A Scale-Free Transportation Network Explains the City-Size Distribution *

Marcus Berliant† and Hiroki Watanabe‡

November 18, 2011

Abstract

Zipf's law is one of the best-known empirical regularities of the city-size distribution. 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. Under the scale-free transport network framework, the chance of observing extremes becomes higher than the Gaussian distribution predicts and therefore it explains the emergence of large clusters. City-size distributions share the same pattern. This paper proposes a way to incorporate network structure into urban economic models and explains the city-size distribution as a result of transport cost between cities.

Keywords: Zipf's law, city-size distribution, scale-free network
JEL classification: R12, R40

1 Introduction

1.1 Cities on a Network

The transaction pattern between any two cities affects both the way cities are populated and the overall city-size distribution. In reality, 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 lots of other cities are likely to grow because they undertake

*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 Professors Sukkoo Kim, John Nachbar, Jody O'Sullivan, and Victor Wickerhauser for their helpful comments and advice.

†Department of Economics, Washington University, Campus Box 1208, 1 Brookings Drive, St. Louis, MO 63130-4899 USA. Phone: (314) 935-8486, Fax: (314) 935-4156, e-mail: berliant@artsci.wustl.edu

‡Department of Economics, Washington University, Campus Box 1208, 1 Brookings Drive, St. Louis, MO 63130-4899 USA. e-mail: watanabe@wustl.edu.

lots of economic activities, whereas cities with sparse connections to a limited number of cities are liable to remain small in size.

When we think through intercity exchange patterns, we can represent cities by a set of vertices and the transport network by edges. Thus, networy theory is important when constructing a model of cities in the national economy.

The recent seminal work by Barabási and Albert [BA99] has revitalized network theory. Classical network theory pioneered by Erdős and Rényi [ER59]'s model (ER network) cannot explain the emergence of a cluster or hub in a network, which we usually observe in most real networks. In a classic random graph, each node is linked with an equal probability to any other and lacks distinctiveness, for the number of pre-existing links does not matter in forming a network. Barabási and Albert (BA) add a dynamic feature and preferential attachment to the classical random graph model so that the nodes are no longer identical. Some nodes gather lots 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 numbers of links rather easily with these networks compared with classical random graph theory.

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 1000 largest cities and so on, a property known as a power law. We will come back to the fat-tail nature of the city size in Section 1.3. In the mean time, the immediate question at hand is: "Which type of network explains the city-size distribution?" We expect that the degree (the number of links each node has) of a city is positively related to its population. And for that reason, we imagine that the economy is based on a BA network rather than an ER network. This turns out to be right, but selection of the appropriate network structure depends on exactly how node degree is related to city size. We decode their relation in Section 3.7.

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. Duranton [Dur05], for example, provides a precise description of the way commodities are produced and distributed, whereas transportation cost is not assumed to depend on the location of a city in its commodity transport network. In contrast, transaction and/or communication between hub cities is much easier than that of cities on peripheries. Our goal in this paper is to bring to the fore the interaction between transportation network structure and the city-size distribution. To address this, we attempt to 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 in a bigger research program.

1.2 Some Transportation Networks Are Scale Free

The United States has seen a number of changes in the mode of transportation over the 20th century. At the turn of the century, we saw trains before the introduction of affordable trucks and the highly organized interstate highway system, which has been partially supplanted by emerging low-budget airline companies now. Figure 1 shows a simplified U.S. interstate system on the left and a typical airline flight system map on the right. The existing literature on the city-size distribution usually does not investigate the structure of an underlying network in which economic activities take place. An organizational pattern generated and developed among cities connected by transport links or network edges is, in the existing literature, not thought to be included in conventional forces (positive and negative externalities) that serve a significant role in determining the city-size distribution. Apparently a network composed of interstates does not share its structure with that of airlines at all. There is not much variance in degree in the Interstate network, whereas the airline network has a limited number of highly connected cities. The BA network explains the latter network better, as it follows a power law.

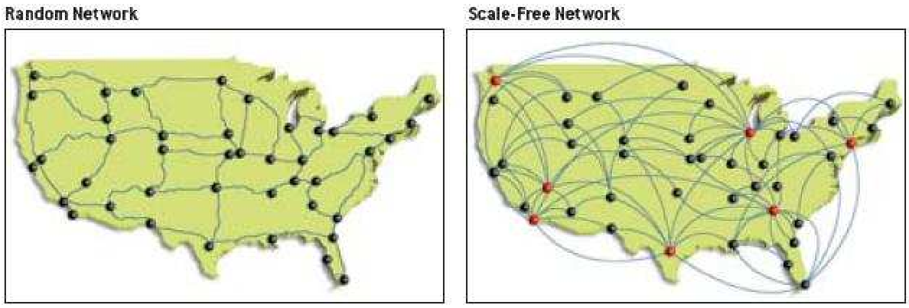


Figure 1. The U.S. interstate state system (left) and the U.S. airline system (right). Source: *Barabási and Bonabeau [BB03]*.

It should be noted, however, that what is geographically visible may not represent the actual network that the economy relies on at work. The interstate highway network, for example, exhibits an ER-network type topology. 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, where the shipments from Memphis, as well as shipments from St. Louis and Kansas City, are processed. In this case, Memphis is connected to Chicago. It looks like Memphis is connected to St. Louis, but that is a mere geographical representation of the visible, physical network. An economically relevant network is buried beneath the easily noticeable surface network.

1.3 City-Size Distribution Is Scale Free Too

As we mentioned earlier, the city-size distribution has a distinct feature. Figure 2 plots the frequency of the city-size distribution from U.S. Census 2000. It is only when we take the log of population that the distribution exhibits resemblance to a Gaussian distribution. We can see the chance of the extremes is high.

The fat-tailed distribution also makes its appearance on a map. Figure 3 illustrates the population density of each metropolitan statistical area in the

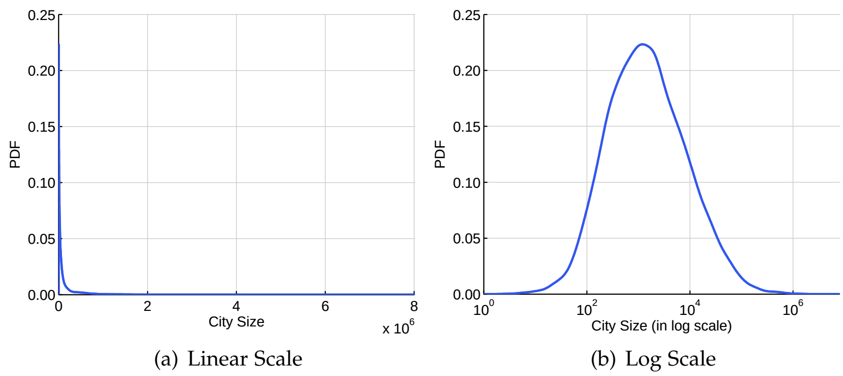


Figure 2. Frequency plot of the city-size distribution. *Data source*: U.S. Census 2000.

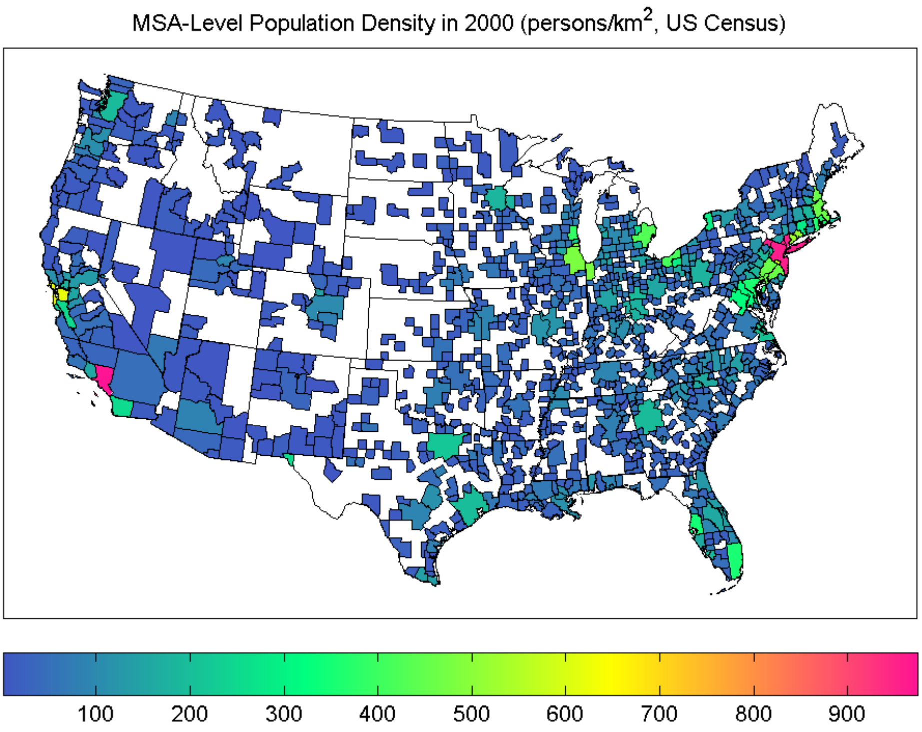


Figure 3. Population density by MSA. *Data source*: Census 2000.

United States in 2000. Most of the cities have a low density and are painted in blue; there are only few cities that are green and only two cities are colored in red. If the city-size distribution followed a Gaussian distribution, most of the cities should be green or its neighboring colors, and only a few should be painted in blue or red. In reality, a few large cities take up the lion's share of urban population and the majority of cities are left in blue.

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 cost makes a 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 new link to the city will not increase accessibility much because the city is readily accessible from other cities through existing links.

We test implications of our model using 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 two types of network structures mentioned above as a preamble to the next section. In Section 3 we introduce and develop a model of spatial equilibrium with a transportation network woven into it. Particularly, in Section 3.7, we connect the network structure to the city-size distribution. In Section 4, we verify the prediction of our model using three kinds of data sets before we draw conclusions from our research in Section 5.

2 Preliminaries

As a reminder, we will briefly review how ER and BA networks are built and examine the qualitative differences in terms of their degree distributions.

2.1 ER Networks

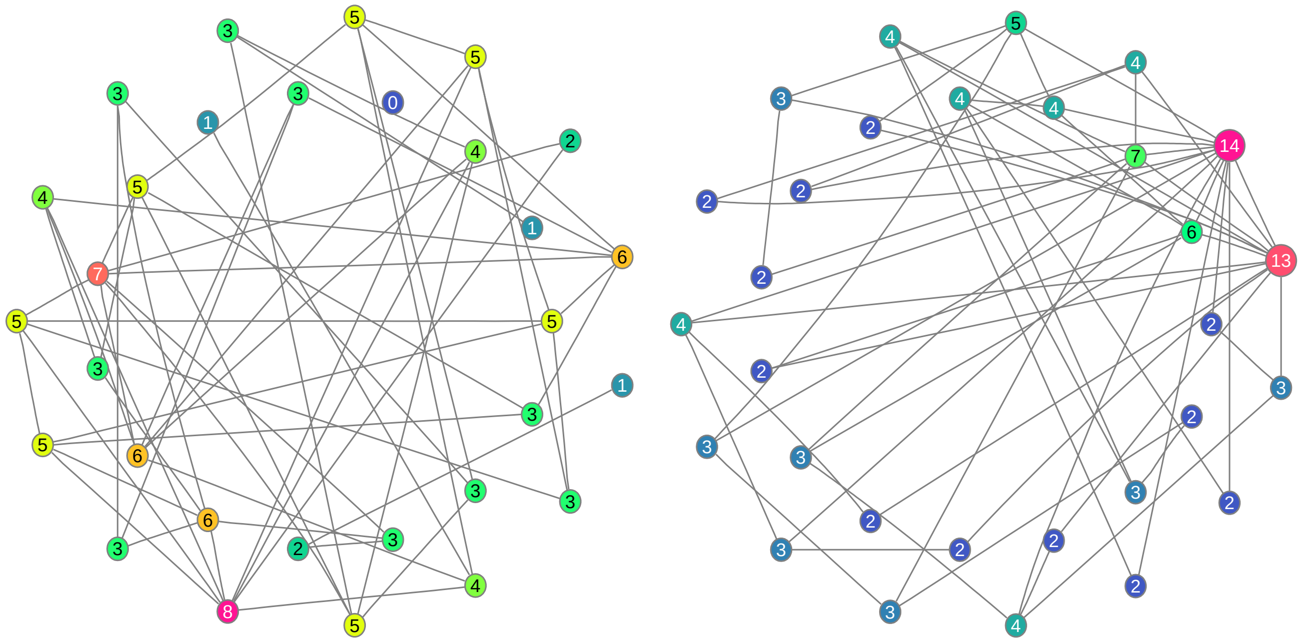
The ER network is the simplest random graph of all. A pair of nodes are connected with a fixed connection probability q. Two special degenerate cases, where $q = 0$ or 1 is called a completely isolated graph and a complete graph, respectively. These are the networks most commonly assumed in the literature on city-size distributions.

The degree distribution of an ER network follows the Poisson distribution (see Appendix A.1 for details). An important feature is that the degree distribution is concentrated around its mean and we rarely observe a city with an exceedingly large degree. All pairs of nodes share the *same* connection probability, which leads to a small variance, and the network is *egalitarian* in that sense.

2.2 BA Networks

As we mentioned in the Introduction, the degree distribution of most real network structures does not follow the 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 [AB02] for review). To get a feeling for how the power-law type behavior emerges, we review a BA model [BA99] as an example. Two major characteristics of BA model is growth and preferential attachment. The model starts with a complete graph of x_0 nodes. A node with $x(\leq x_0)$ edges will be added sequentially to the existing network (growth). The entering node connects to an existing node with a probability proportional to its current degree (preferential attachment). In particular, the probability that node i wins a new edge is $k_i / \sum_j^J k_j$, where k_i counts the number of edges of node i.

As we can see from this mechanism, in general, older nodes are likely to gain an excessively large number of edges. The rich get richer because they are already rich. The rest of the nodes are merely mediocre in terms of degree. They are poor because they are already poor. This type of variance in degree barely arises with ER network. Compare the example networks in Figure 4. BA network is *not* egalitarian, as connection probability depends on the



(a) ER Network

(b) BA Network

Figure 4. Networks with 30 nodes and 57 edges. The number in a node marks its degree. We see that BA network exhibits a closer resemblance to Figure 3 than ER network and we expect that the BA network explains the city-size distribution better. This conjecture will be validated as we will see later.

number of acquired edges, which is path dependent. The BA model predicts the Pareto exponent to be $\alpha = 3$ at the stationary state, which is consistent with the majority of real networks ($\alpha = 2$ to 3).

3 The 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

In what follows, a superscript denotes a city of production or origin, whereas a subscript denotes a city of consumption or destination. There are J cities in the economy, with each indexed by j. 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.

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:[1]

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

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 $\Gamma = (V, B)$, where $V = \{1, \cdots, J\}$ denotes the set of vertices representing each city and $B \subseteq V \times V$ denotes a set of edges. We assume that the network is unipartite (i.e., there is a path between any pair of nodes) to avoid multiple equilibria. While consumers in city j 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, let us define a metric on $V \times V$. A metric $l_j^i : V \times V \to \mathbb{R}_+$ measures a geodesic length (the shortest path length) between node i and j. 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 $\tau(\geq 1)$ marks the iceberg transportation parameter. If you dispatch τ units of commodity to your neighboring city, one unit of it will be delivered and the rest melts en route. The delivered price snowballs as the package travels from one city to another and the initial mill price is inflated by $\tau^{l_j^i}$ by the time the package reaches its final destination.

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

3.3 Equilibrium

Simple calculations yield the Marshallian demand for commodity c_i^j:

$$\varphi_i^j(p_i^1, \cdots, p_i^J, w^i) = w^i(\tau^{l_i^j} p_j^j)^{-1} J^{-1}.$$

The aggregate demand for commodity j is the sum of demand from all the cities in the country: $C^j(p, w) = \sum_{i \in V} s_i \varphi_i^j(\cdot)$. [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

$$s_j = C^j(p, w) = \left(p_j^j\right)^{-1} J^{-1} \sum_{i \in V} s_i w^i \tag{4}$$

The indirect utility function is given by

$$\begin{aligned} v(p_i^1, \cdots, p_i^J, w^i) &= \tfrac{1}{J} \textstyle\sum_{j=1}^J \log \varphi_i^j(\cdot) \\ &= \log w^i - \log J - \tfrac{1}{J} \textstyle\sum_{j \in V} \log p_j^j - a_i \log \tau, \end{aligned}$$

where

$$a_i = \frac{1}{J} \sum_{j=1}^J l_i^j \tag{5}$$

measures accessibility of city i normalized by system size. Note that since a_i counts the number of edges that must be travelled to reach each and every city from city i, the smaller the value, the better the accessibility. We will explore the role of accessibility later.

Free mobility of consumers implies

$$v(p_i^1, \cdots, p_i^J, w^i) = v(p_j^1, \cdots, p_j^J, w^j) \tag{6}$$

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

The equilibrium $(s_1, \cdots, s_J; p_1^1, \cdots, p_J^J; w^1, \cdots, w^J)$ satisfies (1), (2), (4) and (6). Utility equalization (6) leads to

$$\log p_i^i - \log p_j^j = (a_i - a_j) \log \tau. \tag{7}$$

Equation (7), together with (4), implies $s_j = \tau^{a_i - a_j} s_i$. With the population condition (1), we obtain the city-size distribution

$$s_i = \frac{S}{\tau^{a_i} \sum_{j \in V} \tau^{-a_j}}. \tag{8}$$

[2] This expression may seem incredulous at first, for it does not include τ. A large τ discourages demand but it also means that firms have to ship more commodities. A large portion of shipment will melt on its way. They cancel each other in equilibrium. This propitious cancellation may not occur with other preference specifications.

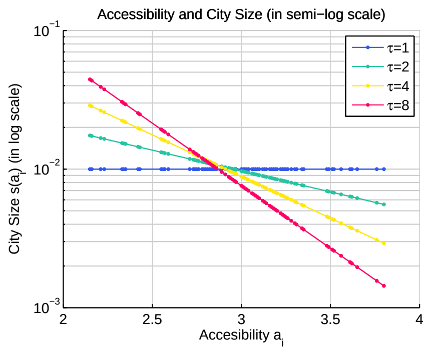
3.4 How Does a Network Break Symmetry?

It is immediately obvious that the better the accessibility of a city is, the larger the city population will be. Naturally, we are tempted to conclude that the entire population will collapse into the city with the best accessibility and the rest of the cities will be completely vacated. As it turns out, this is not the case. The city-size distribution will not become degenerate. Let us break down (8) both mathematically and economically to see why.

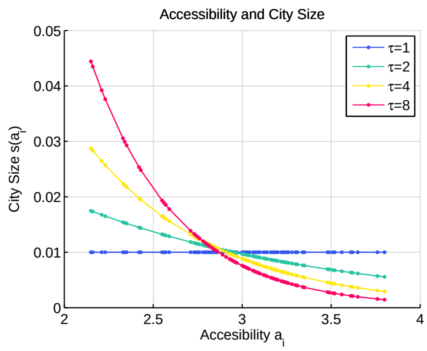
First, let us recast the relationship (8) to explore how accessibility translates to the population of a city. We can rewrite (8) as $s(a_i) = \hat{s}\tau^{-a_i}/\overline{\tau^{-a}}$, where $\hat{s} := S/J$ is a base city size (the size of the city if the city-size distribution were uniform) and $\overline{\tau^{-a}} := \sum_j \tau^{-a_j}/J$ gives the average of t^{-a_j}. In what follows $\overline{x}$ denotes the average of a variable x. The city size spreads around the canonical size $\hat{s}$. A better accessibility (i.e., small a_i) contributes to the city by augmenting the baseline size $\hat{s}$ by a factor of $\tau^{-a_i}/\overline{\tau^{-a}}$. The multiplier is large when τ^{-a_i} is greater than the national average $\overline{\tau^{-a}}$ and vice versa. Put differently, there is a semi-log-linear relationship between size and accessibility:

$$\log s(a_i) = \log \hat{s} - a_i \log \tau - \log \overline{\tau^{-a}},$$

as can be seen in Figure 5(a). Furthermore, the multiplier grows *more than proportionally* as accessibility improves (i.e., a_i gets smaller), for τ^{-a_i} is monotone decreasing and convex in a_i (we will prove convexity in Proposition 3.1). Does this mean New York City sweeps away all the population off the rest of the cities? — Not really. And it calls for an economic exposition of (8).



(a) Semi-Log Scale



(b) Linear Scale

Figure 5. City-szie distributions based on a scale-free network of $J = 100$ cities.

Although restricted accessibility of a city raises its delivered prices, demand for its produced commodity does not cease to exist. Eliminating a commodity from the basket will cost consumers a lot. 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 a high price for imported commodities due to a poor network infrastructure, but they are compensated by a high nominal wage, as indicated by the wage (2) and utility equalization (7). These two equations imply that the mill price (and ultimately, the nominal wage) is positively related to the accessibility parameter in equilibrium, i.e., a sparsely connected city has a high mill price. The prices adjust to make it worth living in cities like St. Louis in equilibrium. The scale of local production is small, but each commodity is sold high to make up for increased cost of living due to costly transport.

Variance in city sizes is solely due to the structure of network. The abovementioned trade-off entails two counteracting forces. One force is the transport network, which tends to spread out the city-size distribution. The opposing force is preference for variety, which tends to push the distribution back to a uniform distribution. Obviously, this trade-off disappears and there will be no variance in city sizes if the first force is removed. This can happen when shipment becomes costless (to be discussed in Proposition 3.1) or network structure becomes redundant (to be discussed next).

Before we throw a couple of asymmetric networks into (8), let us consider how our model compares to the existing models. In most urban models, network structures take simple forms. The assumed structure (be it explicitly stated or not) is either a complete graph or a completely isolated graph. The New Economic Geography engages a complete graph as its transport structure whereas others (for example Eeckhout [Eec04]) use a completely isolated graph. These symmetric networks lead to the uniform distribution of city sizes because all the vertices hold the same accessibility value. commodities. Although we introduced a location-specific technology, commodities are symmetric.. Technology is linear everywhere and 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 city-size distribution is uniform and all the cities share the same size $\hat{s}$ and every household consumes an equal amount of all the commodities available. To break the symmetry, some introduce increasing returns to scale and others introduce random growth. Our plan is to understand the mechanism generating city-size distributions with a more graphic, lifelike transport network structure that affects the state of local economies.

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, τ can nullify or amplify the influence of a network structure over the economy. Figure 5(b) compares the relationship between accessibility and the city-size distribution under different transportation costs. In the extreme situation where shipment is free ($\tau = 1$), the city-size distribution becomes uniform regardless of the network structure The city size $s(a_i)$ becomes constant against a_i (See the blue line in Figure 5(b)). The network becomes a complete graph, in effect, because the delivered price becomes independent of the geodesic length between cities. The distance travelled is irrelevant for the system if $\tau = 1$. For $\tau > 1$, city size (8) becomes a strictly convex function of accessibility.

The transportation network Γ starts to sink in 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, a city with a good accessibility benefits from a low a_i value because high transportation cost amplifies the effect of accessibility. In other words, a high transportation cost reveals the network structure and projects the network Γ onto the city-size distribution in a more defined, clear-cut manner than with a low transportation cost. 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_L and a_H are the lowest and highest 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 $s(a_i)$, which gauges how spread out the distribution of city size $s(a_i)$ is for each τ. See Figure 6. When

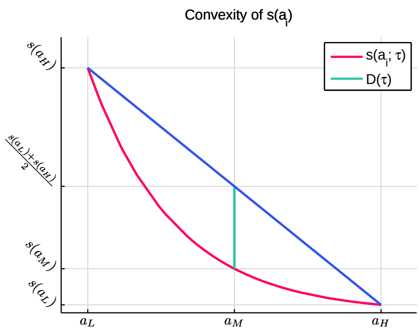


Figure 6. $D(\tau)$ measures the convexity of $s(a_i)$. The midpoint $(a_H + a_L)/2$ is given by a_M above.

$\tau = 1$, $s(\cdot)$ lays flat and $D(\tau) = 0$. As τ grows, $s(\cdot)$ bends more and $D(\tau)$ grows accordingly as can be seen in Figure 5(b).

We confirm the observation above as follows:

Proposition 3.1 Transportation Cost Skews the City Size Distribution

Suppose that the economy has a unipartite network Γ. The city size distribution s_i is a convex function of accessibility a_i for $\tau \geq 1$. Moreover, the degree of convexity measured in the difference $D(\tau)$ between the mid-sized city and the mid-accessible city increases with τ.

Proof. See Appendix A.2. □

3.6 The Geodesic-Length Distribution

The city-size distribution (8) depends on the distribution of accessibility (5), which, in turn, rests on the distribution of geodesic length. While most of the research on network topology is focused on mean intervertex distance, we are more interested in the geodesic length between individual nodes. That is, we would like to derive the city-size distribution, not the average size of cities. Hołyst et al [HSF+05] measure the expected geodesic length between any pair of nodes i and j as follows:

$$l_j^i = A - B\log(k_i k_j), \tag{9}$$

where $A = 1 + \log(J\bar{k})/\log\kappa$ and $B = (\log\kappa)^{-1}$. A number k_i denotes the degree of node i and κ denotes the average branching factor (to be explained later). We briefly repeat their arguments to obtain (9). Consider a geodesic between nodes v_i and v_j. Rearrange the nodes so that we have a tree with v_i as its root.[3] A tree is a sequence of nodes where each node except for the root node has exactly one parent (or ancestor) node. Each node may or may not be followed by (a) child node(s). There are no cycles on a tree. The average number of children is called an average branching factor and denoted κ. If we pick a random tree starting from v_i, we will wind up at v_j somewhere along the tree $k_j/\sum_{r\in V} k_r$ of the time and we will not reach v_i $1 - k_j/\sum_r k_r$ of the time. On average, we will reach v_j within $\sum_r k_r/k_j$ trials. Suppose that the depth (the number of parent nodes that you have to go through before reaching your root node) of v_j is l. There are $k_i\kappa^{l-1}$ nodes whose depth is l. Therefore, on average, we arrive at v_j in l steps if

$$\frac{\sum_r k_r}{k_j} = k_i\kappa^{l-1},$$

from which we obtain (9).

The exact branching factor κ cannot be found until after the graph is generated. We can obtain a good estimate of κ by averaging the degree of adjacent nodes [HSF+05]. Using a probability-generating function, Newman et al [NSW01] compute the average degree of adjacent nodes (see Appendix A.3 for details) and we take this value as an approximation to κ.

3.7 City-Size Distribution

From (9), accessibility (5) is written as

$$a_i = A - B\log k_i - B\overline{\log k}. \tag{10}$$

We observe that accessibility improves as a city acquires more edges, but only on the logarithmic order. Taking the log of (8), we have

$$\log s_i = \log S - (A - B\log k_i - B\overline{\log k})\log\tau - \log\left(\sum_j \tau^{-a_j}\right).$$

[3] 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. As shown in [HSF+05], the resulting error is minimal.

The last term is approximated by $\log J - \bar{a}\log\tau$ [4] so that

$$\log s_i = \log\frac{S}{J} + B\log\tau\left(\log k_i - \overline{\log k}\right). \tag{11}$$

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." The second, and more interesting question is "How so?" The answer is twofold.

In terms of a linear scale, (11) can be rewritten as $s_i = \frac{S}{J}\left(\frac{k_i}{\gamma}\right)^{B\log\tau}$, 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 S/J (mark this size by $\hat{s}$ as before [5]), 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 $\hat{s}$ exactly when its degree matches the national (geometrical) average. The deviation is amplified as shipment becomes costly, which is consistent with our observation 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. The first couple of edges will greatly contribute to the production and consumption level of a city. However, blindly connecting to other cities does not pay. Consider, for example, the economy where city i and j are connected. For some other city n, connection to either one of the cities will sharply increase its accessibility, for it will gain access to not only city i or j but also all the cities that city i or j is connected to. On the other hand, it does not improve a_n much if the city n extends the second edge to connect to both cities. The first edge has already established the path to v_j via v_i (or the other way around). The second edge does reduce the transportation cost but its leverage is marginal compared to the first edge. As we learned from (10), degree improves accessibility only on the logarithmic order.

Based on the degree-size relationship (11) the city-size distribution is given as follows:

Proposition 3.2 City-Size Distribution

Suppose that the economy has a unipartite network Γ with the associated degree distribution $G(k)$. The city-size distribution of this economy follows the distribution function $F(\cdot)$, defined by

$$F(s) = G(k(s)), \tag{12}$$

where $k(s) = \gamma(s/\hat{s})^{\frac{\log\kappa}{\log\tau}}$. Its probability density function (pdf) is

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

[4] Taylor series expansion and law of large numbers imply

$$\begin{aligned}\log\left(\sum_j \tau^{-a_j}\right) = \quad & \log\left(\sum_j \tau^{-\bar{a}}\right) + (a-\bar{a})\cdot D\log\left(\sum_j \tau^{-a_j}\right) + O\left[(a-\bar{a})\cdot(a-\bar{a})\right] \\ \approx \quad & \log J - \bar{a}\log\tau.\end{aligned}$$

[5] The size $\hat{s}$ is what the city size would be if all the cities had the same size. Our predicted city sizes happen to drift around this baseline size.

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 generate our predictions based on different transport network structures. There are two network models of particular interest: Erdős-Rényi (ER) network and Barabási-Albert (BA) network.

The degree distributions are known for both networks. We will obtain the city-size distribution implied by each of the two types of network schemes. Considering the prevalence of the scale-free network in reality, we expect that the BA system fits the city-size distribution better than ER system. And as we will see later, it does.

Note that empirical determination of the transport network relevant to the formation of a system of cities is a tough job. We will discuss this in Section 5. The task at hand is to find a network that is consistent with the real city-size distribution (and we have already discarded complete and completely isolated networks in Section 3.3). The most consistent network structure will give us a clue as to the shape of a network that is germane to the formation of cities.

3.8.1 The City-Size Distribution in the ER Economy

The degree distribution of an ER network is given by the Poisson distribution with $\lambda = qJ$ as its mean. With regard to the city-size distribution (12), we supply two more parameters (κ_{ER} and γ_{ER}) as follows. Using the result of [NSW01], the probability-generating function of the degree of the first neighbor is given by $\psi_1(x) = e^{\lambda(x-1)}$. Then the branching factor is given by $\kappa_{ER} = \psi_1'(1) = \lambda$.

We approximate the geometric mean γ_{ER} by $\lambda e^{\frac{-1}{2\lambda}}$. [6] The degree-size relationship (11) reduces to $k_{ER}(s) = \lambda e^{\frac{-1}{2\lambda}} \left(\frac{s}{\hat{s}}\right)^{\frac{\log\lambda}{\log\tau}}$. The pdf is

$$f_{ER}(s) = k_{ER}'(s) g_{ER}(k_{ER}(s)) = \lambda e^{\frac{-1}{2\lambda}} \frac{\log\lambda}{\log\tau} \left(\frac{s}{\hat{s}}\right)^{\frac{\log\lambda}{\log\tau}} s^{-1} \frac{\lambda^k}{k!} e^{-\lambda}.$$

As we assume a unipartite network, the connection probability q cannot go below the threshold value of $\log(J)/J$ (below which the graph falls apart; see Erdős and Rényi [ER61]). This leads to $\lambda = qJ > \log J$. Therefore, the asymptotic degree distribution follows the normal distribution with mean and variance λ. Hence, the pdf is

$$f_{ER}(s) = \lambda e^{\frac{-1}{2\lambda}} \frac{\log\lambda}{\log\tau} \left(\frac{s}{\hat{s}}\right)^{\frac{\log\lambda}{\log\tau}} s^{-1} \frac{1}{\sqrt{2\pi\lambda}} \exp\left(\frac{-(k_{ER}(s)-\lambda)^2}{2\lambda}\right).$$

[6] Using a Taylor expansion $\log k = \log\lambda + \frac{k-\lambda}{\lambda} + \frac{-1}{2}\left(\frac{k-\lambda}{\lambda}\right)^2 + O[(k-\lambda)^3]$,

$$\begin{aligned} \log\gamma &\approx \sum_{k=0}^{\infty}\left[\log\lambda + \frac{k-\lambda}{\lambda} + \frac{-1}{2}\left(\frac{k-\lambda}{\lambda}\right)^2\right] g(k) \\ &= \log\lambda + E\left[\frac{k-\lambda}{\lambda}\right] + \frac{-1}{2} E\left[\left(\frac{k-\lambda}{k}\right)^2\right], \end{aligned}$$

which leads to the value in the text. Inclusion of the moments of higher order do not seem to improve approximation much in practice (See Young and Trent [YT69] for some empirical experiments). Note that Jensen's inequality is met with this approximation ($\lambda > \lambda e^{\frac{-1}{2\lambda}}$).

3.8.2 The City-Size Distribution in the BA Economy

The degree distribution of BA network follows the power law $g_{BA}(k) \propto k^{-\alpha}$. Empirically, the Pareto index α of most networks (a relevant transportation network that leads the city-size distribution is likely among them) falls between 2 and 3.

We represent the degree distribution by the Pareto distribution $g_{BA}(k) = \alpha\frac{k_m^\alpha}{k^{\alpha+1}}$, where k_m marks the lowest value of k. We use an alternative method to what we used in ER economy, based on Pastor-Satorras et al [PSVV01] to find the branching factor $\kappa_{BA} = \frac{k_m}{\alpha-1}\left(\frac{1}{\alpha-2}+\alpha\right)$ (See Appendix A.4 for details).

If k follows the Pareto distribution, $\log(k/k_m)$ follows the exponential distribution with the reverse scale parameter α. Simple calculation leads to $\log\gamma_{BA} = \int_0^\infty \log(k/k_m)d\tilde{G}(\log(k/k_m)) + \log k_m$ [7] so that $\gamma_{BA} = k_m e^{\frac{1}{\alpha}}$ (note the mean of the exponential distribution is $\frac{1}{\alpha}$). The degree-size relationship (11) becomes

$$k_{BA}(s) = k_m e^{\frac{1}{\alpha}}\left(\frac{s}{\hat{s}}\right)^{\frac{\log k_m - \log(\alpha-1)+\log\left(\frac{1}{\alpha-2}+\alpha\right)}{\log\tau}}.$$

Hence, the pdf is

$$f_{BA}(s) = \frac{\log\kappa_{BA}}{\log\tau}k_m e^{\frac{1}{\alpha}}\left(\frac{s}{\hat{s}}\right)^{\frac{\log\kappa_{BA}}{\log\tau}} s^{-1}\frac{\alpha k_m^\alpha}{k_{BA}(s)^{\alpha+1}}. \tag{13}$$

4 Empirical Implementation

We tried maximum likelihood estimation first, which turns out to be problematic,[8] and then obtained alternative estimates that minimize the Kolmogorov-Smirnov (KS) statistic (the largest gap between the empirical and predicted CDF). We summarize our estimation in Table 1 and present distribution functions for the USA using Soo [Soo05]'s data (Figure 7).

As expected, the BA economy is more consistent with the real city-size distributions. Eeckhout [Eec04] and Berliant and Watanabe [BW11] have testable predictions and we compare our result to their results as representatives of existing models. Note that each model leads to different distributions. Direct comparison among their KS statistics or maximum likelihood values may not be satisfactory to evaluate a model's performance. The model's fit to the data can be made arbitrarily better by adding more parameters to the distribution function. To account for the difference in the number of parameters, we report Akaike and Bayesian information criteria along with the KS statistic in Table 1. These criteria penalize a distribution's fit for having many parameters. A lower value of Akaike or Bayesian information criteria means a better fit. The BA economy is comparable to existing models in terms of these criteria.

Now, let us examine our findings in detail. First, the complete graph as a candidate model is out of the question. Second, the ER network fails to explain the city-size distribution, for it yields an economically unsustainable

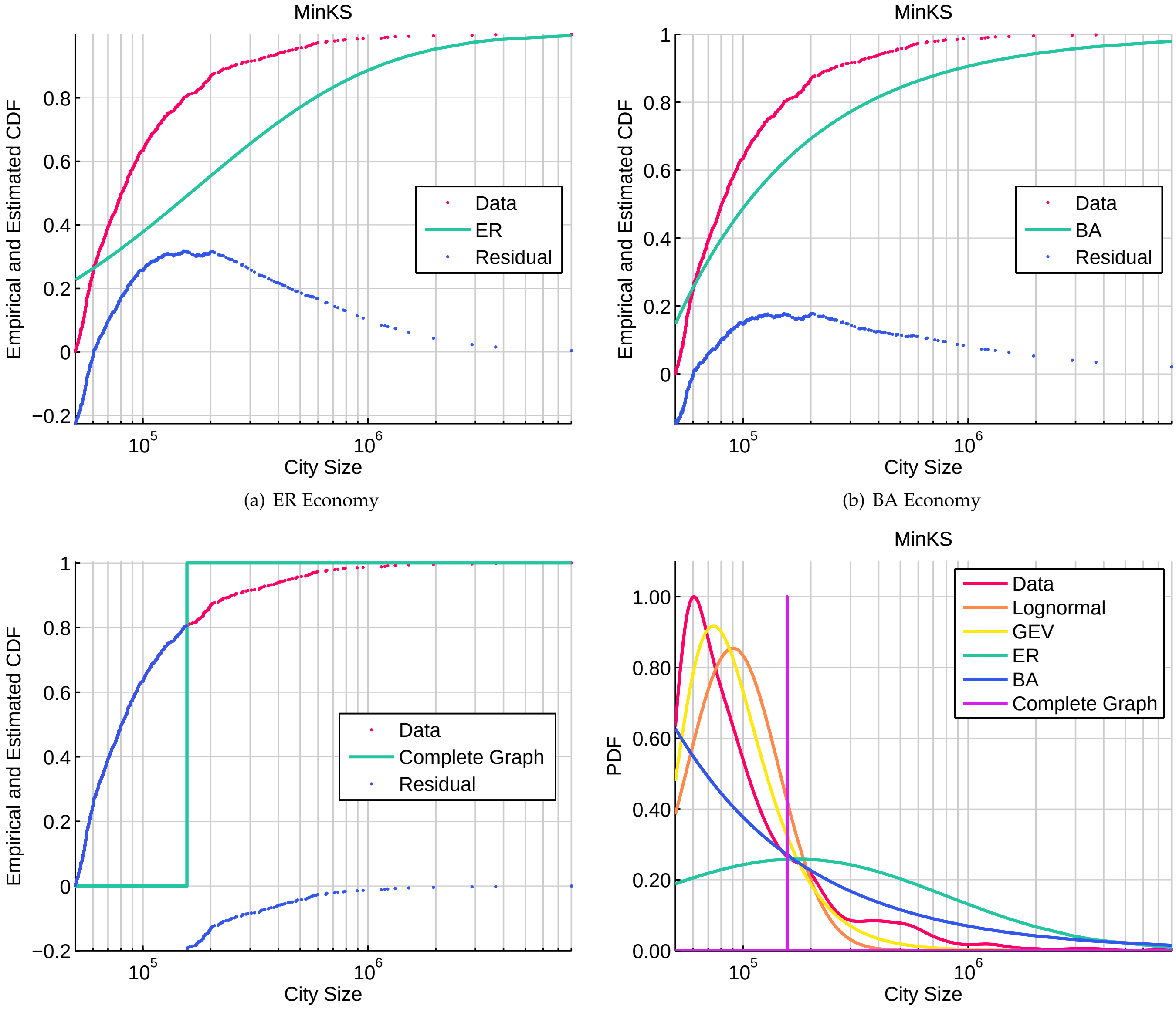
[7] $\tilde{G}(\cdot)$ denotes the cumulative density function (CDF) of the exponential distribution.

[8] The likelihood function tends to explode. For the ER economy, the likelihood is increasing in τ, and for the BA economy, α tends to 2.

Table 1. Parameter Estimates and Related Statistics

Country	Unit	Distribution	Method	Location	Scale	Shape	Other 1	Other 2	KS	AKS [a]	LogLH [b]	AIC [c]	BIC [d]
		Complete Graph					$\hat{s}$						
		Lognormal		μ	σ								
		GEV		μ	σ	ζ	κ	γ					
		ER		λ			τ						
		BA			k_m	α	τ						
USA	Soo[f]	Complete Graph	–	–	–	–	1.569E+05	–	0.8051	229.14	$-\infty$	∞	∞
		Lognormal	MLE	11.52	0.7107	–	–	–	0.1610	54.46	-8402	16808	16817
		GEV	MLE	1.562	0.02067	0.3039	1.005	0.9616	0.0726	21.20	-8165	16339	16362
		ER	MLE[e]	1.457E+03	–	–	6.977E+100	–	0.3564	132.79	-8525	17054	17063
		BA	MLE[e]	–	–	–	–	–	–	–	–	–	–
		Lognormal	MinKS	11.41	0.4669	–	–	–	0.1019	42.54	-8579	17162	17171
		GEV	MinKS	3.056	0.2279	0.05946	1.804E+02	0.8432	0.0744	28.80	-8248	16506	16529
		ER	MinKS[e]	526.5	–	–	2.557E+96	–	0.3183	118.79	$-\infty$	∞	∞
		BA	MinKS	–	0.5089	2.876	1.398	–	0.1780	70.13	-8434	16873	16887
Belgium	Soo	Complete Graph	–	–	–	–	–	–	0.6812	19.70	$-\infty$	∞	∞
		Lognormal	MinKS	10.71	0.4089	–	–	–	0.1472	5.31	-807	1619	1623
		GEV	MinKS	3.188	0.2319	0.1715	156.5	0.8257	0.1152	3.66	-779	1567	1578
		ER	MinKS	49.67	–	–	4.417E+09	–	0.2738	10.67	$-\infty$	∞	∞
		BA	MinKS	–	0.5284	2.874	1.386	–	0.1457	5.59	-794	1594	1600
USA	MSA	Complete Graph	–	–	–	–	2.837E+05	–	0.8362	335.07	$-\infty$	∞	∞
		Lognormal	MinKS	11.32	1.039	–	–	–	0.0614	33.40	-12056	24117	24127
		GEV	MinKS		2.544	0.3285	111.4	0.8681	0.0184	9.03	-11910	23829	23854
		ER	MinKS[e]	1.782E+02	–	–	2.313E+93	–	0.3540	178.24	$-\infty$	∞	∞
		BA	MinKS[e]	–	1.001	41.83	23.71	–	0.2083	100.71	-11369	22744	22758

a. Sum of the residuals.
b. Sum of loglikelihood
c. Akaike information criterion.
d. Bayesian information criterion.
e. The objective function did not converge. We interrupted the algorithm after sufficient number of iterations and report, if available, the obtained estimates as asymptotic values for reference.
f. Soo [Soo05].



(a) ER Economy

(b) BA Economy

(c) Complete Graph

(d) PDF

Figure 7. Soo [Soo05], USA

estimate. In both methods of estimation, the objective function of estimation explodes. The fit becomes marginally better by greatly increasing τ. 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 distribution has (c.f., Proposition 3.1). As a result, τ has to be ludicrously large. We terminated the iteration in the midst of optimization. By the time of termination, τ had reached $2.557E+96$. A one-dollar pen in St. Louis would cost 2.557E+96 dollars in Chicago, which is far beyond the US GDP. An ER network is not capable of reproducing the empirical city-size distribution.

The BA network performs well in the mid range but worse than [BW11] towards the left end of the tail, especially when more cities are included in the data set. Our initial concern was that the BA degree distribution may be too stable to generate city-size distributions, considering the apparent difference in Pareto exponents ($\alpha = 2$ to 3 for most networks, whereas city-size distributions usually have $\alpha = 1$). It turns out to be the opposite. Lognormal [Eec04] is too thin near the top, whereas BA is too thick,[9] possibly for two reasons. First, as mentioned in Appendix A.4, the power law in its purest form is theoretically tractable but deviation near the top of the actual distribution is not negligible. On the other hand, the power law with an exponential cutoff is realistic, but theoretically not manageable. As a result, the tail of the distribution becomes too fat. We computed the expected degree out of the given city-size distribution in Table 2 (note that the upper tail is sensitive to parametarization, so we provide the table only as a reference). It seems that the degree is too thick near the top due to the lack of cutoff.

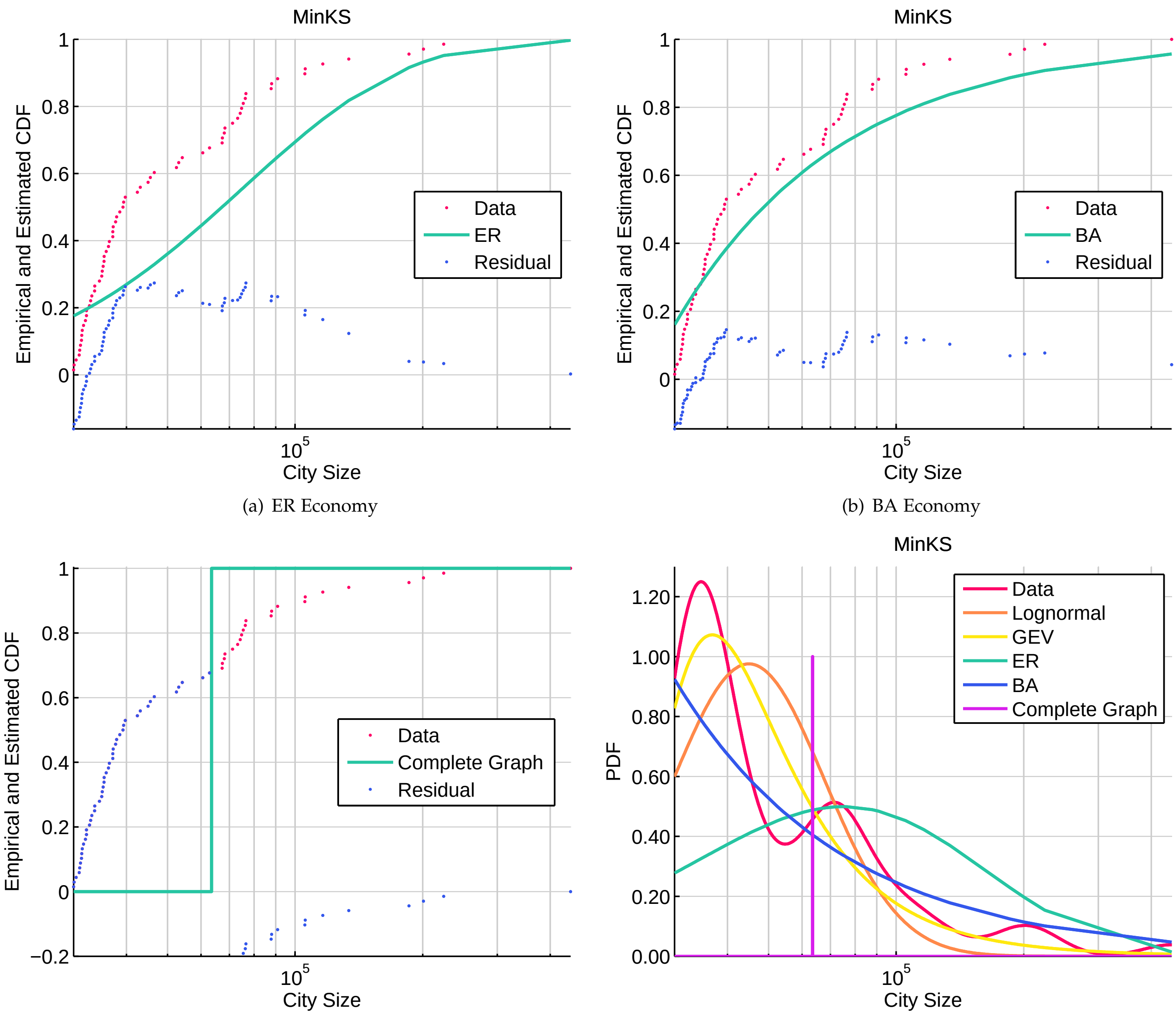
	USA [So005]			Belgium			USA (MSA)		
Rank	City	Size	Degree	City	Size	Degree	City	Size	Degree
1	New York	8,008,278	36.1	Antwerpen	446,525	29.0	New York	18,323,002	26.0
2	Los Angeles	3,694,820	5.1	Gent	224,180	2.3	Los Angeles	12,365,627	5.2
3	Chicago	2,896,016	3.9	Charleroi	200,827	1.8	Chicago	9,098,316	4.1
4	Houston	1,953,631	3.4	Liège	185,639	1.6	Philadelphia	5,687,147	3.5
5	Philadelphia	1,517,550	3.0	Bruxelles	133,859	1.4	Dallas	5,161,544	3.2
6	Phoenix	1,321,045	2.8	Brugge	116,246	1.3	Miami	5,007,564	3.0
7	San Diego	1,223,400	2.6	Schaerbeek	105,692	1.2	Washington DC	4,796,183	2.8
8	Dallas	1,188,580	2.4	Namur	105,419	1.2	Houston	4,715,407	2.7
9	San Antonio	1,144,646	2.3	Mons	90,935	1.1	Detroit	4,452,557	2.5
10	Detroit	951,270	2.2	Leuven	88,014	1.1	Boston	4,391,344	2.4

Table 2. 10 Largest Cities and Their Expected Degree

Secondly, we assumed a unipartite network. Our estimation seems to suggest that the US may be too large to be considered as a unipartite network. And it can be "too large" in two senses: one in a physical sense and the other in terms of truncation. There may be remote cities whose economy remains

[9]This implies that the network based economy with an Eeckhout-type externality may help improve the fit. Note, however, that our research concerns how a network structure affects city sizes. We keep our model to the bare minimum to isolate the network's leverage. More complicated models, such as Eeckhout's with a network, might not have a closed form solution. Moreover, all cities in Eeckhout's model produce the same commodity, so there is no reason to trade.

comparatively unaffected by the network configuration of other large cities. Or there may be a dichotomy among the cities: Large cities are based on a scale-free network and small cities have another kind of network. If this is the case, filtering the data by removing small cities will improve the fit.



(a) ER Economy

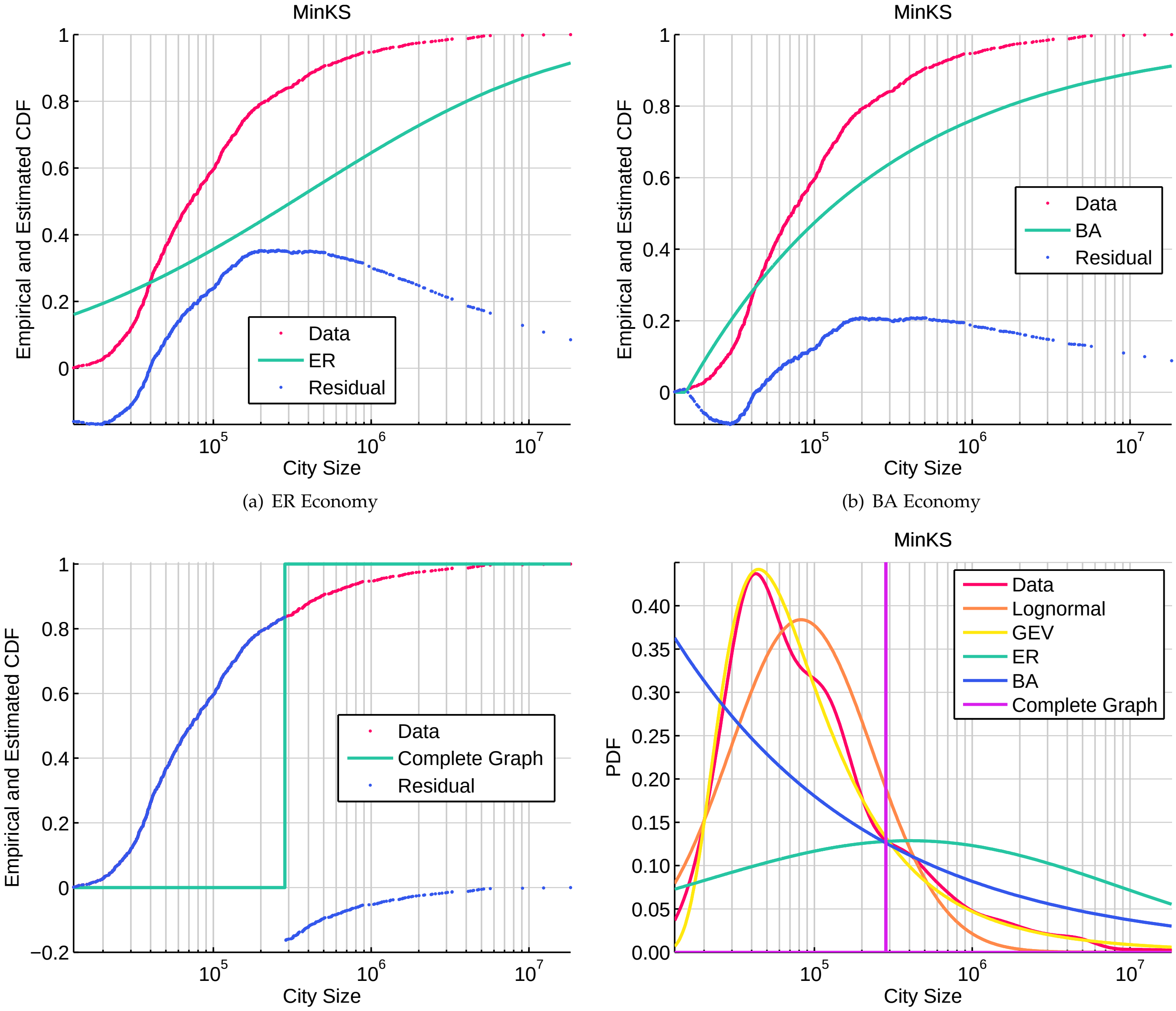
(b) BA Economy

(c) Complete Graph

(d) PDF

Figure 8. Soo [Soo05], Belgium

To confirm our two hypotheses above, we took Belgian city-size data and US metropolitan statistical area (MSA) level data as a follow-up exercise. Distribution functions are represented in Figures 8 and 9. Related statistics are



(a) ER Economy

(b) BA Economy

(c) Complete Graph

(d) PDF

Figure 9. MSA, USA

reported in Table 1.

Belgium is one third of one hundredth of the area size of the US whereas the total population is about 3.5% of the US population. Due to the size of the country, it is harder to form a city off the grid in Belgium than in the US. If physical proximity improves the connectivity of the network, Belgian data would fit our story better, other economic factors being equal. The data seems to support our conjecture on locational contiguity.

If truncation is the issue, then Soo's data for the US (with $J = 667$) would be more suitable to our model than the US MSA data (with $J = 922$). It is likely that cities that are on the list of MSA's but not on the primary network of large cities are filtered out. The data appears to be in favor of our second conjecture as well.

Modelling a multi-component network is challenging for economic and tractability reasons. With our model, there will be multiple equilibria if the cities are grouped into distinct subnetworks.

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 (11), where we have learned that *there is a log-linear relationship between city size and city degree*.

We put two commonly studied network models to the test. The classical ER random graph is too egalitarian to generate gravitationally large cities like New York City or Los Angeles. When translating a degree distribution into the city-size distribution, we need to do so in a way that adds more weight on the extremes, if we were to explain the city-size distribution by an ER model. This was only possible by raising the transportation cost to a very high level, which is not realistic.

The BA model explains the city-size distribution better than ER model, but there is room for improvement. BA model has a degree distribution that follows a power law, and the resulting city-size distribution behaves similarly. While we did obtain economically reasonable estimates, contrary to the BA model, now the predicted city-size distribution is slightly too fat. We expect that an adjusted power law and further inspection of actual transport networks will improve the fit.

We finish the discussion with one last remark. We argued that network structures motivate the population to form a specific distribution of city sizes. The structure of the network is pre-selected. Considering the fact that it is easier to relocate people than to build transport infrastructure, this is not an unreasonable assumption in the short run. New York City would have been much smaller had it not been the port of entry to Europe. However, the degree-city relationship is not a one-way street. It may be the other way around: the relocation of people forces the transportation network to follow a specific pattern. 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. As we have mentioned in Section 1.2, the U.S. has seen a number of drastic changes in its network structure. Tracing the historical co-development of the network structure with the city-size distribution may reveal a clue to identifying the direction of causality. Eventually, it will be important to develop models where both the transport network and city size are endogenous, and co-develop in a fashion where each is dependent on the other.

A Appendix

A.1 ER Degree Distribution

Derivation of the ER degree distribution is twofold. First, we derive the distribution of X_k, the number of nodes with degree k. Suppose any two nodes are connected with probability q. The probability of a node i wired to k other nodes is given by

$$P_i(k_i = k) = {}_{J-1}C_k q^k (1-q)^{J-1-k},$$

where ${}_aC_b := \frac{a!}{b!(a-b)!}$. [10] The probability that X_k is equal to $r \in \{0, 1, \cdots, J\}$ follows:

$$P_r(X_k = r) = {}_JC_r P_i(k_i = k)^r \left[1 - P_i(k_i = k)\right]^{J-r}. \tag{14}$$

We can approximate (14) by a Poisson distribution with mean $\lambda_k = JP(k_i = k) = J\,{}_{J-1}C_k q^k (1-q)^{J-1-k}$ so that $P_r(X_k = r) = e^{-\lambda_k}\frac{\lambda_k^r}{r!}$. The random variable X_k with a Poisson distribution clusters around its mean λ_k and we can approximate X_k by λ_k.

Next, we derive the degree distribution $P_k(k)$: The fraction of nodes that has k' edges is given by $P_k(k = k') = X_{k'}/J \approx \lambda_{k'}/J$, i.e.,

$$P_k(k = k') = {}_{J-1}C_{k'} q^{k'} (1-q)^{J-1-k'}. \tag{15}$$

If J is sufficiently large, we can replace (15) by

$$P_k(k) = e^{-qJ}\frac{(qJ)^k}{k!}.$$

We conclude that the degree distribution of a random network follows a Poisson distribution with mean qJ.

A.2 Proof of Proposition 3.1

Proof. Note that $s(a_i)$ is monotone decreasing in a_i: Suppose $J > 2$ and the network is neither complete or completely isolated. We have

$$s'(a_i) := \frac{ds(a_i)}{da_i} = -(\log\tau)s(a_i)S^{-1}(S - s_i) \leq 0$$

with equality iff $\tau = 1$. The second derivative is, therefore,

$$\frac{d^2 s(a_i)}{da_i^2} = [s'(a_i)]^2 \frac{S - 2s_i}{s_i(S - s_i)} \geq 0,$$

with equality iff $\tau = 1$. Hence $s(a_i)$ is strictly convex in a_i.

To show that $s(a_i)$ bulges as τ grows, first note $\frac{\partial s(a_i)}{\partial\tau} = -\tau^{-1}s(a_i)(a_i - AB^{-1})$, where $A := \sum_j a_j\tau^{-a_j}$ and $B := \sum_j \tau^{-a_j}$. Then

$$\frac{dD(\tau)}{d\tau} = \frac{1}{2\tau}\left\{[s(a_M) - s(a_H)](a_H - AB^{-1}) + [s(a_M) - s(a_L)](a_L - AB^{-1})\right\},$$

[10] The random variable k_i is almost independent from k_j: an edge from a node i to j shares only $1/k_i$ of all the edges from i.

where $a_M := (a_H + a_L)/2$. The first term in the curly braces is positive because $s(a_M) - s(a_H) > 0$ and $a_H - AB^{-1} = B^{-1}\sum_{j \neq H}(a_H - a_j)\tau^{-a_j} > 0$. Likewise, the second term is positive because $s(a_M) - s(a_L) < 0$ and $a_L - AB^{-1} < 0$. Therefore $\frac{dD(\tau)}{d\tau} > 0$, which establishes the claim. □

A.3 Probability-Generating Function

Newman et al [NSW01] took a generating-function approach to derive the degree distribution. We recap their method with the aim of deriving a branching factor.

A probability-generating function is a handy device that enables us to compute a probability or moment of interest by taking derivatives of desired orders. Suppose that the probability that a randomly chosen node has a degree k is $g(k)$. Define the probability-generating function by

$$\psi_0(x) := \sum_{k=0}^{\infty} g(k)x^k.$$

With this function, the probability $g(k)$ is simply given by

$$g(k) = \frac{1}{k!}\psi_0^{(k)}(x = 0).$$

The first moment is also readily available by taking the first derivative of $\psi_0(x)$ and evaluating it at $x = 1$, i.e.,

$$\sum_k kg(k) = \psi_0'(1).$$

To derive the adjacent neighbors' degree distribution, first, pick a random edge and walk towards one end. The probability that the end node has a degree k, $g_1(k)$, is proportional to $(k+1)g(k+1)$ (note that the edge that you just walked on is excluded), that is,

$$g_1(k) = \frac{(k+1)g(k+1)}{\sum_m (m+1)g(m+1)}.$$

A gracious coincidence is that $g_1(k)$ happens to be the same as

$$\frac{1}{k!}\psi_1^{(k)}(0),$$

where $\psi_1(x) := \dfrac{\psi_0'(x)}{\psi_0'(1)}$.

A.4 Branching Factor for a BA Network

Suppose that k follows the Pareto distribution with the Pareto exponent α. Recall from Appendix A.3 that the probability that the nearest neighbor has a degree k (denote this by $g_{nn}(k)$) is proportional to the number of edges whose other end has a degree k, i.e., $g_{nn}(k) \propto kg(k)$ (following the approximation by

Pastor-Satorras et al [PSVV01]). Average degree of the nearest neighbors is given by

$$\kappa = \int k \frac{kg(k)}{\bar{k}} dk = \frac{E[k^2]}{E[k]} = \frac{V[k] + E[k]^2}{E[k]},$$

where $E[\cdot]$ and $V[\cdot]$ denote the mean and variance. In particular, for the Pareto distribution,

$$\kappa_{BA} = \frac{k_m}{\alpha - 1} \left(\frac{1}{\alpha - 2} + \alpha \right) \tag{16}$$

(Note $E[k] = \frac{\alpha k_m}{\alpha - 1}$ for $\alpha > 1$ and $V[k] = \frac{k_m^2 \alpha}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$).

As in the case of ER model, we could use the probability-generating function [NSW01]. However that is not useful for our purpose. To derive the probability-generating function, we use the zeta distribution $\tilde{g}_{BA}(k) = k^{-\alpha}/\zeta(\alpha)$, where $\zeta(\cdot)$ is the Riemann zeta function $\zeta(\alpha) = \sum_{x=1}^{\infty} \frac{1}{x^\alpha}$ (this normalizes the pdf so that it sums up to 1). The probability-generating function of the degree of the immediate neighbors is given by $\psi_1(x) = \frac{\zeta(x\alpha - 1)}{x\zeta(\alpha)}$. Then the branching factor is $\kappa_{BA} = \psi_1'(x = 1)/\psi_1(x = 1) = \frac{\zeta(\alpha-2) - \zeta(\alpha-1)}{\zeta(\alpha)} / \frac{\zeta(\alpha-1)}{\zeta(\alpha)}$. Hence, $B_{BA} = \{\log[\zeta(\alpha - 2) - \zeta(\alpha - 1)] - \log \zeta(\alpha - 1)\}^{-1}$.

We replace the branching factor κ_{BA} above for empirical implementation by (16). While estimation using (13) is technically accurate, it is problematic for two reasons. First, B_{BA} does not have a closed form, which makes its estimation computationally expensive. Second, B_{BA} does not have a real value for $\alpha < 3$, i.e., we have to give up the region of interest $2 \leq \alpha \leq 3$. The actual degree distribution (or any variables that are said to follow the power law in general) are known to have an exponential cutoff. The upper end of the empirical distribution cannot be modeled as the power law. Consider, for example, the city-size distribution. The probability that New York City has size strictly larger than the US population is zero. Or, for a degree distribution, the probability that the largest degree is larger than J should be, and actually is, zero as well (for more on this, see Newman [New05]). The degree distribution of nearest neighbors with exponential cutoff is available [NSW01], but the lack of the closed form will only be aggravated. As an alternative, we approximate the branching factor with the Pareto distribution as in (16). In this formula, α still cannot fall below 2, but the region of search does expand down to 2, which covers most of the real degree distributions. Moreover, the formula has a closed form.

References

[AB02] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 2002.

[BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 1999.

[BB03] Albert-László Barabási and Eric Bonabeau. Scale-free networks. *Scientific American*, 2003.

[BW11] Marcus Berliant and Hiroki Watanabe. Explaining the size distribution of cities: X-treme economies. *Working Paper*, 2011.

[Dur05] Gilles Duranton. Some foundations for Zipf's law: Product proliferation and local spillovers. *Regional Science and Urban Economics*, 2005.

[Eec04] Jan Eeckhout. Gibrat's law for (all) cities. *American Economic Review*, 2004.

[ER59] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae*, 1959.

[ER61] Paul Erdős and Alfréd Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, 1961.

[HSF+05] Janusz A. Hołyst, Julian Sienkiewicz, Agata Fronczak, Piotr Fronczak, and Krzysztof Suchecki. Universal scaling of distances in complex networks. *Physical Review E*, 2005.

[New05] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 2005.

[NSW01] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 2001.

[PSVV01] Romualdo Pastor-Satorras, Alexei Vázquez, and Alessandro Vespignani. Dynamical and correlation properties of the internet. *Physical Review Letters*, 2001.

[Soo05] Kwok Tong Soo. Zipf's law for cities: a cross-country investigation. *Regional Science and Urban Economics*, 2005.

[YT69] William E. Young and Robert H. Trent. Geometric mean approximations of individual security and portfolio performance. *The Journal of Financial and Quantitative Analysis*, 1969.