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

Marcus Berliant† and Hiroki Watanabe‡

October 30, 2008

Abstract

Zipf’s law is one of the most well-known empirical regularities of the city-size distribution and explaining it has long been the Holy Grail of urban economics. There is extensive research on the subject, where each city is treated equally in terms of transactions with other cities. Recent developments in network theory facilitate the examination of asymmetric communication patterns among the cities. Under the scale-free network framework, the chance of observing extremes becomes lower 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 the urban economics with a view to explaining the city-size distribution.

Keywords: Zipf’s law, city-size distribution, fractals, self-organizing economy, Gibrat’s law, random graph, scale-free network.

1 Introduction

1.1 Distribution of City Size and Network Theory

Any economic activity is associated with the location where it takes place. Standard economic theory is sufficient to analyze these activities as long as they take place at the same location, which is not the case when it comes to the location choice of consumers or firms. The distribution of city size is not degenerate in one location nor is it uniform. Your local economy is not independent of your neighboring economies unless your local economy is autarkic. A transaction pattern between any two cities alters the way cities are populated and modifies the overall city-size distribution.

*This project received a grant from the Center for Research in Economics and Strategy (CRES), in the Olin Business School, Washington University in St. Louis. The second author thanks Professors Sukkoo Kim, 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.
Interaction between individual cities has not caught much attention so far. Duranton [8], for example, provides a precise description of the way commodities are produced and distributed, while transportation cost is not assumed to depend on the size of a city. In contrast, transaction and/or communication between hub cities is much easier than that of cities on peripheries.

The recent seminal work by Barabási and Albert [3] has revitalized the network theory. Classical network theory cannot explain the emergence of a cluster or hub in a network, which we usually observe in most networks in reality. Each node is linked with an equal probability and lacks distinctiveness, for the number of links for any node does not matter in forming a network. Their model (BA model) adds dynamic feature and preferential attachment to the classical model so that the nodes are not identical anymore. 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 sciences, and produced a good description of the organization and development of networks. All applied 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 extremely large numbers of links rather easily with these networks than with classical random graph theory.

This is true of the city-size distribution as well. Suppose there are two United States: One with New York City as the largest city and the other featuring Philadelphia as the largest city. To track down the initial cause of their differences, we can trace these two worlds backwards to observe the moment before which they shared the same development pattern and after which they followed different paths. If the moment includes only a negligible and indistinguishable difference, then it is impossible to model the city-size distribution. A relocation of some minor firm may have triggered the growth of New York City, which later on would surpass Philadelphia in size. Another firm or household, which was completely indifferent between New York City and Philadelphia, may choose to move to New York City following the previous firm and the process goes on. In this way, an initial, almost undetectable difference grows exponentially just like a butterfly flap in New Delhi causes a hurricane in New York in two years. Still, development follows certain rules and exhibits fractal patterns. The fractal patterns usually lead to a distribution with a fat tail.

On the other hand, if the moment of divergence between New York City and Philadelphia includes some noticeable incidence, e.g., one with some aggregated regional shock affecting New York City and another without it, then it is easier to identify the cause. We, as economists, tend to believe that the latter is the case, but considering the complexity of city formation and migration stemming from the combination of geographical features and a variety of economic incentives behind them, the truth should be found somewhere in the middle of the two extremes.

1.2 Street Cars Do Not Fly

The U.S. has seen a number of changes in the mode of transportation used over the 20th century. At the turn of the century, we saw street cars on the streets before the introduction of affordable cars and the highly organized interstate highway system, which has been partially supplanted by emerging
low-budget airline companies later on. 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 city-size distribution usually does not investigate the structure of an underlying network in which economic activities have taken place. An organizational pattern generated and developed among cities connected by edges is not thought to be included in conventional forces (positive and negative externalities) that serve a significant role in determining city-size distribution (Starrett [17]). Apparently a network composed of interstates does not share its structure with that of airlines at all.

Figure 1: The U.S. interstate state system (left) and the U.S. airline system (right). Source: Barabási and Bonabeau [4].

These changes in available means of transportation have been always accompanied by changes in the nature of city development as well (Anas, Arnott and Small [2]). Black and Henderson [6] find that the cities near the top of the rank remain there for longer periods of time than mediocre ones. Network theory can give an explanation to the observed persistence as a result of changes in connection probability and/or switching from one regime to another. For example, the rate of completion in the interstates can be handled by the change in connection probability as long as the order of construction is random. We can also think of regime switching including changes from random graph to a scale-free network. Random graphs alone can describe a "jump" despite the smooth increase in connection probability (c.f. Figure 3). It will be interesting to see if the city-size distribution exhibits some discontinuity every time a new mode of transportation or communication is introduced.

1.3 Denver Is Not 7 Feet and 11 Inches Tall

The city-size distribution is known to have a fat-tail distribution. To visualize the nature of the city-size distribution, let us conduct a thought experiment and see how it is compared to human height, which is known to follow the normal distribution. Suppose that the city-size distribution follows the normal distribution as well. The cumulative distribution function (CDF) at Denver, Colorado (population 554,636, 24th in rank) is .99. If we place Denver among US males between the ages of 20 to 29, he will be 7 ft. 11 in. tall, while the mean is 5 ft. 10 in. (Ogden et al. [15]). We do not see such a tall man so often.
and still we have 23 other men who are much taller than Mr. Denver.  

Once we abstain from the Gaussian distribution that economists hold dear and choose any long-tail distribution, a lognormal distribution for example, then the U.S. male equivalent of Denver is 6 ft. 8 in.  

Admittedly it is still on the higher side but not outrageously so.

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 are high.

![Figure 2: Frequency plot of the city-size distribution. Data source: U.S. Census 2000.](image)

### 1.4 Inclusive Data Set

We use the data set that includes more than 25,000 places. The conventional unit of measure of city size is the metropolitan area (MA). There are 276 MAs in the United States and 80% of the U.S. population resides there. Our data set covers the entire population of the U.S. The aim here is not just to find out why the top 276 MAs are larger than others and follow a certain distribution; we also need to explain why the other cities did not make their way to the top of the list.

Let us consider the following: Some counties in the U.S. are regarded as the oracles of presidential elections. Vigo County in Indiana, for example, has been picking the winner since 1960. It sounds as if this county is endowed with a certain vision. While it is newsworthy, the story is subject to survivorship bias. We tend to forget the size of the sample when we think of the survivors or counties that have remained predictors over decades in this

---

1. CDF for a city larger than Denver city is much too close to one to find its U.S. male counterpart.
2. A “New Yorker” will be 7 ft. 1 in. tall.
3. We thank Professor Jan Eeckhout for sharing his data.
example. There are around 3,000 counties in the U.S. and if they have chosen
t heir candidate at random for 40 years, 1.5 counties will predict the winners
correctly just by mere luck.

We can also consider the following: Survivorship bias, as discussed in [18],
needs to be taken into account. Suppose we flip a coin a million times. Along
the way we might observe 13 heads in a row. We cannot pick this particular
segment of observation and discuss the nature of the coin. If we just toss
the coin for 13 times and obtained 13 heads, then it is likely that the coin is
biased, but if we tossed it for a million times, it would not be surprising to see
13 heads in a row. (There is still a tails). Likewise, if we pick the top 276 cities
to discuss the nature of city-size distribution, we will miss a complete picture
for there are lots of other cities, that could have won, in the pecking order
but somehow did not survive and do not appear on the list of the largest 276
cities.

There are two major types of networks: random-graph and scale-free net-
works. We examine these networks one by one to see how they translate to
the city-size distribution in the following sections. Section 4 concludes.

2 Random Graph

In this section, we will review the classic random graph theory and propose
an economic model that can translate the degree distribution to the city-size
distribution.

2.1 Random Graph

Let us begin the section by reviewing the random graph theory initially en-
geineered by Erdős and Rényi [10], to be applied to economic geography later
on. A network is a pair \( G = \{P, E\} \). A set \( P \) is a set of nodes or vertices
connected by an edge or link specified by a set \( E \). For example if you can
connect New York City to St. Louis without an intermediate node (a mode of
connection can be anything at this point), then we say \( \{P_{NYC}, P_{STL}\} \in E \).

Our concern is not an exact description of \( G \) itself, i.e., whether a partic-
ular node is connected to some other particular node, but a property of the
degree distribution of networks under certain parametarizations. The number
of edges that a particular node builds is called its degree. Let us examine a
network \( G_{n,N} \) with \( N \) nodes and \( n \) edges in total. Derivation of the degree
distribution is twofold. First, we derive a distribution of \( X_k \), the number of
nodes with degree \( k \).

\[
\text{Derivation of degree distribution:}
\]

To find the probability of a node \( i \) wired to \( k \) other nodes is given by
\[
P_i(k_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k},
\]
where \( \binom{a}{b} = \frac{a!}{b!(a-b)!} \). The number of nodes with \( k \) edges follows:
\[
P_r(X_k = r) = N C_r p_r(k_i = k) \left( 1 - P_i(k_i = k) \right)^{N-r}. \tag{2}
\]

Note that \( \sum_{k=1}^{N-1} X_k = N \). A number of edges that a node builds does not exceed \( N - 1 \).

\[\]

The random variable \( k \) is almost independent from \( k_i \); an edge from a node \( i \) to \( j \) shares only
\[\]

1.5 counties will predict the winners correctly just by mere luck.

2 Random Graph

In this section, we will review the classic random graph theory and propose
an economic model that can translate the degree distribution to the city-size
distribution.

2.1 Random Graph

Let us begin the section by reviewing the random graph theory initially en-
geineered by Erdős and Rényi [10], to be applied to economic geography later
on. A network is a pair \( G = \{P, E\} \). A set \( P \) is a set of nodes or vertices
connected by an edge or link specified by a set \( E \). For example if you can
connect New York City to St. Louis without an intermediate node (a mode of
connection can be anything at this point), then we say \( \{P_{NYC}, P_{STL}\} \in E \).

Our concern is not an exact description of \( G \) itself, i.e., whether a partic-
ular node is connected to some other particular node, but a property of the
degree distribution of networks under certain parametarizations. The number
of edges that a particular node builds is called its degree. Let us examine a
network \( G_{n,N} \) with \( N \) nodes and \( n \) edges in total. Derivation of the degree
distribution is twofold. First, we derive a distribution of \( X_k \), the number of
nodes with degree \( k \).

\[P_i(k_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k},\]

where \( \binom{a}{b} = \frac{a!}{b!(a-b)!} \). The number of nodes with \( k \) edges follows:

\[P_r(X_k = r) = N C_r p_r(k_i = k) \left( 1 - P_i(k_i = k) \right)^{N-r}. \tag{2}\]

Note that \( \sum_{k=1}^{N-1} X_k = N \). A number of edges that a node builds does not exceed \( N - 1 \).

The random variable \( k \) is almost independent from \( k_i \); an edge from a node \( i \) to \( j \) shares only
\[1/k_i \text{ of all the edges from } i.\]
We can approximate (2) by a Poisson distribution with mean $\lambda_k = NP(k_i = k) = N N_{-1}^{-1} C_k p^k (1 - p)^{N-1-k}$ so that $P_r(X_k = r) = e^{-\lambda} \frac{\lambda^r}{r!}$. The random variable $X_k$ with a Poisson distribution clusters around its mean $\lambda_k$ and we can approximate $X_k$ by $\lambda_k$.

Next, we derive the degree distribution $P_k(k)$ of a network $G_{n,N}$ as follows: A fraction of nodes that has $k'$ edges is given by $P_k(k = k') = X_k / N \approx \lambda k' / N$, i.e.,

$$P_k(k = k') = N_{-1}^{-1} C_k p^k (1 - p)^{N-1-k'}.$$ (3)

In our interpretation where a node is a city, $N$ is large enough ($N = 25,358$). So we may replace (3) by

$$P_k(k) = e^{-pN} \frac{(pN)^k}{k!}.$$ (4)

We conclude that the degree distribution of a random network follows a Poisson distribution with mean $pN$. Figure 3 shows two examples of the above distribution.

**Figure 3:** Simulated random graphs at $N = 5,000$ with $p = .001$ and $p = .5$. Random graphs leave a substantial gap once $p$ goes above $N^{-1}$.

### 2.2 Generative Model 1: Poisson to Lognormal

Now that we have shown that the degree distribution follows the Poisson distribution, we will present a condition that economic models based on random network should satisfy. Our target distribution is either lognormal or generalized extreme value distribution. Both of the distributions give a coherent explanation of the empirical distribution of city sizes (Berliant and Watanabe [5]).

---

6We will not use the Pareto distribution, which is often quoted in traditional research, for we do not truncate the data as discussed in section 1.4. Eeckhout [9] establishes the relationship between the Pareto distributions and the lognormal distribution.
We do not expect to use the degree distribution itself (4) to describe the city-size distribution. The Poisson distribution is not fat-tailed. As we discussed earlier, city-size distributions are not exponentially bounded (e.g., Pareto or lognormal distributions)\(^7\) while the Poisson distribution is. Under the Poisson distribution, observing cities like New York City or Los Angeles is extremely unlikely. An economic model will relate the degree distribution to the city-size distribution. The table below compares degree distribution and city-size distribution (\(f(x)\) denotes probability density function (pdf) of city-size distribution).

<table>
<thead>
<tr>
<th>(k)</th>
<th>Number of edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NP_k(k))</td>
<td>Number of nodes with (k) edges</td>
</tr>
</tbody>
</table>

\[ \text{Economic model: } x = x(k) \]

<table>
<thead>
<tr>
<th>(x)</th>
<th>Number of people</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NFf(x))</td>
<td>Number of cities with population (x)</td>
</tr>
</tbody>
</table>

We propose the following:

**Proposition 2.1.** If an economic model based on random networks predicts \(x = \log(k)\) in its reduced form, where \(x\) denotes city size, it will generate a lognormal city-size distribution.

**Proof.** For a large \(\lambda\), a normal distribution with mean \(\lambda\) and variance \(\lambda\) approximates the Poisson distribution by the central limit theorem. Since the degree distribution follows Poisson distribution, \(\log(k)\) follows a lognormal distribution.

Further research aimed at relating network to economic geography is called for. At this point, we will conclude this section by proposing a sketch of what it will look like. The bottom line is that such a model should be equipped with a mechanism that “fattens” the tail. First we assign a unit of population to each edge. Then people decide at which end of the edge they live according to the number of edges each node has. For example, if you are on the edge between New York City and Chicago, you will move to New York City, which has more nodes than Chicago. Likewise, if you are on the edge connecting Chicago and St. Louis, you will move to Chicago. If it fattens the tail too much, then we can divide population on each edge in proportion to both end in proportion to the number of edges each node has. See Figure 4 for a schematic representation. Either way, the resulting distribution will exhibit dependency on the number of edges from a particular node. The original random network theory misses this feature: two vertices are connected at random regardless of the number of edges each vertex has. An economic model supplies this missing component (preferential attachment) and renders the random-network degree distribution as scale-free city-size distribution. Further discussion will follow in Section 3.1.

### 2.3 Proxy for Connection Probability

In section 2.2 we estimated the connection probability from the data set. To complete the analysis we need to confirm that the estimated connection prob-\(^7\)A distribution is fat-tailed if its survival function \(F(x)\) satisfies \(\lim_{x \to \infty} e^{\lambda x} F(x) = \infty\) for all \(\lambda > 0\).}
ability falls in the range that makes economic sense. Connection probability itself is not an economic indicator. It may represent a number of economic variables: If we focus on the exchange of ideas as a key factor, then a volume of correspondence between places, including phone line usage, or mail delivery, is a place to look for data. An exchange of commodities may explain the city-size distribution as well. In this case the connection probability is estimated by a weighted average of networks of airlines, interstates or railroads. Different industries use different modes of transportation.

3 Scale-Free Networks

3.1 Gibrat’s Law

As we have seen in the previous section, while random network theory provides a pedagogical framework to analyze network structures, it does not lead to a heavy-tailed distribution. Most of the phenomena observed in reality feature scale-free nature: World Wide Web, Hollywood actors, neural networks, human sexual contacts and the list goes on. There are several attempts made to generate a power law out of networks (Kumar et al. [14], Dangalchev [7]). We will take Barabási and Albert [3]’s model (BA model) as an example.

The model has two features that distinguish it from random networks. The first one is that the number of nodes is no longer fixed: a new node with m edges is constantly introduced to the network. This assumption can be compared to the emergence of edge cities as documented by Garreau [11]. The model is also based on the postulate that the added node prefers to establish edges to existing nodes with a large number of edges. In the example of the Web, a newly-launched web page usually includes links to web pages that are already well known. The assumption is usually termed preferential attachment or Gibrat’s law. It is safe to assume that city-size distribution shares the same property. Eeckhout [9] confirms empirical evidence and provides a
3.2 Generative Model 2: Pareto to Lognormal / GEV

The examples of network mentioned above have power-law exponents roughly between one and two, whereas that of city-size distributions usually falls around one (Rossi-Hansberg and Wright [16]). The preferential attachment employed in generating the scale-invariant degree distribution (BA model) is as follows. At any time $t$, probability of an existing node $i$ being connected to a newly-added node is proportional to its degree $k_i$, that is,

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^{m_0} k_j},$$

where $m_0$ is a initial number of nodes. With this dynamic, the predicted exponent is two. The discrepancy in scaling exponents implies that (5) is too weak to reproduce city-size distributions: The tail of the city-size distribution is thicker than what BA model predicts.

Once again, edges are not population. To bridge the gap in scaling parameters, we can adjust the model in the following ways: 1) Starting from BA model, establish a function translating degree to population as we did in Section 2.2, or 2) Intensify the degree of preferential attachment in (5) to put more weight on larger cities, or both.

Let us examine the first option. We propose the following:

**Proposition 3.1.** Suppose $r : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing or decreasing. If an economic model based on a scale-free network predicts $x = x(k)$ satisfying

$$k(x) = k[\Phi(\log x)]^{-1},$$

where $k(x) := x^{-1}(x)$, $k$ is the support of a degree distribution, $\mu$ and $\sigma$ are mean and variance of the log of population, $\Phi(\cdot)$ denotes the CDF of a normal distribution, then the model generates lognormal city-size distribution.

Similarly, if $x(k)$ satisfies

$$k(x) = k\sqrt{\pi} \left[ \int |\log F_{GEV}(x)|^{1+\xi} F_{GEV}(x) dx \right]^{-\frac{1}{2}},$$

where $\sigma$, $\xi$ and $F_{GEV}(x)$ are the paremeters and the CDF of the generalized extreme value distribution, then the model replicates the city-size distribution specified by the generalized extreme value distribution.

**Proof.** BA model predicts

$$P_k(k) = 2k^2k^{-3}.$$ If $k = k(x)$ then the pdf of city-size distribution is given by $f(x) = \frac{dk(x)}{dx} |2k^2k^{-3}$ so that

$$\frac{dk(x)}{dx} |2k^2k^{-3} = x^{-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{\log x - \mu}{\sigma} \right)^2 \right\},$$

which leads to (6). The similar argument follows for the generalized extreme value distribution.
Remark. The empirical values found in U.S. Census 2000 are $\mu = 7.28$ and $\sigma = 1.75$ for a lognormal distribution ([5]).

4 Conclusion

We pointed out how the network of cities affects the city-size distribution. The classical random graph is too weak 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. We have provided the reduced form that the desired economic model has and proposed a sketch of a mechanism that motivates the households or firms to relocate. The optimization problem including the choice of a node needs to be specified to formulate the structural form.

As an alternative, we employed BA model to replicate the city-size distribution. BA model results in a fat-tailed distribution, but the tail is still too thin to match the city-size distribution. We proposed other reduced forms based on BA model. Alternatively, we can take one step back to alter the model’s generative mechanism of adding nodes to increase the edges of a cluster.

We finish the discussion with the following remark. We argued that network structures motivate the population to form a specific distribution. It is, however, not guaranteed that we know the actual causality. It may be the other way around: The relocation of people forces their 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 development of the network structure along with the city-size distribution may reveal a clue to identify the direction of causality.

References


Scale-Free Network Explains the City-Size Distribution


