Quantitative Economics 9 (2018), 1419–1451 1759-7331/20181419

A scale-free transportation network explains the city-size distribution

Marcus Berliant
Department of Economics, Washington University

Axel H. Watanabe
Department of Economics, Concordia University and CIREQ

Zipf's law is one of the best known empirical regularities in urban economics. There is extensive research on the subject, where each city is treated symmetrically in terms of the cost of transactions with other cities. Recent developments in network theory facilitate the examination of an asymmetric transport network. In a scale-free network, the chance of observing extremes in network connections becomes higher than the Gaussian distribution predicts and, therefore, it explains the emergence of large clusters. The city-size distribution shares the same pattern. This paper decodes how accessibility of a city to other cities on the transportation network can boost its local economy and explains the city-size distribution as a result of its underlying transportation network structure. We confirm our model predictions with US and Belgian data. Finally, we discuss the endogenous evolution of transport networks.

Keywords. Zipf's law, city-size distribution, scale-free network.

JEL classification. L14, R12, R40.

1. Introduction

Cities develop in relation to other cities rather than in a vacuum. What we consume in a city differs from what we produce in a city. The gap between the range and scale of production and consumption at the city level is bridged by the transportation network, over which cities trade their products with others. The transportation network, in turn, does not coordinate cities uniformly. Some cities have only limited connections while others receive many links from cities across the country, both large and small, near and far away. The fate of a city's economy, and by extension its population size, is more or less conditioned by how it is positioned (inadvertently or otherwise) in the overall interurban network of cities and how accessible it is from others. We will show that the city-size

Marcus Berliant: berliant@wustl.edu
Axel H. Watanabe: axel.watanabe@concordia.ca
This project received a grant from the Center for Research in Economics and Strategy (CRES) at the Olin Business School, Washington University in St. Louis. The authors thank the Co-editor, three anonymous referees, Wen-Tai Hsu, Sukkoo Kim, John Nachbar, Jody O'Sullivan, and Victor Wickerhauser for their helpful comments and advice.

 https://doi.org/10.3982/QE619

distribution is the result of a particular class of network that our economy installs on itself for interurban trading purposes, namely, a scale-free network.

The city-size distribution is the culmination of various economic decisions made in all walks of life. Virtually any location-related choice changes the distribution, some significantly but others only negligibly. Researchers try to isolate primary determinants of the distribution and validate the strength of causal relationships with data. As such, there are many different angles from which to model the city-size distribution. Eeckhout (2004) used random growth to show that the city-size distribution asymptotically follows the lognormal distribution. Duranton (2007) used a quality-ladder model to gauge the growth of a city to derive the distribution. Glaeser, Scheinkman, and Shleifer (1995) offered yet another angle and focused more on the socioeconomic factors behind urban growth. Berliant and Watanabe (2015) looked into the potential for local technological shocks, used in much of the literature, to drive migration. They found that insurance and saving can serve as substitutes for migration, but that more severe reactions to shocks, namely survival of only the most efficient firm in an industry, can drive migration. This generates a Generalized Extreme Value (GEV) city-size distribution. As in the balance of the literature, the transportation network is in the background.

The existing literature's treatment of the transportation network has been rather naive and simplistic. Most existing models of the city-size distribution implicitly or explicitly assume a completely isolated graph (Figure 1) or complete graph (Figure 2). Each node represents a city and a link represents a route available for shipment in these figures. The number inside a node counts its degree, that is, the number of edges or routes each node has. Commodities cannot be shipped at all on a completely isolated graph, but they can be shipped anywhere in a single step from any city on a complete graph. Either way, neglecting other factors, the resulting equilibrium will be an even split of population among the cities, which does not match the actual city-size distribution. To explain the city-size distribution, researchers have sought a source of variation other than what the nexus of interurban relationships has to offer. Some use a completely isolated graph (e.g., Eeckhout (2004)). Others such as Duranton (2006), Rossi-Hansberg and Wright (2007), or the New Economic Geography (Fujita, Krugman, and Venables (1999)) engaged a complete graph as the transport structure, when in fact, transaction and/or

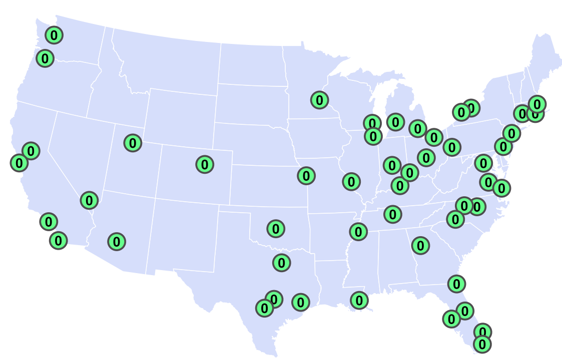

FIGURE 1. The United States according to completely isolated graph with the 50 largest cities.

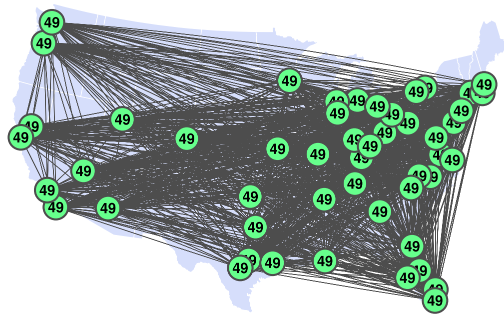

Figure 2. The United States according to complete graph with the 50 largest cities.

communication between hub cities is much easier than between cities on peripheries. Behrens, Mion, Murata, and Südekum (2017) and Eaton and Kortum (2002) introduced a more lifelike representation of transportation cost in that the delivered price depends on a particular city pair. The price differential reflects monopolistic pricing (in Behrens et al. (2017)) or exogenous trade barriers (in Eaton and Kortum (2002)) rather than the underlying transportation network structure, which is still an (ex ante) complete graph and thus network features such as a hub or through traffic are absent. The literature usually introduces a tiebreaker in the form of externalities, random growth, economies of scale, or economies of scope to replicate the actual city-size distribution (cf. Section 3.4).

In practice, transportation cost differs greatly depending on where you are and where you are headed. We will drop the assumption that our economy operates on a complete or completely isolated graph and see how much explanatory power network structure exerts as the engine of local economies of various sizes.

The transaction pattern between any two cities affects both the way cities are populated and the overall city-size distribution. Cities are tied together in various ways both topologically and economically. Some cities function as an intersection of major transportation routes and they trade and process commodities frequently in large volume. Others are less active in the interurban exchange of commodities. Differences among cities in terms of exchange patterns reverberate in the city-size distribution. Cities heavily interrelated to many others are likely to grow due to increased economic activities, whereas cities with sparse connections to a limited number of cities are liable to remain small in size. Those small cities, however, will not be completely wiped off the map.

1.1 *Cities on a network*

Intercity exchange patterns like Figures 1 and 2 are best described by a network with cities as a set of vertices and traffic by edges. In this regard, network theory is indispensable when constructing a model of cities in the nationwide economy.

The recent seminal work by Barabási and Albert (1999) has revitalized network theory. Classical network theory pioneered by Erdős and Rényi (1959)'s model (ER network) cannot explain the emergence of a cluster or hub in a network, which we observe in most real social networks. In a classic random graph, each node is linked with an equal probability to any other and lacks distinctiveness, for the number of preexisting links does not matter in forming a network. Barabási and Albert (BA) added a dynamic feature and preferential attachment to the classical random graph model so that the nodes are no longer ex ante identical. Some nodes gather a lot of links while others are wired to just a few. The model has been applied to many fields, including the emergence of web science, and has produced an improved description of the organization and development of networks. Most real-world networks have one thing in common: the resulting distributions of links are scale invariant, that is, the distributions have fat tails. We can find nodes with an extremely large number of links rather easily with these networks compared to a classical random graph.

The city-size distribution shares the same pattern of scale invariance: the distribution of the 100 largest cities follows the same distribution as the one for the 1000 largest cities and so on, a property known as a power law, and in particular, Zipf's law in the city-size literature. We expect that the degree of a city is positively related to its population. And for that reason, we imagine that our economy is based on a BA network rather than an ER network. This turns out to be correct, but selection of the appropriate network structure depends on exactly how node degree is related to city size. We will decode their relationship in Section 3.8.

The urban economic application of network theory is in its very early stage of development and there is much room for advancement. Interaction between individual cities has not caught much attention so far. Our goal in this paper is to bring to the forefront the interaction between transportation network structure and the city-size distribution. With this goal in mind, we introduce (asymptotic) techniques from network theory and merge them with a tractable economic model in a new way. We do not intend this work to be the last word on this topic, but merely a suggestion of a first step into a bigger research program.

1.2 *Some transportation networks are scale-free*

Our economy operates on various modes of transportation and each mode comes with distinct network structures; take a highway and airline network, for example. Figures 6 and 7 in the Supplementary Material (Berliant and Watanabe (2018)) are schematic representations of the Interstate System and a typical airline route map for the 50 largest US cities. Apparently, a network composed of the Interstates does not share its structure with that of airlines. The Interstate network will remain relatively intact when we take away New York, Houston, and Cleveland. On the other hand, it would prove devastating if we did the same to the airline network (cf. Barabási and Bonabeau (2003)). More broadly, there is not much variance in the degree of nodes in the Interstate network, whereas the airline network has a limited number of heavily wired cities. The BA network (Figure 7) explains the latter network better, as it follows a power law.

It should be noted, however, that what is geographically visible may *not* represent the real network that our economy relies on in effect. The Interstate network exhibits an ER-type topology as in Figure 6. Nonetheless, the economy may operate a transportation network of a scale-free class on it. Shipment from Memphis has to go through St. Louis even if its final destination is Chicago. In this case, Memphis is connected to Chicago in a single step rather than in two steps via St. Louis. For a carrier making Chicago-bound shipment from Memphis, St. Louis (a seeming layover node) is no different from the cornfield they pass through along the way (just a part of the edge), in that neither one of them add anything to the shipment. An economically relevant network is buried beneath the easily noticeable surface network and we do not want to confuse one with the other.

It is also very important to note here a difference between the literature on dynamic social network formation and transportation networks. In the standard economics literature on social networks, for example, Mele (2011) or Christakis et al. (2010), it is the individual agents, represented by nodes, who make decisions about forming links among themselves. In contrast, the nodes of a transport network are cities. Typically, it is not the cities or their agents who make decisions about forming links. Rather, it is another agent who controls an entire networks, for example, the federal government in the case of highways or airlines in the case of an airline system.[1]

1.3 *The city-size distribution is scale-free, too*

The city-size distribution has a distinct feature. Figure 8 in the Supplementary Material (Berliant and Watanabe (2018)) plots the frequency of the city-size distribution from US Census 2000. It is only when we take the log of population (Figure 8(b)) that the distribution exhibits resemblance to a familiar Gaussian distribution. Black and Henderson (2003) and Soo (2005) explained how widespread scale-free distributions are in urban economics.[2] Under the scale-free distribution, the arithmetic mean (Hillsboro, TX in Figure 8) becomes less interpretive and the geometric mean (Sutton, NE) takes over the role of the average in the conventional sense.

The fat-tailed distribution also makes its appearance on a map. Figure 9 in the Supplementary Material (Berliant and Watanabe (2018)) illustrates the population density of each metropolitan and micropolitan statistical area (MSA and μSA, collectively referred to as Core Based Statistical Area, CBSA) in the United States in 2000. Most of the cities have a low density and are painted in blue; there are only a few cities that are green and only two cities are colored in red. If the city-size distribution followed a Gaussian distribution or Poisson distribution with a large mean,[3] most of the cities should be green and only a few should be in blue or red. Just as for the airline network in Figure 7, if we take away the ten largest US cities, we will leave more than a quarter of the urban population unaccounted for.

Our main findings are as follows. City sizes are positively related to their degree. A city with a high degree has good accessibility to other cities. Reduced transportation

[1] See Section 3.10 for further details.

[2] Scale-free distributions are commonplace in the socioeconomic realm. It seems that something of an additive nature presides over natural phenomena, leading to a Gaussian distribution, and something of multiplicative nature (cf. Limpert, Stahel, and Abbt (2001)) is at work among socioeconomic phenomena, leading to a scale-free domain. We study the latter.

[3] as in the degree distribution of an ER network

cost makes the city's product inexpensive and stimulates a large demand. As a consequence, the city creates large-scale employment. However, a marginal increase in degree contributes less to the city size as the degree increases. If a city is well connected, then adding a new link to the city will not increase accessibility much because the city is already readily accessible from other cities through the existing grid.

We test implications of our model with Belgian and US data. The BA network leads to a result comparable to existing models, whereas the ER network fails to replicate the empirical city-size distribution. This confirms that the BA transport network is more consistent with reality.

The rest of the paper is organized as follows. In Section 2, we will go over the two types of network structures mentioned above as a preamble to the next section, where we introduce and develop a model of spatial equilibrium with a transportation network woven into it. Particularly, in Section 3.8, we will connect the network structure to the city-size distribution. In Section 4, we verify the prediction of our model with data before we draw conclusions from our project in Section 5.

2. Preliminaries

We will briefly review how ER and BA networks are built and examine the qualitative differences in terms of their degree distributions before we apply them to transportation networks.

2.1 *ER networks*

The ER network is the simplest random graph of all. A pair of nodes are connected with a fixed connection probability. A completely isolated graph illustrated in Figure 1 and a complete graph illustrated in Figure 2 are the special cases of the ER network where connection probability is zero and one, respectively.

The degree distribution of an ER network follows a Poisson distribution. The important feature is that the degree distribution is concentrated around its arithmetic mean[4] and we rarely observe a city with an exceedingly large degree. All pairs of nodes share the *same* ex ante connection probability, which leads to a small variance, and the network is *egalitarian* in that sense.

Unsophisticated as it may seem, the ER network makes a good entryway to economic applications of network theory. Network theory puts emphasis on interactions, and thus it becomes particularly useful for situations where an economic agent does not interact with all the other agents either at his discretion or due to external restrictions. We would not have to pay any attention to networking if everyone were in direct contact with anyone else. In reality, system-wide interactions are not common. Most economic decisions or interactions are made in reference to limited alternatives available, which we represent by an edge on a network. Ultimately, we would like to know how agents choose their trading or collaborating partners as a result of their optimization.

[4] Recall that arithmetic mean does not mean much for scale-free distributions like the city-size distribution or a BA degree distribution.

However, leaving their choice purely stochastic (as in the construction of ER networks) still proves to be a good reference point to see whether the network is self-organized as a result of decentralized decision making. Kakade et al. (2004) used it as a benchmark for the Arrow–Debreu model with transactions constrained by connected traders on a network. Calvó-Armengol and Zenou (2005) assumed that each worker selects a collection of (randomly selected) direct neighbors to describe the role that a network plays in job matching. In some cases, the ER network *is* the sensible choice to represent real networks. Toulis and Parkes (2011) modeled the kidney exchange program with the ER network to evaluate the efficiency of the program. Any pair of a donor and a patient is compatible with a fixed probability. See Ioannides (2006) for a comprehensive review of economic applications of ER networks.

2.2 *BA networks*

The degree distribution of most real network structures does not follow a Poisson distribution. Rather, it follows a power law. This class of networks is called scale-free. There are a number of proposed generative models that lead to power-law degree distributions (see Section VII of Albert and Barabási (2002) for a review). To get a sense of how power-law type behavior emerges, consider the BA model (Barabási and Albert (1999)), for example. Two major characteristics of the BA model are growth and preferential attachment. The model sets off with a complete graph of a fixed number of nodes as a starting grid. New nodes with edges will be added sequentially to the existing network (growth) with the probability of attachment proportional to the degree of existing nodes (preferential attachment). In general, older nodes are likely to gain an excessively large number of edges. The rich get richer because they are already rich (known as the Matthew effect). The rest of the nodes are merely mediocre in terms of degree. They remain poor because they are already poor. This type of variance in degree hardly arises with an ER network. That is, New York City will not happen if the links are formed uniformly at random. Compare a BA network (Figure 7) to an ER network (Figure 6). A BA network is *not* egalitarian, as connection probability depends on the number of acquired edges, which is path dependent. We shall also employ the network structure of Jackson and Rogers (2007) that contains both the ER and BA types of networks as special cases, the details of which will be provided in Section 3.8.

3. Model

We propose a model where the trading costs of commodities among cities are explicitly specified. The city-size distribution is derived as a result of gains from trade and the underlying transport network configuration.

3.1 *Location-specific commodities*

There are J cities in the economy, with index i or j. A city is defined as a geographic entity within which it produces the same commodity and from within which the geodesic paths (the shortest path on the network) to any other city in the country have the same length. The endogenous population of city j is given by s_j and in total there are

$$\sum_{j=1}^{J} s_j = S \tag{1}$$

households in the economy. Each household supplies a unit of labor inelastically. City j produces consumption commodity c^j in a competitive environment. We assume that technology exhibits constant returns to scale and that one unit of labor produces one unit of commodity. In what follows, a superscript denotes a city of production or origin, whereas a subscript denotes a city of consumption or destination.

The delivered price of commodity j in city i is denoted by p_i^j. The value of marginal product $p_j^j \cdot 1$ coincides with the local wage w^j in equilibrium:[5]

$$p_j^j = w^j. \tag{2}$$

Consumer preferences are represented by a Cobb–Douglas utility function of the form $u(c_i) = \frac{1}{J} \sum_{j=1}^{J} \log(c_i^j)$. The set of consumption bundles is constrained by the budget $w^i \geq \sum_{j=1}^{J} p_i^j c_i^j$.

3.2 *Network infrastructure and delivered price*

The economy has a network infrastructure $\Gamma = (V, E)$, where $V = \{1, \dots, J\}$ denotes a set of vertices representing cities and E denotes a set of edges. All the traffic flow will follow Γ. We assume that Γ is connected, that is, there is at least one path between any city pairs, to avoid multiple equilibria. Whereas consumers in city i can consume any commodity in the economy, they have to incur an extra iceberg transport cost to consume commodities brought in from other cities. Transportation cost piles up as a commodity travels from city to city along the path. To describe the exact transport cost structure, we define a metric $l_i^j : V \times V \to \mathbb{R}_+$ to measure a geodesic length between nodes j and i given Γ. The delivered price of commodity j shipped to city i is given by

$$p_i^j = \tau^{l_i^j} p_j^j, \tag{3}$$

where τ (≥ 1) marks the iceberg transportation parameter. We use the iceberg transport technology, standard in urban economics, for tractability reasons.[6] If you dispatch $\tau^{l_i^j}$ units of commodity j to city i, one unit of it will be delivered. Consequently, the delivered price snowballs as the package travels from one city to another and the initial mill price is inflated by τ raised to the l_i^jth power by the time the package arrives at its final destination l_i^j steps over.[7]

[5]Note that p_j^j denotes the mill price.

[6]For detailed discussion, see McCann (2005).

[7]We adopt the exponential form of iceberg transportation cost for the remainder of the paper. The linear form yields approximately the same results (see Appendix A.1 for the case of the linear transportation cost).

We assume that all the links share the same value of τ. The large fraction of transportation cost is a location-invariant fixed cost. Having τ dependent on each link will not add much to our analysis but will make our equilibrium analytically insolvable.

3.3 *Equilibrium*

Marshallian demand for commodity c_i^j at destination i is $\varphi_i^j(p_i^1, \dots, p_i^J, w^i) = \frac{w^i}{\tau^{l_i^j} p_j^j J}$, and accordingly, at origin j is $\psi_i^j(\cdot) := \tau^{l_i^j} \varphi_j^i(\cdot) = \frac{w^i}{p_j^j J}$.[8] The aggregate demand for commodity j at its origin is the sum of demand from all the cities in the country,

$$\Psi^j(p, w) := \sum_{i \in V} s_i \psi_i^j(\cdot) = \frac{\sum_i s_i w^i}{p_j^j J} = \frac{\sum_i X^i}{p_j^j J} = \frac{\langle X \rangle}{p_j^j}, \tag{4}$$

where $X^i := s_i p_i^i$ is the value of output inclusive of transportation sector in city i. In what follows, $\langle x \rangle$ denotes the average value of x, for example, $\langle X \rangle := \sum_i X^i / J$. The third equality in (4) holds when labor market is in equilibrium as in (2). Recalling that each household supplies one unit of labor inelastically and one unit of labor produces one unit of output, the commodity market j clears when

$$s_j = \Psi^j(p, w). \tag{5}$$

From (2), (4), and (5), we obtain the equilibrium price and wage as follows:

$$w^j = p_j^j = \frac{\langle X \rangle}{s_j}. \tag{6}$$

The indirect utility function is given by

$$\begin{aligned} v(p_i^1, \dots, p_i^J, w^i) &= \frac{1}{J} \sum_{j=1}^{J} \log \varphi_i^j(\cdot) \\ &= \log w^i - \log J - \langle \log p_j^j \rangle + a_i \log \tau, \end{aligned}$$

where

$$a_i := -\langle l_i \rangle = -\sum_k l_i^k / J \tag{7}$$

measures accessibility of city i. We will examine the role of a_i shortly. Free mobility of consumers implies

$$v(p_i^1, \dots, p_i^J, w^i) = v(p_j^1, \dots, p_j^J, w^j) \tag{8}$$

for all $i, j \in V$ in equilibrium.

[8] This expression may seem incredulous at first, for it does not include τ. We will explore the reason in Sections 3.9 and 3.10.

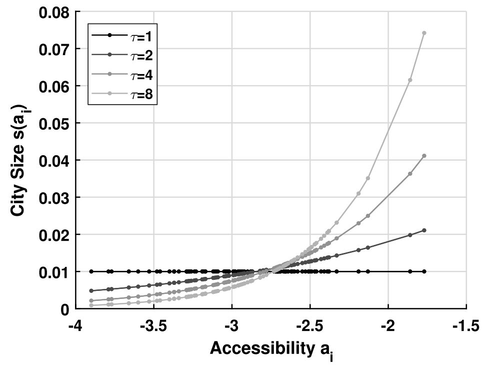

FIGURE 3.

The equilibrium $(s_1, \dots, s_J; p_1^1, \dots, p_J^J; w^1, \dots, w^J)$ satisfies (1), (2), (5), and (8). Equation (8), together with (5), implies $\log s_i - \log s_j = a_i \log \tau - a_j \log \tau$. With the population condition (1), we obtain the city-size distribution

$$s_i = \langle s \rangle \frac{\tau^{a_i}}{\langle \tau^a \rangle}, \tag{9}$$

where $\langle s \rangle := S/J$ is the size of a city if the population were split evenly.

Since $\langle l_i \rangle$ is an average geodesic length from city i to anywhere in the nation, a high value of a_i as defined by (7) implies that on average, city i is easy to get to, and vice versa if a_i is low. A better accessibility increases a city size: The ratio of s_i to $\langle s \rangle$ matches the ratio of τ^{a_i} to $\langle \tau^a \rangle$. Therefore, the city size grows *more than proportionately* with accessibility as can be seen in Figure 3.

3.4 *Interplay between network structure and convex preferences*

The relationship we derived in (9) begs one question: If an accessible city attracts workers, what is stopping the city-size distribution from becoming degenerate, that is, would not the entire population collapse into the city with the best accessibility and the rest of the cities be completely vacated?

That actually will not happen. The economy faces a trade-off between accessibility and convex preferences, with the former pushing the city-size distribution toward a degenerate distribution as above but with the latter dragging it back to a uniform distribution. The equilibrium distribution will be somewhere in between the two as a result of the balancing act, which we will describe below.

Although restricted accessibility of a city raises its delivered prices, demand for its product does not cease to exist. Eliminating a commodity would be vindictive to consumers. They appreciate variety and missing a single variety will push the utility level down to negative infinity. Workers in a poorly connected city will have to pay high prices for imported commodities due to its poor transportation infrastructure, but they are compensated with a high nominal wage: (6) indicates that the nominal wage grows as the city becomes small.[9] Furthermore, (6) and (9) imply $w^j = p^j_j = \frac{\langle X\rangle\langle\tau^a\rangle}{\langle s\rangle\tau^{a_j}}$, that is, the nominal wage increases as accessibility to the city becomes restricted. The prices adjust to make it worth living in small cities in equilibrium. In particular, (6) implies that GDP in each city $X^j := p^j_j s_j$ levels out to

$$X^j = \langle X\rangle \tag{10}$$

across the country. The scale of local production is small, but each commodity is sold high to make up for an increased cost of living due to remoteness and the resulting costly transport.

Variance in city sizes is solely due to the structure of the network. The aforementioned trade-off entails two counteracting forces. The agglomerative force is heterogenous accessibility, which tends to create heterogeneity in the city-size distribution. The dispersion force is preference for variety, which tends to push the distribution back toward a collection of equal-sized cities.

There are alternative ways to derive city size with a tractable economic model, particularly for the dispersion force. In this model, location-specific commodity production drives dispersion, as a bundle of all goods is desired by consumers. An alternative model would use another natural dispersive force, say housing or land markets. If we had just a few produced commodities (say one for illustration), then Starrett's spatial impossibility theorem (Fujita and Thisse (2002, Chapter 2)) applies, and we would have an autarkic equilibrium where no commodity is transported.[10] Yet another alternative is to introduce a congestion externality, but then the model begins to look more complicated and, at the same time, arbitrary.

Obviously, this trade-off disappears and there will be no variance in city sizes if the agglomerative force is removed. This can happen when shipment becomes costless (to be discussed in Proposition 3.1) or network structure becomes redundant, that is, if it turns into a complete graph. Although we introduced a location-specific technology, commodities are symmetric. Technology is linear everywhere. Consumer preferences are identical and they put the same weight on each commodity. If we take the network structure out of the equation, the resulting equilibrium is such that all the cities share the same size $\langle s\rangle$ and every household consumes an equal portion of all the commodities available.

[9]Whereas this implication may not sound realistic, we emphasize that a small city earns a high wage *only in a nominal sense*. The delivered prices are also high in a small, wage-rich city, and thus its utility level will work out to the same level as a large, wage-poor city's in the end.

It is possible to make wages increase with size but that will create another problem. One way to do so is to allow a city to produce multiple commodities by exogenously limiting the employment in, and thus the scale of, each industry. (We thank an anonymous referee for this suggestion). For this alternative model to work, we make an individual industry size increase with the city size (otherwise, the equilibrium wage would depend on the location-invariant individual industry size rather than the location-variant city size, and thus the equilibrium would support any city-size distribution). Starting from this assumption, we can secure a positive relationship between the wage and size as desired. However, we now have to face another unwanted consequence: The city size *declines* with its degree because a large city comes with a wide range of commodities, which compensates for its low accessibility.

[10]Starrett's theorem makes no assumption about the transport network or transport cost.

3.5 *Transportation cost skews the city-size distribution*

Along with accessibility a_i, transportation cost τ plays a leading role in the determination of the city-size distribution. Depending on its magnitude, shipment cost can nullify or amplify the influence of a network structure over the economy. Figure 3 compares the relationship between accessibility and the city-size distribution under different transportation costs.

In the extreme situation where shipment is free ($\tau = 1$), all cities will be of an equal size regardless of the network structure. The city size $s(a_i)$ becomes constant against a_i (see the line for $\tau = 1$ in Figure 3). The network becomes a complete graph in effect, because the delivered price will be the same no matter how long the geodesic length is. For $\tau > 1$, city size (9) becomes a strictly convex function of accessibility.

The agglomerative force mentioned in Section 3.4 becomes more potent as τ grows. A large τ implies that the geodesic length exerts a more dominant influence on the size of a city. With a small value of τ, a city with good accessibility does not distinguish itself well from other cities because the effect of path length is limited due to low transportation cost. On the other hand, if shipping is costly, a city with a good accessibility benefits from a high a_i value because high transportation cost amplifies the effect of accessibility. As a result, holding the accessibility distribution constant, large τ skews the city-size distribution and makes the emergence of disproportionately large hubs more likely. To measure how the cost of transportation τ bends the city-size distribution, consider a measure

$$D(\tau) = \frac{s(a_H) + s(a_L)}{2} - s\left(\frac{a_H + a_L}{2}\right),$$

where a_H and a_L are the highest and lowest accessibility of a given network. The first term is the average of the smallest and the largest city whereas the second term is the city size of average accessibility. For a given distribution of accessibility a_i, $D(\tau)$ measures the convexity of $s(a_i)$, that is, it gauges how spread out the distribution of city size $s(a_i)$ is for each τ. See Figure 4. When $\tau = 1$, $s(\cdot)$ lays flat and $D(\tau) = 0$. As τ grows, $s(\cdot)$ bends more and $D(\tau)$ grows accordingly as can be seen in Figure 3.

We confirm the observation above as follows.

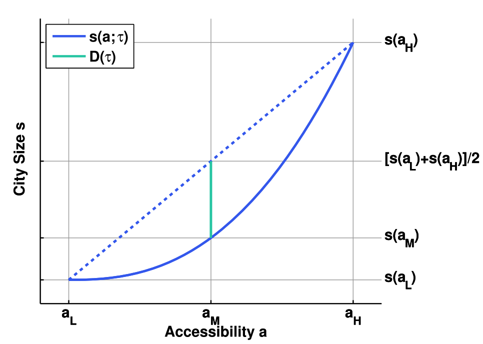

FIGURE 4. $D(\tau)$ measures the convexity of $s(a_i)$. The midpoint $(a_H + a_L)/2$ is given by a_M above.

Proposition 3.1 (Transportation Cost Skews the City-Size Distribution). *Suppose that the economy operates on a connected network* Γ. *The city-size distribution* $s(a_i)$ *is a convex function of accessibility* a_i *for* $\tau \geq 1$. *Moreover, the degree of convexity measured by the size difference* $D(\tau)$ *between the average of the highest and lowest size cities and the city of average accessibility increases with* τ.

Proof. See Appendix A.2. □

It is also worth pointing out the difference in the role of transportation cost between the current paper and New Economic Geography (NEG, Fujita, Krugman, and Venables (1999)). As Proposition 3.1 suggests, *high* transportation cost concentrates production and population in a few cities. NEG predicts otherwise: It is *low* transportation cost that leads to a core-periphery pattern. The root cause of these conflicting implications lies in the production function adopted in each model. Whereas we assume constant returns to scale, NEG assumes increasing returns to scale due to a fixed cost. In NEG models, it pays to concentrate production in one city to capitalize on scale economies so long as transportation cost is not too high. Here, in contrast, there is no reason to concentrate production in a few places just for the sake of a scale economy (because there is none). Low transportation subdues the effect of transportation structure, which is our source of variation in city sizes. Unlike NEG, we leave the production technology neutral and ascribe the difference in city sizes solely to the underlying network structure. We could similarly adopt increasing returns to scale in our model but that would blur the very role of network structure that we are interested in.

3.6 *Geodesic-length distribution*

The city-size distribution (9) depends on the distribution of accessibility (7), which, in turn, rests on the distribution of geodesic length. There is not much research that looks into the geodesic length between each pair of nodes.[11] At the time of writing, the analytical form of geodesic length between individual nodes is yet to be discovered.[12] Hołyst, Sienkiewicz, Fronczak, Fronczak, and Suchecki (2005) took a different approach to derive an intuitive solution for a wide range of network types. They measured the expected geodesic length between any pair of nodes i and j as follows:

$$l^i_j = A - B\log(k_i k_j), \tag{11}$$

[11]While most of the research on network topology is focused on *mean* intervertex distance (Newman, Strogatz, and Watts (2001), Fronczak, Fronczak, and Hołyst (2004), Zhang, Lin, Gao, Zhou, and Guan (2009)), what we need here is the geodesic length between *individual* nodes.

[12]The one for the average intervertex separation has already been brought out into the open. Newman and Watts (1999), Newman, Moore, and Watts (2000), Zhang et al. (2009). Zhang et al. (2009) provided an analytical background for the mean intervertex distance for a special case. There has also been an attempt to track down the geodesic length by guessing the analytical form from sequentially generated, fractal-like networks reverse-engineered from a Pareto degree distribution (Dorogovtsev, Mendes, and Oliveira (2006)), which we cannot use because our distribution (16) is not a Pareto distribution.

where $A := 1 + \log(J\langle k \rangle)/\log \kappa$ and $B := (\log \kappa)^{-1}$. The number k_i denotes the degree of node i and κ is a mean branching factor. The branching factor of a node is the number of children that the node branches off on a tree. See Appendix A.3 for a full description of κ.

Although Hołyst et al. (2005) does not provide a formal proof of (11), but rather is based on a heuristic,[13] it appears to be the best we can do given the current state of network theory. We hope that its extension to individual distances will become available in the near future.

Meanwhile, (11) proves to be quite useful in translating a network structure into economic context without loss of generality. A geodesic length l_j^i is a *global* property whereas a degree k_i is a *local* property.[14] We cannot compute the individual geodesic path unless we compare all possible paths between a city pair of interest and pick the shortest one, which calls for a systemic search all across the board. The geodesic path thus obtained is too specific to the particular network in question and does not have wide implications beyond the specific network under study itself. Degree is much easier to compute because we do not have to launch a nationwide search for it, and the degree distribution is readily available for a wide range of networks. Equation (11) succinctly writes a global property (a geodesic path length) in terms of the analytically manageable local property (a degree). It implies that the path length will be short if your city and/or your destination city have many edges to choose from to begin with and/or to end with. This abundance in selection should save you from being thrown to circuitous paths, and vice versa when your degree is small. Absent this conversion of the global property into the local property, we would not be able to describe a general relationship between degree and city size, when in fact, there is an obvious symbiotic interaction between them waiting to be investigated.

3.7 *City-size distribution*

From (11), accessibility (7) is written as

$$a_i = -A + B \log k_i + B\langle \log k \rangle, \tag{12}$$

where $\langle \log k \rangle := \frac{1}{J} \sum_j^J \log k_j$. We observe that accessibility improves as a city acquires more edges, but only on the logarithmic order. Taking the log of (9), we have

$$\log s_i = \log S + \big(-A + B \log k_i + B\langle \log k \rangle\big) \log \tau - \log\Bigg(\sum_j \tau^{a_j}\Bigg).$$

[13] In a manner similar to Simon (1959).

[14] In fact, both a_i (closeness centrality) and k_i (reach centrality) are specific examples of network centrality, and we unite them via (11) (cf. Freeman (1978)).

The last term is approximated by $\log J + \langle a \rangle \log \tau$[15] so that

$$\log s_i = \log\langle s \rangle + B \log \tau \big(\log k_i - \langle \log k \rangle\big). \tag{13}$$

A couple of observations are in order. The equation above answers two questions concerning the relationship between a network structure and a system of cities. The first one is "Does construction of an edge boost the local economy?" The answer is "Apparently, it does." The second, and more interesting question is "How so?" The answer is twofold.

In terms of a linear scale, (13) can be rewritten as $s_i = \langle s \rangle (\frac{k_i}{\gamma})^{B \log \tau}$, where $\gamma := \prod_{i=1}^{J} k_i^{1/J}$ is the geometric mean of the degree. It indicates that city size is anchored around the base city size $\langle s \rangle$ multiplied by the deviation $(k_i/\gamma)^{B \log \tau}$. If a city has a large degree, then its size becomes larger than the standard city size by a factor of $(k_i/\gamma)^{B \log \tau}$ and vice versa for a city with a small degree. The city size coincides with the cornerstone size of $\langle s \rangle$ exactly when its degree matches the national (geometric) average.[16] The deviation is amplified as shipment becomes costly, which, in turn, confirms our observation made in Proposition 3.1.

We also note that adding an edge to a city increases its size, but the change in size is inversely proportional to the current degree provided $B \log \tau < 1$. If city i is highly wired already, then the introduction of a new edge to city j does not add much to city i. The geodesic length to city j is already short before the establishment of the new edge. You can go to many cities in a single step and city j is likely to be linked to at least one of those many neighboring cities already, making the geodesic length to city j just two. The added edge will only reduce the geodesic length by one. On the other hand, if the current degree of city i is low, then the link to city j will not only reduce the geodesic length to city j greatly but also reduce the geodesic lengths to the cities in city j's neighborhood. Consequently, city i will see significant improvement in its accessibility.

Based on the degree-size relationship (13), our main theoretical result gives the city-size distribution as follows.

Proposition 3.2 (City-Size Distribution). *Suppose that the economy operates on a connected network Γ with the associated degree distribution $G(k)$. The city-size distribution of this economy follows the distribution function $F(s)$, defined by*

$$F(s) = G\big(k(s)\big), \tag{14}$$

[15] Let $\vec{a} := (a_1, a_2, \ldots, a_J)$ and $\langle \vec{a} \rangle := (\langle a \rangle, \langle a \rangle, \ldots, \langle a \rangle)$. The Taylor series expansion about $\vec{a} = \langle \vec{a} \rangle$ tends to

$$\begin{aligned}\log\Big(\sum_j \tau^{a_j}\Big) &= \log\Big(\sum_j \tau^{\langle a \rangle}\Big) + \big(\vec{a} - \langle \vec{a} \rangle\big) \cdot D \log\Big(\sum_j \tau^{a_j}\Big)\Big|_{\vec{a}=\langle \vec{a} \rangle} + O\big[\big(\vec{a} - \langle \vec{a} \rangle\big) \cdot \big(\vec{a} - \langle \vec{a} \rangle\big)\big] \\ &= \log J + \langle a \rangle \log \tau + O\big[\big(\vec{a} - \langle \vec{a} \rangle\big) \cdot \big(\vec{a} - \langle \vec{a} \rangle\big)\big].\end{aligned}$$

[16] This examination begs one question: If my city has the average number of edges, is my city larger or smaller than the national average in size? The answer is "larger." Since transportation cost and the branching factor are both greater than one, $\frac{\log \tau}{\log \kappa}$ is positive. Plus, the geometric mean is smaller than the arithmetic mean. To score a national average $\langle s \rangle$, you only need γ edges. It should be noted, however, that in a scale-free world, the arithmetic mean does not carry much information. The lognormal is the new normal (or any heavy-tailed distribution is for that matter) and the geometric average is the new average in this world as we saw in Figure 8(b).

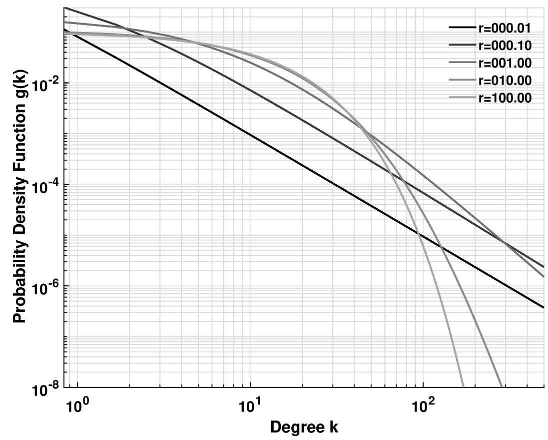

FIGURE 5. Probability density function of degree with $k_0 = 0$ and $m = 10$.

where $k(s) := \gamma(s/\langle s\rangle)^{\frac{\log\kappa}{\log\tau}}$. *Its probability density function (PDF) is*

$$f(s) = k'(s)g\big[k(s)\big] = \frac{\log\kappa}{\log\tau}k(s)s^{-1}g\big[k(s)\big], \tag{15}$$

where $g(\cdot)$ *denotes the PDF of degree* k.

3.8 *City-size distribution under different network systems*

Now that we have the city-size distribution based on the city's degree, we can make our predictions based on different transport network structures. There are two network models of particular interest: ER and BA networks. Jackson and Rogers (2007) construct a degree distribution of a directed[17] dynamic network as follows:

$$G(k) = 1 - \left(\frac{k_0 + rm}{k + rm}\right)^{1+r} \quad \text{for } k \geq k_0, \tag{16}$$

where k_0 denotes the in-degree with which an entering node is endowed. This value is shared across all the nodes. The parameter r plays a crucial role in our analysis. It locates where the existing network stands on the spectrum of networks ranging from an ER to a BA network. In particular, it is the ratio of the number of links formed by an ER-like random connection to a BA-like network-based connection. The average out-degree of a node is given by m. Five PDFs of (16) are depicted in Figure 5 as a visual cue. In the figure, parameter r ranges from 0.01 (over 99% network-based and less than 1% random

[17]Commodities can flow either way on an edge. We take an arrowhead on a directed edge just as a decorative memorabilia indicating from which end the edge was constructed, but nothing more. We represent degree distribution by an in-degree distribution. It is impossible to tell different networks apart with an *out*-degree distribution due to the way a network is constructed in Jackson and Rogers (2007). Any network comes with a degenerate out-degree distribution.

links) to 100 (the other way around). A predominantly random PDF (with large r) tapers off quickly whereas a mostly network-based PDF (with small r) only gradually dissipates with degree. We expect that our economy operates with a small r. BA network's degree distribution is (16) with $r=0$, in which case, (16) turns into a Pareto distribution. ER network calls for $r \to \infty$, in which case (16) is no longer well defined and the degree distribution turns into an exponential distribution.[18]

What is left to do is write the mean branching factor κ in terms of other parameters in (16) before we can fully identify the city-size distribution.[19] The actual mean branching factor cannot be computed until after the network is formed. Hołyst et al. (2005) provided a good approximate to κ:

$$\kappa = \sum_{k=1}^{J} k \frac{kg(k)}{\sum_{x=1}^{J} xg(x)} - 1 = \frac{\sum_{k}(2k-1)G(k)}{\sum_{x} G(x)} - 1 = \frac{\mu_k^2 + \sigma_k^2}{\mu_k} - 1, \tag{17}$$

where μ_k and σ_k^2 denote the mean and variance of k. See Appendix A.4 for details.

3.9 *The gravity equation*

Before we compare our theoretical prediction to actual data, let us briefly turn aside to discuss our model in the context of the gravity model (cf. Bergstrand (1985)). In fact our model is a special case of it. Our consumer preferences are represented by a Cobb–Douglas utility function, a limiting case of CES utility function with the elasticity of substitution approaching one. Due to the absence of cross-price effect, our gravity equation is less involved than its generic CES counterpart.

Consider a trade flow from producing city j to consuming city i. The delivered volume of good is $s_i \psi_i^j(p_i^1, \dots, p_i^J, w^i) = \frac{X^i}{p_i^j J}$ so that sales value X_i^j in city j is $p_i^j \frac{X^i}{p_i^j J}$. Therefore, the gravity equation takes a simple form $X_i^j = \frac{X^i}{J}$. In this case, the gravity is one-sided (X^j does not have any gravitational pull) and transportation cost does not appear in the equation. Under the current preference specification, the expenditure share on good j is always one Jth of the budget X^i regardless of X^j. Transportation cost does not affect the trade flow because two opposing factors that underlie the gravity equation offset each other: *A high transportation cost reduces demand but it also requires more to be shipped out of the origin*. This contrasts with the generic CES case, where the former exceeds the latter, and thus the iceberg transportation parameter makes an explicit

[18]The original ER network (Erdős and Rényi (1959)) comes with a Poisson degree distribution rather than an exponential degree distribution. The differences in the distribution arise from the way the network is constructed: Jackson and Rogers (2007)'s network is dynamic, whereas Erdős and Rényi (1959)'s network is static.

[19]The branching factor is not a free parameter and it cannot be directly estimated from the data, because the estimation algorithm will either explode or create indeterminacy. It is dependent on the shape of the network, which, in turn, is characterized by the other parameters via (17).

appearance in the gravity equation. Furthermore, (10) implies

$$X_i^j := s_i \psi_i^j \big(p_i^1, \dots, p_i^J, w^i\big) = p_i^j \frac{X^i}{p_i^j J} = \frac{X^i}{J} = \frac{\langle X \rangle}{J} \tag{18}$$

after all. We did not use the CES function for its lack of a closed-form equilibrium solution to address our question at hand. We shall leave the case of more complicated situations for future work.

3.10 *Endogenous transportation networks*

To this point, we have assumed that the transportation network is exogenous and the city-size distribution is contingent on the underlying network. Considering the fact that it is easier to relocate people than to build intercity transportation infrastructure, this is not an unreasonable assumption in the short run. New York City would have been much smaller had it not been the entrepôt to Europe. However, the degree-city relationship is not a one-way street and in fact, it may be the other way around: The relocation of people forces the transportation network to follow a specific pattern particularly in a long-term setting. It can also be the case that the network structure and its associated city-size distribution are in fact a product of some common underlying causes. We discuss these issues next.

Consider a commodity shipping firm that arranges a transport network to accommodate commodity flow (18). They will maximize their profit by choosing degree $\{k_i\}_{i \in V}$ given the city-size distribution and iceberg transportation parameter.[20] We shall assume that the expected degree m of a new node is predetermined so as to concentrate on network choice of r rather than on the selection of a total number of edges $|E|$. The firm will maximize their profit calculated as

$$\pi(r) = \langle X \rangle J \big(1 - \tau^{-A} \big\langle k^{B \log \tau} \big\rangle^2 \big) - \big\langle h(k, r) \big\rangle \tag{19}$$

with respect to r. We will derive the firm's revenue first (the first term) and then examine the cost (the second term) afterwards.

3.10.1 *Revenue* The choice of network structure acts on firm revenue in two ways. First, it modifies the equilibrium price and changes the trade flow accordingly, which constitutes the revenue base for the firm (the first effect). Then out of the trade flow thus calculated, the fraction that melts en route, namely $\tau^{l_i^j} - 1$ will be the shipper's cut, which also hinges on the selection of network by way of geodesic length l_i^j (the second effect).

In our case, the first effect is actually absent. Our commodity flow (18) is independent of transportation cost and by extension, network configuration as we demonstrated in Section 3.9. If the shipping firm raises the degree in city j, then demand for good j increases thanks to improved accessibility to city j and resulting lower delivered prices of

[20]Alternatively, we could model these parameters as endogenous variables, but it is hard to imagine one shipping firm single-handedly affecting the entire distribution of cities. By leaving them predetermined, the firm behaves competitively.

good j, which in turn increases their revenue generated in city j. On the other hand, also due to improved accessibility, good j travels a shorter distance than before, which reduces their revenue from city j. Shipping volume in total will increase but each unit shipped will bring in less and the firm's revenue will remain the same as a result. Thus, the firm can ignore the first effect and only needs to take the second effect into account for network optimization.

To be more specific, take shipment from city j to i. From (18), city i pays $t^{l_i^j} p_j^j \times \Phi_i^j(p^i, w^i) = X_i^j = \langle X \rangle / J$ to city j in total (inclusive of shipping charges). As we examined in Section 3.9, city i pays one Jth of its income X^i $(= \langle X \rangle)$ for each commodity regardless of transportation cost $\tau^{l_i^j}$. Therefore, the first effect is cancelled out and irrelevant to optimization. Out of city i's payment, producers in city j take $p_j^j \Phi_i^j(p^i, w^i) = \langle X \rangle / (J \tau^{l_i^j})$, leaving the transportation sector with the remainder $(t^{l_i^j} - 1) p_j^j \Phi_i^j(p^i, w^i) = (1 - \tau^{-l_i^j}) \langle X \rangle / J$. On a national scale, the shipping firm's revenue works out to

$$\sum_j \sum_i \bigl(1 - \tau^{-l_i^j}\bigr) \langle X \rangle / J \tag{20}$$

$$= \langle X \rangle J \left(1 - \tau^{-A} \left[\int_{k>0} k^{B \log \tau} \, dG(k; r)\right]^2\right). \tag{21}$$

The equality follows from (11). $\langle X \rangle J$ is the urban GDP. The last term $\tau^{-A} J \langle k^{B \log \tau} \rangle^2$ is the fraction of the GDP that goes to the nonshipping sector and the remainder $1 - \tau^{-A} J \langle k^{B \log \tau} \rangle^2$ is the shipper's revenue. This constitutes the first term in (19).

The crux of the profit maximization problem lies in the value of $B \log \tau$. It is positive but it may or may not be greater than one. If it is one, then the revenue becomes constant because $\langle k^{B \log \tau} \rangle = \langle k \rangle = 2|E|/J$ is independent of the network choice of r by assumption. In general, the revenue increases as r drops if $B \log \tau < 1$ and vice versa if the inequality is in reverse. The degree distribution $G(k; r')$ strictly second-order stochastically dominates $G(k; r)$ if $r' > r$ (see Theorem 6 on p. 903 in Jackson and Rogers (2007)). If $B \log \tau < 1$, that is, $k^{B \log \tau}$ is concave, then low r improves revenue.[21] Low r concentrates the degree to a limited few, which tips the scale of the second effect ($\tau^{l_i^j}$ to $1 - \tau^{l_i^j}$ for all i and j) in the shipper's favor.

An in-depth analysis is required to see why. Looking at one particular city pair, the shipping firm's cut $1 - \tau^{l_i^j}$ will increase when they take away degrees from the two cities on purpose to raise l_i^j, thanks to the second effect (in the absence of the first effect). The firm wants to keep their degree as low as possible to raise their revenue—that is, *if* they earn revenue only from this particular city pair. On the national level, if they take away some edge from city i or j, that edge needs to be reallocated somewhere else and their cut will decline from the city to which the edge is reassigned. They need to closely monitor this trade-off and distribute their degree in the way that maximizes their *overall* (not just one particular city pair's) second effect.

[21] The value of B decreases as r drops (cf. (17)) but $k^{B \log \tau}$ will still be concave.

In particular, their overall revenue is the weighted sum of their cut from every city pair as in (20). The weight would normally include the location-variant trade flow X_i^j so that a city pair along a busy transportation corridor would weigh in more on the revenue calculation than a barely trodden city pair. In our case, however, due to the lack of the first effect, the revenue does not depend on the trade flow and will be just a weighted sum of a much simpler term $k^{B\log\tau}$ as in (21), calculated with the second effect alone. Furthermore, if $B\log\tau = 1$, then (21) becomes just a simple, unweighted sum of degree and the shipping firm can allocate their edges any way they see fit only to minimize their cost: If they remove an edge from some city, their share of revenue from that city will go up, but they will lose the exact same amount of share from the city to which they redistributed the edge. Therefore, the revenue will be the same no matter which r the firm chooses. If, on the other hand, $B\log\tau < 1$, then they want to concentrate the degrees to a limited few cities to increase their weighted sum of the split. If $k_i = k_j$ (≥ 2) for some i and j, then (21) will increase by switching to $k_i + 1$ and $k_j - 1$. Their cut drops in city i but an increased revenue from city j will more than make up for the loss because $k^{B\log\tau}$ is concave. This can be achieved by setting $r = 0$, and vice versa if $B\log\tau > 1$.

3.10.2 *Cost* Turning to the cost end of optimization (19), assume that cost is additively separable over cities as well. First, consider when cost is concave in degree. As above, Theorem 6 in Jackson and Rogers (2007) applies and the shipping firm will bring r down to zero to minimize the cost. Intuitively, they want to spread the degree distribution to take advantage of substantial cost reduction in large hub cities in exchange for lost cost-effectiveness in small cities as the former surpasses the latter when their cost is concave. Thus, a BA network will minimize the cost; and vice versa, if cost is convex in degree, then they will form an ER network. In this case, cost savings from building a large hub do not cover the loss from lowering degrees of other cities. They would rather even out the degree distribution so as to avoid efficiency loss from making degrees too small.

3.10.3 *Profit* Putting the two sides together, the shipping firm will

$$\max_r \pi(r) = \langle X\rangle J\left(1 - \tau^{-A}\left[\int_{k>0} k^{B\log\tau}\, dG(k;r)\right]^2\right) - \int_{k>0} h(k,r)\, dG(k;r),$$

or equivalently (19), where $h(k,r)$ is cost incurred in a city of degree k. If both $k^{B\log\tau}$ and $h(k,r)$ are concave, then the shipping firm will lower r as far as possible to maximize their profit, and the resulting optimal network configuration will be BA. This is supported by the empirics. On the other hand, if they are both convex, then the optimal network will be ER. If one is concave and the other is convex, then the firm will settle with some medium value of r at which the marginal change in revenue offsets that of cost, leading to a network that is part BA and part ER.

Empirical validation of the framework above may be hard to come by. On the revenue front, we have estimates for the critical parameter $B\log\tau$ in Section 4. For a BA network, estimates barely top the threshold value of one, ranging from 0.3943 (Belgium) to at most 1.009 (US Places). The shipping firm can increase their revenue by lowering r for the most part, which is consistent with the existing network configuration. Note that this only proves that *if* they go for a BA network, then $k^{B \log \tau}$ will be concave, and thus they should stick to a BA network. We know from Section 4 that τ will be exorbitantly high if the underlying network is ER, which does make $k^{B \log \tau}$ *convex*. Thus, an ER network may well be a solution if the exogenous transportation cost parameter τ happens to be prohibitively high. Furthermore, if the estimate asymptotically converges to one with data size, which can potentially be the case here as can be seen in Section 4 (cf. Footnote 26), then r makes no difference to the revenue side of decision making and profit maximization reduces to cost minimization.

On the cost front, let us take the airline industry for illustration. Considering recent mega-mergers between network carriers, such as United and Continental, Air France and KLM or Delta and Northwest, and subsequent hub consolidations (e.g., dehubbing of Cleveland of Continental or Memphis of Northwest), it seems that degree exhibits scale economies among airlines. Airliners are taking advantage of them by trimming down r to cut the loss from underperforming small hubs and redirect degrees to a select few large hubs, leading to a BA network as a result of optimization. A problem with this methodology is that transportation networks are not unique, in that there are generally multiple modes of transport and multiple companies providing services in each mode.

Further investigation should attempt to gauge the magnitude of reverse causality from the city-size distribution to networks. In the meantime, we shall return to the forward causality that we are interested in and pitch our model against the actual city-size distribution to identify what class of transportation network governs the city-size distribution.

4. Empirical implementation

By and large, the empirical results are in full support of our initial inkling that a scale-free network explains the city-size distribution but ER or other network structures commonly adopted do not.

All told, we have four sets of data on our plate: Belgium, Metropolitan Area (MA), CBSA, and Places.[22] Descriptive statistics for each data set are in Table 1. The Belgian data are included to see if our model's predictive value is subject to both the area and population size of a country under study. (It was not.) MA and CBSA are the popular choices in the literature. In addition, the use of Places is becoming increasingly commonplace following the publication of Eeckhout (2004).

There is no agreement on appropriate city-size data to validate the model's relevance to the reality. Whereas none of the available data sets are flawless, we believe that they complement each other. Each data set has its own pros and cons, and we do not intend to favor one over another. Rather, we include all the data sets commonly in use so that we

[22] The Belgian data is provided courtesy of Soo (2005) and the remainder are from US Census 2000. MA is an umbrella term encompassing metropolitan statistical areas (MSA's), consolidated MSA's and primary MSA's. For more on definitions of MA and CBSA, see https://www.census.gov/programs-surveys/metro-micro/about.html and for Places, see http://www.census.gov/geo/reference/gtc/gtc_place.html, https://www.census.gov/content/dam/Census/data/developers/understandingplace.pdf and Section 8.5 of Ioannides (2013, pp. 371–372). We thank Jan Eeckhout for sharing his data used in Eeckhout (2004).

Table 1. Descriptive statistics. The statistics above the middle line are related to a linear scale and those below are related to a log scale. Mean of $\log(s)$ is same as the log of geometric mean.

Data	Belgium	MA	CBSA	Places
Data size J	69	276	922	25,358
Total urban population S	4,344,222	225,981,679	261,534,991	208,735,266
Population covered	42.38%	80.30%	92.93%	74.17%
Largest city	Antwerp	New York CMSA	New York MSA	New York city
Largest size	446,525	21,199,865	18,323,002	8,008,278
City near arithmetic mean	Genk	Oklahoma, OK MSA	Green Bay, WI MSA	Hillsboro city, TX
Arithmetic mean	62,960	818,774	283,661	8,232
Median city	Beringen	Anchorage, AK MSA	Hinesville-Fort Stewart, GA MSA	Harristown village, IL
Median size	39,261	259,600	71,800	1,338
Smallest city	Arlon	Enid, OK MSA	Andrews, TX µSA	New Amsterdam, IN
Smallest size	24,791	57,813	13,004	1
Standard deviation	61,240	1,968,621	974,190	68,390
Skewness	4.183	6.682	10.98	75.53
City near geometric mean	Mouscron	Huntsville, AL MSA	Sunbury, PA µSA	Sutton city, NE
Geometric mean	50,809	342,844	94,373	1,447
Mean of $\log(s)$	10.84	12.75	11.46	7.278
Standard deviation of $\log(s)$	0.5697	1.119	1.191	1.754
Skewness of $\log(s)$	1.498	1.048	1.187	0.2091

can compare the predictive power of network structures with that of other explanatory variables identified by researchers, which are tested with various data sets of researchers' choice.

The smallest unit of measurement in MA and CBSA is a county. Counties preserve the contiguity of a city well especially when it involves a wide expanse of land. To be listed under MA or CBSA, the population needs to exceed the lower bound set by the Census Bureau. Any settlements that do not reach the threshold are not included in the data.

On the contrary, Places offer a finer data set than MA or CBSA. They are a city, town, or Census-designated equivalent if the area is not incorporated. A county typically has multiple municipalities in it and as such, Places may dissect the contiguously populated area in multiple nodes when it may better be considered as one node, particularly in a large MA or CBSA. Nevertheless, we tested our model with Places for two vantage points that we cannot have with other data. The first point is that Places are a truncation-free measurement. Places impose no threshold on the size, whereas MA and CBSA do. The threshold creates a survivorship bias and contaminates empirical validation (cf. Eeckhout (2004)).

The second point is the resolution that Places have to offer. Some counties cover a very large area and are too coarse for our analysis. As we defined in Section 3.1, a city is an expanse of land within which the firms produce the same commodity and from within which the transportation cost is the same to the rest of the country. In some counties, it takes several hours to get from the principal city of MA to cities on the periphery. In this case, the transportation cost will not be the same and they should be treated as separate nodes.

To conclude, MA and CBSA could be too coarse in some areas and suffer from survivorship bias. Places are truncation free but too fine in some areas. The data set that offers the perfect resolution in the entire range of the distribution without survivorship bias does not exist as of now. Meanwhile, we believe that inclusion of various data helps us verify that our findings are not sensitive to the city units selected.

We examined how well our model predictions (14) and (15) fit these data by feeding an ER/BA, ER, and complete graph into the model, whose degree distributions are (16), exponential and degenerate, respectively. Along with these networks' performance, we also checked how two of the predicted city-size distributions from the existing literature do as a point of comparison. We estimate each in three ways: maximum spacing estimation (MSE), minimum Kolomogorov–Smirnov estimation (minKS), and maximum likelihood estimation (MLE).

In what follows, a hat on parameter x indicates its estimate $\hat{x}$.

4.1 *Estimation methods employed*

The first choice is to go for MLE, which does not work with (16). The likelihood function is monotone increasing in k_0. Thus, MLE will imply $\hat{k}_0 \to \infty$, which makes no sense. As a workaround to MLE, we calculated the estimates by MSE. Whereas its use is limited in the city-size literature so far especially when compared to MLE, it is more robust and easier to handle than MLE. The problem we have with MLE is exactly the one exemplified in Ranneby (1984) and we used his solution. The MSE estimator maximizes the geometric mean of the gap or step between two adjacent CDF values

$$F(s_i;\theta) - F(s_{i-1};\theta), \tag{22}$$

where θ is a vector of parameters to be estimated and data sequence s is rearranged in the ascending order $s_1 \leq s_2 \leq \cdots \leq s_J$.[23] The idea here is to split the interval $[0, 1]$, the range of a CDF, in J steps in the way that none of the assigned $F(s_i;\theta)$ will create a disruptively large gap with its neighbors and the gaps should be evenly spaced as much as possible on the *logarithmic* scale. Maximizing the *arithmetic* mean does not work here because it will always be $1/J$ no matter what estimates we toss in. This actually works as a cap on our geometric mean in turn, by Jensen's inequality. Thus, we can safely rule out the possibility that the maximand tends to infinity, which is exactly why we had to discard MLE. For more on MSE, see Appendix A.5.

4.2 *A scale-free transportation network explains the city-size distribution*

Estimation with four different data sets unanimously chooses BA over ER as the underlying transport network in our economy. We report our results in Table 2 and Figures 10 to 13 in the Supplementary Material (Berliant and Watanabe (2018)).

Since the transport cost and average branching factor only come into the equation in the form of a quotient of their logarithmic values $\frac{\log \kappa}{\log \tau}$, we will denote this by δ for estimation purposes, in which case, (15) becomes $f(s) = \gamma\delta\langle s\rangle^{-\delta}s^{\delta-1}g[k(s)]$. As we have already seen, a small δ stretches out the distribution and a large δ does the opposite.

We evaluated each network's performance with a number of different statistics. In Table 2, $\langle \log \mathsf{LH}\rangle$ is the maximand of the log likelihood value, normalized by system size J. KS stands for the Kolomogorov–Smirnov statistic, which measures the maximum gap between the predicted and empirical CDFs. On the other hand, $\langle \log \mathsf{step}\rangle$ is the log of the maximand of MSE, normalized by system size J (see Appendix A.5 for the relationship between KS and $\langle \log \mathsf{step}\rangle$), and geo/arith is the ratio of the maximand of MSE (the geometric mean of steps in (22)) to the arithmetic mean of the steps, which is the highest value that the geometric mean can take.

In Table 2, ER/BA corresponds to (16). As low values of $\hat{r}$ indicate, edges are formed predominantly through networking rather than completely at random. We cross-checked estimates with minKS and MLE[24] and we obtained a similar result. To be doubly sure of our findings, we ran estimation with $r \to \infty$ (ER in Table 2). The statistics of ER seem to be comparable with other distributions except that the estimated transportation cost is astronomically high.[25] Thus, we dismissed the ER network. All in all, we

[23]The first and last gap are defined by $F(s_1;\theta) - F(-\infty;\theta)$ and $F(\infty;\theta) - F(s_J;\theta)$, respectively.

[24]We constrained k_0 to zero for MLE. We know from the results of MSE and minKS that $\hat{k}_0$ tends to zero.

[25]A one-dollar pen will cost more than the US GDP five towns over on the ER network. There is not enough variance in the ER degree distribution, certainly not power-law type behavior. To generate the empirical city-size distribution, the ER economy has to amplify and capitalize on what little variance its degree

conclude that a scale-free transportation network explains the city-size distribution but a scale-variant network does not.

Estimated $\hat{\hat{\delta}}$ ranges from 0.9911 to 2.536.[26] As we discussed in reference to (13), we confirm that in most cases, the impact of adding an edge on city size wears off as degree itself becomes saturated (it cannot exceed $J-1$), or put differently, New York has more edges, size for size, than any other cities as it takes more edges to raise the city size as the city grows further.

Along with ER/BA and ER above, we ran MSE with three other distributions representative of the existing city-size models to compare with our model. Eeckhout (2004)'s model lead to a lognormal distribution and Berliant and Watanabe (2015) predicted a GEV distribution as the city-size distribution. A complete graph will result in a degenerate probability distribution.

As we discussed earlier in Section 1, due to a multitude of variables involved in location choice, no one factor can single-handedly explain the city-size distribution. Our model's fit to the data lags slightly behind Berliant and Watanabe (2015), but it is still comparable to the the existing testable models, based on all the statistics we computed in Table 2. ER/BA comes in second behind GEV on all fronts except Places. A different transportation network results in a different city-size distribution and one type of network is particularly coherent with the empirical distribution. We believe that the network structure weighs in on the matter as much as, if not more than, the existing explanatory variables identified by other researchers such as random growth or other socioeconomic factors. Eventually, we should be able to merge our current model, that focuses on the transportation network, with the existing models like Berliant and Watanabe (2015), that focus on technological shocks. This will involve dynamic rewiring of edges in response to technological shocks and the coevolution of the resulting distribution of city sizes and edges.[27] For now, we are examining the explanatory power of the network structure in isolation to see if it adds anything to our current understanding of the city-size distribution. And it does, as it turns out.

Figures 10 to 13 represent PDF and PP plots of the five distributions tested with four data sets. PP plots, which we used as a substitute for usual CDF plots,[28] sketch the estimated CDF against the empirical CDF. If the fit is perfect, then the PP plot will become a 45° line. ER/BA and two existing distributions (lognormal and GEV) are almost indistinguishable in Figure 13, indicating that the network structure is just as effective as the existing models. Once again, the three distributions become almost identical to a 45° line as J grows.

distribution has to offer (cf. Proposition 3.1). As a result τ has to be ludicrously large to make things work. On the other hand, if the transportation infrastructure is in its early stage of development without any hubs, then the country's transportation cost will probably be higher than more BA-like countries because Zipf's law is a universally observed phenomenon. We will comment on this in Section 5.

[26]The estimate tends to decrease as data size J increases.

[27]We made a brief mention of reverse causality in Section 3.10 as a preview of a future line of network-oriented research.

[28]PP plots are more revelatory than regular CDF plots, as they place data points at equal intervals, whereas CDF plots usually dump lots of data points in the middle and make it hard to see the fit in the cramped midsection, especially when the distribution in question is very skewed like city-size distributions.

The value of $\langle \log \mathsf{LH} \rangle$ can be made arbitrarily large by increasing the number of parameters $|\theta|$. Bayesian and Akaike Information Criteria (BIC and AIC) are based on the likelihood value but penalize increased use of parameters to detect overfitting. Since GEV and ER/BA use as many as five parameters, these two distributions' performance should be discounted on the BIC and AIC front. However, due to the large size of data sets, BIC and AIC barely overturn the primary evaluations made with $\langle \log \mathsf{LH} \rangle$. With the exception of Belgian data (whose system size is the smallest among the four data sets), there is no disagreement among those three statistics.

In addition, we put two other fat-tailed degree distributions to the test. The network structure is exogenous in our model. We used Jackson and Rogers (2007) to represent a scale-free network. While Jackson and Rogers (2007)'s model is microfounded and sufficient to generate a fat-tailed degree distribution, it is not the only degree distribution which a scale-free network gives rise to. There is a chance that our economy's transportation network may have come around from a different mechanism than Jackson and Rogers (2007). In this light, we picked the lognormal and GEV distributions for use as examples of a fat-tailed *degree* distribution, from which to derive the city-size distribution[29] (in comparison, Eeckhout (2004), Berliant and Watanabe (2015) cited lognormal and GEV as a *city-size* distribution). The results (the last two rows in Table 2) seem to indicate that the network formation does not necessarily have to be of Jackson and Rogers (2007) type. Regardless of how it came about, a network with a fat-tailed degree distribution results in a city-size distribution that closely resembles the actual distribution.

5. Conclusion and extensions

We examined how the network of cities affects the city-size distribution. We built a simple economic model with an explicit transport network. The bridge between network structure and city size is represented in (13), where we learned that *there is a log-linear relationship between city size and city degree*.

We put two commonly studied networks to the test. The classical ER random graph is too egalitarian to generate gravitationally large cities like New York City. The BA model explains the city-size distribution better than the ER model and bears very close comparison with other proposed city-size models. The BA network has a scale-free degree distribution and the resulting city-size distribution behaves similarly via (13). In fact, it would be odd if the city-size distribution were *not* scale-free under a BA network. Large nodes with a high degree like Chicago attract a large mass of people because A) goods produced in Chicago are in high demand for its inexpensive delivered price owing to its high degree and B) goods available for consumption in Chicago are also inexpensive thanks to its high degree. The exact opposite applies to small cities. But there are still some people knowingly living in small cities because we cannot afford to wipe them off the map due to preference for variety. This gives rise to a few cities of an overwhelming size and a myriad of small cities. The actual city-size distributions (we tried Belgium and the United States in particular) unanimously opt for a BA network.

[29]To our knowledge, these degree distributions are not yet microfounded.

From this point on, it would be reasonable to combine GEV to determine firm productivity as in Berliant and Watanabe (2015) and BA for transportation network structure by way of simulations, but we will not have an analytical solution due to the added complexity.

Tracing the historical codevelopment of the network structure with the city-size distribution may reveal a clue to identifying the direction of causality, but the result may still be inconclusive due to multiple factors involved. We briefly explored the possibility of the network structure conforming to a given city-size distribution in Section 3.10. The United States has seen a number of drastic changes in its modes of transportation. Despite falling transportation cost, however, the city-size distribution in the United States has been stable at least since 1900 (Black and Henderson (2003)). It is then tempting to conclude from this observation that the transportation network used to be close to the ER network back in 1900: As we discussed in Proposition 3.1, falling τ makes the city-size distribution less skewed. If the city-size distribution remained the same throughout in the United States, then the transportation network must have been closer to the ER network than the BA network in 1900—that is, *if* we hold everything else constant. The reality is that total population S and the total number of cities and commodities J have increased over the same period as τ drops. Our city-size distribution (14) depends on the base city size $\langle s \rangle = S/J$ as much as it does on τ. And $\langle s \rangle$ is a scale parameter in (14), that is, an increase in $\langle s \rangle$ spreads out the distribution. Thus, even when the transportation network has not changed, the city-size distribution will still be robust against falling τ if the total population increases to compensate for reduced variance. We cannot tell whether the network structure has changed since 1900 for certain without the data on the degree distribution in 1900, which are unavailable.

It has been suggested that other networks be implemented in our framework, for example, the optimal transport network for a given population distribution (assuming a cost function) rather than the choice of r, which is a less precise control variable. This would require the geodesic length or degree distribution for the optimal network. We are not aware of any results addressing this issue.

Appendix

A.1 *Linear transportation cost*

We consider two possible transportation cost structures: The first case is exponential transportation cost with parameter τ (≥ 1). The second one is a less steep, linear transportation cost with parameter τ_L (≥ 0). In comparison to the first case, the linear transportation cost structure deducts τ_L *units* (rather than *fraction* $\tau - 1$) of shipment on each leg of the travel. Thus, $1 + l_i^j \tau_L$ units of shipment are required at origin j to deliver one unit to destination i.

For a sufficiently small τ_L, delivered price will be approximately identical under two different transportation cost structures if $\log\tau = \tau_L$.[30] All the analyses in the main text apply to a linear case as well with τ replaced with e^{τ_L} for small τ_L.[31]

A.2 *Proof of Proposition* 3.1

Proof. Suppose $J > 2$ and the network is neither complete or completely isolated. Then

$$\frac{ds(a_i)}{da_i} = (\log\tau)s(a_i)\left[1 - \frac{s(a_i)}{S}\right] \geq 0$$

with equality iff $\tau = 1$. Furthermore,

$$\frac{d^2 s(a_i)}{da_i^2} = (\log\tau)s'(a_i)\left[1 - 2\frac{s(a_i)}{S}\right] \geq 0$$

for $i < \operatorname{argmax}_{j\in V} s(a_j)$ with equality iff $\tau = 1$. Hence $s(a_i)$ is increasing and strictly convex in a_i.

To show that $s(a_i)$ bulges as τ grows, first we define a weighted accessibility $h(a_i) := \frac{\sum_j \tau^{a_j}(a_i - a_j)}{\sum_k \tau^{a_k}}$. Note $h(a_H) - h(a_M) = a_H - a_M > 0$ and $h(a_M) - h(a_L) = a_M - a_L > 0$. Then

$$\begin{aligned}
\frac{dD(\tau)}{d\tau} &= \frac{1}{2\tau}\big\{\big[s(a_H)h(a_H) - s(a_M)h(a_M)\big] + \big[s(a_M)h(a_M) - s(a_L)h(a_L)\big]\big\} \\
&> \frac{1}{2\tau}\big\{\big[s(a_H)h(a_M) - s(a_M)h(a_M)\big] + \big[s(a_M)h(a_L) - s(a_L)h(a_L)\big]\big\} \\
&> 0,
\end{aligned}$$

which establishes the claim. □

A.3 *Idea behind geodesic length (*11*)*

We briefly repeat Hołyst et al. (2005)'s arguments to obtain (11) in our context. Consider a geodesic between nodes i and j. We ignore loops. The probability that a child node traces back to its ancestors via some circumvention is proportional to $1/J$. It becomes negligible as the system size J grows (our system size ranges from 69 to 25,358 in Section 4). As shown in Hołyst et al. (2005), the resulting error is minimal. A tree is a sequence of nodes where each node except for the root node has exactly one parent (or

[30] Delivered price on exponential and linear iceberg will be identical if

$$\begin{aligned}
l_i^j \log\tau &= \log\big(1 + l_i^j \tau_L\big) \\
&= l_i^j \tau_L + O\big(\tau_L^2\big).
\end{aligned}$$

[31] Our model is *multiplicative* in nature just as much as the city-size distribution and scale-free networks are. A linear (or *additive*) form of iceberg transportation cost is not readily compatible for our purposes unless we convert it into a multiplicative form by, for example, approximation in Footnote 30.

ancestor) node. Each node may or may not be followed by (a) child node(s). There are no cycles on a tree. If we pick a random tree starting from node i, we will wind up at node j somewhere along the tree $k_j / \sum_{x \in V} k_x$ of the time and we will not reach node j the remaining $1 - k_j / \sum_x k_x$ of the time. On average, we will reach node j within $\sum_x k_x / k_j$ trials. Suppose that the depth (the number of parent nodes that you have to go through before reaching your root node) of node j is l. There are $k_i \kappa^{l-1}$ nodes whose depth is l. Therefore, on average, we arrive at node j in l steps if

$$\frac{\sum_x k_x}{k_j} = k_i \kappa^{l-1}, \tag{23}$$

from which we obtain (11). In other words, if, on average, it takes more than $k_i \kappa^{l-1}$ trials to reach city j, that is, $\frac{\sum_x k_x}{k_j} > k_i \kappa^{l-1}$, then it is likely that city j is more than l steps away from your city i. You would try $k_i \kappa^{l-1}$ times to find city j, when in fact you would need additional $\frac{\sum_x k_x}{k_j} - k_i \kappa^{l-1}$ trials to reach city j, meaning that city j is not in the group of cities l steps away from you but actually located somewhere farther down. On the contrary, if it takes less than $k_i \kappa^{l-1}$ trials to reach city j, then city j should be less than l steps away from you. You would not need that many trials to find a city j, the implication being that, once again, you are looking at a wrong group of cities. Thus, city i and j are l steps apart from each other exactly when (23) is satisfied with equality.

A.4 *Branching factor*

Take a random edge and walk toward one arbitrarily selected end. Call where you arrived at a neighboring node. The average degree of neighboring nodes thus reached approximates the mean branching factor κ. In effect, we will take one degree off the average degree found above because the edge we just walked on cannot be used to reach the destination city. We are climbing up a tree, not down (recall how goods find their destination city in Section 3.6). Also note that the mean branching factor is not just a mean degree $\langle k \rangle$. We are not hopping from one city to another but climbing a tree from one neighbor to the next to reach the destination city. Thus, a city charged with lots of links is more likely to be a neighbor of some city than a poorly connected city, and cities are duly weighted when fed into the mean branching factor. In other words, Houston is rare while there are quite a few mid-sized cities but that does not mean Houston is hard to reach at random for its rarity. Houston has far more edges than mid-sized cities and we are likely to travel through Houston at some point or another (cf. Figure 7). In particular, a node of degree k has a chance proportional to $kg(k)$ of being at one end of an arbitrary direction on a randomly chosen edge, where $g(k)$ is a probability density function of (16). Or put differently, if we parachute into a random edge and then flip a coin to decide which direction to go in, we will arrive at a kth degree city $kg(k)$ out of $\sum_{x=1}^{J} xg(x)$ times. Thus, the mean branching factor is given by (17).

A.5 *Maximum spacing estimation*

It might be easier to make sense of the use of geometric mean in MSE if we recast it as an analogue of a more familiar, linear regression. The geometric mean of steps here corresponds to ordinary least squares and the arithmetic mean corresponds to a plain sum of residuals. Say we are trying to regress $y = (-1, 0, 1)$ on $x = (-1, 0, 1)$. If we aim to minimize the sum of residuals, *any* real estimate that makes the regression line run through the origin $(0, 0)$ will work, just as much as any estimate will make the arithmetic mean of gaps $1/J$. We will end up with infinitely many estimates because residual at $x = 1$ always offsets the one at $x = -1$. To ward off this cancellation problem, we usually try to minimize the sum of *squared* residuals, which leads to a unique estimate, a 45° line. Similarly, the use of *geometric* mean will solve the indeterminacy problem that comes with arithmetic mean and will promise us sensible estimates.

The geometric mean also comes in handy here. The gap tends to get tighter near the top and/or the bottom of most distributions as the CDF creeps up to one and/or bears down on zero. However, this does not mean New York or New Amsterdam, IN counts less than other cities as a sample. The geometric mean offsets this general tendency and duly stretches small gaps so that these extremities will receive no less attention than the ones in the middle. There is no particular reason to let the mid-sized cities punch above their weight.

On a related matter, we report Kolomogorov–Smirnov (KS) statistic. MSE is similar to KS in that both KS and the maximand of MSE are a power mean. KS statistic is a power mean of the form

$$\left\{\frac{1}{J}\sum_i \left|\text{Empirical } F(s_i) - F(s_i)\right|^\rho\right\}^{\frac{1}{\rho}} \tag{24}$$

with $\rho \to \infty$ (i.e., the maximum of the residuals, the L^∞ norm), whereas the maximand of MSE is a power mean of the form

$$\left\{\frac{1}{J}\sum_i \bigl(F(s_i) - F(s_{i-1})\bigr)^\rho\right\}^{\frac{1}{\rho}} \tag{25}$$

with $\rho \to 0$ (i.e., the geometric mean of the gaps). The way they aggregate the data is where their difference comes in. KS statistic only picks up a single city where the predicted value deviates from the actual value the most. It does not tell us anything about the selected model's performance over the remainder of cities other than the fact that their gap is tighter than the KS value (but *not* by how far). On the other hand, the maximand of MSE is determined by the step gap log-averaged over the entire range of the cities, and probably a better measuring tool to gauge the model's performance in that respect.

To get a sense of what MSE hunts for, consider what happens if we pull out the estimate that *minimizes* the geometric mean instead. Minimum spacing estimator would dump the entire interval [0, 1] on one particular city i (any city will do) so that $F(s_j; \theta) = 0$ for all $j < i$ and $F(s_j; \theta) = 1$ for all $j \geq i$, in which case, the geometric mean would be zero, the smallest value possible (practically the same result when you try to maximize the arithmetic mean as we mentioned above, in the sense that *any* estimate will be as good as any other). This would make such a pointless estimator. MSE does the exact opposite.

References

Albert, R. and A.-L. Barabási (2002), "Statistical mechanics of complex networks." *Reviews of Modern Physics*, 74 (1), 47. [1425]

Barabási, A.-L. and R. Albert (1999), "Emergence of scaling in random networks." *Science*, 286 (5439), 509–512. [1421, 1425]

Barabási, A.-L. and E. Bonabeau (2003), "Scale-free networks." *Scientific American*, 288 (5), 50–59. [1422]

Behrens, K., G. Mion, Y. Murata, and J. Südekum (2017), "Spatial frictions." *Journal of Urban Economics*, 97, 40–70. [1421]

Bergstrand, J. H. (1985), "The gravity equation in international trade: Some microeconomic foundations and empirical evidence." *The Review of Economics and Statistics*, 67, 474–481. [1435]

Berliant, M. and H. Watanabe (2015), "Explaining the size distribution of cities: Extreme economies." *Quantitative Economics*, 6 (1), 153–187. [1420, 1443, 1444, 1445]

Berliant, M. and H. Watanabe (2018), "Supplement to 'A scale-free transportation network explains the city-size distribution'." *Quantitative Economics Supplemental Material*, 86, https://doi.org/10.3982/QE619. [1422, 1423, 1442]

Black, D. and V. Henderson (2003), "Urban evolution in the USA." *Journal of Economic Geography*, 3 (4), 343–372. [1423, 1445]

Calvó-Armengol, A. and Y. Zenou (2005), "Job matching, social network and word-of-mouth communication." *Journal of Urban Economics*, 57 (3), 500–522. [1425]

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010), "An empirical model for strategic network formation." Report. [1423]

Dorogovtsev, S. N., J. F. F. Mendes, and J. G. Oliveira (2006), "Degree-dependent intervertex separation in complex networks." *Physical Review E*, 73 (5), 056122. [1431]

Duranton, G. (2006), "Some foundations for Zipf's law: Product proliferation and local spillovers." *Regional Science and Urban Economics*, 36 (4), 542–563. [1420]

Duranton, G. (2007), "Urban evolutions: The fast, the slow, and the still." *The American Economic Review*, 97 (1), 197–221. [1420]

Eaton, J. and S. Kortum (2002), "Technology, geography, and trade." *Econometrica*, 70 (5), 1741–1779. [1421]

Eeckhout, J. (2004), "Gibrat's law for (all) cities." *The American Economic Review*, 94 (5), 1429–1451. [1420, 1439, 1441, 1443, 1444]

Erdős, P. and A. Rényi (1959), "On random graphs." *Publicationes Mathematicae*, 6, 290–297. [1421, 1435]

Freeman, L. C. (1978), "Centrality in social networks conceptual clarification." *Social Networks*, 1 (3), 215–239. [1432]

Fronczak, A., P. Fronczak, and J. A. Hołyst (2004), "Average path length in random networks." *Physical Review E*, 70 (5), 056110. [1431]

Fujita, M., P. Krugman, and A. J. Venables (1999), *The Spatial Economy*. The MIT Press, Cambridge. [1420, 1431]

Fujita, M. and J.-F. Thisse (2002), *Economics of Agglomeration*. Cambridge University Press, Cambridge. [1429]

Glaeser, E. L., J. Scheinkman, and A. Shleifer (1995), "Economic growth in a cross-section of cities." *Journal of Monetary Economics*, 36 (1), 117–143. [1420]

Hołyst, J. A., J. Sienkiewicz, A. Fronczak, P. Fronczak, and K. Suchecki (2005), "Universal scaling of distances in complex networks." *Physical Review E*, 72 (2), 026108. [1431, 1432, 1435, 1446]

Ioannides, Y. M. (2006), "Random graphs and social networks: An economics perspective." In *IUI Conference on Business and Social Networks, Vaxholm, Sweden, June*. [1425]

Ioannides, Y. M. (2013), *From Neighborhoods to Nations: The Economics of Social Interactions*. Princeton University Press, Princeton. [1439]

Jackson, M. O. and B. W. Rogers (2007), "Meeting strangers and friends of friends: How random are social networks?" *The American Economic Review*, 97 (3), 890–915. [1425, 1434, 1435, 1437, 1438, 1444]

Kakade, S. M., M. Kearns, L. E. Ortiz, R. Pemantle, and S. Suri (2004), "Economic properties of social networks." In *Advances in Neural Information Processing Systems*, 633–640. [1425]

Limpert, E., W. A. Stahel, and M. Abbt (2001), "Log-normal distributions across the sciences: Keys and clues." *BioScience*, 51 (5), 341–352. [1423]

McCann, P. (2005), "Transport costs and new economic geography." *Journal of Economic Geography*, 5 (3), 305–318. [1426]

Mele, A. (2011), "A structural model of segregation in social networks." Report. [1423]

Newman, M. E., C. Moore, and D. J. Watts (2000), "Mean-field solution of the small-world network model." *Physical Review Letters*, 84 (14), 3201. [1431]

Newman, M. E., S. H. Strogatz, and D. J. Watts (2001), "Random graphs with arbitrary degree distributions and their applications." *Physical Review E*, 64 (2), 026118. [1431]

Newman, M. E. and D. J. Watts (1999), "Renormalization group analysis of the small-world network model." *Physics Letters A*, 263 (4), 341–346. [1431]

Ranneby, B. (1984), "The maximum spacing method. An estimation method related to the maximum likelihood method." *Scandinavian Journal of Statistics*, 11, 93–112. [1442]

Rossi-Hansberg, E. and M. L. Wright (2007), "Urban structure and growth." *The Review of Economic Studies*, 74 (2), 597–624. [1420]

Simon, H. A. (1959), "Theories of decision-making in economics and behavioral science." *The American Economic Review*, 49 (3), 253–283. [1432]

Soo, K. T. (2005), "Zipf's law for cities: A cross-country investigation." *Regional Science and Urban Economics*, 35 (3), 239–263. [1423, 1439]

Toulis, P. and D. C. Parkes (2011), "A random graph model of kidney exchanges: Efficiency, individual-rationality and incentives." In *Proceedings of the 12th ACM Conference on Electronic Commerce*, 323–332, ACM, New York. [1425]

Zhang, Z., Y. Lin, S. Gao, S. Zhou, and J. Guan (2009), "Average distance in a hierarchical scale-free network: An exact solution." *Journal of Statistical Mechanics: Theory and Experiment*, 2009 (10), P10022. [1431]

Co-editor Karl Schmedders handled this manuscript.

Manuscript received 21 September, 2015; final version accepted 28 September, 2017; available online 23 October, 2017.