Determination of the complexity of distance weights in Mexican city systems

Igor Lugo *

Centro Regional de Investigaciones Multidisciplinarias, Universidad Nacional Autónoma de México, Av. Universidad s/n, Cto. 2, Col. Chamilpa 62210, Cuernavaca, Mor., Mexico

* Corresponding author.
E-mail address: igoriugo@correo.crim.unam.mx.

Abstract

This study tests distance weights based on the economic geography assumption of straight lines and the complex networks approach of empirical road segments in the Mexican system of cities to determine the best distance specification. We generated network graphs by using geospatial data and computed weights by measuring shortest paths, thereby characterizing their probability distributions and comparing them with spatial null models. Findings show that distributions are sufficiently different and are associated with asymmetrical beta distributions. Straight lines over- and underestimated distances compared to the empirical data, and they showed compatibility with random models. Therefore, accurate distance weights depend on the type of the network specification.

Keywords: Economics, Information science, Geography, Computational mathematics

1. Introduction

In recent years, the computational progress presented in the analysis of big data has decreased the gap among disciplines. In particular, economics, geography, and complexity use geospatial data to express interactions among locations in urban systems, but these fields apply different approaches to measure them. In the economic geography literature, distances are given by a straight line between each
pair of points that determine their potential for interaction (Anselin, 2003; Fischer and Getis, 2010). However, this formulation oversimplifies distances without considering the context and effects of the terrain on urban systems. Economists have assumed such magnitudes for simplicity in modeling because of the computational restrictions to quantify interrelations in complex transport structures (Fujita et al., 2001; Krugman, 1996). For example, distances have been made implicit in the iceberg transport costs—the price of a good decreases a fraction when shipping between two locations (Krugman, 1991; McCann, 2005). This version of costs was operationalized by weights based on a flat surface without spatial constrains across locations. Even though this approach has showed significant developments to understand and analyze spatial agglomerations, it hides the complexity of real-world systems. On the other hand, the complexity approximation—self-organized systems with increasing, multiple and changing interconnections among their elements—investigates attributes and mechanisms that can explain the structure and dynamics behind distance weights, but they cannot describe the economic principles. Empirical studies have showed singular distance arrays related to different scales and underlying processes that depend not only on the surface specification, but also on the selected transport mode (Newman et al., 2006; Sparks et al., 2010; Barthélémy, 2011; Burgoine et al., 2013; Batty, 2014). Therefore, we studied the complexity behind distance weights in a system of cities related to the economic geography and the complex network approximations by exploring geospatial graphs and identifying their statistical underlying process of connectivity. We then determined the best weights specification.

We used the notion of a system of cities to generate the weighted matrices (Berry, 1964; Pumain, 2006). This system was represented by network graphs in which each city was related to others based on abstract and physical connections—nodes and edges associated with a geographic coordinate system. We selected the road network as the transport mode to connect cities because it shows nontrivial geometric structures (Ducruet and Lugo, 2011; Lugo, 2015). The main problem is to test how similar or different distances are, relying on the economic geography assumption and the complexity of empirical data. In addition, we want to know when we can assume each of such distances. To answer these questions, we characterized the probability distribution functions of distances in those graphs and contrasted with those distributions relaying on null spatial models—random and minimum spanning tree (MST) structures (Clauset et al., 2009; Newman, 2005; Sornette, 2012). This method was exemplified by using geospatial data of road lines and urban polygons in Mexico. Therefore, to compute distance weights, we stated that the graph based on empirical data is more accurate than the straight-line interactions among cities. These interactions are imprecise, thereby creating errors in the statistical analysis. Hence, before using distances in empirical works, it is important to identify the types of spatial networks behind them and their implications. This study offers
some important insights into the spatial analysis field because it mixes the economic geography and the complexity methods to model spatial networks and to improve distance measures.

The document is divided into four sections. The first section describes the geospatial data selected for the study. The second section explains the method to quantify distances in a spatial system of cities, showing the connection between the economic geography and the complex network views. In addition, we describe the computational process to obtain, generate, and compare spatial graphs. The third section shows results based on inferential data analyses of probability distributions—the probability density function (PDF) and the cumulative distribution function (CDF)—particularly the two-side Kolmogorov–Smirnov (KS) test and its goodness-of-fit test computed by maximizing a log-likelihood function (Massey, 1951; ScipyStatsKstest, 2016). Furthermore, we used the Monte Carlo method (MCM) of integration to test differences in the straight-line and random distances (Kroese et al., 2011, 2014). The last section discusses outcomes, limitations, and extensions.

2. Materials

Representing and analyzing spatial agglomerations as urban systems require large geospatial databases and computational tools. The former is collected and analyzed by the geographic information system in which digital data are divided into vector and raster models. In this case, we used vector data provided by the National Institute of Statistics and Geography of Mexico: The layer of lines represents the road network, and the layer of urban polygons indicates cities as urban centroids. The latter was generated using the catalog of the national urban system and the urban locations (INEGI, 2014; SEDESOL et al., 2012). Finally, we projected these data to the International Terrestrial Reference Frame 1992—ITFR92.

Further, we applied computational tools relying on the Python programming language to manage and analyze the data. In particular, we used the Osgeo, Networkx, Matplotlib, and Scipy libraries to write a program to handle the computational processing (Matplotlib, 2016; Networkx, 2016; Osgeo, 2016; Scipy, 2016). In addition, these materials—data and source code—follow the Open Science Framework to contribute to the effort to increase the quality of results in sciences (Nosek et al., 2015). Our findings can be replicated directly and independently, “OSF:ShareProject.”

3. Methods

The spatial agglomeration theory describes urban systems in which agents present strong collective preferences to locate closer to each other because of increasing
returns based on positive externalities (Henderson and Thisse, 1991). This is possibly expressed into models of economic geography and complex networks. The economic geography has defined a system of cities as a network in which nodes represent cities, and edges are connections—physical or social—among them. Depending on the city attribute—qualitative or quantitative data—and its neighborhood—global or local—these connections form diverse networks—for example, ring or fully connected graphs. Furthermore, adding geospatial coordinates into nodes, we can study geospatial effects on interactions. For example, the “race-track” model sets a circular graph in which cities are located on the circumference and connected to their close neighbors (Batty, 2005; Krugman, 1996). Even though it was a constrained and an abstract representation of the urban system with local interactions, its dynamics produced important aggregate results according to the city growth distribution (Wilson and Dearden, 2011; Zhonga et al., 2014). That is, similar initial conditions—the same number of habitants per city—reproduce, in the long run, uneven population growth described by bias distributions. Therefore, both well-defined geometric structures and nongeometric configurations can produce emergent patterns.

Despite this, little progress has been made in connecting this formulation with complex networks. In particular, spatial networks—graphs with singular geographic coordinate systems—represent a good approximation for modeling urban systems. Their computational and analytic tools complement methods in the economic geography because the interactions might be modeled by using empirical evidence and compared with statistical tests. The key factor in both approximations is the connectivity because it indicates a type of interaction depending on the distance among locations. It captures the idea of hierarchical interrelations—for example, close locations have lower transportation costs than distant ones. Therefore, distances have been operationalized by the spatial weights matrix or the spatial interaction matrix (Rodrigue, 2013). This formulation is the basis for computing distances among locations, thereby defining the connectivity in spatial structures.

### 3.1. Instrumentation

Understanding interactions in spatial systems is the key to quantify and analyze distance weights. The economic geography has used the classic view in which a system is considered as a simplified version of a machine structure based on homogeneous parts, linear relationships among them, and unique basins of attractions. In this approach, distances are measured by a straight line between two points, even though there are other options to estimate it based on arrays—for example, Manhattan and Chebyshev formulations (Fujita et al., 2001; Celikoğlu and Silgu, 2016; Silgu and Çelikoğlu, 2014). Following the spatial analysis notation
(Smith and Longley, 2015), we defined the simplest interaction measure, \( W_{ij} \), from location \( i \) to \( j \), with coordinates \( (x_i, y_i) \) and \( (x_j, y_j) \) respectively, as a function of the Euclidean distance:

\[
W_{ij} = f(S_{ij}) = f(d_{ij})
\]

(1)

\[
d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}
\]

(2)

where \( S_{ij} \) is the attribute of physical proximity; it is also known as a transport friction in spatial interaction models (Rodrigue, 2013). Expression (2) shows explicitly the function in distance units, \( d_{ij} \). An important example in economics is the computation of transport cost, \( TC_{ij} = P_i W_{ij} \), where \( TC_{ij} \) is the transport cost in monetary units, \( P_i \) is the price to move a good produced in location \( i \), and \( W_{ij} \) is the interaction measured in distance units (Fujita et al., 2001). The problem with this formulation is the assumption that distances are measured on a flat surface without geospatial constrains between two points. In addition, it did not state that those distances depend on transport modes—air, land, or maritime—to connect locations. Spatial econometrics is a good illustration for using this approach (Anselin and Rey, 2014). Therefore, distances have been predefined by a linear proximity between two points without geospatial restrictions and types of mobility.

Furthermore, in spatial networks, these distances are unknown because they depend on the type of transport modes and their spatial patterns. This approach regards the system as a living entity with heterogeneous elements, nonlinear dynamics, and multiple or far-from-equilibrium characteristics. Then, the goal is to obtain a distance array by computing algorithms based on empirical evidence. Distance weights are obtained by processing and analyzing geospatial data. A good illustration is the case of the road system due to its nontrivial network specification (Ducruet and Lugo, 2013). According to this idea, we computed distances in a spatial system of cities as follows.

\[
W_{\nu\nu'} = \sum_{\nu=1}^{n-1} d_{\nu,\nu+1}
\]

(3)

where \( W_{\nu\nu'} \) is the sum of road segments in the shortest path from node or city \( \nu \) to \( \nu' \), and \( d_{\nu,\nu+1} \) is the magnitude, computed by formula (2), of the distance per line segment or edge. In this case, we used the Dijkstra’s algorithm for computing the shortest path based on the igraph Python library (PythonIgraphSP, 2016). Therefore, this measure considered each road segment between the source and target locations instead of assuming a straight line. The next section describes the computational process to manage geospatial data and to generate spatial graphs for calculating shortest paths.
3.2. Calculation

The economic geography and the network science increasingly need computational tools and models to manage and analyze big data. We used computational-intensive tasks to produce different types of distance weights relying on empirical and theoretical graphs that describe a system of cities connected by physical infrastructures. These graphs represent spatial, weighted, and bidirectional networks relying on the road topology and city attributes. Nodes represent two types of data: one is associated with road lines intersections and dead-end points, and the other relies on urban polygon centroids. Edges indicate sections of physical road infrastructures, weighted by their Euclidean distances.

To form the empirical graph, we used the layer of road lines as a baseline and the urban polygon data to add city node attributes on the road geometry. We translated this data to a planar graph adding coordinates to nodes and the distance into edge attributes. In particular, we wrote a function that takes as inputs shapefiles of lines and points and returns a spatial graph, “OSF:ShareProject.” We used the Geospatial Data Abstraction Library (GDAL), provided by Osgeo, and the Networkx library.

Next, we generated the random and MST models. First, we created random points within a spatial object inside the layer of administrative areas of Mexico. Second, we created the triangulation layer, selecting a Delaunay triangulation algorithm provided by the scipy library (ScipySpatial, 2016). Third, we worked with the QGIS software to clean this triangulation based on geographic objects, deleting lines that intersect water polygons—ocean limitations (QGIS, 2016). Fourth, we associated city points—centroids—with the closest triangulation points. Fifth, using a different distribution of random points, we repeated this process and generated the MST graph. We applied the MST algorithm using the Networkx library (NetworkxMst, 2016). Finally, we applied the function, described above, to translate these data to spatial graphs.

With respect to the economic geography assumption, we formed a graph with straight-line connections. We used the layer of city points to link each city. This connection can cross each line in a flat surface. Finally, to compute distances in these graphs, we used the expression (3) and analyzed them by the two-side KS test and its goodness-of-fit test (PythonIgraphSP, 2016; ScipyStats, 2016).

4. Results

For distance weights in a system of cities, different complex graphs have to be considered. We compared distances in the shortest paths that connect cities using the straight-line assumption and the road segment connections, and we explored their attributes related to random and MST structures.
**Figure 1.** Distance approximations based on graph connectivity. **Figure 1** (a) is the straight-line approach to compute distances in a system of cities. Its number of nodes and edges are 383 and 73,153, respectively. Colors in this graph were based on a sequential colormaps (MatplotlibColor, 2016). **Figure 1** (b) is the road segment connectivity based on the empirical data. The number of nodes and edges are 537,740 and 540,220, respectively.

The first result illustrates the visual differences between both graphs. **Figure 1** shows the two spatial networks with dissimilar distance approximations, thereby network topology. The straight-line distances show a complete graph where it does not consider geospatial constrains and transportation mode. On the other hand, the road-segment distances connect cities based on the national road infrastructure. Both graphs represent a system of cities with unlike connectivity, even though they have the same city location.

In addition, we showed the statistical difference in distance measures between these graphs. **Figure 2** shows the two CDFs that are best described by beta distributions even though their KS goodness-of-fit tests were inconclusive (Table 1). The two-side KS test rejected the hypothesis that the two samples are coming from the
Figure 2. Difference between two CDFs of shortest paths among cities in a spatial system. We show distances fewer than 2,000 km. Three-hundred and eighty-three cities are in the system; the number of nodes and edges in the straight-line graph are 356 and 73,153, respectively, and in the empirical one are 537,740 and 540,220, respectively. The solid line represents the economic geography assumption, and the dashed line is related to empirical data of the road infrastructure. By using the two-side KS test, we rejected the hypothesis that the two samples are the same (D = 0.145, p-value = 0.0). Compared with different continuous distributions—normal, power law, exponential, gamma, and beta—we found that the KS goodness-of-fit tests resulted in asymmetric beta distributions, even though they are inconclusive (Table 1).

Table 1. Best-fit, estimated parameters, and moments of the straight-line and empirical beta distributions.

| Statistics | Straight line | Empirical |
|------------|---------------|-----------|
| D          | 0.006         | 0.012     |
| p-value    | 0.006         | 1.955 × 10^{-10} |
| shape1     | 1.719         | 1.902     |
| shape2     | 10.884        | 215 582.532 |
| location   | 4.657         | 3.477     |
| scale      | 5 796.322     | 123 256 056.121 |
| mean       | 795.450       | 1 090.923 |
| var        | 290 973.219   | 621 716.601 |
| skew       | 1.070         | 1.450     |
| kurt       | 1.222         | 3.154     |

same distribution (i.e., similar shape parameters). Computing the difference between the total distances in both systems, we found that the straight-line connectivity underestimates measures in the empirical one by 27%. However, if we count the number of shortest paths less than 1,000 km, the straight-line connectivity overestimated the empirical data (i.e., 70% and 55% of shortest paths, respectively). These results are best described by their estimated PDFs (Figure 3). This figure provides the distance likelihood to take on a particular range of values (i.e., the probability for finding short distances based on the straight-line paths is higher...
Interestingly, play the distributions behaviors of the PDFs, which may under- or overestimate the random weights because they omit geospatial information of road transportation, while the empirical distribution is more accurate than the empirical, while large distances display an inverse behavior. Therefore, it was evident that these distributions were dissimilar and biased to larger values. The underlying mechanism that generates them was not additive as in the Gaussian process. These results suggest that the economic geography assumption distorts weights because they omit geospatial information of road transportation, while the empirical data quantify them accurately.

Comparing these CDFs with their random and MST data, we can see similar behaviors indicating high probabilities for finding paths with low distances and low probabilities for observing paths with extremely high distances. Figure 4 shows the four CDFs, in which the empirical and the random structures produce similar distributions because of their first moments and fitting results. Both were best described by beta distributions with close shape parameters, even though their best-fit tests were inconclusive (Table 2). However, we observed an important difference between them in Figure 5. The PDF of the random is similar to the straight lines (i.e., over and underestimate distances compared with the empirical data). On the basis of the MCM, we found that 94% of the simulated random distances fall below the PDF of the straight-line beta distribution. This suggests that the straight line can play the role of a random null model. Furthermore, the straight-line and the MTS configurations show almost the same CDF results. Both are best described by beta distributions with near shape parameters, even though their tests were inconclusive. Interestingly, Figure 6 shows contrasting PDFs between the straight-line and the MST distance. The latter under- and overestimated distances were comparable to the empirical evidence. This is caused by the different number of nodes and edges.
Figure 4. CDFs of shortest paths among cities in a spatial system and null models. The figure shows distances fewer than 2,000 km. The number of nodes and edges in the random graph are 536,512 and 1,605,094, respectively, and in the MST graph are 536,475 and 536,474, respectively. The random and MST values are best described by a beta distribution (Table 2).

Table 2. Best-fit, estimated parameters, and moments of the random and MST beta distributions.

| Statistics | Random | MST |
|------------|--------|-----|
| D          | 0.006  | 0.016 |
| p-value    | 0.004  | $1.212 \times 10^{-18}$ |
| shape1     | 1.759  | 1.510 |
| shape2     | 6,040  | 183,080 |
| location   | 3.571  | 4.951 |
| scale      | 3,028  | 202,626 |
| mean       | 885,518 | 4,208,668 |
| var        | 442,157 | 860,748 |
| skew       | 1.507  | 0.861 |
| kurt       | 3.410  | 0.397 |

in both networks when minimizing the travel distances. Therefore, distance weights based on these data indicate opposed likelihoods of weights, thereby exemplifying extreme cases to model a system of cities.

Taken together, these results imply important differences in the straight-line and the road segment graphs. They showed different and asymmetric distributions. Distance weights were based on two dissimilar networks—topological and statistical attributes—in which the empirical displayed accurate measures and the straight line showed higher and lower values. According to Alvarez-Martínez et al. (2011), the fitting parameters of the beta distribution suggest scale invariance and order-disorder properties; therefore, the shortest paths among close cities confirm the first property, while larger paths show complex structures. The straight-line distances then exhibit a low level of complexity compared with the empirical data, and they are closely
Figure 5. Estimated PDF of the straight-line, empirical, and random data. The PDFs of the straight-line and the random distances over- and underestimate the empirical data. On the basis of the MCM, we simulated 10,000 random variables using the estimated parameters of the random data and compared with the PDF of the straight-line.

Figure 6. Estimated PDF of the straight line, empirical, and MST. Compared with the straight-line and the empirical distances, the MST data under- and overestimates weights. The PDF of the MST shows a completely different shape, suggesting an extreme spatial network behind it.

related to random spatial variations. Therefore, the road network data are important to estimate distance weights in a system of cities connected by transport modes.

5. Discussion

In this study, distance weights associated with the economic geography assumption and the geospatial data were found to be statistically different. They were best described by asymmetric beta distributions showing biases to extreme distance
values, thereby affecting weight measures. This indicated that the economic assumption overstates the power of simplicity while computing distances. These measures were inaccurate, and their distribution was similar to that of the random data. Therefore, the complexity of distance weights depends on the type of connectivity among cities, in which the empirical data showed the more significant key attributes.

In addition, considering the current geographic information system development in geospatial materials and computational technology, we still cannot assume distance weights in this kind of systems. They have to be explicitly based on transport modes to improve interdisciplinary methods and techniques. In particular, before specifying, estimating, and testing causal relationships among variables in a system of cities, scientists have to define the scale, identify the spatial network, and recognize statistical characteristics, thereby decreasing errors in spatial econometric models—for example, incorrect model specifications and biased estimators. Furthermore, in the network science, scientists have to consider the economic theory to support their results and improve its data analysis. The method in economics guides where and what to look for when studying big data. For example, to examine the economic impact of the spatial friction in a large-scale system of cities, we need to select the framework—the urban and regional economics or the economic geography—and its possible expression associated with it—transport costs defined by distances and prices. Therefore, the increasing empirical data and their computational analysis can favor a connection between empirical and theoretical approaches.

This study’s major limitation was the algorithm design to generate and analyze the null models. It is not trivial to generate large-scale spatial networks and to compute their shortest paths because of the computational time. An interdisciplinary approach is fundamental to code efficient algorithms for processing and testing big spatial data. In addition, considerably more work will need to be done to determine the practical implications for using geospatial data in the economic geography and the complex networks. In particular, questions that further research should ask seek to answer what are the spatial weights in diverse scales of a system of cities connecting by the rail, air, and maritime modes, and what are their underlying network structures and dynamic mechanisms.

Finally, we highly recommend using geospatial data and developing computational tools to measure distance weights instead of assuming them as straight lines. A definite progress is needed for understanding urban systems as spatial networks because they can shed light on their underlying processes to explain ancient conditions and complex dynamics.
Declarations

Author contribution statement

Igor Lugo: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Data associated with this study has been deposited at the Open Science Framework and is available at https://osf.io/2qgj4/?view_only=cdb6e74a3d8e4cb79c54d93b82f76173 or http://dx.doi.org/10.17605/OSF.IO/2QGJ4.

Acknowledgements

I would like to thank UNAM for the supercomputing resources and services provided by DGTIC (project number SC15-1-S-46).

References

Alvareza-Martínez, R., Martínez-Mekler, G., Cocho, G., 2011. Order-disorder transition in conflicting dynamics leading to rank-frequency generalized beta distributions. Physica A 390, 120–130.

Anselin, L., 2003. Spatial econometrics. In: Baltagi, B.H. (Ed.), A Companion to Theoretical Econometrics. Blackwell Publishing Ltd. Chapter Fourteen.

Anselin, L., Rey, S.J., 2014. Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySal. GeoDa Press LLC. ISBN 978-0986342103.

Barthélemy, M., 2011. Spatial networks. Phys. Rep. 499 (1–3), 1–101.
Batty, M., 2005. Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. MIT Press. ISBN 9780262524797.

Batty, M., 2014. The New Science of Cities. MIT Press. ISBN 9780262318228.

Berry, B.J.L., 1964. Cities as systems within systems of cities. Pap. Proc. Reg. Sci. Assoc. 13, 147–163.

Burgoin, T., Alvanides, S., Lake, A.A., 2013. Creating ‘obesogenic realities’; do our methodological choices make a difference when measuring the food environment? Int. J. Health Geogr. 12, 33.

Celikoglu, H.B., Silgu, M.A., 2016. Extension of traffic flow pattern dynamic classification by a macroscopic model using multivariate clustering. Transp. Sci. 50 (3), 966–981.

Clauset, A., Shalizi, C., Newman, M., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703.

Ducruet, C., Lugo, I., 2011. Structure and dynamics of transportation networks: models, methods, and applications. In: The SAGE Handbook of Transport Studies. SAGE Publications, pp. 347–364.

Ducruet, C., Lugo, I., 2013. Cities and transport networks in shipping and logistics research. Asian J. Shipping Logist. 29 (2), 145–166.

Fischer, M.M., Getis, A., 2010. Handbook of Applied Spatial Analysis. Springer-Verlag, Berlin.

Fujita, M., Krugman, P., Venables, A., 2001. The Spatial Economy: Cities, Regions, and International Trade. MIT Press. ISBN 9780262320061.

Henderson, V., Thisse, J.F., 1991. Handbook of Regional and Urban Economics. Elsevier. ISBN 978-0-444-50967-3.

INEGI, 2014. Datos de relieve. Cem 3.0. http://www.inegi.org.mx/geo/contenidos/datosrelieve/continental/Descarga.aspx (accessed 25 February 2014).

Kroese, D.P., Taimre, T., Betev, Z.I., 2011. Handbook of Monte Carlo Methods. Wiley Series in Probability and Statistics. ISBN 978-0-470-17793-8.

Kroese, D.P., Brereton, T., Taimre, T., Botev, Z.I., 2014. Why the Monte Carlo method is so important today. Wiley Interdiscip. Rev.: Comput. Stat. 6 (6), 386–392.

Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Econ. 99 (3), 483–499.

Krugman, P., 1996. The Self-Organizing Economy, first edition. Blackwell Publishers Inc. ISBN 978-1557866981.
Lugo, I., 2015. Interplay between maritime and land modes in a system of cities. In: Ducruet, C. (Ed.), Maritime Network: Spatial Structures and Time Dynamics. Routledge, pp. 322–329.

Martínez-Mekler, G., Álvarez-Martínez, R., Beltrán del Río, M., Mansilla, R., Miramontes, P., Cocho, G., 2009. Universality of rank-ordering distributions in the arts and sciences. PLoS ONE 4 (3), e4971.

Massey, F., 1951. The Kolmogorov–Smirnov test for goodness of fit. J. Am. Stat. Assoc. 46 (253), 68–78.

Matplotlib, 2016. http://matplotlib.org/ (accessed 6 April 2016).

MatplotlibColor, 2016. Source. http://matplotlib.org/examples/color/colormaps_reference.html (accessed 15 April 2016).

McCann, P., 2005. Transport costs and new economic geography. J. Econ. Geogr. 5, 305–318.

Networkx, 2016. https://networkx.github.io/ (accessed 6 April 2016).

NetworkxMst, 2016. Source. https://networkx.github.io/documentation/networkx-1.10/_modules/networkx/algorithms/mst.html#minimum_spanning_tree (accessed 4 March 2016).

Newman, M.E.J., 2005. Power laws, Pareto distributions and Zipf’s law. Contemp. Phys. 46, 323–351.

Newman, M., Barabási, Watts, D., 2006. The Structure and Dynamics of Networks. Princeton University Press. ISBN 9780691113579.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupía, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E., Levy, Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T., 2015. Promoting an open research culture. Science 348 (6242), 1422–1425.

Osgeo, 2016. http://gdal.org/python/ (accessed 6 April 2016).

Pumain, D., 2006. Alternative explanation of hierarchical differentiation in urban systems. In: Pumain, D. (Ed.), Hierarchy in Natural and Social Sciences. Springer, pp. 169–222.

PythonIgraphSP, 2016. Source code for package igraph. http://igraph.org/python/doc/igraph-pysrc.html#Graph.shortest_paths_dijkstra (accessed 4 March 2016).
QGIS, 2016. Vector geometry. https://docs.qgis.org/2.6/en/docs/user_manual/processing_algs/qgis/vector_geometry_tools/delaunaytriangulation.html (accessed 4 March 2016).

Rodrigue, J-P., 2013. The Geography of Transport Systems, third edition. Routledge, New York. ISBN 978-1138669574.

Scipy, 2016. http://docs.scipy.org/doc/scipy/reference/ (accessed 26 September 2016).

ScipySpatial, 2016. Spatial algorithms and data structures. http://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.spatial.Delaunay.html (accessed 7 March 2016).

ScipyStats, 2016. Statistical functions. http://docs.scipy.org/doc/scipy-0.14.0/reference/stats.html (accessed 30 March 2016).

ScipyStatsKstest, 2016. Statistical functions. scipy.stats.kstest. http://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.kstest.html#scipy.stats.kstest (accessed 6 April 2016).

SEDESOL, SEGOB, CONAPO, 2012. Catálogo (2012) Sistema Urbano Nacional 2012. SEDESOL, SEGOB, and CONAPO, México. ISBN 978-607-427-167-6.

Silgu, M.A., Çelikoğlu, H.B., 2014. K-means clustering method to classify freeway traffic flow patterns. Pamukkale Univ. Muh. Bilim. Derg. 20 (6), 232–239.

Smith, M., Longley, P., 2015. Geospatial Analysis—A Comprehensive Guide, 5th edition. Winchelsea Press. ISBN 978-1906221522.

Sornette, D., 2012. Probability distributions in complex systems. In: Meyers, R.A. (Ed.), Computational Complexity. Springer, New York, pp. 2286–2300.

Sparks, A.L., Bania, N., Leete, L., 2010. Comparative approaches to measuring food access in urban areas: the case of Portland, Oregon. Urban Stud. 48, 1715–1737.

Wilson, A., Dearden, J., 2011. Tracking the evolution of the populations of a system of cities. In: Stillwell, J., Clarke, M. (Eds.), Population Dynamics and Projection Methods. In: Understanding Population Trends and Processes, vol. 4. Springer, Netherlands, pp. 209–222.

Zhonga, C., Arisonaab, S., Huangc, X., Batty, M., Schmitt, G., 2014. Detecting the dynamics of urban structure through spatial network analysis. Int. J. Geogr. Inf. Sci. 28 (11), 2178–2199.