Travel Networks and US City Prosperity

by

S Joshua Mendelsohn

Department of Sociology
Duke University

Date: __________________________
Approved:

James Moody, Supervisor

Charles Becker

Kieran Healy

Lisa Keister

Peter Mucha

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Sociology in the Graduate School of Duke University
2014
ABSTRACT

Travel Networks and US City Prosperity

by

S Joshua Mendelsohn

Department of Sociology
Duke University

Date: __________________
Approved:

________________________
James Moody, Supervisor

________________________
Charles Becker

________________________
Kieran Healy

________________________
Lisa Keister

________________________
Peter Mucha

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Sociology in the Graduate School of Duke University 2014
Abstract

Does the movement of people between cities influence the economic prosperity of those cities? I examine highway and air travel network data for continental US cities between 2000 and 2010 to argue that it might. My argument consists of three analyses. The first uses network ARMA modeling to show that there is an association between the US travel network and city median income. However, the explanatory contribution of the network varies from 2% to 16%, depending on the model specified. The second uses two-stage dyadic linear modeling to establish directionality, showing that the association is less likely to reflect the influence of prosperity on travel networks than vice versa. However, these models explain only 17% of the variation in traffic volumes. The third addresses causality. It uses a natural experiment design to demonstrate that natural disasters in distant countries correspond to diminished city median incomes for the US cities with connections to them, but not for a propensity-matched sample of unconnected cities. However, the finding are not statistically significant at the .05 level (p=.09), and the estimated size of the effect is implausibly large. Together, the analyses examine the association, directionality and causation, raising the possibility that the movement of people between networks influences the prosperity of those cities.
# Contents

Abstract v

List of Tables viii

List of Figures ix

1 Introduction 1

2 Conceptual Context 4

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

2.2 Influences on Travel Networks . . . . . . . . . . . . . . . . . . . . . 6

2.2.1 Distance – Traffic Attenuated . . . . . . . . . . . . . . . . . . . . 6

2.2.2 Technology – Distance Rescaled . . . . . . . . . . . . . . . . . . . 7

2.2.3 Demographics – Travel Easier for Some . . . . . . . . . . . . . . . 9

2.2.4 Political Considerations – Facilitation and Perpetuation . . . 10

2.3 Network Influences on Economic Prosperity . . . . . . . . . . . . . . . 12

2.3.1 State-Centered Research . . . . . . . . . . . . . . . . . . . . . . . 12

2.3.2 Organizations-Centered Research . . . . . . . . . . . . . . . . . . 14

2.3.3 City-Centered Research . . . . . . . . . . . . . . . . . . . . . . . 18

2.4 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

2.5 Operationalization . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

3 Core-Periphery Structure of the US Travel Network 26

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
3.2 Data ................................................................. 27
  3.2.1 City Data ............................................. 27
  3.2.2 Network Data ....................................... 32
3.3 Methods: Exploratory Network Analysis .......................... 35
3.4 Results: US Travel Networks Has a Core-Periphery Structure .... 42
  3.4.1 Inter-City Traffic Volumes .............................. 42
  3.4.2 Structure of the travel network ......................... 44
  3.4.3 Hierarchy of Central Places in the US ................. 45
3.5 Conclusion ..................................................... 48

4 City Prosperity Flows Through Travel Networks ............... 50
  4.1 Introduction .............................................. 50
  4.2 Data ......................................................... 51
  4.3 Methodology: Network Auto-Regressive Model .................. 52
    4.3.1 Auto-Regressive Model ............................... 52
    4.3.2 Network Simplification ............................... 53
  4.4 Results: Traffic Has a Strong Association with Prosperity ... 54
    4.4.1 Analysis 1: An Annual Examination of 78 City “Constellations” 54
    4.4.2 Analysis 2: A Decade-Wide Examination of 504 Cities .......... 56
  4.5 Conclusions .................................................. 58

5 Travel is Not Tightly Endogenous with Prosperity ............ 60
  5.1 Introduction .............................................. 60
  5.2 Data ......................................................... 61
  5.3 Methodology: Two-Stage Dyadic Model .......................... 61
    5.3.1 Specification Details ................................. 63
    5.3.2 Functional Form ....................................... 65
5.3.3 Error Estimation ............................................ 66

5.4 Results: Traffic Sluggish in Responding to City Influence ............. 66
  5.4.1 Static Model: Geography and History Influence Travel Volumes 66
  5.4.2 Dynamic Model: Prosperity Does Not Predict Annual Change 70
  5.4.3 Variable Group Total Effects ................................ 71

5.5 Conclusions .................................................. 73

6 Travel Exerts Causal Influence on Prosperity .......................... 74
  6.1 Introduction .................................................. 74
  6.2 Data .......................................................... 75
  6.3 Methodology: Matched Sample Natural Experiment ..................... 76
     6.3.1 T-Test of Matched Samples .............................. 76
     6.3.2 Propensity Matching .................................... 77
  6.4 Results: Traffic Transmits Exogenous Shocks .......................... 77
  6.5 Conclusions .................................................. 79

7 Closing Thoughts ................................................. 81
  7.1 Travel Networks Influence City Prosperity .......................... 81
  7.2 Potential Applications: Improving US Economic Prosperity ............. 84
  7.3 Future Work and Implications ................................... 86

A Contours of the Field ........................................... 88

B Excluded Cities ................................................. 90

Bibliography ...................................................... 92

Biography ......................................................... 113
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Basic Statistics (2000-2010)</td>
<td>28</td>
</tr>
<tr>
<td>3.2</td>
<td>Network Statistics</td>
<td>33</td>
</tr>
<tr>
<td>3.3</td>
<td>Hawaiian Air Traffic Matrix Example</td>
<td>36</td>
</tr>
<tr>
<td>4.1</td>
<td>OLS Model of City Prosperity</td>
<td>55</td>
</tr>
<tr>
<td>4.2</td>
<td>Network Auto-correlation Model of Constellation Prosperity</td>
<td>55</td>
</tr>
<tr>
<td>4.3</td>
<td>Pooled Prosperity Model</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Static Model</td>
<td>67</td>
</tr>
<tr>
<td>5.2</td>
<td>Dynamic Model</td>
<td>71</td>
</tr>
<tr>
<td>6.1</td>
<td>Connection Propensity Model</td>
<td>78</td>
</tr>
<tr>
<td>6.2</td>
<td>Mean Propensity of Qualifying Cities</td>
<td>78</td>
</tr>
<tr>
<td>6.3</td>
<td>Income Growth After Exogenous Shock</td>
<td>79</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Conceptual Model</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Primary Topographic Impediment</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>US Travel Network</td>
<td>32</td>
</tr>
<tr>
<td>3.3</td>
<td>Example Air Network</td>
<td>35</td>
</tr>
<tr>
<td>3.4</td>
<td>Major Traffic Routes</td>
<td>43</td>
</tr>
<tr>
<td>3.5</td>
<td>Eigenvector Centrality</td>
<td>44</td>
</tr>
<tr>
<td>3.6</td>
<td>Travel Network Space</td>
<td>45</td>
</tr>
<tr>
<td>3.7</td>
<td>78 Constellation Solution</td>
<td>46</td>
</tr>
<tr>
<td>3.8</td>
<td>The Nested Structure of the US Traffic System</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>City Self-Influence vs Network Influences on City Prosperity</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Sum/Dif Term Interactions</td>
<td>69</td>
</tr>
<tr>
<td>5.2</td>
<td>Combined Variable Group Effect</td>
<td>72</td>
</tr>
<tr>
<td>6.1</td>
<td>T-Test Probability Dist.</td>
<td>80</td>
</tr>
<tr>
<td>7.1</td>
<td>Proposed Intervention</td>
<td>85</td>
</tr>
<tr>
<td>A.1</td>
<td>Structure of the Field</td>
<td>89</td>
</tr>
<tr>
<td>B.1</td>
<td>% Local Pop. Excluded from Data</td>
<td>90</td>
</tr>
<tr>
<td>B.2</td>
<td>Exclusions by Pop. and Income</td>
<td>91</td>
</tr>
</tbody>
</table>
Introduction

Does personal travel between cities influence city prosperity?

The 2010 eruption of Iceland’s Eyjafjallajkull volcano created the largest air traffic shutdown in the history of civil aviation. By the time the ash settled, 107,000 flights had been cancelled. Global financial experts estimate worldwide economic costs in the billions of dollars (Erlanger and Ewing, 2010; PWC, 2010; Jamieson, 2010). Eyjafjallajkull poignantly demonstrates the importance of globe-connecting networks for sustaining the modern economy. The eruption caused almost no physical damage, but fresh produce, high value documents, business professionals, just-in-time factory inputs, technical specialists and many other economically relevant items suddenly could not reach European economies in a timely manner. If the disruption had continued, it would have caused major economic losses, as corporate supply chains faltered and long-distance commerce came to a halt. This event illustrates that economies in one place depend on resources and performance in other places. The goal of my dissertation is to contribute to our understanding of how place-connecting networks\(^1\) shape prosperity.

This topic is large, and has a long history. Scholars find that geography and history strongly influence the shape of travel networks, especially relative to the ability

\(^1\) For consistency, I generally will use “travel network” to denote the network formed among cities from people traveling between them.
to traverse large distances, and these travel networks influence prosperity. Nation-state research supports this conclusion at the macro scale, and organizational research supports this conclusion at the mechanism level. However, a discipline-centric “silo” effect has kept many of these insights from building on each other, and both lines of research suffer from empirical problems: lack of system-wide data for organization-centered research, and over-generalizations in state-centered research. Riding on a wave of newly available data and methods, city research is emerging as a compromise unit of analysis – able to speak more directly to the mechanism-centric research on organizations without sacrificing the systemic perspective of nation-states. This dissertation adds to that wave, making at least three contributions.

First, this dissertation makes important empirical contributions to a field that has struggled to parse cause from effect. Economic performance and resource flow are conceptually endogenous. We expect a feedback process: affordable, easy connectivity between places spurs trade between them, which fuels growth, which increases demand for connectivity. However, it is unclear which (if either) is the dominant driver, and even to what extent the observed association is causal. My dissertation employs a three-project examination of association, directionality and causality.

Second, this dissertation makes a scholarly contribution in that the research design synthesizes key concepts and insights from several different communities of research. Interdisciplinary designs aid efforts to unite a fragmented field. The study of economic networks is stronger when the unique insights of sociology, economics, urban studies, organizational behavior, and network science are leveraged in tandem.

Third, this dissertation makes a practical contribution because it applies cutting edge methods and higher quality data than ever used before. Data have only recently become widely available on network connections among cities. While case study work has traced links within industries, full system-level information has never been available, nor would it have been comprehensible without recent developments in
dynamic network analysis. A key contribution of this dissertation is compiling the data and building the networks necessary to test these ideas. For example, while air traffic data is federally compiled, road traffic remains in the hands of local agencies. Its decentralized collection (51 agencies), unstandardized format (4 incompatible formats), massive size (many data points per tie), and erstwhile non-digitized format have kept it beyond the reach of previous scholarship.

I have divided this dissertation into seven chapters. In chapter 2, I examine what previous scholarship has uncovered about social connectivity, cities and prosperity. The literature is vast, but deeply divided. In chapter 3, I describe the data, and the basic empirical characteristics of the US travel network. The US travel network is a core-periphery system comprised of a highway-based geographic lattice, and an airplane-based system of regional bridges between core cities. In chapter 4, I ask, does personal travel between cities influence city prosperity? I find that the travel network is associated with city median income, but the explanatory power of that association varies widely across model specifications. In chapter 5, I ask, do city characteristics influence the number of people traveling between them? I find that, in the short term, the travel network is sluggish and inertial in responding to changes in city characteristics, even if they influence each other in the long term. However, model fit is generally poor. This analysis sheds some light on the directionality of the association between travel networks and income. In chapter 6, I ask, can travel networks transmit economic shocks? I find that cities experience diminished median income growth when network connected alters experience exogenous economic shocks. However, the effect is not statistically significant at the .05 level (p=.09), despite it’s credulity-straining size. This speaks to whether the association between travel networks and income has a causal component. In chapter 7, I reflect on my findings as a whole. The evidence is mixed, but weakly hints that possibility that travel networks may influence city prosperity.
2

Conceptual Context

2.1 Introduction

What Does Previous Scholarship Say About the Relationship between Travel Networks and City Prosperity?

Travel networks, and their relationship to prosperity, inspired the writings of many of the founding giants of social science. When Adam Smith was making the case against mercantilism, international trade was key to his plan for prosperity (Smith, 1776). When Marx was writing on capitalism, the global pursuit of resources and markets was a key feature of his bourgeois (Marx, 1848; Engels, 1892). Scholars long ago realized that travel networks move history. From the Inca Road to the Spice Routes to the Golden Fleet to the Information Highway, economic networks cannot be ignored. While classical scholars focused on trade in goods, rather than travel itself, more recent scholars have called attention to the economic importance of personal connectivity itself (Powell, 1990; Lin et al., 1981). Personal travel between cities likely has important economic consequences. In this chapter, I review the literature pertaining to the relationship between travel networks and prosperity. There are
three key points.

First, travel networks are heavily constrained, and these constraints give them inertia and consistency over time (2.2). Foremost among those constraints is distance, which attenuates traffic through travel costs and cultural dissimilarity (2.2.1). Technology can loosen those constraints, but technology requires infrastructure and shapes the built environment. Those built structures create constraints that persist long after new technologies have emerged (2.2.2). Even with the aid of technology, the ability to travel is not equal for all members of the population. From life course to discrimination to resources, some members of society are less mobile (2.2.3). Institutions also matter for travel networks, making travel easier through public safety and standardization, but also reproducing existing hierarchies (2.2.4).

Second, a wide array of scholarship confirms that network connectivity and structure are association with economic success (2.3). When networks enhance prosperity, they either enrich the environment in which economic activity takes place, or facilitate mechanisms for enhancing the productivity of economic actors. The former is well represented among nation-state studies (2.3.1), while the later is well represented among organizational case studies (2.3.2). City studies are experiencing a renaissance in popularity, because they are large enough to be relevant for economic environment (state) research, but small enough to leverage mechanism (org.) research (2.3.3). Section 2.4 provides a conceptual model for how these literatures relate to each other.

Third, while typical research operationalizes economic networks as trade networks between nation-states, there is growing recognition alternative operationalization choices would yield new insights (Alderson and Beckfield, 2004). Cities provide a lower-level geographic unit of analysis, allowing more more finely grained models. Personal travel captures social connectivity, which remains an underexplored component of economic production.
2.2 Influences on Travel Networks

2.2.1 Distance – Traffic Attenuated

Distance is the most important constraint on travel. 70 years of research has repeatedly confirmed that the strongest predictor of the travel network is the location of people, and the distance between them (Reilly, 1931; Levy, 2012) – history and geography. This applies equally to world trade (Pulliainen, 1963; Zhou, 2010; Eaton and Kortum, 2002), migration (Stewart, 1960), organizational alliances (Owen-Smith and Powell, 2004; Hedstrom, 1994), and intercity movement (Zipf, 1946). Distance discourages friendships (Lee et al., 2011), diminishes social ties (Wellman, 1996) and shapes many features of social networks (Butts et al., 2012; Papachristos et al., 2013).

Two main mechanisms account for the power of distance: cost and culture. In terms of costs – traveling requires both time and resources, costs that are strongly proportional to distance (Combes and Lafourcade, 2005). Because it is costly to overcome distance, it only makes sense to do so if the trip is more valuable than its resource costs (Neal, 2010), and if distance isn’t shielding actors from the competitive pressures of the market (Braha et al., 2011). The greater the distance, the fewer trips will qualify (Losch, 1938; Christaller, 1933). In terms of culture – daily habits, customs, and tastes heavily correspond to location.\(^1\) Even in the purportedly distanceless online economy, culture still gives consumption patterns a distinctly local structure (Blum and Goldfarb, 2006; Mok et al., 2007). Culture also conditions travel behavior (Murdie, 1965), shapes international trade (Zhou, 2010), facilitates knowledge transfer (Tagliaventi and Mattarelli, 2006; Owen-Smith and Powell, 2004), and creates barriers to interaction (Stoller, 2007; Johnson et al., 2006).

Based on these findings, I expect distance to exert a negative influence on travel volumes, and thereby, constrain the structure of the travel network. Furthermore,\(^1\) So does language, but the culture effect is more than just language barriers

---

\(^1\) So does language, but the culture effect is more than just language barriers
since geography and history shape distance, I expect distance to imbue the travel network with consistency over time.

2.2.2 Technology – Distance Rescaled

Technology and Infrastructure Change the Experience of Distance

The meaningful scale of distance has changed rapidly over the past 100 years – an effect known as “globalization.” People respond to the perception of distance much more than the literal physical distance (Schivelbusch, 1978). The more effortlessly that distance can be traversed, the more distance is necessary to discourage travel.\(^2\) Transportation technology and infrastructure are key to travel sensibilities. As travel becomes more rapid, cheap, and effortless, the amount of distance that can be covered increases. In other words, technological advance rescales distance (Arribas et al., 2011), and has been rescaling distance since, at least, the domestication of the horse (Levine, 1999). However, the last century has seen unusually rapid change in the scale of distance. Since 1900, some estimate that transportation costs have dropped over 90% (Hummels, 2007; Estevadeordal et al., 2003). As a result, we are currently witnessing a massive increase in local traffic connectivity – a rescaling of territory (Brenner, 1999a).

Travel technologies make places closer/cheaper in travel network space than they actually are in geographic space (Knowles, 2006; Limao and Venables, 2001). However, different technologies have different kinds of effects on distance. Roads create a line of rescaled distances along their paths (Lay, 1992), while airports only rescale distance between the two endpoints (Graham, 1999; Castells, 1999). However, endpoint-centric technologies have raised more scholarly concern, because they

\(^2\) Marchetti (1994) goes so far as to argue instinct – most primates travel about an hour per day, and this travel pattern produces territories of defensible sizes. Biological or not, this time frame is consistent with observed commuting patterns. The median US commute is 20-25 minutes (each way), and people tend to move after a job switch to keep commute times within this frame (USCB, 2009; Clark et al., 2003).
exclude all of the places that lie between, amplifying inequality (Castells, 1999).

Transportation technology shapes the built environment, but the built environment remains long after that technology has advanced. As such, historical traffic patterns can cast a long shadow over the shape of modern travel networks (Marquis, 2003), the placement of new cities\(^3\) (Murdie, 1965; Dobkins and Ioannides, 2001), average distances traveled (Pan et al., 2009), and the dominant mode of transportation (Giuliano and Narayan, 2003; Schwanen et al., 2004). They can also produce significant mismatches. For example, most US urban areas have excess road capacity, given current transportation (Samaniego and Moses, 2008).

*Caveat: Naysayers of Globalization*

There has been speculation about the death of distance in a globalized world (Cairncross, 1997). However, skeptics claim the evidence does not support that conclusion, outside of limited, specific circumstances like high finance (Bordo et al., 1999). For the sake of completeness, the rest of this section gives voice to those critics. The main point is this: there is a great deal of hype surrounding globalization, and this hype can lead to overstatement. Geography still matters, if less than it once did, and it shapes many supposedly distance-less things, such as electronic communication.

Skeptics have argued that remoteness from markets continues to diminish regional GDP (Boulhol and Serres, 2010) and distance continues to attenuate international trade (Zhou et al., 2011; Brun et al., 2005). While some scholars have argued that the decline of distance can be observed with certain complex model specifications (Coe et al., 2007; Buch et al., 2004), others concede that the case may be been overstated – lacking historical perspective (Erikson and Bearman, 2006; Estevadeordal

\(^3\) eg Buffalo NY (Boats, by 1820), Oakland CA (Trains, by 1870), Phoenix AZ (Trains, by 1910), Anchorage AK (Airplanes, by 1950) all owe their CBSA status to transportation infrastructure. Emerging cities, such as Glenwood Springs CO (Car, CBSA by 2020), attest that this process continues.
et al., 2003), and disregarding key socio-economic mechanisms like agglomeration economies (Morgan, 2004; Taylor, 2009). Even inter-governmental organizations, the institutional heart of globalization, show distinct regional bias in their influence networks (Beckfield, 2008; Ingram and Torfason, 2010)

A related speculation concerns the death of distance in a digitalized world (Castells, 1998; Graham, 1999). While information technology is changing the way we live, it is not clear that it is displacing the physical world (yet?). Research shows that usage correspond to increased travel (Wang and Law, 2007; Lenz and Nobis, 2007), and increased variety in leisure activities (Mokhtarian et al., 2006). When digital activities do displace other activities, they displace media consumption, more than social interaction (Thulin and Vilhelmson, 2006). Moreover, distance strongly shapes digital communication patterns. Most digital communication, from calls to tweets, occurs between people in the same urban area, and can be predicted from traffic patterns (SCL, 2011; Takhteyev et al., 2012). Digital communication tends to support, not replace, travel and proximate interaction (Hampton and Wellman, 2000; Logan, 2012).

Based on these findings, I expect travel network to reflect its technology, with air traffic connections spanning greater distances than road traffic connections. However, I also expect these connections to be much shorter in distance than the typical range of automobiles and airplanes – since most US cities were founded before these technologies rose to prominence, they were founded in places that were sensible for slower transportation modes.

2.2.3 Demographics – Travel Easier for Some

Cost is key to how distance affects travel (Hanson and Hanson, 1981). Consequently, the socio-economic characteristics of people are important to travel networks, as they moderate travel costs (Hine and Mitchell, 2001). Tiebout (1956) conceptualized
residential location choices as a trade-off between the desired public goods, and the taxation/rent/ownership costs of those goods. As a result, those with less wealth tend to end up in less desirable locations (Downey, 2003). In addition, wealthier individuals are able to shoulder higher transportation costs (Giuliano and Narayan, 2003; Dieleman et al., 2002), use more/faster modes of transportation (Chen et al., 2011; Beaverstock et al., 2004), and sacrifice a short commute for better economic opportunities (Wheeler, 1967).

Socio-demographic characteristics also heavily constrain location choice (Boarnet and Sarmiento, 1998). Disadvantaged social groups are pushed away from advantageous residential locations, leading to the need to traverse more distance to accomplish basic tasks – a spatial mismatch (Fernandez and Su, 2004). This can many forms, such as “red lining” discriminatory home loan practices (Tootell, 1996; Ross, 2004), “white flight” discriminatory home buying behavior (Schelling, 1971; Clark and Fossett, 2008), and “sundown town” discrimination against minority workers (Loewen, 2013). Life stage is also an important influence on travel. The aged are less likely to migrate/travel (Schwartz, 1973; Soboleva, 1980), as are people with children (Dieleman et al., 2002; Boarnet and Sarmiento, 1998).

Based on these findings, I expect that the demographics of a city will partly predict its traffic patterns, with younger, affluent, educated populations corresponding to increased travel volumes and longer average distances.

2.2.4 Political Considerations – Facilitation and Perpetuation

Political factors also influence the ease of travel, especially at the international level. Travel networks become denser and farther reaching when war violence is low (Estevadeordal et al., 2003). This process can become self-reinforcing if travel networks increase wealth and economic integration, which (in tandem) encourage peace (Mousseau et al., 2003; Drezner, 2009). Non-governmental organizations play a key
role in this process, promoting peaceful conflict resolution (Maggi and Staiger, 2011; Halliday and Carruthers, 2007), encouraging pro-trade norms (Torfason and Ingram, 2010; Henisz et al., 2005), and serving as powerful sources of symbolic capital (Polillo and Guillen, 2005; Liu, 2006). Travel policy also matters, as official movement controls can create severe barriers to travel networks (Paul, 2011).

While much of this research does not directly apply to a study of US cities 2000-2010, there are two key insights to keep in mind. First, US traffic is likely higher because public safety is generally quite good. This is a marked contrast to terrorism in Israeli (Stecklov and Goldstein, 2010), Mexican crime, and Syrian civil war, each of which curtailed domestic travel at points in the 2000-10 period. Second, US travel is likely higher because of the standardization of US society, economy, and traffic grids. Both reflect the strength and ubiquity of US institutions. Third, US traffic is likely higher because the US has few internal movement controls.

However, that stability comes at a price. Institutions perpetuate existing power systems. For example, global trade has thrived in peace, but the foundation of this peace is military might. The UK and US have exerted significant military dominance for most of the last 250 years. Both are naval powers – deriving wealth from global trade, rather than territorial conquest. Much of the global proliferation of trade rests on these efforts to secure the seas, and protect trade routes (Levy and Thompson, 2010; Case-Dunn et al., 2000). As such, global trade tends to favor and protect the existing order of nation-states (Mahutga and Smith, 2011; Mahutga, 2006), as does globalized media (Janssen et al., 2008). In contrast, developing countries can experience mobility in the global trade hierarchy (Clark, 2010), but do not always reap the expected benefits of neo-liberal policies (Kick et al., 2000). Applied to the

---

4 While this is increasingly standard governance around the world, movement restrictions were common place in many countries for much of the last century. As recently as 1990, there was Propiska in the Soviet Russia (Becker et al., 2012), Hukou in China (Cheng and Selden, 1994), Huji in South Korea, and Apartheid in South Africa.
US context, US institutions likely perpetuate the existing hierarchy of cities. For example, the only high-speed rail line in America was federally subsided, and connects its most economically powerful city (New York) to its most politically powerful city (Washington).\textsuperscript{5}

Based on these findings, I expect the US to be a good test case, because a US domestic study controls for many of the political factors that influence travel, and because US traffic is relatively free political constraints on travel.

2.3 Network Influences on Economic Prosperity

Across disciplines, the unit of analysis is a convenient way to organize past research on the relation between economic performance and social connectivity. Nation-state research (2.3.1) provides a rich, theory-centric literature. It excels at providing “big picture” views on big issues like global systems, inequality, and institutions. Organizational research (2.3.2) provides a detailed, process-centric literature. It excels at providing practical insights into the mechanisms of economic prosperity and decision-making. City research (2.3.3) is the emerging hybridization of the two. It excels at combining the high-level theoretical and systemic insights of state research with the grounded, mechanical insights of organizational research.

2.3.1 State-Centered Research

While globalization has done much to erode the “hard shell of state”, nation-states continue to matter (Brady et al., 2007). State borders are formidable barriers to the global movement of financial capital (Yeung, 1998), and shape firm ownership (Kogut and Walker, 2001). Activist (“developmental”) states play a tremendous role in creating economic advantages for domestic firms (Gereffi and Fonda, 1992; \textsuperscript{5}This is changing. However, the time lag between the federally-funded 2000 NYC-DC route (Amtrak, 2014), and the state-funded 2029 SF-LA route (CHSRA, 2014) is illuminating
Johnson, 1999), and have significant power to redistribute the largess of the global economy (Huber et al., 2006).

International migration and trade networks influence the economic fortunes of nation-states. In terms of migration, many national economies rely heavily on remittances from expatriates (Skeldon, 2008; Levitt and Jaworsky, 2007), even though they may suffer from the loss of skilled labor (Feliciano, 2005; Johnston, 1991). Regarding trade, scholars have argued for the existence of an international trade hierarchy. States with favorable locations in the trade network accrue benefits from their position (Case-Dunn and Grimes, 1995; Gereffi and Fonda, 1992). Research supports this claim. Position in the trade hierarchy corresponds with levels of economic development (Kick et al., 2000), and advancement in the hierarchy corresponds with developmental growth (Clark, 2010), especially for middle-rank countries (Mahutga and Smith, 2011; Mahutga, 2006). States with the least dense trading networks tend to have significantly higher infant mortality rates (Moore et al., 2006).

However, international economic networks do not raise benefit everyone. Since at least Kuznets (1955), scholars have noted (with alarm) that economic development tends to increase inequality, especially in the early stages. Global economic networks exacerbate this kind of inequality (Alderson and Neilsen, 2002; Cornia, 2001) both because they weaken the demand for local labor (Beckfield, 2006; Miller, 2001), and connect globalized elites to a vastly richer economic pool (Sassen, 2001; Castells et al., 2006). The increased prominence of global economic networks over the past two decades has had profound influences on inequality. While the inequality between countries is shrinking, the inequality within countries has grown tremendously (Firebaugh and Goesling, 2004; Clark, 2011). In 2010, the Gini coefficient for population-weighted, nation-state, average incomes was .543, and the Gini coefficient for US income was .541 (Fessenden and McLean, 2011; World Bank, 2011; USBEA, 2010).
Based on these findings, I expect three things. First, I expect that economic actors indirectly benefit from networks that connect their geographical location to other geographical locations, even if they do not directly enact the network. Second, I expect nation-states, especially large ones, to be too internally diverse to be an effective unit of analysis for an economic network study. Third, I expect that movement of people to matter for economic prosperity, even if the mechanisms are more indirect than the mechanisms related to trade networks.

2.3.2 Organizations-Centered Research

Organizations have inspired the largest and most diverse literature on economic networks. Partly, this reflects the nature of organizations as networks of people with (somewhat) defined roles and relationships, and partly, this reflects the level of qualitative detail that can be achieved with an organizational study. At least four inter-related topics are prominent among the organizational network literature: clustering, innovation, trust, and power. Each is deeply rooted in organizational theory, and each offers a distinct mechanism by which networks influence productivity.

Clusters and Agglomeration Economy

Firms often tend to locate in geographic clusters, especially if they are in the same organizational field (Duranton and Overman, 2005; Porter, 2000). While the geographic location of these clusters is often a historical accident, the persistence of these clusters reflects their economic benefits. Geographic clustering means that firms can enjoy the benefits of high quality infrastructure and amenities, while splitting the costs with many other firms (Porter, 1998; Duranton and Puga, 2002). Firm clustering also enhances the available labor pool, since all of those firms attract potential employees of their industry to that location (Keeble and Nachum, 2002; Rosenthal and Strange, 2004). As a result, geographic clusters foster the right conditions for
the diffusion of innovations through social networks (Crevoisier, 2004; Reagans and McEvily, 2003). This occurs because geographic proximity strongly encourages social ties (McPherson et al., 2001; Hipp and Perrin, 2009), and the enhanced local labor market ensures that social actors have the technical sophistication to spread useful knowledge through those ties (Owen-Smith and Powell, 2004; Tagliaventi and Mattarelli, 2006). In this sense, geographic clustering is a network-based strategy for benefiting from the innovations of other organizations (Suire and Vicente, 2009).

Based on these findings, I expect that geographic clusters, like cities, create economic advantages for organizations because they enable the pooling of resources. I also expect that dense interconnections among a group of organizations create economic advantages for similar reasons.

**Innovation and Diffusion**

Innovation is a key topic in this field because of the important role of networks. Organizations derive much of their super-human effectiveness from the predictability (Thompson, 1967), and efficiency (Taylor, 1911) of their systemic process (Weber, 1922; Barnard, 1938). Thus, the two key challenges for a prototypical organization are how to make the system efficient (relative to competitors), and how to prevent the unexpected from disrupting the system (Barnard, 1938; Thompson, 1967). For both challenges, the search for innovative solutions is key to organizational survival (March and Simon, 1958; Cyert and March, 1963).

While organizational actors may generate these innovations *de novo*, it is often the case that many viable solution strategies already exist, but are scattered among organizational actors (Cohen et al., 1972). Social and inter-organizational networks facilitate the spread of innovation to the actors in need of them (Brass et al., 2004; Strang and Soule, 1998). In fact, Burt (2004) argues that innovations are generated when actors create bridges between distant parts of the network, because this
combines diverse pools of knowledge and strategies. Research supports this argument. Inter-firm alliances, workgroups, and managers all tend to perform better if their constituents have a wide range of diverse specialties (Cummings, 2004; Rodan, 2010). The firms with the most diverse collaborators play bigger roles in shaping the evolution of their field (Powell et al., 1996, 2005).

Based on these findings, I expect that organizations produce more innovations when their connections expose members to ideas from diverse parts of the network – network diversity will enhance productivity. I further expect that cities will enhance this process if their travel connections expose the city population to ideas from a diverse pool of other cities.

*Trust and Transaction Costs*

The lowering of transaction costs is a key feature of organizational efficiency, and often a byproduct of organizational predictability. Transactions costs are the costs of governing production, especially the costs of coping with the unknown, and preventing opportunism (Williamson, 1975; Ouchi, 1980). Structures, such as organizations, create formal systems to diminish these cost by creating stability, enhancing information, and providing mechanisms for addressing grievances (Simon, 1991; Williamson, 1991).

However, formal structure is often insufficient to deal with uncertainty (Meyer and Rowan, 1977). Rather, rapid organizational learning and adjustment tends to happen at the level of informal, ground-level action (Chan, 2002; Brown and Duguid, 1991). For this reason, a logic of autonomy, trust, and loose structure can produce more effective organization, especially for highly dynamic (“high velocity”) environments (Weick, 1976; Eisenhardt and Martin, 2000).

Networks play a key role in this process. In fact, a large body of scholarship argues that networks are a distinct form of production governance, where trust and
relational commitment provide stable, diminished transaction costs (Powell, 1990; Podolny and Page, 1998). Organizations that pursue a trust network strategy tend to strive for mutually beneficial solutions with partners (Uzzi, 1997), and allow formal organizational boundaries to be porous (Morton et al., 2004). Research has shown the effectiveness of trust-based strategies of network organization. Social ties correspond with cheaper sales prices (DiMaggio and Louch, 1998), favorable loan rates (Uzzi, 1999), and a strategy of overcoming uncertainty through transaction partner exclusivity (Podolny, 1994). In fact, some have argued that the liability of newness is actually the liability of not yet having established many inter-organizational ties (Baum and Oliver, 1991; Powell et al., 1996).

Based on these findings, I expect that organizations benefit from interpersonal networks, because interpersonal trust and relationships correspond to diminish risk, uncertainty, and opportunism. While this dissertation will not directly measure trust or social capital, they are likely one of the mechanisms underlying how travel networks improve economic productivity, and thereby improve income. Rises in city median income are a culmination of interpersonal processes like this one.

**Power and Unequal Exchange**

Because they are networks of linked, specialized actors, organizations often entail a network of dependence – each actor needs the cooperation of others to perform a role, and performs a role that others need. Actors control organizations to the extent that their cooperation is essential to the survival of the organization (Pfeffer, 1981). Resource dependence dynamics are important for two reasons. First, they imply that the powerful actors within an organization are not necessarily those with official rank (Mechanic, 1962), and, in fact, may not even be members of the organization.

---

6 This is the observation that organizations have a much higher rate of “death” in their first 2-5 years (Hannan and Freeman, 1984; Amburgery et al., 1993)
(Pfeffer, 1978). Second, they imply that economic actors will obtain more favorable exchanges to the extent that they are asymmetric within the dependence network – others depend on them more than they depend on others (Willer, 1999; Walker et al., 2000). This is often operationalized as A having more alternative trading partners (high power) to B, than B has trading partner alternatives to A.

Research supports the importance of dependence networks in economic exchange. In experimental studies, asymmetric dependence among actors raises the inequality of exchanges (Cook and Witmeyer, 1992), and lowers resistance to unfavorable exchange offers (Borch and Girard, 2009). In organizational studies, dependence asymmetry corresponds to higher profits, and faster expansion (Gereffi et al., 2005; Craig and von Peter, 2010).

Based on these findings, I expect that organizations will benefit if other organizations depend upon their unique network connections, because dependency is the inverse of power (Emerson, 1962). Actors benefit from being the “middlemen” between organizations, and the resources/other actors that those organizations need.

Based on a preponderance of all the organizational research, I expect networks to improve organizational productivity at multiple scales – inter-personal, inter-organizational, and inter-geographic. In particular, I expect diversity in networks connections, density in network groups, and lack of dependencies in network connections to correspond with organizational productivity. While this dissertation will not directly measure power, it is likely one of the mechanisms underlying how travel networks improve economic productivity, and thereby improve income. Rises in city median income are a culmination of interpersonal processes like this one.

2.3.3 City-Centered Research

Distance creates costs – resources are used less efficiency because it is difficult for actors to access them. Cities are sites of spatial economy in at least four ways. First,
they contain actors and resources in proximity, which entails low distance barriers to resource access (Marshall, 1890; Moses, 1958). Second, they host communities of economic actors, which collectively can afford to build advanced infrastructure that would be beyond the resources of any individual actor. In turn, infrastructure enhances productivity and diminish distance costs to other cities (Brenner, 1999b; Rosenthal and Strange, 2004). Third, cities contain large populations, which makes it easier for specialists to find a sufficiently large market for their services (Christaller, 1933; Losch, 1938). Fourth, since proximity strongly shapes social ties (Butts et al., 2012; Hipp and Perrin, 2009; Entwisle et al., 1996), cities are focal points of massive information connectivity (Takhteyev et al., 2012; Calabrese et al., 2011). Consequently, cities are an important milieu of innovation and culture (Dvir and Pasher, 2004; Arbesman et al., 2009).

In addition to being major sources of economic prosperity, cities are adept at luring valuable capital away from other places (Polese, 2005). Major cities offer a highly desired environment (Wang et al., 2011), and a high quality of life for those who can afford it. As a result, migrants with high financial and human capital (Johnston, 1991; Rogerson, 1999), especially young, educated professionals (McCormick and Wahba, 2005; Ritchey, 1976), are disproportionately more likely to migrate to major cities.\(^7\) In fact, given the increasing importance of intellectual work for economic productivity, Florida (2008) argues that economic competitiveness is contingent on the ability of cities to attract members of the “creative class” (Felock et al., 2008). Cities can be viewed as the economic “champions” of regions. Cities link regional economies into the world system, and reflect the power of those territories as they compete for global standing (Friedmann, 2001). Some cities are national champi-

\(^7\) eg The average pre-move income for all US cross-county migrants in 2010 was $45,258. In contrast, the average pre-move income for someone moving to Manhattan in 2010 was $74,597, and the average for San Francisco was $60,504 (USIRS, 2011). Domestic migrants to America’s top cities average a much higher income bracket than the average American domestic migrant.
ons (Mann, 1997; Benton-Short et al., 2005), supported through explicit state policy (Brenner, 1998; Hill and Kim, 2000). In trade studies, the world nation-state hierarchy does not necessarily correspond with the hierarchy of cities. However, the hierarchy of national “champion” cities\(^8\) does correspond to the nation-state hierarchy (Mahutga et al., 2010).

*Global City Hierarchy*

Hierarchy has been a major focus of city network scholarship, both in the Marxist tradition of dominance studies, and in the geo-economic tradition of city primacy. The city hierarchy concept originates with central place theory – one of the bedrock theories of urban economics (Christaller, 1933; Losch, 1938). Central place theory builds on the basic insight that businesses will profit from minimizing the travel costs of their customers,\(^9\) and therefore, traditionally seek out centralized locations. The theory holds that the maximum distance a consumer will travel to acquire a good/service is proportional to the value of that good. Businesses that sell less valuable goods/services will need to locate closer to consumers, while more valuable goods/services can locate in the center of much larger market areas. The result is a city hierarchy, where higher-order specialists locate in cities at the center of larger market areas, while lower-order businesses locate nearer the center of small segments of that market area. Many central place theory predictions have been validated over the years (Partridge et al., 2008; Openshaw and Veneris, 2003; Batty, 2008), and it continues to be a essential pillar of urban theory, even in the globalization era (Neal, 2011a; Taylor et al., 2010; Derudder and Witlox, 2004).

The findings of global hierarchy studies are fairly consistent. At the top of the hi-

---

\(^8\) For example, New York City for the US, Toronto for Canada, and Mexico City for Mexico

\(^9\) Specifically, CPT holds that the value of the cost of a good/service plus the cost of traveling to acquire it must be less than the value of the good. Since travel costs are essentially part of the price of the good, businesses can engage in spatial competition to have lower travel costs, much as they could engage in traditional price competition.
erarchy is one city from North America, Europe, and Asia. Generally, these are New York City, London/Paris, and Tokyo/Hong Kong, with Los Angeles, Chicago, San Francisco, Shanghai, and Singapore close behind (Foreign Policy et al., 2010; Price-WaterhouseCoopers, 2007; Derudder et al., 2010; Alderson et al., 2010). Together, these cities form the core of the global city network. They have been called “global cities” because they unite regional transportation hubs into a global transportation network, contain the world’s most powerful multi-national corporations, and are home to many of the “tertiary service” firms that make global command and control possible (Sassen, 2001; Alderson et al., 2010). Uncoincidentally, they are cities with prominent places in colonial history (Friedmann and Wolff, 1982; Friedmann, 1995).

Below the global cities are a group of regional cities. These cities connect the various world regions to the global core (Rozenblat et al., 2006). Regional cities experience much more mobility within the hierarchy. Most recently, hub cities in the global periphery (e.g., Johannesburg) have been losing ground to clusters of hub cities in the EU, NAFTA, and East Asian regions (Smith, 2001). Falling transportation costs allow these cities to form a network of tightly linked city-regions, and experience region-wide agglomeration economic benefits (Scott, 2001; Townsend, 2001).

While there is generally stable consensus on the broad ordering of cities – at least in terms of core, semi-periphery and periphery distinctions – there has been significant evolution in the methodology. The earliest studies measured global connectivity indirectly, using corporate headquarters, and tertiary service firm locations as proxies for global reach (Friedmann and Wolff, 1982; Moss, 1987). The next generation kicked off with the realization that global city measures needed to be connection-oriented – measures of how the city interacts with the world. This generation of studies focused on the number and variety of international immigrants (Beaverstock, 1994), the amount of investment in telecommunications (Warf, 1995), and content analysis of business newspapers around the world (Taylor, 1997). Gen-
eration three (ongoing) stemmed from the recognition that formal network analysis presented the opportunity to measure inter-city connectivity directly (Smith and Timberlake, 1995). Most studies either use air traffic to measure the movement of people between cities (Smith, 2001; Mahutga et al., 2010), or branch office networks to measure how multi-national corporations connect cities (Alderson and Beckfield, 2004; Derudder et al., 2010).

Cities and Networks

Modern network analysis has revolutionized city prosperity research. The network framework is able to incorporate all of the key insights of central place and world systems theories, while overcoming many of the limitations in those approaches – especially regarding globalization.

In terms of central place theory, classic theory views cities as the key summation points for regions. Since more specialized goods and services draw from larger market areas, each city is hierarchically embedded in the region of a higher ranked city, which will contain higher order specialists (Christaller, 1933; Losch, 1938). While some had argued that globalization would mean the “death of cities”, the importance of agglomeration economies and face-to-face contact have firmly prevented that from happening (Kolko, 1999; Hall, 1999). Instead, globalization has made network centrality a key complement to traditional central place geographic centrality (Taylor et al., 2010; Amin, 2002). Cities can now reap greater benefits from network position for at least three reasons. First, information increasingly serves as a raw material in the modern economy, allowing cities with high quality information flow to benefit (Hall, 1997; Derudder and Witlox, 2004). Second, there are now more opportunities to benefit from facilitating regional networks (Derudder et al., 2007). Miami’s links to Latin America and Denver’s links to the Northwest are prime examples (Neal, 2011b). Third, enhanced connectivity between regional cities creates new opportuni-
ties for specialized labor markets, such as the researcher market near Raleigh, or the programmer market near San Francisco (Neal, 2011b; Derudder and Witlox, 2004). This is a departure from the central place theory tradition of rank equivalent cities having equivalent specialists (Taylor, 2009; Derudder and Witlox, 2004).

In terms of world systems, the geography of power has become less state-centric, and more city-centric, as network connectivity weakens the unity of regional economies. Cities can come more to depend on global competition for capital, and global economic networks (Sassen, 2002; Cox, 1995), weakening the coupling between city and regional /national economy (Alderson and Beckfield, 2004; Hung and Kucinskas, 2011). They can also be “cherry-picked” for global economic networks, while adjacent cities are left out (UN, 2001; Graham, 1999). The resulting geography of power now favors elite cities in elite countries, instead of all cities in elite countries.

While much of the high concept theory around city networks remains subject to debate (Friedmann, 2001; Derudder et al., 2007), research suggests that networks do, in fact, influence city prosperity. Brueckner (2003) finds that a 10% increase in air traffic passengers corresponds to increased employment, and Neal (2011a) finds that the relationship likely has a causal component.

Based on these findings, I have three expectations: First, I expect cities to enhance the economic productivity of the actors they contain. Second, I expect city networks to be highly hierarchical, with New York at the top and Los Angeles/Chicago/San Francisco close behind. Third, I expect network connectivity and hierarchical position to influence city prosperity.

2.4 Conclusion

Figure 2.1 places these literatures in context. Slow-changing forces shape travel networks, imbuing them with consistency over time, and resistance to change (blue dashed line). Location attenuates travel through costs and cultural dissimilarity.
Technology reshapes the costs of distance, but its influences on the built environment can persist long after that technology changes. Biography – ability, constraints, resources, and privilege – grants some segments of society more mobility than others. Institutions influence public safety and standardization, facilitating travel.

These “slow” forces shape the economically relevant connections between cities, including the movement of people between cities. These movements may include migration, commuting, and business travel, among others (red dotted line).

Travel networks exert two broad kinds of influences of city prosperity: environmental and mechanical. Environmental effects are ways that geographically bounded entities, like nation-states and cities, enhance the productivity of the economic actors contained within its boundaries. Mechanical effects are network-driven mechanisms by which economic actors, primarily organizations, enhance their own productivity. City prosperity emerges from the prosperity of those economic actors.

2.5 Operationalization

This dissertation examines core-based statistical areas. CBSA are the official census bureau “cities,” defined based on economic and commuter integration. Cities share many of the same properties of nation-states, the traditional unit of analysis, but allow for finer grained models. Like nation-states, cities (CBSA defined) are geo-
graphically bounded collections of people, with strong internal economic and travel cohesion. However, the median nation-state contains over 4 million people, spread across 43 thousand square miles, while the median US city contains 126 thousand people, spread across 127 square miles. In examining cities, this dissertation joins an emerging wave of research applying city-level precision to an erstwhile nation-state research program (Alderson et al., 2010; Mahutga et al., 2010).

Personal travel deviates from the traditional focus on trade, but provides insight into social connectivity – a key component of economic activity that remains underexplored in macro-scale studies. Organizational research has highlighted the role of social networks in generating innovation, social capital10 and trust (Burt, 2004; Nahapiet and Ghoshal, 1998; Powell, 1990). Each enhances economic production. Innovation transforms knowledge, particularly tacit knowledge, into more efficient and effective ways of accomplishing goals (Powell et al., 2005; Brass et al., 2004). Social capital provides access to economic opportunity (Marin, 2012; Adler and Kwon, 2002). Trust diminishes uncertainty, reduces the risks of opportunism, and enables freer flow of information (Morton et al., 2004; Levin and Cross, 2004; Baker et al., 2002). All three require significant and ongoing face to face contact (Kolko, 1999; Owen-Smith and Powell, 2004; Morgan, 2004).

In contrast, most work on connectivity between places has focused on trade networks (Zhou et al., 2011; Clark, 2010). While international trade networks have important consequences (Moore et al., 2006; Mahutga and Smith, 2011), this dissertation joins an emerging wave of research exploring the consequences of social connectivity for geographic places (Neal, 2012; Brueckner, 2003).

10 ie Personal connections that can be leveraged in times of need (Lin et al., 1981)
3

Core-Periphery Structure of the US Travel Network

3.1 Introduction

How Does the Travel Network Connect US Cities?

The continental US is well connected. The average US citizen lives within an hour drive of 7 airports, and 3 interstates. However, the first wisdom of network science is that it is not enough to be connected, it also matters how those connections fit into the larger US travel network. This raises the intriguing possibility that US cities might occupy unique positions with the network, and derive unique opportunities from those positions.

Previous literature suggests that network connectivity matters. Scholars have observed that influential cities tend to be traffic hubs – key intermediaries that connect smaller cities to each other (Alderson et al., 2010; Neal, 2011a), and also that cities with better connectivity experience higher rates of employment (Neal, 2012) Nation-state scholars have noted that groups of countries that all trade with each other tend to have better health outcomes (Moore et al., 2006).

In this chapter, I ask, how does the travel network connect US cities? I find that
the US travel consists primarily of two networks. The interstate highway system, which carries 96\% of all passenger volume, and the air traffic system, which connects 78\% of all city pairs that have any passenger traffic between them. Together they move 46 million people each day between 696 cities containing 95\% of the US population. New York, Chicago, Los Angeles, and their surrounding cities are exceptionally central (measured by eigenvector) to this network, but Minneapolis, Denver, St Louis, Dallas, Houston, and San Antonio play a key role in bridging the divide between the east and west. Speaking more generally, major cities are crucial to the connectivity of the US travel system. Just 51 cities make 75\% of the connections between cities. These cities form a dense (0.8), inter-connected core, but also exhibit a strong hierarchical organization.

3.2 Data

3.2.1 City Data

Table 3.1 provides statistics on the city variables contained in these data. The cities examined in this document are core-based statistical areas (CBSA), which are census bureau defined urban areas with high commercial and commuter integration. This definition is superior to considering only the city proper as most cities are tightly integrated with a ring of commuter communities just outside city limits (Logan, 2012; Frey and Zimmer, 2001). Since this project examines the influences of travel networks, the commuter-centric definition of a CBSA also makes it most consistent with the focus of this project. There are 933 CBSA’s in the continental US, of which 696 are examined here. The remaining 237 (which collectively possess only 5\% of the US population) have statistically negligible ties to the national traffic grid.\(^1\)

There are four variables that are purely longitudinal. Year refers to the calendar year to which the data pertain, ranging between 2000 and 2010. This time

\(^1\) See Appendix B for more information on excluded cities
range follows the massive data digitization efforts of the late 1990’s, but precedes the recession-induced data collection cutbacks that took effect around 2010. **GDP (PPP)** refers to the income per capita generated in the US in a given year. All dollar values are adjusted to account for inflation, and reported in 2010 dollars ($_{10}$). Since GDP is a nation-level variable, there is exactly one value for each year. GDP figures originate with the International Monetary Fund (IMF) database. **Kerosene price** is the national average spot market price of kerosene, the principle ingredient in jet fuel. It is priced in 2010 dollars, and comes courtesy of the US Energy Information Agency (EIA). **Inflation** measures the value (%) of the US dollar, relative to the 2010 dollar, based on the change in the consumer price index. It comes courtesy of the US Bureau of Economic Analysis (BEA).

<table>
<thead>
<tr>
<th>Table 3.1: Basic Statistics (2000-2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Year</td>
</tr>
<tr>
<td>Year</td>
</tr>
<tr>
<td>GDP</td>
</tr>
<tr>
<td>Kerosene</td>
</tr>
<tr>
<td>Inflation</td>
</tr>
<tr>
<td>Node</td>
</tr>
<tr>
<td>Elevation</td>
</tr>
<tr>
<td>Temperature</td>
</tr>
<tr>
<td>Rain</td>
</tr>
<tr>
<td>Snow</td>
</tr>
<tr>
<td>Longitude</td>
</tr>
<tr>
<td>Latitude</td>
</tr>
<tr>
<td>Node x Year</td>
</tr>
<tr>
<td>Average Age</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Gasoline Price</td>
</tr>
<tr>
<td>Housing Price</td>
</tr>
<tr>
<td>Net Migration</td>
</tr>
<tr>
<td><strong>Income(DV)</strong></td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Edge x Year</td>
</tr>
<tr>
<td>Same Race</td>
</tr>
<tr>
<td>Same Vote</td>
</tr>
<tr>
<td>Traffic(IV)</td>
</tr>
</tbody>
</table>

28
The US is a rich nation that experienced a wide variety of economic conditions over the 2000-2010 period. Between 2000-2002, US GDP was relatively consistent at about $50,000 per capita (2010 dollars). During 2003-2007, GDP rose to $56,000, only to decline to $53,000 by 2010. Throughout the decade, inflation proceeded at about 3% per year until the recession, at which point it sharply declined to under 1% per year. Kerosene prices were immensely volatile, experiencing three price spikes of 50% or more. However, this likely overstates the effects of price swings, since most airlines use long-term contracts and extensive hedging to smooth out fuel costs.²

There are six variables that vary by city, but are consistent over time because they are geographic features of the city. These variables come from the US National Oceanic and Atmospheric Administration (NOAA), except for latitude/longitude, which come from the US Census Bureau. These include elevation (feet above sea level), temperature (annual average degrees Fahrenheit), rainfall (inches per year), longitude (degrees), and latitude (degrees).

Rain, snow, and elevation serve as proxies for the difficulty of travel and construction on terrain (ECJRC, 2000; Nelson, 2008). Temperature (along with coastal distance) serves as a proxy for natural amenities (Wang and Wu, 2011). It is measured as the absolute deviation of the annual mean temperature from room temperature.

² Note: while oil spot prices are national, price variation may influence regions differently. In general, high prices benefit oil producing regions, while low prices benefit oil consuming regions.
Most US cities are near sea level (elevation), and experience seasonal weather patterns with generally mild winters. Figure 3.1 maps out which terrain features slow the pace of travel and construction in parts of the US. In the west, mountains (red) turn highways into feats of modern engineering. In the south, thick vegetation (green) elevates the cost of clearing land for transit use, due to significant rainfall. In the north, snow (blue) slows down traffic, increase vehicle/road wear, and periodically grounds planes.\(^3\) Six of the ten states with the largest number of interstates lie in the zone where no terrain feature is notably adverse.

There are seven variables that vary by both city and year. **Population** is the average annual residential population of each core-based statistical area (CBSA, ie city). **Housing price** estimates the median price of homes within each CBSA. Prices are in 2010 dollars, and data comes from the US Bureau of Economic Analysis (BEA). **Av. Income** estimates the median income of CBSA residents **Net Migration** is equal to the total number of immigrants minus the total number of emigrants – the net population change attributable to migration. **Gasoline Price** reports the annual average gas price for various regions of the US. The figures are adjusted to 2010 dollars, and come from the EIA. **Av. Education** reports the mean years of education for residents of the CBSA. **Av. Age** reports the average age of the residents of the CBSA.

Substantively, there is wide variation in all of these control variables across US cities. There are small cities (CBSA min is 10,000), and large (over 10,000,000). Some have mostly cheap mobile housing, and while others are dominated by million dollar homes. In some, the average resident has little more than a middle school education, while the average resident of others has completed graduate work. American cities are highly unequal. Moreover, the landscape is shifting rapidly, with many cities

\(^3\) These terrain maps are based on NOAA terrain feature data, and the “drag coefficient” interpretation of them is consistent with Nelson (2008), among others
losing population to migration, while a few grow significantly from it. The total population is also getting older, but some cities remain young due to this migration. These findings are consistent with global trends (UN, 2008; McCormick and Wahba, 2005).

There are two variables that vary by year, and describe a relational characteristic of a pair of cities. **Same Race** denotes the percentage of residents in two CBSA that share a racial/ethnic identity, and **Same Vote** denotes the percentage of residents in two CBSA who voted for the same presidential candidate. Data on the racial and political composition of each CBSA comes from the Census Bureau and Associated Press respectively. I calculate the percentage of residents who share an ethnicity/vote using the network homophily formula (Wasserman and Faust, 1994), shown in equation 3.1:

\[
\frac{\sum_{k} X_{ik} \cdot X_{jk}}{X_i \cdot X_j}
\]

It states that the similarity between cities i and j is equal to the population in city i with identity k times the population of city j with identity k, summed for all k, and then divided by the total population of city i times the total population of city j. It is the probability that two randomly chosen people (one each from i and j) would have the same identity.

**Latitude and longitude** mark the coordinates of a CBSA population centroid, which is point that is closest to the residential addresses of all CBSA residents. In essence, it is the average location of every home in the city. The CBSA centroid serves as the basis for distance calculations. At various points, I measure distance from coastlines, distance from borders, and distance from cities, to name a few. These distance calculations serve as key variables in all of the models in this dissertation.4

---

4 Distances on an ellipsoid (ie the earth) are far different than Euclidean distances, so it is wrong
When calculating the distance between cities, I calculate the distance between their population centroids. In essence, this is the average distance from all residents of the city. When calculating the distance to linear things like highways, coastlines, borders, etc., I calculate the distance to the nearest point.

3.2.2 Network Data

Traffic is the primary network variable in this study. It counts the number of people moving between a pair of cities on an average day through both interstate highway and air traffic systems (i.e., the sum of the two). Data on air traffic come from the US Bureau of Transportation Services (BTS). These data report annual origin-to-destination passenger volumes between all US airports. The data do not suffer from the layover vs. final destination problem that plagues many other sources, because these data list departure to final destination passenger counts.\(^5\) I divide the counts by the number of days per year to ascertain the average daily passenger traffic (AADT). Since airports do not correspond exactly to cities, I assign the traffic from each airport to the nearest CBSA. Sensitivity analysis reveals that this is reasonable – there are few

\(^5\) This is an important feature, as airline hubs can artificially inflate the importance of cities in air travel networks.
ambiguous cases, especially among high traffic airports.

Data on interstate highway traffic comes from 46 separate state Bureau’s of Transportation (DOT). Some states make these data available online. For others, it is necessary to negotiate with DOT traffic engineers for access.

The highway traffic data examined here are the geometric averages of all interstate highway traffic point estimates connecting each pair of cities, controlled for differences in the spacing of traffic counters. Two states (KS, MI) were unable to provide data, and two (AK, HI) are not part of the continental US. The KS and MI traffic data used in these analyses are interpolations\(^6\), in which the population of the area and traffic data from neighboring areas are used to estimate missing traffic measurements.

This project does not include data on railroad ridership at all. There is currently no dataset that provides US ridership data at the CBSA-level of detail. The current BTS dataset uses very large “primary market areas” that are not detailed enough for use here.

\(^6\) To be precise, they are Kriges spatial interpolations (Krige, 1951). This methodology originates with the mining industry, where it was used to predict the location of new mineral deposits, based on the locations and mineral purity of existing veins. A Kriging interpolation assumes that traffic (or any other variable) changes gradually, so surrounding measurements can be used to make guesses about missing measurements.
Table 3.2 provides statistics on the travel network, and figure 3.2 depicts its structure. These data contain average annual traffic volume (AADT) estimates combined road and air traffic between 696 cities in the continental US over the course of the 2000-2010 period.

Each AADT indicates the average number of person-trips moving between two cities on an average day. The network has a strong core-periphery structure. Just 7% of cities are the origin/destination for over 75% of all the trips between cities. In terms of degrees of separation, average pairs of cities sit no more than 3 degrees from each other, with the farthest pair being only 8 degrees removed.

That is, it is only necessary to pass through two cities to reach most cities from most other cities. The air traffic and road traffic components of the network have distinct structural properties that complement each other. Air traffic forms a transitive hub structure – if A is connected to B, and B is connected to C, then A is likely also connected to C. Air traffic is highly transitive because air traffic connects a small group of large cities to each other, where each large city serves as the regional air transit hub for the other cities in its region. This also makes it highly disassortative, meaning that most big city airports primarily connect to a large number of surrounding smaller ones. In contrast, the road traffic forms a non-transitive lattice structure that connects cities within about 83 miles (Median 41 mi) of each other. Since car travel is far slower, there are generally only highways between cities that are close to each other. This cuts down the number of opportunities for transitivity, as most A/C pairs will not be within range of each other, even if A/B and B/C are close enough. This also makes the interstate highway traffic system neither assortative nor disassortative. It connects geographic neighbors, regardless of network

7 I will describe the travel network in much greater detail in the next section, so here I will only discuss its structural properties from a data description perspective.

8 Note: this reflects the economic dependence of smaller cities on larger cities. It is not related to the use of layover hubs in air traffic.
3.3 Methods: Exploratory Network Analysis

This chapter relies heavily on descriptive network analysis methods. Network science examines “edges,” which are connections / ties/ links between “nodes.” Nodes can be anything which can be affected by its edges, and how those edges position it relative to other nodes within the “network” – the linked system that emerges from edges binding nodes into a connected whole.

Figure 3.3 provides an example illustration of a network. The network shown is the Hawaiian air traffic network – a network of air traffic passengers moving between airports in Hawaiian cities. The nodes (airport cities) are represented as circles, and the edges (passengers) are represented as lines between the cities. Edge thickness and color indicates the total number of passengers, which thick, bright, red lines representing many passengers, and thin, dark, blue lines representing few.\(^9\)

**Mathematical Representation**

Networks can be represented as a matrix. Table 3.3 represents the data behind figure 3.3. Each cell represents the total number of passengers moving between the city

---

\(^9\) The airports with no passenger traffic between them are either on the same island, or had so few passengers that I rounded it down to zero for this illustration. For example, 16 passengers flew between Maui and Lahaina. For simplicity, directionality is ignored, so the passenger volume is the total of passengers moving from i to j and from j to i.
Table 3.3: Hawaiian Air Traffic Matrix Example

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kamuela</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>900</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1400</td>
<td>0</td>
</tr>
<tr>
<td>Hana</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1400</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2500</td>
<td>0</td>
</tr>
<tr>
<td>Kona</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>83600</td>
<td>1200</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>767800</td>
<td>19700</td>
</tr>
<tr>
<td>Maui</td>
<td>900</td>
<td>900</td>
<td>106300</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>200</td>
<td>14900</td>
<td>1119100</td>
<td>95800</td>
</tr>
<tr>
<td>Lahaina</td>
<td>0</td>
<td>0</td>
<td>1300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>51800</td>
<td>0</td>
</tr>
<tr>
<td>Lanai</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45700</td>
<td>0</td>
</tr>
<tr>
<td>Kalaupapa</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1600</td>
<td>0</td>
</tr>
<tr>
<td>Hoolehua</td>
<td>0</td>
<td>0</td>
<td>14500</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>50400</td>
<td>0</td>
</tr>
<tr>
<td>Honolulu</td>
<td>1500</td>
<td>2500</td>
<td>759200</td>
<td>1091300</td>
<td>54600</td>
<td>46200</td>
<td>1600</td>
<td>50900</td>
<td>0</td>
<td>835100</td>
</tr>
<tr>
<td>Kauai</td>
<td>0</td>
<td>0</td>
<td>18500</td>
<td>82300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>869600</td>
<td>0</td>
</tr>
</tbody>
</table>

listed in the row and the city listed in the column during the given year. For example, over a million passengers moved between Honolulu and Maui\(^{10}\) during the example year. In contrast, no passengers flew between Kalaupapa and Hoolehua during the example year – they are 10 miles apart on the same island. Representing networks as matrices is advantageous because there is a well-developed body of mathematics that can be used to investigate matrices. The rest of this section applies mathematical and algorithmic methods to the matrix representation of the network. For equation purposes, \(X_{ij}\) represents the network matrix, where \(i\) indexes the row city, and \(j\) indexes the column city.

**Centrality**

A node is central if it is well connected to the rest of the network. In figure 3.3, node centrality is represented with color. Bright red indicates high centrality, and dark blue represents low centrality. Honolulu is the most central node. It has air traffic connections with every other node, and, in fact, serves as origin/destination for the vast majority of Hawaiian air traffic.

There are actually many different definitions of centrality, because there are many ways that a node can be central to the network. This paper relies on eigenvector

\(^{10}\) The numbers are high, given the population of Hawaii. This most likely reflects the enormous amount of tourists passing through Hawaii in a given year.
centrality. By the eigenvector criteria\textsuperscript{11}, a node is highly central if it connects to many nodes, and its connections (called “alters” in network terminology) also connect to many nodes. Conceptually, think of edges like water pumps between tanks. Drop enough food coloring in one tank, and, eventually all tanks will change color. The most central node has large pump connections to tanks that also have larger pump connections to other tanks. Drop food coloring in the most central tank, and all tanks will change color in the least amount of time.

Speaking at a more technical level, eigenvector centrality has ideal conceptual properties for this dissertation, with its focus on income flow between cities. A matrix (ie a network) is a system of transformations that can be applied to a vector. For example, the tank-pump system is a system that transforms how much food coloring is in each tank. The original “vector” in this metaphor is that all food coloring is in one tank, and none is in the others. The pump system transforms that vector into one where there is food coloring in all tanks. If the pump system was permitted to operate for a long time, the first eigenvector of that system describes where food coloring would end up, regardless of where that food coloring was initially placed. It is a purest measure of the how that flow system influences the distribution of something. Since this dissertation examines how the flow of travel networks (ie the pumps) influence the prosperity (ie the food coloring) of cities (ie the tanks), eigenvector centrality most accurately describes how the travel network influences cities over time.

\textsuperscript{11} Formally, an eigenvector satisfies this equation: $\lambda v = Xv$, where $\lambda$ denotes an eigenvector, $X$ denotes a matrix (ie network), and $v$ denotes a scalar. Given certain conditions, matrices have as many eigenvectors ($\lambda$) as they have rows/columns. However, one eigenvector will have the largest eigenvalue ($\nu$). At the risk of over-simplifying a complex and important mathematical concept, the first eigenvector reveals the random walk tendencies of the matrix – if something were to randomly move through a network, the first eigenvector reveals which nodes it will pass through most often (Borgatti, 2005)
Community

Community is the idea that some groups of nodes have stronger ties to each other than to the rest of the network. For example, a group of adolescents who hang out together will have more friendship ties with each other than with non-group members. Moreover, they will share more friends in common.

In city networks, larger cities serve as hubs for smaller cities. As such, communities in city networks are generally related to having mutual hub cities in common. For example, the dashed gray lines in figure 3.3 indicate the two major groups of Hawaiian cities. The network divides into two groups: cities that send more traffic to Honolulu (Kauai, Kona, Lahaina, Lanai) and cities that send more traffic to Maui (Kalaupapa, Kamuela, Hana). In the Honolulu group, Kauai, Kona, and Lahaina also send traffic to each other, which strengthens their ties to that group.

Figure 3.3 also highlights the difficulties in dividing nodes into communities. Every node in the Maui group sends some air traffic to Honolulu, and the majority of Honolulu group members send traffic to Maui. The case of Hoolehua is particularly difficult, as it sends similar amounts of traffic to both Honolulu and Maui. To deal with these ambiguities, scholars have developed formula for calculating how well communities fit the observed data – the modularity of the network (Newman, 2006; Fortunato, 2010). Equations 3.2 and 3.3 report the two components of the Newman (2006) formula for modularity. $X_{ij}$ represents a network (such as the matrix in previous table), where i and j index individual nodes (ie an individual row and column of the matrix).

\[
P_{ij} = \frac{\sum_i X_{ij} * \sum_j X_{ij}}{\sum_i \sum_j X_{ij}}
\]  

(3.2)

The first component (3.2) defines our expectations of the network. If there were
no groups, how would we expect the network to look? According to 3.2, the expected volume of traffic between nodes $i$ and $j$ ($P_{ij}$) is equal to the amount of traffic passing through node $i$ ($\sum_i X_{ij}$) times the amount of traffic passing through node $j$ ($\sum_i X_{ij}$), divided by the total amount of air traffic passing through all nodes combined.\footnote{This is the same as the logic underlying the chi-square formula for expected value} Conceptually, we would expect that nodes would be most likely to connect to nodes with large volumes of traffic. For example, in figure 3.3 we would expect that nodes have a higher chance of connecting to Honolulu because Honolulu is the origin/destination of a massive amount of traffic.

\[
Q = \frac{X_{ij} - P_{ij} \cdot \delta(C_i, C_j)}{\sum_i \sum_j X_{ij}}
\]  

(3.3)

The second component (3.3) defines how different the actual network is, compared to our expectations of it in the absence of groups. $P_{ij}$ is the outcome of formula 3.2 – our expectations for how the network would look if there were no groups. $\frac{X_{ij} - P_{ij}}{\sum_i \sum_j X_{ij}}$ scores each edge according to how non-random it is, given how much traffic each node sends ($P_{ij}$) and how much total traffic there is in the network ($\sum_i \sum_j X_{ij}$). $\delta(C_i, C_j)$ is a function which yields 1 if $i$ and $j$ are assigned to the same group and 0 otherwise. $Q$ scores the network according to how unusually well-connected each group is, compared to a model of group-less, random connectivity – the modularity score. By calculating this modularity scores for different group assignments, computer algorithms can find nearly optimal group assignments – the “fastgreedy” method. The aforementioned Honolulu and Maui groups were detected with the Csardi and Nepusz (2006) implementation of “fastgreedy” modularity maximization strategy, based on the Newman (2006) modularity formula.

However, much like centrality, some modification is necessary to leverage modularity in specific contexts. For example, scholars have developed modularity formulas
to partition networks of shared affiliations (Barber, 2007), and networks with a temporal component (Mucha et al., 2010), among others. For this project, the challenge is that cities can be proximate in both geographic space and the network. A useful community has heavily intertwined traffic flows and is geographically contiguous. Equation 3.4 reports how I factor proximity into modularity calculations.

\[ X_{ijd} = X_{ij} \cdot \min(1, \frac{D_{ij}^{-\beta}}{2s}) \]  

(3.4)

\( D_{ij} \) is the distance between cities i and j. \( s \) is the average commuter distance, which reflects the current scale of distance (See 2.2.2). \( X_{ij} \) is the total traffic between cities i and j. \( X_{ijd} \) is the adjusted traffic counts, now weighted to reflect geographic information. The \( \min \) function ensure that geography can attenuate traffic, but cannot increase traffic counts. Otherwise, this formula would increase the traffic counts of any city pair that lay closer than \( 2s \) from each other. This is consistent with previous scholarship on commuter patterns, which establishes that travelers become increasingly insensitive to distances for short distances (Giuliano and Small, 1993). \( \beta \) is a tunable parameter. When \( \beta \) is 0, \( X_{ijd} \) reports the traffic counts alone. When \( \beta \) is a large number, \( X_{ijd} \) reports almost entirely geographic position information. Between these extreme values lies a range of \( \beta \) coefficients that will detect increasingly small urban communities.

In essence, I downweight the traffic between cities in proportion to the distance between them. Then, I use the standard modularity maximization procedure (equation 3.3) to find modular groups in \( X_{ijd} \), instead of \( X_{ij} \).13 This step is necessary to prevent the standard community detection procedure from overlooking the hierar-

\[ Q = \frac{X_{ijd} - P_{ij}}{\sum \sum X_{ijd}} \cdot \min(1, \frac{D_{ij}^{-\beta}}{2s}) \cdot \delta(C_i, C_j) \]

However, here it is used to weight \( X_{ij} \):  

\[ Q = \frac{(X_{ij} \min(1, \frac{(D_{ij}/2s)^{-\beta}}{\sum \sum X_{ij}}) - P_{ij}) \cdot \delta(C_i, C_j)}{\sum \sum X_{ij}} \]

13 This procedure could also be done to the modularity matrix itself:
chical aspect of the US city system. There are 51 prime cities that are key economic focal points for their regions. Smaller cities connect to prime cities, and prime cities connect to each other to form a cohesive national system. The standard procedure will (understandably) lump all prime cities together. However, there is more insight to be gained from understanding how prime cities integrate regions, than simply lumping them into an elite group – we already know that New York, Los Angeles, etc are superstars, but what regions fall under their sway?

This is not the only approach to geographically-sensible community detection. For the purposes of this project, I treat geographic contiguity as a importance feature, because of the role of regions in classic central place theory (Christaller, 1933; Losch, 1938), an idea that remains conceptually important to modern work on cities and regions (Partridge et al., 2008; Taylor et al., 2010). However, other research explorations may find geography to be less of a feature, and more of a confounder – making connectivity easier some actors, regardless of their propensity to connect with each other. Expert et al. (2011) modify the modularity null model to control for geography. Equation 3.5 reports how that approach would translate into a potential alternate null model in this project. In the Newman-Girvan approach, edges between nodes are more likely when both i and j have many edges, relative to the total number of edges in the network. In 3.5, edges between nodes are more likely when both i and j have many edges, relative to the total number of edges between nodes at that distance ($\sum_{ij} d_{ij} X_{ij}$) over the total number of edges we’d expect if distance had no effect ($\sum_{d_{ij}=d} \sum_i X_{ij} \cdot \sum_j X_{ij}$). In essence, 3.5 calculates the discrepancy between the observed number of ties among all i and j at distance d, and the number we’d expect at random, given the number of ties incident to i and j. It then diminishes our expectation for the number of ties between i and j, based on that discrepancy.
\[ P_{ij} = \sum_i X_{ij} \cdot \sum_j X_{ij} \cdot \frac{\sum_{d_{ij}=d} A_{ij}}{\sum_i X_{ij} \cdot \sum_j X_{ij}} \] (3.5)

The rest of this chapter presents descriptive analyses of the US travel network. While the analyses are varied, they all rely on combinations of the methods presented here: visual inspection, centrality and community analysis.

3.4 Results: US Travel Networks Has a Core-Periphery Structure

3.4.1 Inter-City Traffic Volumes

While Americans move by many means, the interstate highway system and the national air traffic system are its primary transportation networks. They knit the vast continental US together into a highly integrated nation. The interstate system provides cheap, easy transit among cities within a hundred miles of each other, while the air traffic system provides fast, adaptive connectivity among more distant cities. Of the 46 million Americans moving through these systems each day, 96% will pass through the interstates, which facilitate the overwhelming majority of inter-city trips. However, the air traffic system also plays a crucial role in travel, facilitating over 75% of the connections between all connected pairs of cities. In other words, the highway system accounts for most travel volume, while the air traffic system accounts for most of the network connectivity. Between them, the inter-city travel network connects 696 (of 933) distinct urban areas, representing 95% of the American population and 99.9% of all trips between cities.

Most traffic links convey a relatively small number of passengers. Only a small

\footnote{That is to say, there are 3,198 pairs of cities that have traffic moving between them. Of these pairs, over 75% of them have only air traffic moving between them.}

\footnote{Scholars have also argued that air traffic is disproportionately important because of its high price point – people who move by air either have greater personal wealth or represent wealthier organizations. As such, they may be disproportionately important to economic processes.}
number of inter-city traffic links average more than (for instance) 10,000 passengers per day, and almost all of those high volume links occur between large cities within 80 miles of each other. The remainder connects major cities to New York City and/or Los Angeles. The NY-LA flight route is particularly massive, hosting traffic volumes on par with a rural interstate (≈ I-80 near Cheyenne WY). Figure 3.4 depicts the US traffic system. The urban areas are depicted as interlocking tiles, where each tie depicts the area that is closer to that city than to any other city. For reference purposes, urban areas with more than a million inhabitants (mega-cities) are outlined in black, while all other cities are outlined in gray. Traffic connections that traverse more than 300 mi are showed in blue, while shorter links are rendered in red. Traffic connections with less than 10,000 passengers per day are represented with faded coloring.

\footnote{ie a Voronoi decomposition (Okabe et al., 2000).}
Using the eigenvector centrality score described above, which focuses on the flowsystem nature of the network, three regions of the US are extraordinarily important – the Los Angeles-Las Vegas-Phoenix triangle, the Boston-New York-Philadelphia-Washington corridor, and Chicago-Milwaukee-Minneapolis triangle. About half of America’s other mega-cities are also important, primarily because they connect local regions to the big three. Figure 3.5 reports the importance (red) of US cities according to their eigenvector centrality. The urban areas are depicted as interlocking tiles, where each tile depicts the area that is closer to that city than to any other city. Urban areas with more than a million inhabitants (mega-cities) are outlined in black, while all other urban areas are in gray. Most high centrality cities are mega-cities, but not all mega-cities have high eigenvector centrality.

3.4.2 Structure of the travel network

Cities with high volumes of traffic likely experience higher levels of economic interdependence and have more cross-city social ties. If so, this implies a socio-economic geography, analogous to physical geography, in which cities are “closer” when they have stronger traffic ties. In such a socio-economic space (estimated via MDS), America’s mega-cities are nearly all close neighbors – often closer to each other than to their own regions. However, the arid, mountainous terrain in the center-west effectively splits the country into Eastern and Western blocks. Texas plays a key role in bridging east and west. In addition to Texan cities, San Diego, Minneapolis, and St. Louis also play key bridging roles.

Figure 3.5: Eigenvector Centrality
Figure 3.6 arranges US cities according to their traffic “closeness,” using an MDS transformation of the traffic between cities. Urban areas with more than a million inhabitants (mega-cities) are outlined in black, while all other urban areas are outlined in light gray. Each urban area is colored to reflect the general region from whence it came. “Cooler” colors, like purples and blues, indicate cities from west of the Mississippi river, while “warmer” colors, like red, green, and yellow, indicate cities from east of the Mississippi.

For reference, I have placed the socioeconomic space visualization (left panel in Figure 3.6) next to the core-periphery sociogram from 3.2. The red nodes in the right panel correspond to the major cities (regional primes) outlined in black in the left panel. The cities in the center of the left panel correspond to the blue nodes connected to red nodes through purple lines.

3.4.3 Hierarchy of Central Places in the US

Due to the dominance of regional prime cities, the US traffic system has a distinctly hierarchical structure.\(^\text{17}\) In fact, only a handful of cities are responsible for the

\(^{17}\) By hierarchy, I am making reference to the classic central place theory (Christaller, 1933; Losch, 1938) conception of city hierarchy. The idea is that smaller cities reply on bigger cities for advanced/rare services (e.g., brain surgery), and, often, employment as well. These bigger cities tend to have a similar relationship with even larger cities. For example, Durham NC residents rely on Raleigh NC for advanced services. Raleigh NC relies on Charlotte NC for even more advanced...
tremendous inter-connectivity of US cities. These regional prime cities tend to be the central point for a group of surrounding cities. For convenience, I will refer to these groups as “constellations.”

Using the modified modularity formula described in equations 3.3 and 3.4, I examined how the US travel network breaks into these distinct constellations. The specific number of constellations depends on $\beta$, which controls how much weight is given to geographic proximity vs network proximity. When geography matters less, the formula reveals the large regions surrounding the cities at the top of the US city hierarchy. When geography matters more, the formula reveals smaller local areas that are tightly integrated with the local hub city.

For example, at $\beta = 1.5$, there are 28 distinct constellations, while at $\beta = 6.2$, there are 78. These constellations can be treated as nested. For example, the Atlanta constellation at $\beta = 2$ breaks down into Atlanta GA, Chattanooga TN, Montgomery AL, Birmingham AL, and Macon GA constellations at $\beta = 6.2$.

Figure 3.7 reports the 78 constellation solution. City network scholars have been engaging with central place theory for years (Neal, 2011b; Taylor et al., 2010), but the approach presented here may be fairly unique. Classic central place theory formally predicts a system of “honeycombed” market areas, where no honeycomb in the same hierarchical level overlaps, but each level has non-monotonic overlap with services.
the next/previous level (Openshaw and Veneris, 2003). Equation 3.4 may be one of the only network science methodologies that can replicate this market area spacing and overlap.

For the rest of this section, I will discuss the 28 city solution, because it provides the “high-altitude” view of this network. These 28 cities serve as the key central points for large groups of US cities. Along with the border cities of each constellation, they also play a major role in connecting their constellation to other constellations. The interstate system is especially important in this process. Many urban constellations have a distinctly elongated shape because they spread along one of the primary interstates. Figure 3.8 presents the 28 constellation network.

The inset panel of 3.8 presents a further level of abstraction. Inter-constellation
traffic binds the 28 areas into 6 distinct regions: The West, The Midwest, Greater Texas, The South, The Steel Belt, and the East Coast. The connections between these regions overwhelmingly reflect their geography, with Texas and the Eastern Midwest playing a key role in uniting the nation. However, East Coast - West Coast ties are surprisingly strong, given their distance.

3.5 Conclusion

How Does the Travel Network Connect US Cities?

As a core-periphery system composed of a highway lattice between nearby cities (96% of all traffic), and a nearly complete air traffic clique of air traffic between regional prime cities.

Using descriptive network analysis techniques, I examine the structure of the US travel network. City hierarchy is a key feature of this network – smaller cities connect to larger cities that are central to the regional economy and travel network. These larger cities have a similar relationship with even larger cities that are central to even larger regions. At each level of the hierarchy, the network splits the US into geographically contiguous regions. In general terms, New York and Los Angeles organize the US into two counter-balancing blocks. In terms of centrality, the greater New York, Los Angeles, and Chicago areas enable massive amounts of travel flow between US regions. In terms of bridging regions (ie “brokering”), Minneapolis, Denver, St Louis, Dallas/Houston play key roles in bridging the East-West divide. In terms of local integration, 51 cities make 75% of the connections (ie unique edges) between US cities, integrating location constellations of cities into a coherent whole. In terms of central places, 78 cities serve as commercial hubs for local economic activity. At higher levels of the commercial hub hierarchy, traffic passes through airports, which make 78% of the connections between US cities. At lower levels,
traffic passes primarily through highways, which carry 96% of all traffic volume.

While many scholars have examined the air traffic network in isolation (Lansing et al., 1961; Shin and Timberlake, 2000; Smith, 2001; Rozenblat et al., 2006; Neal, 2011a), far fewer have examine the full network of people moving between cities because of the difficulty of acquiring road traffic data.\textsuperscript{18} Since air and highway travel perform complementary roles,\textsuperscript{19} and since highways carry 96% of all travel, this is an important addition.

Speaking more broadly, these finding suggest that there is a great deal of structural differentiation with the US travel network, and that important processes happen at different levels of that structure – oscillation, flow centrality, bridging, and local summation, just to name a few. Each process creates opportunities for network structure to influence city outcomes at multiple levels of the hierarchy.

... 

Chapter 3 explores the connections between cities, but offers no insights into whether and how those connections matter. The subsequent chapter (4) will explore the association between the city network and a consequential outcome – economic prosperity.

\textsuperscript{18} A notable exception is the Balcan et al. (2009) epidemiological study, which uses commuter data.

\textsuperscript{19} Essentially, air traffic “fixes” the long geodesic paths created by the road lattice – it is a “reply” to the road network.
City Prosperity Flows Through Travel Networks

4.1 Introduction

Is Personal Travel Between Cities Associated With Prosperity Within Them?

On an average day, over 46 million people move between US cities. These movements often involve economically relevant activity. Migrants carry their labor and human capital to a new city labor market. Commuters transport their wages from sites of economic production to sites of residence and recreation. They also transport their labor to sites of production, creating economic value at those sites. Business travelers conduct transactions on behalf of commercial organizations. All travelers carry information and experience, which are increasingly valuable in the modern “knowledge economy,” and also social capital, which creates access to resources through relationships. Moreover, since the travel network involves many actors conducting economically valuable activities many times over the course of an average day, economic value could potentially travel much farther per day than any individual person. This raises the intriguing possibility that the movement of people could exert influence on the prosperity of cities, because it facilitates the rapid flow of
economically relevant resources like information, money, experience, trust, and labor outputs.

Previous literature suggests that this is plausible. Scholars have theorized that local economies benefit from geography-spanning exchange networks, because they make resource use more efficient (Krugman, 1979), improve the quality of the local information supply (Castells, 1999), and supply higher quality labor (Beaverstock, 1994). Empirical research supports these expectations. Recent studies have concluded that highway construction bolsters local economies (Michaels, 2008), increased air traffic bolsters employment (Neal, 2011a), and dense trading networks correspond to lower national infant mortality, a generally recognized proxy for poverty and low economic development (Moore et al., 2006; CIESIN, 2000).

In this chapter, I ask, does the movement of people between cities influence city prosperity? Using US highway/air travel and city census data (2000-2010), I estimate the influence of network flow on city median income via two variations of network ARMA linear models. Both models show that there is a statistically significant association between the travel network and median incomes. However, they disagree over how much the travel network matters. The first set of models suggests that network flow may be responsible for up to 16% of the variation in city median incomes, and that percentage has been growing every year since 2003. The second set disagrees, explaining only 2% of the variation.

4.2 Data

The city data used here are described in 3.2.1, and the network data are described in 3.2.2. Some network simplication is necessary to estimate stable models, due to the structure of this particular network. In the first analysis, I aggregate the data into 78 city constellations using the community detection procedure from 3.3. In the second, I dropped the 192 cities with less than 50,000 residents.
4.3 Methodology: Network Auto-Regressive Model

4.3.1 Auto-Regressive Model

To measure network influences on prosperity, one must simultaneous account for how cities influence themselves, and how cities influence each other. The auto-regressive model family was developed for exactly this purpose. Auto-regressive models take all the analytical benefits of linear modeling, and apply them to situations where the outcome in one location is at least partly dependent on the outcomes in other locations (Butts, 2008; Ward and Gleditsch, 2008; Ord, 1975). Networks create such dependency. In this section, I use Butts (2010) implementation of the well-established network auto-regressive, moving average (ARMA) model.

The ARMA posits a system in which the outcome for nodes (Y) is a function of both their internal attributes (X) and the outcome of their neighbors, connected through networks (W):

\[ Y_t = \rho W \cdot Y_{t-1} + X \beta + \epsilon \quad (4.1) \]

In this case, Y is a vector of city average incomes. Rho (\( \rho \)) is a coefficient that indicates how much the network contributes to neighboring outcomes. \( X \beta \) measures how the internal attributes of Y influence its outcomes, and is equivalent to the standard linear regression formula. Epsilon (\( \epsilon \)) models the error term. This equation is iterative. As each node influences its neighbors, those neighbors go on to influence their neighbors, creating pathways of indirect influence with each successive iteration.

To solve this equation, we find the equilibrium of the system by setting \( Y_t \) to be equal to \( Y_{t-1} \), a condition which is only true when the system is in steady state.\(^1\)

\(^1\) This solution only applies if the system can be said to reach equilibrium at a faster rate than the rate of network change. The subsequent chapter demonstrates the high degree of inertia in traffic volume change, so this is a reasonable assumption.
\[ Y = \rho \mathbf{W} \cdot Y + \mathbf{X}\beta + \epsilon \]  

(4.2)

However, this equation needs to be reformulated before it can be solved, since it places \( Y \) on both sides of the equation. Butts (2008) offers this solution:

\[ Y = (I - \rho \mathbf{W})^{-1} \cdot (\mathbf{X}\beta + \epsilon) \]  

(4.3)

With the auto-regressive error term modeled similarly:

\[ \epsilon = (I - \psi \mathbf{W})^{-1} \cdot \mathbf{v} \]  

(4.4)

These models use log-linear variables. There are two reasons for this. First, the log-linear framework is one of the few that can successfully process a distribution as severely skewed as passenger traffic – one that ranges from dozens of passengers to millions. Second, this is highly consistent with previous literature (Zipf, 1946; Stewart, 1960; Brown et al., 1970; Bergstrand, 1985; Zhou, 2010).

4.3.2 Network Simplication

Where 696 cities have significant connections to the national travel network, some cities are far more connected than others. Many of the smallest cities only connect to the rest of the network through ties to the local prime cities. Working in the ARMA modeling framework, it is difficult to estimate models when there are so many nodes with poor connectivity to the rest of the network – some network simplification is necessary. This chapter contains two analyses, each of which overcomes this problem in a different way. The first analysis uses the community detection methods from section 3.3 to simplify the network into 78 “constellations” of densely interconnected cities, and then examines the network influences of those 78 for each year between
2000-2010. The second analysis drops the 192 cities with populations under 50,000, and the estimates network influences across the 2000-2010 period as a whole. The first analysis sacrifices spatial resolution (n= 78), but preserves more temporal detail with a model for each year. The second sacrifices temporal resolution (2000-2010 pooled), but preserves more spatial variation (n= 504).

4.4 Results: Traffic Has a Strong Association with Prosperity

4.4.1 Analysis 1: An Annual Examination of 78 City “Constellations”

Table 4.1 reports a model of city prosperity for each year, based on only the attributes of that city. This model has no network component. It is an ordinary least squares (OLS) model, since the ARMA model uses OLS for its non-network component. Since the model is essentially the ARMA model without the network effects, it serves as a baseline against which the ARMA models can be measured. Table 4.1 reports the best fitting OLS model that can be achieved with these data, given constraints on the total number of independent variables. 4.1 serves as the standard against which to measure the ARMA. The ARMA must outperform it to demonstrate that the network matters for median income. 4.1 is a strong fit for the data, accounting for 60% of the variance for most years. However, model fit declines somewhat after the financial crisis, leveling off around 50%.

The three strongest predictors of city median income are total population, median education and number of workers per capita, which are undoubtedly endogenous. Total population exerts a strong, consistent influence across all years. Average education predicts median income before the financial crisis, but the coefficient for education drop after the crisis. In contrast, the number of workers per capita – in essence, the true employment rate – increases in predictive power after the crisis.

While these data are generally unkind to the assumptions that underlie OLS, log-linearizing the variables partly ameliorates those issues.
Table 4.1: OLS Model of City Prosperity

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons</td>
<td>0.041</td>
<td>0.036</td>
<td>0.035</td>
<td>0.031</td>
<td>0.026</td>
<td>0.025</td>
<td>0.024</td>
<td>0.022</td>
<td>0.022</td>
<td>0.024</td>
<td>0.027</td>
</tr>
<tr>
<td>Education</td>
<td>1.205</td>
<td>1.343</td>
<td>1.279</td>
<td>1.02</td>
<td>1.13</td>
<td>1.212</td>
<td>1.339</td>
<td>1.358</td>
<td>1.05</td>
<td>0.921</td>
<td>0.77</td>
</tr>
<tr>
<td>Gasoline</td>
<td>0.667</td>
<td>0.337</td>
<td>0.709</td>
<td>0.527</td>
<td>0.767</td>
<td>1.047</td>
<td>1.17</td>
<td>0.76</td>
<td>1.315</td>
<td>0.696</td>
<td>0.551</td>
</tr>
<tr>
<td>Migration</td>
<td>0.002</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>Race</td>
<td>0.37</td>
<td>0.288</td>
<td>0.249</td>
<td>0.251</td>
<td>0.094</td>
<td>0.033</td>
<td>0.086</td>
<td>0.069</td>
<td>0.02</td>
<td>0.139</td>
<td>0.071</td>
</tr>
<tr>
<td>Workers</td>
<td>0.684</td>
<td>0.457</td>
<td>0.688</td>
<td>1.11</td>
<td>1.014</td>
<td>1.052</td>
<td>0.975</td>
<td>1.436</td>
<td>1.326</td>
<td>1.518</td>
<td>0.518</td>
</tr>
</tbody>
</table>

r^2 | 0.615 | 0.604 | 0.635 | 0.648 | 0.628 | 0.577 | 0.593 | 0.56 | 0.512 | 0.537 | 0.5 |
N  | 78    | 78    | 78    | 78    | 78    | 78    | 78    | 78    | 78    | 78    | 78    |

Together, these three indicators drive the model, predicting median income across city constellations. These predictions are consistent with previous scholarship (Bettencourt et al., 2004; Moller et al., 2009). Table 4.2 adds network auto-regressive influence to the 4.1 model. For all years, the network term improves the model fit by a statistically significant margin.

Table 4.2: Network Auto-correlation Model of Constellation Prosperity

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>0.115</td>
<td>0.115</td>
<td>0.095</td>
<td>0.098</td>
<td>0.09</td>
<td>0.072</td>
<td>0.071</td>
<td>0.073</td>
<td>0.067</td>
<td>0.078</td>
<td>0.086</td>
</tr>
<tr>
<td>Gasoline</td>
<td>0.754</td>
<td>0.359</td>
<td>1.108</td>
<td>0.383</td>
<td>0.917</td>
<td>2.356</td>
<td>2.219</td>
<td>1.847</td>
<td>2.723</td>
<td>1.041</td>
<td>1.268</td>
</tr>
<tr>
<td>Migration</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>Diversity</td>
<td>0.148</td>
<td>0.224</td>
<td>0.204</td>
<td>0.192</td>
<td>0.252</td>
<td>0.339</td>
<td>0.35</td>
<td>0.48</td>
<td>0.303</td>
<td>0.389</td>
<td>0.563</td>
</tr>
<tr>
<td>Employment</td>
<td>0.205</td>
<td>0.027</td>
<td>0.122</td>
<td>0.03</td>
<td>0.246</td>
<td>0.807</td>
<td>0.911</td>
<td>0.626</td>
<td>1.371</td>
<td>0.544</td>
<td>0.902</td>
</tr>
<tr>
<td>Travel</td>
<td>.006</td>
<td>.006</td>
<td>.005</td>
<td>.006</td>
<td>.005</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
<td>.005</td>
<td>.006</td>
</tr>
<tr>
<td>r^2</td>
<td>0.721</td>
<td>0.722</td>
<td>0.732</td>
<td>0.722</td>
<td>0.722</td>
<td>0.695</td>
<td>0.696</td>
<td>0.68</td>
<td>0.653</td>
<td>0.679</td>
<td>0.658</td>
</tr>
<tr>
<td>N</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
</tr>
</tbody>
</table>

Figure 4.1 summarizes the findings of the network sensitive analysis. The addition of a network auto-correlation term improves the fit of this model of median income. This improvement is statistically significant, and upwards of 10%. This supports the assertion that networks make an important contribution to the economic prosperity of US cities. Moreover, the estimated influence of networks appears to be growing steadily, as reported in the second panel of 4.1.\(^3\) The rise is consistent with previous

\(^3\) A spike in fuel prices is a possible culprit for the pronounced dip in 2003. The impending invasion of Iraq caused world oil prices to spike in late 2002, and they remained elevated through
The "information economy" may be making the network more important for productivity (Hall, 1997; Derudder and Witlox, 2004; Neal, 2010).

4.4.2 Analysis 2: A Decade-Wide Examination of 504 Cities

Table 4.3 reports the results of the second analysis. Compared to the previous, city attributes predict a larger share of the variation in 4.3. The "Minimal" model (1st column) predicts inflation-adjusted city median income with just the total population of the city. Con-

the early years of the war. While there were several fuel price spikes during this decade, this one was exogenous (ie political, not economic in cause), and sudden.
sistent with the economies of size literature, population size is a strong predictor of economic prosperity. It accounts for over half of the variation in city median incomes.

City size is, in essence, shorthand for a wide variety of economic processes that all happen to get more efficient when more people are involved. The “Economic” model disentangles three of these effects. Gas price, employment, and land price, speak to travel costs, economic opportunity, and the economic value of being located in that city. As expected, the effect of population diminishes when these economic factors have been controlled, and land value/employment correspond to higher median incomes. However, the effect of gas price is puzzlingly positive, and remains so throughout the models. This may reflect the strong influence of taxation policy on gas prices – gas taxes fund the maintenance of most US transportation infrastructure.

The “Social” model adds in socio-demographic factors – the median age and education of the city population. These coefficients are positive and significant, as would be expected from the earnings literature. They also diminish the effects of all variables in the “Economic” and “Minimal” models.

The “Network” model incorporates the influences of the travel network, both as model terms, and as a component in the error estimation process. In 4.3, the travel network is broken into two components. The first measures only the influence of short-distance network connections. The second measures all connections, both long and short. The short-distance network, which carries the vast majority of inter-city traffic is statistically significant and positive. This supports the idea that the travel network is a positive influence of median income. However, the effect of the longer-distance network is small, and statistically insignificant – the influence of networks between regions is dwarfed by their influence within regions. The addition of network

4 Age is modeled with one coefficient, instead of two, because city (pop. > 50,000) median ages do not have the extreme variation that individual ages have. The vast majority of median ages fall between 32 and 44.

5 Racial diversity does not converge well in this model, so it is excluded.
terms does improve model fit, but the improvement is much more modest than in the previous analysis.

4.5 Conclusions

Is the movement of people between city constellations associated with prosperity?
   Likely.

Using US highway/air travel and city census data (2000-2010), I estimated the influence of network flow on city median income via network ARMA linear modeling. Some network simplification was necessary to create stable model estimations, so I conducted two analyses, using different simplification strategies.

In the first, I examined the estimated influence of the network on clusters of densely connected cities. As a baseline, I first calculated the influence of standard independent variables on city median incomes. As expected, the OLS model reveals that total population, median education, and employment are among the strongest predictors of city median income. The OLS model explained half of the variation in city median incomes. Next, I factored network flow into the model. The network term changed the significance of other variables – making employment insignificant, and gasoline price significant. Network flow was also significant in its own right, and explained an additional 10-16% of the variance in median incomes. In addition, network flow explains an increasingly large amount of variation over time. In 2003 it explained 10%. By 2010, that figure had risen to 16%.

In the second analysis, I estimated the influence of the network on medium and large cities (50,000+), pooling data across years. In this model, city attributes accounted for a much larger portion of the variation in income, and the network only accounts for an additional 2%. This finding is an important cautionary tale – the scale of analysis really matters for estimation of the network effect (Neal, 2014).
a field where the unit of analysis can range from interpersonal to international, scale choices have substantive consequences. That said, the travel network in the second analysis still exerts a statistically significant effect on median incomes, especially the massive, local travel connections between propinquitous cities.

These findings extend the findings of previous literature. Neal (2011a) finds that increased air traffic corresponds with increased employment rates in US cities. Examining air and highway travel, this chapter adds that travel networks also shape the median income of employees in those cities. It also suggests that air traffic may not be the best travel network to use in these studies – most US travel connectivity passes through local highways, not long distance air traffic.

These findings also offer important insight into the economics of cities. People are not necessarily the intuitive choice for a study of economic networks. Generally, scholars turn to trade networks for such studies. However, organizational studies have repeatedly shown the interpersonal connectivity has economic consequences. This chapter suggests that social connectivity between places might be economically consequential, and may provide a proxy measure for the total economic connectivity between cities.

Chapter 4 demonstrates that there is an association between the travel network and economic prosperity. However, directionality remains ambiguous – do travel networks predict economic prosperity, or does prosperity predict travel networks? The next chapter (5) examines whether travel networks are more likely to be a product or source of prosperity.
Travel is Not Tightly Endogenous with Prosperity

5.1 Introduction

Do City Characteristics Influence the Number of People Moving Between Them?

On an average day, 46 million people move between US cities. These movements reflect the necessity, accessibility and desirability of cities. Migrants move to new cities in pursuit of better opportunity, or more desirable urban amenities. Commuters connect cities with better commercial opportunities to cities with better consumption opportunities. Both groups make these decisions in a complex decision-making environment, where geography, culture, social ties, biography and history each play a role. While some of these factors are idiosyncratic, many of them exert influence across the population, and are directly related to the characteristics of the city. This raises the intriguing possibility that the attributes of cities may play a significant role in shaping the movement of people between them.

Previous literature is uncertain. Some scholars have argued that city residence is a life choice with significant ramifications, and that migrants move accordingly (Ravenstein, 1885; Florida, 2008). They have also posited that people choose lo-
cations based on trade-offs between tax burdens, and services provided (Tiebout, 1956). Empirical work supports these suppositions. Recent studies have concluded that people move to pursue employment opportunities, especially during recessions (Neal, 2011a), and that young, educated graduates tend to flock to big cities, due to their superior amenities and opportunities (McCormick and Wahba, 2005; Johnston, 1991). They have also confirmed that Americans commute to arbitrage between areas of higher wages and areas that are advantageous for consumption (Reeder, 1956; Wachs et al., 1993). However, other scholars have pointed out that American society is not particularly mobile, and has become less mobile over time (Wolf and Longino, 2005; Cooke, 2011). Geography, especially distance, remains a strong deterrent to all forms of travel (Schwartz, 1973; Logan, 2012; Levy, 2012).

In this chapter, I ask, do city characteristics influence the number of people moving between them? Using US highway/air travel and census data (2000-2010), I estimate the influence of various city characteristics on the volume of travel between cities via two-stage dyadic regression. The models suggest that city attractiveness may play a role in shaping traffic volumes somewhat, but geography places constraints on travel networks. However, the case can easily be overstated – no model predicts more than 17% of the variation in traffic volumes.

5.2 Data

The city data used here are described in 3.2.1, and the network data are described in 3.2.2. The models in this chapter use all 696 cities in the data, for all years of the data, without any aggregation.

5.3 Methodology: Two-Stage Dyadic Model

First, I model of the number of people moving between pairs of cities as a function of the attributes of those cities. Then, I model the annual change in traffic between pairs
as a function of change in city attributes. Formula 5.1-5.2 depict these functions, where t indexes the year, and i/j index cities. Conceptually speaking, there are two kinds of effects in this model. Some effects are attributes of the connection between two cities. These are denoted in 5.1 as $\text{Attribute}_{ijt}$, and would include variables such as the amount of snow fall on the terrain between cities i and j in time t. Other effects are interactions between an attribute of i and an attribute of j in time t. These are denoted in 5.1 as $f(\text{Attribute}_{it}, \text{Attribute}_{jt})$. For example, previous scholarship suggests that people tend to commute from areas with lower median incomes to areas with higher median incomes. This effect would be a function of the income of i and the income of j at time t. The $f$ in this case would be the difference between i and j: $|\text{Income}_{it} - \text{income}_{jt}|$. I discuss the specifications in further detail in the next subsection.

$$Traffic_{ijt} = f(\text{Attrib}_{it}, \text{Attrib}_{jt}) + \text{Attrib}_{ijt} \quad (5.1)$$

$$\Delta Traffic_{ij(t_2-t_1)} = f(\Delta \text{Attrib}_{i(t_2-t_1)}, \Delta \text{Attrib}_{j(t_2-t_1)}) + \Delta \text{Attrib}_{ij(t_2-t_1)} \quad (5.2)$$

Both 5.1 and 5.2 model the potential for city influence on travel networks. However, they model city influence on different time scales. Equation 5.1 models the long-term, cumulative association, while equation 5.2 models the short-term, immediate association. The difference between these two models reveals the time-scale of endogeneity. If the association in the long-term model is significantly stronger, it would suggest that travel networks are slow to respond to changes in city income in the short-term. In turn, this would suggest that travel networks could still influence city income in the short-run, even if the two are endogenous in the long-term.
5.3.1 Specification Details

As with the previous models, these models use log-linear variables. First, 70 years of gravity model scholarship have demonstrated that log linearizing is a highly effective, reliable way to model traffic data (Zipf, 1946; Stewart, 1960; Bergstrand, 1985; Brun et al., 2005; Zhou, 2010). Second, log linearizing is one of the few ways to successfully process a distribution as severely skewed as passenger traffic – one that ranges from dozens of passengers to millions. Third, using log-linear variables here maintains consistency with chapter 4. Fourth, as shown in equation 5.3, the log form is also the most natural way to model the non-linear scaling that occurs in cities, and the non-linear decline that occurs with distance.

\[ \beta \cdot \log X = X^\beta \] (5.3)

For example, due to the various economies and diseconomies of scale, size exerts non-linear effects (Bettencourt et al., 2004; Arbesman et al., 2009). For example, a city that is twice as large tends to receive less than twice the amount of total incoming traffic. This may be because citizens of larger cities have more opportunities inside the city, and therefore less reason to travel outside. Conversely, twice the distance tends to correspond to less than half the travel volume. This modeling choice is highly consistent with the way sociologists have traditionally modeled traffic (Zipf, 1946; Stewart, 1960; Zhou, 2010), and is also supported in economic theory (Duranton and Puga, 2002; Rosenthal and Strange, 2004).

To emphasize the point, equation 5.4 estimates the actual log-linear regression model of population size and distance on non-zero traffic volumes for these data. This model explains 50% of the variance \( R^2 = 0.498 \) in non-zero traffic volumes.

\[ \log Traffic_{ij} = 0.16 \log (Pop_i + Pop_j) - 1.6 \log Dist_{ij} + 23.33 \] (5.4)
Incidentally, this is a bare-bones model of traffic as a function of geography and history. It models traffic as a function of where cities are located (Pop) and how much distance (Dist) stands between them. It serves as our first clue that city prosperity might not be the dominant influence on traffic volumes.

However, this model only holds for non-zero traffic. An additional modeling step is necessary, because 98% of all city pairs have zero traffic between them, even though many of them are relatively proximate.

A small part of this reflects the fixity of infrastructure. Both airports and highways are expensive to build, and tend to remain in operation once established. However, this is a surprisingly small part of the story, because America has far more transportation infrastructure than it uses. In terms of air traffic (which accounts for the vast majority of all traffic-connected dyads), approximately 1,900 continental US airports are rated for international traffic, but 9% of those airports carry 96% of all air traffic. This is not simply a matter of capacity. Between 1990 and 2011, 281 airports saw their total traffic volumes fall by more than half, while 1,123 saw their total volumes rise to more than double. In fact, 241 of the airports that at least doubled in the 1990’s, went on to lose more than half after 2000. The wide fluctuations in traffic volume suggest that there is far more air traffic capacity than air traffic. Thus, air traffic infrastructure is probably not a cause of the high prevalence of zero traffic city pairs. Infrastructure is even less important for highway traffic. 90% of all cities within 25mi (twice the average commuting distance) of each other share an interstate, as do 80% of all cities within 40mi (over three times the average). The vast majority of cities that are close enough for a car commute are able to use the interstate system. Previous research strongly supports this argument (Samaniego and Moses, 2008; Youn et al., 2008; Ishii et al., 2009).

Instead of infrastructure, the large volume of zero-traffic city pairs is a reflection of randomness – some cities end up connected over equivalent others through his-
tory and happenstance. While not selection bias in the traditional sense, this is close enough to canonical selection bias to justify the use of a Heckman correction strategy (Puhani, 2000; Heckman, 1979). First, I predict the chances of zero-traffic between cities, based on the distribution of the population in geographic space, using a logit model.\(^1\) In essence, I predict the chances of there being any traffic, given city placement. Next, I incorporate the predicted values from the city placement model into the main log-linear models, producing Heckman-corrected estimates of city attribute influence on travel.\(^2\)

### 5.3.2 Functional Form

In addition to methodological challenges, this model is substantively challenging because these variables interact with traffic in complex ways. For example, previous research informs us that prosperous locations receive more traffic. However, other research suggests that commuters arbitrage between areas with unequal economic opportunity, and a third body of research notes that traffic volumes are higher among areas with similar socio-economic status. All of these imply different predictions for how prosperity will affect traffic.

The solution is to insert model terms for each applicable effect. If previous research indicates that traffic increases whenever there is more of variable X, then the traffic between city i and city j is modeled as a function of \(X_i + X_j\). If previous research suggests that traffic increases when there are unequal amounts of X between cities, then the traffic between city i and city j is modeled as a function of \(|X_i - X_j|\). If research indicates that traffic increases when most residents of i and j belong to the same social / political groups (ie homophily), this is modeled as \(\sum X_{ik}X_{jk} / X_iX_j\). If research implicates a factor that is specific to the edge, it is modeled as \(X_{ij}\).

---

\(^1\) Essentially, the logit version of equation of 5.4

\(^2\) For consistency, I will generally use the term “city placement” to refer to the first stage (Heckman correction) model of city traffic, since it is based on the location and size of city populations.
example, a traffic volume model based on all functional forms of one variable might look like equation 5.5.

\[
Y_{ij}^{OLS} = (X_i + X_j)^{\beta_2} \cdot (X_i - X_j)^{\beta_3} \cdot \frac{\sum_k X_{ik} \cdot X_{jk}^{\beta_4}}{X_i \cdot X_j} \cdot X_{ij}^{\beta_5} \cdot Y_{Heckman}^{\beta_6} 
\] (5.5)

Note: Log-linearized before coefficient estimation

5.3.3 Error Estimation

This model has a number of features that violate classical error estimation assumptions. As such, I bootstrap the standard errors from 80 resamples\(^3\) of the data. First, I sample (with replacement) 10,000 traffic connections. Next, I calculate the model on each resample. Third, I calculate the mean for each coefficient across models – the bootstrapped coefficient. Fourth, I calculate the standard deviation across model coefficients – the standard error. This procedure produces much more rigorous tests of the significance of each coefficient. Many coefficients that are significant by classical significance tests are not significant by this methodology.

5.4 Results: Traffic Sluggish in Responding to City Influence

5.4.1 Static Model: Geography and History Influence Travel Volumes

Tables 5.1 and 5.2 (further below) report the resulting static (traffic levels) and dynamic (annual change in traffic) models. The first column reports the model coefficients. The second column reports the semi-standardized coefficients – the coefficients multiplied by the standard deviation of X.

The third column reports the bootstrapped standard errors (*1.96). In both models, the city placement correction is the single strongest predictor by a wide margin. Since the indicator is based solely on population and distance, these models

\(^3\) Previous iterations of the model used much larger numbers of resamples. However, it turned out that few resamples were needed to produce accurate, stable estimates, because of the large size of the data
indicate that the geographic location of people is the single most important predictor of their movements. This is important, because geography does not change, and thereby imbues that US travel network with significant stability over time.

While differential birthrates will somewhat influence how many people live in each city, the most important determinant of the city placement variable is geographic. If a non-changing variable strongly shapes the travel between cities, then it is more likely that travel networks influence economic prosperity than that economic prosperity influences travel networks.

Table 5.1 contains several other patterns of note. In the “Geographic (Travel Route)” variable group, we see the influence of geo-spatial barriers on traffic. Routes that pass through areas with higher levels of rainfall (and thereby thicker vegetation) have lower average traffic volumes (negative “Rain” term), and US state

Table 5.1: Static Model

<table>
<thead>
<tr>
<th>Variable Group</th>
<th>b</th>
<th>( \beta )</th>
<th>1.96se</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-4.315</td>
<td>8.873</td>
<td></td>
</tr>
<tr>
<td>City Placement</td>
<td>5.634*</td>
<td>0.923</td>
<td>1.488</td>
</tr>
<tr>
<td>Geographic (Sum)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border Distance</td>
<td>0.023</td>
<td>0.017</td>
<td>0.024</td>
</tr>
<tr>
<td>Inclement Weather</td>
<td>-0.02</td>
<td>-0.021</td>
<td>0.043</td>
</tr>
<tr>
<td>Distance Inland</td>
<td>0.039*</td>
<td>0.033</td>
<td>0.033</td>
</tr>
<tr>
<td>Geographic (Difference)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border Distance</td>
<td>-0.011*</td>
<td>-0.014</td>
<td>0.008</td>
</tr>
<tr>
<td>Inclement Weather</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Distance Inland</td>
<td>-0.028*</td>
<td>-0.038</td>
<td>0.011</td>
</tr>
<tr>
<td>Geographic (Travel Route)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>-0.034*</td>
<td>-0.04</td>
<td>0.026</td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.017</td>
<td>-0.022</td>
<td>0.022</td>
</tr>
<tr>
<td>Rain</td>
<td>-0.095*</td>
<td>-0.038</td>
<td>0.066</td>
</tr>
<tr>
<td>Snow</td>
<td>0.009</td>
<td>0.009</td>
<td>0.041</td>
</tr>
<tr>
<td>State</td>
<td>0.543*</td>
<td>0.177</td>
<td>0.24</td>
</tr>
<tr>
<td>Social (Sum)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Age</td>
<td>-0.387*</td>
<td>-0.019</td>
<td>0.216</td>
</tr>
<tr>
<td>Education Years</td>
<td>0.668*</td>
<td>0.032</td>
<td>0.293</td>
</tr>
<tr>
<td>Gasoline Price</td>
<td>-0.208</td>
<td>-0.037</td>
<td>0.487</td>
</tr>
<tr>
<td>Home Price</td>
<td>-0.013</td>
<td>-0.006</td>
<td>0.064</td>
</tr>
<tr>
<td>Median Income</td>
<td>0.395*</td>
<td>0.065</td>
<td>0.112</td>
</tr>
<tr>
<td>Net Migration</td>
<td>0.003*</td>
<td>0.022</td>
<td>0.002</td>
</tr>
<tr>
<td>Total Persons</td>
<td>0.018</td>
<td>0.029</td>
<td>0.027</td>
</tr>
<tr>
<td>Income Inequality</td>
<td>-0.019</td>
<td>-0.005</td>
<td>0.056</td>
</tr>
<tr>
<td>Workers Per Capita</td>
<td>-0.17*</td>
<td>-0.011</td>
<td>0.138</td>
</tr>
<tr>
<td>Social (Difference)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Age</td>
<td>-0.014*</td>
<td>-0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>Education Years</td>
<td>-0.045*</td>
<td>-0.015</td>
<td>0.021</td>
</tr>
<tr>
<td>Gasoline Price</td>
<td>0.472*</td>
<td>0.031</td>
<td>0.304</td>
</tr>
<tr>
<td>Home Price</td>
<td>-0.018*</td>
<td>-0.024</td>
<td>0.008</td>
</tr>
<tr>
<td>Median Income</td>
<td>-0.021*</td>
<td>-0.024</td>
<td>0.005</td>
</tr>
<tr>
<td>Net Migration</td>
<td>0.001</td>
<td>0.002</td>
<td>0.008</td>
</tr>
<tr>
<td>Total Persons</td>
<td>-0.032*</td>
<td>-0.067</td>
<td>0.008</td>
</tr>
<tr>
<td>Income Inequality</td>
<td>-0.008</td>
<td>-0.009</td>
<td>0.009</td>
</tr>
<tr>
<td>Workers Per Capita</td>
<td>-0.111</td>
<td>-0.007</td>
<td>0.117</td>
</tr>
<tr>
<td>Social (Homophily)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race Homophily</td>
<td>-0.292*</td>
<td>-0.026</td>
<td>0.117</td>
</tr>
<tr>
<td>Political Homophily</td>
<td>0.408*</td>
<td>0.008</td>
<td>0.389</td>
</tr>
<tr>
<td>Social (Time)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflation</td>
<td>0.292</td>
<td>0.012</td>
<td>0.459</td>
</tr>
<tr>
<td>GDP (PPP)</td>
<td>0.093</td>
<td>0.003</td>
<td>0.757</td>
</tr>
<tr>
<td>Kerosine</td>
<td>0.159</td>
<td>0.036</td>
<td>0.344</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.177^* \]

N (Edges) 10,000 x 80 0

\footnote{For this to not be true, the annual change in the population would have to be larger than the size of the population. The average annual population change in these data is 1%, so the historical population of a city effectively predicts its current population. The fastest growing city in these data was the Palm Coast FL during the real estate boom. It grew at a rate of about 10% per year for 3 years.}
boundaries prove surprisingly non-porous (positive “State” term). Likewise, cities that lie close to international borders and waters have lower domestic traffic volumes (“Border Distance” and “Distance Inland”). This hints at the non-porousness of international borders. Also inland cities likely have more opportunities to serve as “gateways” to more distal cities, so it makes sense that they would experience more traffic flow.

Table 5.1 also contains some notable surprises. Even though social homophily is a well-documented influence of social interaction, racial homophily actually corresponds to lower average traffic volumes. Cities with diverse and distinct racial compositions appear to reap a not insubstantial traffic benefit. This may reflect migration. Diverse cities tend to have higher immigrant (domestic and foreign) populations. Immigrants tend to travel to maintain hometown ties (McCann et al., 2010; Levitt and Jaworsky, 2007), and may even migrate back at a later point (Wyman, 1993; Portes et al., 1999). However, the US population is not particularly prone to high rates of domestic migration, so it is unclear if migration is responsible for this coefficient (Wolf and Longino, 2005). An alternative explanation may be that affluent, white communities tolerate minorities occupying service sector jobs, but make it difficult for them to take up residence (Loewen, 2013; Fernandez, 2008; Mouw, 2000). Conversely, residents of primarily white exurbs may be commuting into integrated city centers.

Many of these variables have complex interactions between their function forms. Figure 5.1 maps how these interactions play out for systematic variations in the attributes of cities i and j. The Y-axis indicates the total volume of predicted traffic. The X-axis indicates how the predicted volume of traffic changes when all values for city i are multiplied by a given coefficient. The lines indicate how the predicted volume changes for multiples of city j values.

For example, the top left box is labeled income. It maps out how the predicted
traffic between two cities pairs changes when the income of ego and alter change. The top line (red) is labeled 2j. It answers the question, “how much traffic would we expect if the median income of city j doubled?” This can be compared to the middle line 1j, which leaves the income of j at its current value. We see that line 2j is much higher than line 1j, indicating that, based on the model in table 5.1, we would expect a doubling of the median income of j to correspond to a significant increase in the traffic between i and j. The x-axis indicates how the income of city i would influence that prediction. We see that the highest predicted incomes occur when the
The income interaction is simple: more income predicts more traffic. However, many of the other variables have complex interactions between their different functional forms, such that predictions may become non-linear. For example, table 5.1 predicts more travel when gas prices differ between i and j, but less when both i and j are high. The result is row 3, column 3 of figure 5.1. The model predicts the most traffic when prices are low, but also lower in one city than the other.

For education (and income), higher values correspond to more traffic. Cities with highly educated populations tend to average higher volumes of traffic (Wheeler, 1967; Beaverstock et al., 2004). For age, net migration, and workers per capita, inequalities between cities tend to correspond to higher traffic volumes. However, only age is statistically significant for all terms. Travel is higher between cities with different median ages. This is consistent with research on how travel and migration relate to life course (McCormick and Wahba, 2005; Johnston, 1991). Similar dynamics apply to total population and home price, albeit with slightly more complex interactions.

The other variables have strong model effects for both totals and inequalities. For these variables, the highest traffic volumes tend to flow between large equals. For example, in the case of income inequality and inclement weather, the highest predicted traffic volumes tend to flow between cities that are equally mild and egalitarian, especially if those cities are very mild and egalitarian. Over all, model 5.1 explains parts of traffic volume, and yields many of the associations that we would expect. However, the model only predicts 18% of the observed variation in traffic volumes, casting doubt on its adequacy.

5.4.2 Dynamic Model: Prosperity Does Not Predict Annual Change

The association between traffic and city attributes has likely culminated from a complex and causally murky series of pathways over centuries of history. Table 5.2
presents a more rigorous test of the association between traffic and city attributes.

It models the annual change in traffic volumes as a function of the annual change in various attributes of cities. Not only does this factor out the influences of static variables (like geography), but it also filters out many of the historical legacies that shaped correlations in the previous model. The result is devastating. Not only does 5.2 predict less than 1% of the variances in traffic volumes, but most variable coefficients are reduced to near zero levels. This suggests that much of the association between traffic volumes and city attributes is inertial – reflecting geography and history, not the influence of cities.

### 5.4.3 Variable Group Total Effects

Both the dynamic and static model suggest that the placement of city populations is sufficient attain much of the total variance explained in these models. None of the socio-economic variables in this model, prosperity included, is able to achieve a comparable effect size. However, this may be an unfair test – there are many correlated socio-economic indicators in this model, struggling to explain the same variance. Figure 5.2 sums up the total semi-standardized model effect ($\sum_k |\beta_k|$) of three groups of variables in the static (5.1) and dynamic (5.2) models. This provides a fairer test. When we consider the combined effect of socio-economic factors, are they stronger than the effects of city placement?

### Table 5.2: Dynamic Model

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>$\beta$</th>
<th>1.96se</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.029*</td>
<td>0.029</td>
<td>0.013</td>
</tr>
<tr>
<td>City Placement</td>
<td>0.715*</td>
<td>0.117</td>
<td>0.257</td>
</tr>
<tr>
<td><strong>Social (Sum)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Age</td>
<td>-0.025*</td>
<td>-0.003</td>
<td>0.018</td>
</tr>
<tr>
<td>Education Years</td>
<td>0.005</td>
<td>0</td>
<td>0.096</td>
</tr>
<tr>
<td>Gasoline Price</td>
<td>0.027</td>
<td>0.011</td>
<td>0.034</td>
</tr>
<tr>
<td>Home Price</td>
<td>0.0001*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Median Income</td>
<td>0.001*</td>
<td>0.007</td>
<td>0.001</td>
</tr>
<tr>
<td>Net Migration</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Persons</td>
<td>0.001*</td>
<td>0.006</td>
<td>0.001</td>
</tr>
<tr>
<td>Income Inequality</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Workers Per Capita</td>
<td>0.161*</td>
<td>0.003</td>
<td>0.116</td>
</tr>
<tr>
<td><strong>Social (Difference)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Age</td>
<td>-0.003</td>
<td>0</td>
<td>0.007</td>
</tr>
<tr>
<td>Education Years</td>
<td>-0.013</td>
<td>0</td>
<td>0.078</td>
</tr>
<tr>
<td>Gasoline Price</td>
<td>-0.002</td>
<td>0</td>
<td>0.042</td>
</tr>
<tr>
<td>Home Price</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Median Income</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Net Migration</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Persons</td>
<td>-0.001*</td>
<td>-0.007</td>
<td>0.001</td>
</tr>
<tr>
<td>Income Inequality</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Workers Per Capita</td>
<td>-0.026</td>
<td>0</td>
<td>0.115</td>
</tr>
<tr>
<td><strong>Social (Homophily)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race Homophily</td>
<td>-1.241</td>
<td>-0.003</td>
<td>0.724</td>
</tr>
<tr>
<td>Political Homophily</td>
<td>-0.044</td>
<td>0</td>
<td>0.432</td>
</tr>
<tr>
<td><strong>Social (Time)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflation</td>
<td>1.246*</td>
<td>0.015</td>
<td>0.363</td>
</tr>
<tr>
<td>GDP (PPP)</td>
<td>0.001</td>
<td>0.006</td>
<td>0.001</td>
</tr>
<tr>
<td>Kerosine</td>
<td>0.003</td>
<td>0.001</td>
<td>0.038</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.006*</td>
<td></td>
<td>0.003</td>
</tr>
<tr>
<td>N (Edge)</td>
<td>10,000</td>
<td>x80</td>
<td></td>
</tr>
</tbody>
</table>
The first group, “City Placement,” describes the predicted influence of random movement in geographic space on the extent to which cities have any traffic between them. It is based on population sizes and the distances between them. The second group, “Geography,” describes how topography impedes travel, likely by raising the costs of construction and the inconvenience of movement. The third group, “Social,” describes how social, economic, political, and demographic characteristics of the city population relate to the volume of traffic between them.

The static model in figure 5.2 reveals that the combined effects of socio-economic factors are relatively substantial. In fact, they have the largest combined effect on total traffic volumes. This is the feared endogeneity between city prosperity factors, and total travel volumes. However, even here, the influence of the location of city populations – city placement – is nearly as strong. Since city placement factors do not change quickly or in tandem with prosperity, they likely give travel volumes stability over time. Thus, even when considering a cross-sectional model, travel volumes are, at least somewhat, constrained in their ability to change.

The dynamic model is consistent with this interpretation. Annual change in city socio-economics does not effectively predict travel volume annual change. The influence of city placement is three times stronger. This suggests that the influence of socio-economic factors on the travel network is slow, indirect, and inertial.
5.5 Conclusions

Do City Characteristics Influence the Number of People Moving Between Them?

Somewhat, but geography and history are stronger.

Using US highway/air travel and census data (2000-2010), I estimated the influence of various city characteristics on the volume of travel between cities via two-stage dyadic regression. Since 98% of all possible city dyads have no traffic between them, the first stage estimates whether there is any connectivity between cities, based on the location and size of cities. The second stage predicts traffic volume based on the various characteristics of city populations. I calculate two models: one which measures the influence of city attributes on total traffic volumes, and the other measures the influence of annual changes in city attributes on annual changes in traffic volumes. In the long run, socio-economic factors may exert some influence of traffic volumes. This is especially true of median income. However, the location and size of cities is relatively powerful in both models, and utterly dominates the annual change model. Travel networks may have a degree of location-induced stability and inertia, and therefore, cannot rapidly adapt to city changes. However, no model explains more than 18% of the variation in travel volumes, so it is not clear that they effectively capture the dynamics of US travel volumes.

Chapter 5 hints at the most probable direction of the association described in the previous chapter (4). However, neither 4 nor 5 provide any insight into causality. The chapter 6 will test the likelihood that the relationship between the travel network and economic prosperity contains causal components.
6

Travel Exerts Causal Influence on Prosperity

6.1 Introduction

Can Travel Networks Transmit Economic Shocks?

On an average day, nearly 200,000 people fly into and out from the United States. These movements embody a wide range of travelers: tourists, migrants, organizational representatives, families, etc. Regardless of motive, these movements have economic consequences. Travelers carry their skills, consumption patterns, social capital, information, experience and organizational affiliations with them. Through a variety of direct and indirect means, they connect their economy of origin to the US economy. This raises the intriguing possibility that the movement of people could, in effect, transmit exogenous shocks from their economy of origin to the US economy.

Scholars have suggested that the air traffic network plays a major role in facilitating the world economy (Smith and Timberlake, 1995). Recent history suggests that this idea is not far-fetched. The 2010 eruption of Iceland’s Eyjafjallajkull volcano created the largest air traffic shutdown in the history of civil aviation. By the time the ash settled, 107,000 flights had been cancelled. Global financial experts esti-
mated worldwide economic costs in the billions of dollars, even though the volcano caused almost no physical damage (Erlanger and Ewing 2010, PWC 2010, Jamieson 2010). Other empirical scholarship also supports this idea. Air traffic is key to the maintenance of transnational social capital (McCann et al., 2010; Stockdale, 2004), which has been implicated as a key factor in American entrepreneurship (Portes et al., 2002).\(^1\) It is also key to the enactment of multi-national corporations (Sassen, 2001), which is why non-stop air traffic connectivity attracts corporate headquarters (Bel and Fageda, 2008).

In this chapter, I ask, can travel networks transmit economic shocks?\(^2\) I use a natural experiment framework to gain causal insight into this possibility. Focusing on two economically devastating Latin American earthquakes, I examine how connected US cities fared in the subsequent year, compared to unconnected cities with similar connection propensities. I found that the US cities with connections to the disaster site experienced almost no growth in median income in the years following the disaster. In contrast, the unconnected, propensity-matched cities averaged .74% growth in median income. The difference is significant at the .1 level, and hints that air traffic networks might transmit exogenous economic shocks. However, the sample is small, the findings are not significant at the .05 level, and the massive size of the estimated effect strains credulity.

6.2 Data

The city data used here are described in 3.2.1, but include only the cities that have air traffic connections to selected disaster sites and their propensity-matched

---

\(^1\) Roughly 1 in 10 legal immigrant workers own a business, and nearly 70% of those relied on start-up capital from home country social ties (Fairlie, 2012)

\(^2\) To be more precise, I ask whether travel networks directly transmit economic shocks or, at least, serve as proxy indicator for connections that do transmit them. Think of travel as denoting a flow of economic value from one economy to another. When one economy is disrupted, the flow of economic value would likely also be disrupted.
comparator cities. The air traffic network data comes from the same source as the air traffic described in 3.2.2. While most of this dissertation focuses on travel that originates and arrives in the continental US, the BTS data actually includes all flights that either originate or arrive in the US.

The shocks in this study occurred in distant countries (Ecuador and Peru). While plenty of natural disasters occurred in the US during this time period, it would be difficult to disentangle the influence of network connectivity from other connections in the US. For example, when hurricanes Katrina and Rita struck Louisiana, congress was forced to raise the borrowing limit on the National Flood Insurance Program (NFIP) from $1.5 billion to an unprecedented $20.8 billion. NFIP’s massive increase in debt has repercussions for other major US flood zones, such as the northern Mississippi flood plain, regardless of network connectivity (GAO, 2006). In contrast, a natural disaster in a distant country will have fewer indirect pathways for influencing US cities, making it easier to disentangle the influences of network connectivity.

6.3 Methodology: Matched Sample Natural Experiment

6.3.1 T-Test of Matched Samples

Only a small subset of US cities have the level of international air presence required to be eligible for this test, and only a handful of national disasters have the necessary characteristics (international, severe, and occurring between 2000-2010) to be useful here. As a result, the statistics employed here are simple and robust. I compare the mean percentage of income growth for connected cities to the mean percentage of income growth of disconnected match cities in the year following the exogenous shock. Then, I use a one-sided, paired t-test to establish whether connected cities underperform, relative to unconnected matches.
6.3.2 Propensity Matching

A key challenge for this project is generating the proper reference group. An ideal reference group would be exact duplicates of the affected cities, but with experimentally manipulated connectedness. Since that is not feasible, the next best option is to match cities according to their propensity to connect to the affected region. I accomplish this with a logistic regression, which predicts whether there is any air traffic between a US city and a disaster site (in the year before the disaster) as a function of the attributes of the US city. The model sample is limited to the 17 cities (CBSA) with greater than 2.5 million residents, but less than 7 million. Below 2.5 million, no city has air traffic linkages to the affected areas. Above 7 million, the cities (NYC, LA, Chicago) are unmatchable, due to their outlier status. The final sample size amounts to 34 (17 cities x 2 sites = 34).

The propensity score for each city (for each site) is the predicted probability of that city having traffic connections with that site, as shown in equation 6.1. I pair each connected city to the unconnected city with the closest propensity score, allowing cities to match multiple connected cities where needed.

\[
\text{Propensity} = \frac{e^Y}{1 + e^Y}
\]  

(6.1)

6.4 Results: Traffic Transmits Exogenous Shocks

Table 6.1 presents the propensity model. The results are fairly obvious: richer, more populous cities with large Latino populations are most likely to connect to distant places in Latin America (ie Peru and Chile). Cities with older populations that are further away from South America are less likely to connect. All cities are more likely to connect to Peru, which is about 1,000 mi closer, and 10 million more citizens.
Table 6.1: Connection Propensity Model

<table>
<thead>
<tr>
<th></th>
<th>$e^b$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>11.805</td>
<td>0.276</td>
</tr>
<tr>
<td>Person</td>
<td>1.415</td>
<td>0.23</td>
</tr>
<tr>
<td>Age</td>
<td>0.059</td>
<td>0.069</td>
</tr>
<tr>
<td>Education</td>
<td>0.019</td>
<td>0.259</td>
</tr>
<tr>
<td>Income</td>
<td>5.282*</td>
<td>0.046</td>
</tr>
<tr>
<td>Latino</td>
<td>1.727</td>
<td>0.48</td>
</tr>
<tr>
<td>Distance</td>
<td>0.2*</td>
<td>0.012</td>
</tr>
<tr>
<td>Site (Peru)</td>
<td>0.617</td>
<td>0.074</td>
</tr>
</tbody>
</table>

$BIC = 45.4 \quad \rho^2 = 0.514 \quad N = 34$

Table 6.2 presents a stem plot of the connective propensities for the 17 cities, averaged across the two sites. The heavily Latino, heavily populated cities of Houston, Dallas, and Miami have the highest propensities by far, ahead of the highly international Washington DC. They are followed by America’s auxiliary global cities: Atlanta, Boston, Philadelphia, and San Francisco. The remaining cities have lower propensities. They are generally either the secondary cities of their region (Riverside, San Diego, Tampa), major cities in long-term decline (Baltimore, Detroit, St. Louis), or prime cities of sparsely populated regions (Phoenix, Minneapolis, Seattle). Houston, Dallas, and Miami had traffic to all disaster sites. Washington and Philadelphia were the most common match cities.³

Table 6.2: Mean Propensity of Qualifying Cities

<table>
<thead>
<tr>
<th></th>
<th>Baltimore</th>
<th>Detroit</th>
<th>Minneapolis</th>
<th>Phoenix</th>
<th>Riverside⁴</th>
<th>San Diego</th>
<th>St Louis</th>
<th>Tampa</th>
<th>Seattle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Houston</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miami</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>70%</td>
<td>65%</td>
<td>60%</td>
<td>55%</td>
<td>50%</td>
<td>45%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Of the 34 possible city-site pairs, 9 reported traffic with the disaster site during the year before the earthquake. Table 6.3 reports the mean percentage change in median income for connected cities and their matches in the year followed the ex-

³ A cautionary note: there is a very limited number of cities with international air traffic ties, and this places severe constraints on the effectiveness of the propensity matching procedure.
ogenous shock. On average, median income remained approximately flat for cities with connections to the disaster sites. It decreased by a minor 0.08%. In contrast, the average matched city saw median income rise by 0.74% over the same period. Compared to match cities, the cities with connections to disaster sites tended to underperform by an almost statistically significant amount (P Value= 0.09). Power calculations (at 90% power) reveal that 11 more cases are necessary to make the observed difference statistically significant at the .05 level.

The t-test estimates that 91% of all possible samples drawn from this population would support the assertion that connected cities grew less than unconnected matches following these major disasters. Based on these findings, it is, at least, plausible that traffic moving between cities exerts causal influence on the median incomes of those cities.

### Table 6.3: Income Growth After Exogenous Shock

<table>
<thead>
<tr>
<th></th>
<th>Connected</th>
<th>Unconnected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean:</td>
<td>-0.08%</td>
<td>+0.74%</td>
</tr>
<tr>
<td>Example Match:</td>
<td>Boston</td>
<td>Philadelphia</td>
</tr>
<tr>
<td>t=</td>
<td>-1.462</td>
<td>df= 9</td>
</tr>
<tr>
<td>p=</td>
<td>0.09</td>
<td></td>
</tr>
</tbody>
</table>

6.5 Conclusions

**Can Travel Networks Transmit Economic Shocks?**

Likely. If not, they are an effective proxy for networks that do.

Using international air travel and US city census data, I investigate whether travel networks transmit economic shocks from distant foreign economies. In the early months of 2001 and 2010, major earthquakes devastated the Peruvian and Chilean economies respectively. Some US cities had air travel connections to these distant places. I compare growth in the median income of these cities to growth in the median income of cities with no connections, but a similar propensity to connect. I found
that the US cities with connections to the disaster site experienced almost no growth in median income in the years following the disaster. In contrast, the unconnected, propensity-matched cities averaged .74% growth in median income. The difference is significant at the 0.1 level, and suggests that travel networks transmit exogenous economic shocks, or are an effective proxy indicator for networks that do. Figure 6.1 depicts the probability distribution for the t-test.

Previous literature highlighted the importance of international social ties for American business (Portes et al., 2002), the importance of travel ties for global business operations (Bel and Fageda, 2008), and the economic damage that can result from the air traffic disruption (PWC, 2010). This chapter takes the next logical step: if travel networks are conduits for economic benefits, than disruptions to remote economies will also influence economic prosperity of connected US cities.

The implications of these findings are significant to the field as a whole because of the methodology. The natural experimental framework gives these findings stronger causal weight, compared to the associational models that predominate in this field. However, these findings can easily be overstated. The large size of the estimated effect strains credulity, and the limited number of qualifying cities constrains the effectiveness of the propensity matching procedure.
7.1 Travel Networks Influence City Prosperity

Do travel networks influence the prosperity of US cities?

Possibly.

This dissertation examines the US travel network and possibility that it exerts influence on city prosperity. Chapters 2-6 reveal the results.

Chapter 2 suggests that networks influenced prosperity in previous studies. Studies of organizational actors reveal many mechanisms by which networks – and not just trade networks – aid production through more efficient resource use, diminished transaction costs, enriched labor markets, and enhanced innovation production. Studies of cities and countries reveal that geography-spanning networks can improve economic environments, and that cities/states benefit from facilitating them.

Chapter 3 suggests that travel networks are highly structured, and that structure creates opportunities for influence. At the top of that hierarchy are two systems, centered on New York City and Los Angeles, which tend to counter-balance each others influence. Below that are successive levels of hub/bridging structures, by
which the geographically isolated regions of the US become a well-integrated system. There are opportunities for cities to benefit from the traffic network.

Chapter 4 hints that travel network structure may correlate with prosperity. Using network auto-regressive modeling, it demonstrated that a network flow model could account for variation in city median income. In one analysis, it added 10%-16% additional predictive power ($r^2$) to linear models based on city characteristics alone. However, consistent with previous research, model specification significantly influences the result. A second analysis, based on different network specification choices, suggests network ties may only add 2% additional predictive power.

Chapters 5 elaborates on the directionality of the association found in chapter 4. Using a two-stage, dyadic linear model, chapter 5 demonstrates that geography and history heavily constrain travel network volumes, imbuing them with consistency over time, and inertia to change. Consequently, the association from chapter 4 is more likely due to the influence of travel networks on city prosperity, than the influence of city prosperity on travel networks. However, the models only predict 17% of the variation in traffic volumes, so these findings are suggestive, rather than definitive.

Chapter 6 elaborates on the causality of the association found in chapter 4. Using a natural experiment framework, 6 examines cities with connections to distant, foreign economies that recently suffered a major natural disaster. These cities experienced far less growth (0.08%) in median incomes than propensity-matched cities (0.74%) that did not have connections to the distressed economies. The difference was significant at the 0.1 level, despite the small number of cases. This suggests that the travel network matters for economic prosperity, because cities prospered less when their network alters experienced economic shocks. Consequently, the association from chapter 4 may have causal components, because disruptions to network alters corresponds to less prosperity. However, the limited number of qualifying cities heavily constrains the effectiveness of the propensity matching procedure, and the
estimated effect size strains credulity.

Considering chapters 2-6 as a whole, there is moderate evidence that travel networks may influence economic prosperity. This is important for three reasons.

First, these are travel networks, not trade networks. With trade networks, the link to prosperity is tangible – trade involves the exchange of actual inputs to the production process. However, network scholars have suggested a wider variety of ways that networks influence economic production: trust, innovation, and structural position being among them (Powell, 1990; Burt, 2004; Alderson and Beckfield, 2004). Travel networks speak to those mechanisms in a way that trade networks do not.

Second, there has long been a gap between global network theory and global network empirics. This dissertation adds to an emerging empirical reply to global network theory. Scholars were theorizing about global networks long before the data or methods existed to truly test those theories. Many of the key ideas of dependency theory (Cardoso, 1977) and world systems theory (Wallerstein, 1976) rest on the idea that global connectivity and position provide economic advantage. The world cities (Friedmann and Wolff, 1982), global cities (Sassen, 1991)\(^1\), network society (Castells, 1996) concepts elaborated on those ideas, suggesting that city networks played a key role in economic production and power. However, empirical work lagged far behind, with scholars lamenting the poor quality of available data (Short et al., 1996) and calling for better methodology (Smith and Timberlake, 1995). However, quality electronic data, powerful computing, and advanced methodology have extended the bounds of empirical work (Kaluza et al., 2010; Balcan et al., 2009; Daragnova et al., 2012; Butts et al., 2012). This dissertation adds to the new wave of studies (Moore et al., 2006; Neal, 2011a; Zhou, 2011) that finally have the data, methods, and computational power to reply to theory with empirics.

Third, globalization has increased the importance of geography-spanning net-

---

works. Geography is still an important constraint on connectivity (Logan, 2012; Butts et al., 2012) and an important factor in economic production (Wang et al., 2011; Florida, 2008; Rodriguez-Pose, 2011). However, networks are increasingly important components of economic production, prompting a re-visitation of classic city theory (Neal, 2011a; Taylor et al., 2010; Derudder and Witlox, 2004). The findings of this study are consistent with the current round of theory revisions: cities are still Christaller-esque central places (Christaller, 1933; Losch, 1938), but centrality in network space has joined centrality in geographic space as a factor in city prosperity.

7.2 Potential Applications: Improving US Economic Prosperity

In this section, I work through an example of how my findings might be applied towards an intervention strategy. It is not developed to the level of a true policy recommendation – I do no calculations to assess feasibility, potential negative impacts, etc – but serves as an illustration of the potential applications of this work.

For this example, I examine the vitality of the US travel network. Using the auto-regressive model framework from 4, I examine how model predictions for US prosperity change when I half the volume of traffic passing through each city.

In general, I find that prosperity flow is robust. Of the 696 cities examined, only seven significantly change the predicted median income of cities. Moreover, they appear to be linked into two systems. When the traffic from any city in the system is halved, all of the cities in that system experience similar drops in predicted median income, and the other system experiences proportional rises in city income. The two systems each contain one of America’s top two cities: New York and Los Angeles. The New York system (red) contains Miami, Denver and San Francisco. The Los Angeles system (blue) contains Houston and Atlanta. Two systems are linked primarily through the massive NY-LA air traffic link, and are consistent with findings in chapter 3, which notes that the city hierarchy breaks into two oscillating
systems, centered on New York City and Los Angeles.

It also suggests an avenue for intervention. Since these cities have opposite reactions to network disruption, linking them might provide additional stability to the system. The bottom panel of 7.1 proposes three connections that would accomplish this, especially if those connections were high-speed, minimal-stop connections. The first route would connect Los Angeles to San Francisco, California’s most prosperous cities.\footnote{Coincidentally, a high speed rail line between them is under construction (CHSRA, 2014).} The second route would connect Miami to Atlanta, with a stop in Tampa.\footnote{The shorter Tampa-Miami link would allow the route to earn revenue to alleviate the costs of building the longer Tampa-Atlanta route. Similar logic underlies the Houston-Dallas route.} This would connect the Atlanta-centered Southern economy directly to the Floridian economy, and would indirectly connect the Mid-Atlantic and Caribbean economies. The Tampa-Miami link would also more tightly connect South Florida’s two most prosperous urban areas. The third route would connect Houston to Denver, with a stop in Dallas. This would connect the Denver-centered mountain states economy directly to the Texan economy, and indirectly to the Mexican economy. The Houston-Dallas link would also more tightly connect Texas’ two most prosperous urban areas. Of all these routes, the Denver-Dallas link is, perhaps, the most important. It straddles the North-South, East-West, and New York-Los Angeles divides in the travel network.

It is unclear whether it is feasible to build these routes, or what negative externalities might result. However, this section illustrates a strategy for applying this
research, or developing an application centered research agenda from it.

7.3 Future Work and Implications

This dissertation examines a small piece of a large, complex, and nuanced phenomenon. However, each of its limitations suggests an avenue for future work.

First, this dissertation only examines the (contiguous) United States. In effect, it holds political system constant. However, comparable data exists in the European Union, Japan, and China. Taken together, these datasets would encompass over two thirds of global GDP, and could yield new insights into globalization.

Second, this dissertation only examines the years between 2000-2010, because most of the proceeding traffic data has not yet been digitized. However, even in this time period, there is a clear trend of networks becoming increasingly important for economic production. If proceeding records were digitized, it would be possible to measure precisely when networks began to increase in importance, and what changed to make it increasingly more important. This may offer unique insights into globalization, which is, perhaps, the most well known manifestation of the increasing importance of economic networks.

Third, this dissertation conducts a natural experiment, and finds that it may be a promising way to measure the causal influence of social connectivity on city populations. However, the experiment is small in scale, because there (thankfully) are not many natural disasters with all the needed characteristics. If this study were expanded to a wider time framework, it would make a stronger case for causality, and open up more precise avenues for measuring the magnitude and time frame of causal effect.

Fourth, this dissertation centers on social connectivity, but does not directly

---

4 Location is geographically distant from the US, but has network connections to multiple US cities. Disaster is severe, and strikes a major city – creating an exogenous shock to the economy.
measures the mechanisms (trust, innovation, social capital) that link social connectivity to city median incomes. Future work could explore avenues for directly linking mechanism to macro process.

Taken as a whole, this dissertation offers a different way of thinking about economic networks between cities. While most studies focus on networks of direct economic inputs, this study examines social connectivity between places. Organizational research finds that social networks enhance economic processes. This dissertation is one of the first to examine the implications of that social connectivity for cities. While cities do not have social networks, personal networks aggregate up to create social connectivity between cities. This enriches the city economic environment, enhancing productivity among its economic actors, even those not personally involved in those networks. These findings open up new avenues for research into place connecting networks. Specifically, it creates opportunities to ponder how personal networks, in aggregate, produce population consequences.
About 350 distinct pieces of scholarship are cited in this dissertation, and this represents only a small fraction of the field. While these works include 95 distinct academic journals, just 10 journals account for 45% of the works cited. Four of the journals come from the sociological paradigm (red), while three come from the economic paradigm (blue). The remaining three (purple) are either subject or methods focused journals that draw from practice-oriented communities of scholars. The top panel of A.1 reports the proportion of citations in this document that come from each journal. The wedges are colored to reflect whether the journal is sociological, economic, or practice-oriented.

The bottom panel of A.1 arranges the ten most cited journals according to the word similarity of the articles cited, using the same color scheme as the top panel. Over all, disciplines strongly shape the conceptual frameworks and terminology used in the journals on this topic. However, if we disregard discipline-wide terminology differences, text analysis reveals a deeper ordering to the prosperity networks field. Research divides roughly by the unit of analysis.
For each unit, 1-2 journals represent each discipline for each unit of analysis. *Economic Geography*, *Urban Studies*, and *Social Forces* publish much of the city-centric literature in economic, practical, and sociological journals respectively. They tend to use terms like work, household, asymmetry, and exclusion to discuss prosperity within and between cities. *American Journal of Sociology*, *Social Networks/Administrative Science Quarterly*, and *Journal of Economic Geography* perform similar disciplinary functions for organization-centric research. They use terms like collaboration, learning, contagion and pressure to describe how the connections within and between organizations influence those organizations. *American Sociological Review* and *Quarterly Journal of Economics* contain the sociological and economic literatures on nation-states.\(^1\) They use terms like integration, culture, trade, and theory to examine networks within/between nations.

The tragedy of this field is rampant duplication. Putting aside disciplinary jargon, three separate scholarly communities are conducting the same kinds of analysis on the same subjects, and have been doing so for decades. The true opportunities in this field (“the holes in the literature”) lie in tearing down disciplinary walls.

\(^1\) *International Studies Quarterly*, would probably best fit as the third journal here. However, ISQ networks articles are tend to cover topics with only an indirect connection to economic prosperity, so I have omitted it. Similar sentiments apply to runner-up *Journal of Conflict Studies*
Appendix B

Excluded Cities

On the 933 official, core-based statistical areas in the contiguous US, this dissertation examines 696. The remaining 237 collectively contain only 5% of the US population, and have statistically negligible ties to the national travel network.

Figure B.1 reports where parts of the local population have been excluded from this study. Most of the excluded population lives in the southern half of the middle of the US — an area were incorporation occurred relatively late. These areas tend to lie in large gaps in the US Interstate highway system. They also tend to fall along state borders, especially the New Mexico/Texas, Georgia/Alabama, Tennessee/Kentucky, and Oklahoma/Kansas borders.
There are two notable exceptions to this pattern. First, Fresno, CA is the largest US city with no connection to the interstate. Primarily an agricultural city, Fresno relies on California’s route 99 to ship produce to the San Francisco Bay area, which lies 175 miles to the North. Second, the Morgan City, LA region also has no interstate connection, because it has not yet been finished. 85 miles west of New Orleans, Morgan City is small, impoverished, lacking in significant industry, and prone to severe flooding. Its small, poor economy is prototypical of the kinds of cities excluded from the US travel network, even if it is much closer to a major city than most.

In fact, figure B.2 reports the chances that a city was excluded, given its size and median income. The majority of excluded cities are very small – generally falling below the minimum size for metropolitan statistical areas. Excluded cities also tend to have lower median incomes than included cities. However, there is a number of high income cities that are far removed from the passenger travel network. These remote but rich cities tend to be strongly tied to resource extraction industries.
Bibliography


Engels, F. (1892), *Socialism: Utopian and Scientific*, -.


Marx, K. (1848), Manifesto of the Communist Party, -.


PWC, P. (2010), “After the dust has settled... Financial fallout from the Icelandic Volcano,” *Hospitality and Leisure Client Briefing*.


Reilly, W. (1931), The Law of Retail Gravitation, -. 


Weber, M. (1922), *Economy and Society*, -.


Biography

Originally from Southern California, S Joshua Mendelsohn has been enthralled with the social sciences since middle school. As part of a science fair project, he carefully recorded defiant and compliant reactions to an inconvenient sidewalk detour sign.

At UC Berkeley, that interest blossomed into a fascination with complex human systems. He triple majored in sociology, psychology and political science – seeking to understand human behavior from its smallest manifestations to its largest. Joshua completed his honors thesis in sociology. His thesis examined how organizational structure influences purportedly personal choices. After graduating from Berkeley, Joshua continued his studies at Duke Sociology, earning university accolades for his research and data visualizations.

Joshua likes to think of himself as a data-driven problem solver combining complex data, quantitative analysis, and multi-disciplinary teams to provide practical insights and solutions for clients. As a researcher and consultant for the Duke Network Analysis Center, he has had the privilege of applying those skills to topics as diverse as Chinese epidemics, Soviet urbanization, Canadian politics, American traffic, Jordanian economics, and Carolinian public health.

If Joshua isn’t gleefully playing around with data, he is probably out on a dance floor. Contra and swing dancing are his personal favorites. So far, he has resisted the impulse to program an agent-based simulation of contra dance. It’s a daily struggle.

For more information, visit his website: http://www.sjoshuam.com