

Transport routes include: (A) the nearest ocean or navigable river and (B) the nearest trunk road (considered separately). Thus, for a given district, its connectivity scores are equivalent to the travel-time required to reach (A) or (B). We count only ground transportation infrastructure. (Air transport becomes a factor in development only in the late twentieth century, where it is more or less ubiquitous with cities, and thus captured in the following set of indicators.)

Urbanization measures include: (C) the nearest city, (D) the nearest city in the country, (E) the nearest city in the region, and (F) all global cities (mean). Again, it is the travel-time to these destinations that determines a district’s connectivity score. (The definition of a “city” will have to worked out; it will presumably depend upon population, though the threshold may vary across different historical epochs.)

Country measures include: (G) all country districts (mean) and (H) all country districts weighted by their respective population (so that connections with densely settled districts are given greater weight). Countries are defined according to contemporary definitions and boundaries, and include both sovereign and semisovereign states (e.g., Puerto Rico). (At a later date, we may choose to provide historical codings of political units.)

Sub-continental measures are identical to country measures except that the territory is defined in purely geographic terms. A sub-continent is understood to include all land areas lying within 3,000 miles of the district being coded. Sub-continental measures are thus devised as: (I) all districts within a sub-continent (mean) and (J) all districts within a sub-continent weighted by their respective populations.

Global measures are similar except that in this instance no boundaries are placed upon the analysis. Thus, global measures are devised from: (K) all global districts (mean) and (L) all global

---

1 While several similar studies have been conducted within countries or across countries, we know of no studies that combine these two levels of analysis.
2 One might also construct a measure of connectedness based on the cost of transport, and this could be further disaggregated into cost as a percent of GDP in a given country -- thus capturing the fact that it is easier for Americans to travel to Kenya than for Kenyans to travel to the US. However, this requires data that are virtually impossible to gather for eras prior to the late twentieth century. Even so, in a purely cross-sectional analysis (focused on 2000) it might be worthwhile.
districts weighted by their respective populations. In addition, we construct a final variable (M) that is identical to L except that the weighting principle is assigned by the connectivity of each destination. This captures the intuition that a connection to a highly-connectivity district is more valuable than a connection to a low-connectivity district. It might also be thought of as second-order connectivity. (In principle, the same weighting could be assigned to connectivity at a country and sub-continental levels; however, we do not anticipate that these would greatly alter a district’s score.)

It will be seen that these thirteen connectivity variables, all of which are listed in Table 1, capture different aspects of the general theory. Each is therefore regarded as a sub-hypothesis, to be independently (and perhaps jointly) tested. We anticipate that this initial list of variables will diminish over time. It seems likely, for example, that some of these variables will be highly correlated, and therefore empirically redundant. Even so, a great deal may be gained by looking at small variations in empirical performance across measures that are constructed in different ways; this provides a crucial clue into causal mechanisms that may be at work. We need to know not only whether connectivity matters for development, but also what kinds of connectivity matter most.
Table 1:
Connectivity Variables: A Preliminary List

<table>
<thead>
<tr>
<th>Travel-time to...</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>District</strong></td>
<td>Country (district mean)</td>
<td>Country (pop-weighted district mean)</td>
<td>Individual</td>
<td></td>
</tr>
<tr>
<td><strong>Transport route measures:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urbanization measures:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Country measures:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sub-continental measures:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Global measures:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Variables may be observed at four historical points in time: a) 300 BC, b) 1820, c) 1950, and d) 2000. These are understood as approximate dates, data to be drawn from the nearest available year. Since data is not available for all periods, and since the relevance of some factors (e.g., “country”) is primarily contemporary, not all variables will be coded in all periods, as signaled in Table 1.

Information necessary to code districts during the four periods may be understood as geographic (the topography of the land, the existence of navigable rivers and oceans), infrastructural (transport networks), or demographic (the location and density of human population). Because geographic factors are relatively constant through time (with the exception of important changes like the construction of canals), codings are easily extended back in time. Because we know something about the changing shape and location of transport infrastructure, this too should be amenable to historical coding, as discussed below. The density of populations across the globe is known with a high degree of accuracy in the latter twentieth century, and historical estimates are available for 1820. Prior to that, populations were fairly diffuse everywhere; hence, we do not bother to measure the distribution of populations in 300 BC.

300 BC is chosen as a point of departure because it marks a time prior to the invention of an important transportation technology, the triangular “lateen” sail. This allowed ships to tack (sail into the wind) and thus vastly enhanced the range of sailing vessels, leading to cross-ocean voyages. Up until this signal invention (estimated to be about 200 BC), differences in relative connectivity around the world were largely the product of natural features of the topography.3 Travel-times over land routes are estimated at walking speed for areas that are tractable and at half-walking speed for areas that are “rugged” (Nunn, Puga 2006). Large bodies of water are assumed to pose insuperable transport barriers, and thus are entirely unconnected.

The second point of measurement occurs just prior to another transport revolution -- the advent of the railroad and steamship – which greatly diminished transport costs within and across countries (Bairoch 1990: 142; Crafts, Venables 2001; Harley 1988). In 1820, travel-time over water (navigable rivers, lakes, and oceans) will be calculated according to the average speed of sailing ships at that time. Travel-times over land routes will again be estimated at walking speed for areas that are tractable and at half-walking speed for rugged areas.

In 1950 (arguably, another watershed moment in transport technology [Dollar 2001]), travel-times over water (river, lake, or ocean) will be calculated according to the average ocean shipping speed at the time. Travel-times over land will be calculated using maps that locate the existence of railroads and paved roads and estimates of the typical rate of transport over these facilities at that time. Where no transport infrastructure is in place, travel-time will be calculated at walking speed for areas that are tractable and half-walking speed for rugged areas. For the contemporary era (2000), the same procedure is followed, with adjustments for changes in infrastructure and transport technology. (We may also draw upon the infrastructure dataset developed by Canning [1998] for codings in 1950 and 2000.)

Thus, thirteen variables (A-M) will be measured for all districts on earth (except Antarctica) at several points in time (column 1, Table 1). Given the relatively small number of decision-rules noted above, GIS programming should be fairly simple.

Once coded, these thirteen district-based variables can be transformed easily into connectivity scores for larger geographic units such as the nation-state, so as to allow for country-level analysis. Two aggregation techniques seem appropriate. First, we average the values for all

---

3 To be sure, the presence (absence) of domesticable pack animals enhanced the carrying capacity of traders and warriors in areas where such animals were present (Diamond 1997). At some point, we may choose to incorporate this sort of information into our analysis.
districts lying within a contemporary nation-state (column 2, Table 1).

Our focus is on developments over the course of the twentieth century, we expect that the effect of transport will be more important – or at any rate, be highly correlated with – communications networks. Second, transportation architecture is easier to measure; we know the general location of major roads, navigable rivers, and deep-water harbors. And because we know the density of population settlement relative to these features, we can intuit the strength of physical connectivity. Measuring the density of communications technology is much more difficult in the recent era (and, as we have said, much less important in previous eras, since the latter depended upon physical connectivity).

**Testing the Causal Effects of Connectivity**

The effect of connectivity on development will be tested at individual and country levels. Where data is sufficient, these two levels will be combined with the use of hierarchical (multi-level) models. Development outcomes of concern include quality-of-life measures (e.g., life expectancy, infant mortality, maternal mortality, malnutrition), fertility (a key element of the demographic transition), and income (as measured by GDP per capita for countries and assets or consumption for individuals). Each of these outcomes may be independently analyzed. In addition, we propose to subject all outcomes at the individual and all outcomes at the country level to factor analysis so as to identify a few dimensions that may be said to represent in summary form the complex process of development. Thus, the left-side variable in the following analyses may consist of a single outcome measure or a composite indicator of development.

Country-level data is usually available on an annual or quasi-annual basis throughout the late twentieth century (World Development Indicators 2007), allowing for both cross-sectional and panel analyses from 1950 to 2004. Note that the theory supposes that connectivity will affect both long-term as well as short-term outcomes, i.e., the relative status of countries at a particular point in time as well as the rate of improvement in development indicators over some period of time such as the last half-century.

Individual-level data will be drawn from merged DHS surveys, along with additional surveys and census data wherever DHS surveys are unavailable. These surveys provide a highly detailed

---

4 We may also consider analyzing several measures of political development at the national level, e.g., the six World Bank indicators compiled by Daniel Kaufmann and colleagues.
picture of development across the contemporary world. However, there is probably not sufficient historical data to conduct meaningful time-series analysis, given the sluggishness of our measures of connectivity. Hierarchical (multi-level) modeling techniques will be employed in these cross-sectional analyses so as to model individual and national-level phenomena simultaneously.

In both country- and individual-level analyses it is important to differentiate between the effects of connectivity on economic outcomes (as measured by GDP per capita at the national level and personal assets at the individual level) and quality-of-life outcomes (health, mortality, education, and so forth). Since these different species of development are mutually constitutive, and not easily modeled, this is no mean feat. One approach is to run identically specified regression models with alternative outcomes, comparing model-fit across the models as a clue to the varying strength of the underlying relationships. Another approach is to include economic measures as controls in the analysis of noneconomic outcomes, and vice-versa, thereby partialing out their effects. Does connectivity affect quality of life when controlling for income or personal assets? Does connectivity affect income or assets when controlling for quality of life?

A larger problem encountered in all of these regression models is known generically as an assignment (or selection) problem. In the absence of random assignment, it is difficult to ascertain whether regressors are exogenous to the outcome and uncorrelated with possible (unmeasured) confounders. Note that many of the connectivity measures in Table 1 are affected by infrastructure policies. Travel-time depends upon the quality of public transport, and this in turn is a product of investments. The problem is that infrastructural policies are not randomly assigned, and may co-vary with other factors that also affect development outcomes. Thus, investments in infrastructure may be concentrated in regions that are, for some intrinsic reason, suited to development. Or such investments may be correlated with other (non-infrastructural) policies that exert an independent effect on development outcomes. Both of these scenarios would create a spurious result; a district’s connectivity score might be correlated with developmental performance for reasons having nothing to do with connectivity per se.

Several approaches may be taken to this identification problem. First, we intend to lag independent variables one or several periods behind the outcome of interest. Thus, for an outcome measured in 2000, lags of 50, 200, and 2,300 years will be introduced. Note that all measures constructed for 300 BC are purely geographical, and thus entirely exogenous. These may be also employed as instruments in a two-stage least squares analysis, providing a second approach to the identification problem. Here, we must be attentive to the potential problem of chosen instruments that correlate with other factors that have independent effects on long-term development (through pathways other than connectivity). For example, navigable rivers provide not only an avenue of natural transport but also a source of water and rich topsoil for settled agriculture – the material basis for all pre-modern civilizations. Similarly, isolated regions are likely to be characterized by poor agricultural land with few exploitable resources, and perhaps epidemiological dangers.

Clearly, all regression analyses (whether one-stage or two-stage) must be attentive to alternative explanations of development, particularly those which are likely to be correlated with our connectivity measures. This includes (but is not limited to) measures of agricultural fecundity, domesticable plants and draft animals, mineral resources, and epidemiological threats. Fortunately, these factors have been a primary focus of recent research (see previous citations), and are easy to incorporate with GIS data since they rest largely on the geographic features of a territory, features that have changed relatively little over the past century and are not usually amenable to policy intervention. In some cases, it will be necessary to reconstruct these measures so that they describe districts rather than nation-states. Thus, we will need to integrate maps showing the quality of soil, climate, topography, mineral resources, and other factors that can be measured at a district level. Other factors, such as the presence of domesticable plants and animals in pre-historical times, need
not be measured at the district level (district-level variables may be constructed using country-level data). 

To reprise, the core empirical strategy is to regress measures of development against measures of connectivity, along with a wide range of controls, in one- or two-stage models. Note that in models with purely geographic measures of connectivity (e.g., connectivity measured in 300 BC) the set of relevant controls is presumably limited to other geographic controls. All other factors in development – e.g., technology, population, trade – may be viewed as endogenous. In models where connectivity variables draw on infrastructure and demography, a larger and more complicated identification problem is faced, requiring a larger set of controls and suggesting a two-stage approach. We view these two types of models as playing a complementary role. The purely geographic measures of connectivity should establish the plausibility of the causal argument. The more refined – but also more endogenous – measures of connectivity should allow us to clarify which sorts of connections matter most for development.

One final empirical problem must be dealt with. In recent decades, sub-Saharan Africa has been devastated by the HIV/AIDS epidemic. The regional concentration of this disease – which, thus far, has not had a substantial effect on mortality rates in other parts of the world – is accidental with respect to the theoretical interests of this study. To be sure, the diffusion of HIV/AIDS bears directly on the theory of connectivity, since more connected districts are presumably at higher risk. HIV/AIDS thus offers a good example of a negative connectivity effect, and in this respect is similar to previous epidemics (e.g., smallpox, cholera). However, at the present time the effects of HIV/AIDS are heavily concentrated on one region of the world – an accidental feature of the disease that has no bearing on our theory. Its effects on human welfare are so enormous that it cannot be absorbed as part of the error term. Two approaches might be taken to address this problem. The first option would be to endogenize it by including a measure of HIV prevalence as a regressor in each analysis. (This would probably have to be a country-level indicator, since disaggregated analyses are not yet available, to our knowledge.) A second option would be to measure developmental outcomes of interest in eras prior to the spread of the disease – i.e., to back-date the last observations in each regression to the mid-to-late 1980s.

In addition to the “global modeling” approach it may also be possible to exploit “natural experiments” that have taken place via immigration history across the globe. I am thinking of situations in which an immigrant group settles in several territories within a new country, where those territories exemplify high and low connectivity, and – crucially -- where the settlement decision is the product of exogenous factors (e.g., family or village connections) having no bearing on development outcomes of interest, and thus where ceteris paribus conditions across groups is plausibly maintained. Further research on this possibility seems warranted.

Alternative Theories of Long-Run Development, Briefly Considered

Technology and technical knowledge (which we subsume under the rubric of technology) is generally acknowledged as a critical factor in long-run development (Landes 1969; Mokyr 1992). While most of the literature on this subject is focused on economic development it is easy to see how technology might also contribute to human development (e.g., through improved knowledge about the sources of public health, the availability of contraceptives, medicine, and so forth). Indeed, technology is a critical factor in our speculative discussion of causal mechanisms (see above). Unfortunately, technology is difficult to measure prior to the late twentieth century (and even now is rather problematic). Its precise role in development thus resists empirical testing (but see Comin,

5 For examples of this kind of analysis at the country level see Escobal and Torero (2005), Ravallion (2005).
Easterly, Gong 2006). In lieu of direct tests, we make two critical assumptions about the development and spread of technology. First, we assume that innovation is fostered by connectivity, areas with high connectivity being more likely to serve as incubators for new ideas. Second, we consider technological development to be primarily a product of diffusion rather than original creation (Barro, Sala-i-Martin 1997; Keller 2004). Few inventions are truly independent of one another, and the diffusionist feature becomes even more evident in contemporary eras of world history. Our measures of connectivity should account for the likelihood that a given district will benefit from the knowledge of humankind, regardless of where that knowledge was originally created. Thus, technology is looked upon as largely endogenous to the theoretical framework.

Population is also regarded as a crucial factor in long-run development, since it may stimulate innovation, political development, and other intermediate causal factors (Boserup 1965; Galor, Weil 1999; Goodfriend, McDermott 1995; Jones 2001; Kremer 1993). While we do not dispute the general argument, we view demographics as secondary to geography. It is no accident that populations in some parts of the world have grown, while others have stagnated, over the millennia of human history prior to the industrial revolution. Long-run demographic variation may be accounted for by some combination of intrinsic (non-relational) and extrinsic (relational) features of the landscape. While our theoretical interest is in the latter, our empirical challenge is to measure both aspects adequately such that long-run population growth is accounted for with strictly exogenous variables. If this can be accomplished, then population may be regarded as an endogenous feature of the narrative.

Commercial exchange of goods and services is a central mechanism of contact across human populations. As such, trade and trade policy serves as important causal pathway in any story of long-term development that hinges upon connectivity. Importantly, we assume the effect of trade on development is best understood as a long-run relationship; short-term effects, in our view, are much more difficult to theorize and to test. Thus we do not view cross-country regressions of trade openness and growth during the postwar era (e.g., Yanikkaya 2003) as offering strong support for or against the connectivity thesis. In any case, trade flows are viewed as largely endogenous to a district’s (or a country’s) level of connectivity (Frankel, Romer 1999), and thus are not incorporated explicitly into the analysis (except perhaps as a probe into causal mechanisms).

Objectives and Prospects

The objectives of this project may now be re-stated in more specific terms. We hope, first of all, to provide a reasonably comprehensive map of human welfare as it varies spatially across the globe in the contemporary era (based on individual-level data drawn from DHS surveys and other surveys and censuses). In particular, we want to examine the variation that occurs (we presume) as one moves a) from major transport arteries to more remote areas, b) from densely settled (“urban”) regions to diffusely settled regions, and – more generally – c) from highly connected regions to regions of low connectivity. These are descriptive questions, and perhaps best approached visually (with maps of the world, color-coded to reveal possible causal processes).

The main tool of causal analysis is a series of regression models, with countries and/or individuals as units of analysis. First, and most generally, we want to test whether connectivity matters to development and whether the impact (if any) is significant. Second, we want to test whether the strength of this relationship has changed significantly over the past two centuries. Does connectivity increase, or diminish, in predictive power from 1820 to 2000? Third, we want to test which of the manifold measures of connectivity demonstrate the strongest empirical relationship to development. Are the effects of connectivity realized because of proximity to a) an urban area, b) a transport hub, c) a dominant political unit such as a country, d) a larger region, or e) the rest of the
world? Fourth, we want to test whether these various types of connectivity have differential effects over different time-periods. Is the force of connectivity on development increasingly global, and decreasingly national and regional? Fifth, we want to test a variety of different development outcomes – economic, demographic, and quality-of-life -- to see which are most sensitive to levels (and types) of connectivity. Does connectivity have a greater impact on human development than on economic development in the contemporary era? Sixth, we want to test whether certain regions of the world are more sensitive to variations in connectivity. (This can be accomplished by conducting split-sample regression tests or by constructing interactive variables in the full, global sample.) Seventh, we want to test whether connectivity has a redistributive effect (transferring wealth to poorer areas on the periphery or allowing poor persons to migrate to wealthier areas at the core), leading to a more equal distribution of income. Finally, want to piece together the results from this wide range of tests in order to shed light on the vexing question of causal mechanisms. Why and how do connectivity factors affect varying developmental outcomes?

In order to fulfill these ambitious goals, the research must move forward on two tracks. The first track involves the construction of a GIS database with which the key travel-time variables will be calculated. In order to do this, several things need to happen. First, further research needs to be carried out on travel-times usual to varying terrains and different transport types so that the assumptions underlying each connectivity variable can be verified. Second, further coding issues must be resolved, such as how to code travel-time between two districts that are impossible to reach, given extant technology. Third, the locations of cities in the four time-periods need to be determined, and entered into the database. Fourth, estimates of population density across the world in 1950 and 2000 must be identified (presumably from historical atlases) and entered into the database.

The second track involves the construction of a standard rectangular database (using Stata or some other software) that integrates variables from DHS and other relevant surveys from around the world. (Not sure what sort of software will handle such a large number of observations; fortunately, the number of variables will not be overwhelming.) Relevant variables include: a) development outcomes (health, malnutrition, et al.), b) the location of the respondent (so that the individual's district and country can be coded), c) all other factors that might affect the individual's level of development. Next, country-level variables of relevance to the analysis must be merged into this dataset. Finally, the travel-time variables constructed with GIS software must be merged. At this point, the data collection process will be complete and analysis may begin.

**Data sources**

CEPII: [http://www.cepii.fr/anglaisgraph/bdd/distances.htm](http://www.cepii.fr/anglaisgraph/bdd/distances.htm)
References


Gerring, John; Strom C. Thacker; Ruben Enikolopof; Julian Arevalo. 2008. “Public Health Performance: A Model-Based Approach.”


Iyigun, Murat. 2007. “Monotheism (From a Sociopolitical and Economic Perspective).”


