Ann Oper Res (2018) 263:529–549
https://doi.org/10.1007/s10479-016-2358-2

Spatial associations in global household bicycle ownership

Olufolajimi Oke[1,6,7] · **Kavi Bhalla**[2] ·
David C. Love[3,4] · **Sauleh Siddiqui**[1,5,6]

Published online: 8 November 2016

Abstract The interest in bicycling and its determining factors is growing within the public health, transportation and geography communities. Ownership is one factor affecting bicycle usage, but work is still ongoing to not only quantify its effects but also to understand patterns in its growth and influence. In recent work, we mined and discovered patterns in global bicycle ownership that showed the existence of four characteristic country groups and their trends. Building on these results, we show in this paper that the ownership dataset can be modeled as a network. First, we observe mixing tendencies that indicate neighboring countries are more likely to be in the same ownership group and we map the likelihoods for cross-group mixings. Further, we define the strength of connections between countries by their proximity in ownership levels. We then determine the weighted degree assortative coefficient for the network and for each group relative to the network. We find that while the weighted degree assortativity of the ownership network is statistically insignificant, the highest and lowest ownership groups exhibit disassortative behavior with respect to the entire network. The second and

✉ Olufolajimi Oke
oke@jhu.edu; oke@mit.edu

1 Department of Civil Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

2 International Injury Research Unit, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

3 Center for a Livable Future, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

4 Department of Environmental Health Sciences, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

5 Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218, USA

6 Center for Systems Science and Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

7 Present Address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave Bldg 1–178, Cambridge, MA 02139, USA

third ranked groups, however, are strongly assortative. Our model serves as a step toward further work in studying the relationship between proximity and bicycle ownership among nations and unearthing possible patterns of influence. Considering further developments, this work can inform policy-relevant recommendations toward regional planning. This effort also contributes to expanding research in assortativity analyses, especially in weighted networks.

Keywords Bicycle ownership · Spatial associations · Networks · Assortativity

1 Introduction

As efforts are being renewed toward sustainable urban communities and safe and clean streets, a resurgence in bicycle ownership and usage would be a welcome development (Jacobsen 2013; Roseland 2012; Midgley 2011). In keeping with this pattern, policymakers are promoting schemes to further improve cycling conditions in various locales (Dales and Jones 2014; Vandenbulcke-Plasschaert 2011). Some of these include increasing the level of cycling infrastructure, investing in bicycle-sharing programs and raising awareness toward bicycle safety (Vandenbulcke-Plasschaert 2011). An important factor affecting the usage of bicycles is availability. For many, especially in places where sharing programs are inaccessible, household ownership becomes a determining factor of bicycle access and activity. In the first undertaking of its kind on a global scale, we mined household ownership data, discovering patterns in global bicycle ownership trends from 1989 to 2012 (Oke et al. 2015). While this work has laid an important foundation, more questions have arisen, some of which revolve around the nature of the relationships between these patterns and other possible contributing factors.

In this paper, we investigate spatial associations with regard to bicycle ownership among countries building on our earlier work. This is the first time this has been done in the academic literature, to the best of our knowledge. We demonstrate here that a network model of the countries in our database can provide further insight into geographic factors affecting household bicycle ownership patterns. Using graph theoretic properties, we weight the proximity of neighboring countries by a measure of separation in ownership (as given by prior cluster analyses, Oke et al. (2015)) and observe assortative behavior. Our results indicate that neighboring countries tend to share similar ownership levels. However, high ownership countries appear to have a positive influence on their neighbors, as well. Our effort also contributes to the growing body of work in assortativity, which would benefit from further investigation in the area of weighted networks (Noldus and Van Mieghem 2015).

Spatial analysis has proven to be an effective means of understanding and analyzing networks across a wide variety of applications, namely: infrastructure (Vandenbulcke-Plasschaert 2011; Páez and Scott 2005), reliability modeling (Li et al. 2014), ecology (Szmyt 2014), epidemiology (Dijkstra et al. 2013; Caprarelli and Fletcher 2014), social sciences (Darmofal 2015), among many others. To these ends, a variety of network models and approaches have been developed and the resulting associations have been useful for predictions, simulations and other aspects of decision making. Zhukov and Stewart (2013) tested various proximity criteria to infer an appropriate network structure to simulate the diffusion of political ideology. In their work, they provide valuable insights into the choice of an appropriate network model. Vieira et al. (2010) employ a modified "small world" network model (Milgram 1967) to analyze the spread of HIV over time. Ma et al. (2014) use a linear model to perform spatio-temporal investigations in scientific cooperation between cities in China. As will be discussed, methods for analyzing these patterns in networks were notably established by Newman (2002, 2003, 2004), among others.

The rest of the paper is organized as follows: In the next section (Sect. 2) we detail our previous efforts, with attention given to the data mining tools we used in tracking global household bicycle ownership, as they highlight the challenges of dealing with a sparse dataset. Section 3 motivates the network theory via a brief background and then proceeds to describe our approach. We then present our results (Sect. 4) and discuss their implications (Sect. 5). The paper is concluded by a summary and outline of future work (Sect. 6).

2 Review of prior cluster analysis of household bicycle ownership

We collected household bicycle ownership data from various national, regional and international surveys, spanning 1989 to 2012 and 150 countries. The unifying question of interest in all of the surveys was whether the household owned at least one bicycle. The ownership data only amounted to 540 unique country-years. We also obtained and estimated national household populations in the years of interest. The dataset represented 1.25 billion households, and a 42% household weighted average ownership indicates average indicates that there at least half a billion bicycles in homes around the world.[1] Details on the data collection process can be found in Oke et al. (2015).

Hierarchical agglomerative clustering (HAC) was the initial step in pattern discovery beyond geographical proximity. HAC is unsupervised (Jain et al. 1999; Tsui et al. 2006) but it requires a means for calculating the separation between all the nodes at initialization. The HAC method of choice informs the rule governing how the clusters are "grown," ultimately to form a tree. The natural number of clusters in the dataset must then be determined in order to truncate the tree at the proper height.

2.1 Dynamic time warping

Since ownership data were only available for a few different years in each country, finding a suitable point pairing for each ownership vector in time was not a trivial task. Thus, in order to find the optimal pairwise alignments, we applied the dynamic time warping (DTW) algorithm (Giorgino 2009), which uses the "average accumulated distortion" as the objective metric for finding the best matching path.[2] DTW computes a warping curve $\phi(k)$ with M elements, each mapped from a pair of time series A and B, each with P and Q observations, respectively. Thus,

$$\phi(k) = (\phi_a(k), \phi_b(k)) \tag{2.1}$$

where the selection functions $\phi_a \in \{1, \ldots, P\}$ and $\phi_b \in \{1, \ldots, Q\}$ are constrained accordingly to preserve continuity and monotonicity. The optimal alignment gives the minimum deformation D between pairwise sets of observations:

$$D(A, B) = \min_{\phi} d_{\phi}(A, B) \tag{2.2}$$

where the distortion d_ϕ is given by

$$d_{\phi}(A, B) = \frac{1}{C_{\phi}} \sum_{k=1}^{M} d(\phi_a(k), \phi_b(k)) c_{\phi}(k), \tag{2.3}$$

[1] The data and supporting code are available at www.ce.jhu.edu/sauleh/obls-gbu.

[2] The dynamic time warping algorithm was first introduced by Bellman and Kalaba (1959). Its application was subsequently furthered by Sakoe and Chiba (1978). See Giorgino (2009) for more on its execution in R.

Table 1 Fitness test for agglomerative clustering

Method	Greatest singular value λ
Single linkage	2723.7
Complete linkage	4551.8
UPGMA	883.2
WPGMA	966.1

The test quantity λ is given by $\lambda = ||DM - UM||_2$, where DM is the original dissimilarity matrix, while UM is the ultrametric (separation matrix) for each of the clustering structures considered

with $c_\phi(k)$ the weighting coefficient in each step, C_ϕ the normalization constant, and d a distance metric of choice (we used the Euclidean distance in this case). In the end we obtain a normalized dissimilarity matrix DM.

2.2 Clustering procedure

The next step we took was finding the HAC method of choice that worked best for our data. We considered four candidates: the weighted pair-group method with arithmetic means (WPGMA), the unweighted pair-group method with arithmetic means (UPGMA, also referred to as the method of averages), the complete linkage method and the single linkage method. The fitness measure is the Euclidean norm (or greatest singular value) λ of the difference matrix between the unclustered and clustered pairwise separations (Mérigot et al. 2010). The method of averages produced a λ value of 883.2, which was the smallest of the four (see Table 1).

The gap test (Tibshirani et al. 2001) provides a well-defined method for finding the optimal group number in a dataset. The test quantity is defined as

$$\mathrm{Gap}_n(k) = \frac{1}{B}\sum_b \log(W^*_{kb}) - \log(W_k), \tag{2.4}$$

where W_k is the within-cluster sum of pair-wise distances. The first term on the right-hand side of Eq. 2.4 is the expectation of $\log(W_k)$, obtained by a Monte Carlo simulation of B samples on a uniform distribution over each row of the pairwise separation matrix; this term can also be written as $E^*\{\log(W_k)\}$). The optimal number of clusters $\hat{k}$ is chosen as the smallest k such that

$$\mathrm{Gap}(k) \geq \mathrm{Gap}(k+1) - \varepsilon_{k+1} \tag{2.5}$$

where ε_{k+1} is the simulation error.[3] The value of $\hat{k}$ can therefore be formally given as:

$$\hat{k} = \arg\min_k \mathrm{Gap}(k){:}\quad \mathrm{Gap}(k) \geq \mathrm{Gap}(k+1) - \varepsilon_{k+1} \tag{2.6}$$

The intuition behind this method is as follows: If we assume the elements of the dataset are uniformly separated, then the expectation of $\log(W_k)$ decreases at a rate of $\frac{2}{p}\log(k)$, where p is the dimensionality of the dataset (Tibshirani et al. 2001). However, if a certain clustering method identifies well-defined groups, $\log(W_k)$ must be observed to decrease at a much faster rate, up to the point of maximum separation. Beyond this point, $\log(W_k)$ decreases at a slower rate than expected (see Fig. 1). For our dataset, the optimal cluster number $\hat{k}$ is thus 4.

[3] The error term is given by $\varepsilon_k = \sqrt{\left(1+\frac{1}{B}\right)\frac{1}{B}\sum_b\left\{\log(W^*_{k,b}) - \frac{1}{B}\sum_b \log(W^*_{k,b})\right\}^2}$, where b is the index number of the generated sample (Tibshirani et al. 2001).

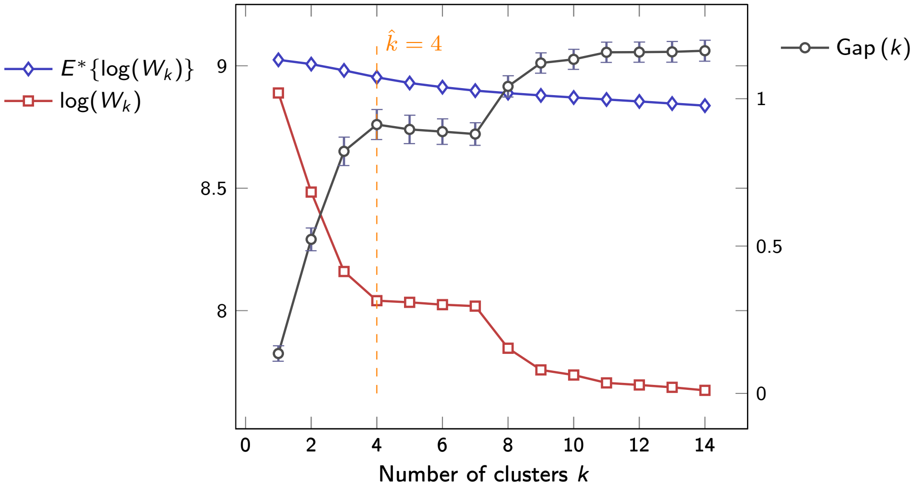


Fig. 1 This plot shows the expectation $E^*\{\log(W_k)\}$, the observation $\log(W_k)$ and the gap statistic Gap(k) as functions of the number of clusters obtained via the clustering method of best fit. The *error bars* on Gap(k) indicate the size of the simulation error ε_k. As an indicator of how well-separated the clusters are, $\log(W_k)$ decreases more rapidly than its expectation as the number of clusters chosen approaches optimality ($k < \hat{k}$), while it decreases less rapidly beyond this value ($k > \hat{k}$). In this case, $\hat{k} = 4$

Table 2 Country groups ranked by weighted average percentage bike ownership PBO

Group	PBO (%)	Countries
1	81	9
2	60	34
3	40	45
4	20	62

Table 2 summarizes the results of our cluster analyses. The group membership or rank of each country is given in Table 6. Household bicycle ownership is concentrated among a few countries, as membership increases with decreasing rank.[4] In this paper, "group rank" precisely refers to the level of bicycle ownership in the corresponding group. Group 1 therefore has the highest average rate of ownership, while Group 4 (the last group) has the lowest rate of ownership.

3 A brief introduction to networks and assortativity analysis

3.1 Graphs and networks

Graph theory has developed a host of tools for describing groups and associations within groups. These have been applied to problems such as route finding (shortest path algorithms), network flow problems (e.g. cost-minimization, resource allocation) and matching. By correctly formulating the network representation of a system, graph theoretic approaches

[4] For a complete discussion on the global and group trends, please refer to Oke et al. (2015). Also, trends in bicycle ownership for each country in the dataset set can be viewed in Appendix B (Ibid).

Table 3 Basic graph notation (as used in this paper) and definitions

Symbol	Description
Parameters	
u, v	Node
d_u	Degree of node u
e	Edge
Sets	
$G(V, E)$	Graph
V	Nodes
E	Edges
$N(u)$	Adjacent nodes of u
D	Unique degree values

can provide analyses for unearthing and understanding patterns of association (Small 1973; Newman 2004), growth (Teichmann and Babu 2004; Capocci et al. 2006), selection (Steglich et al. 2010; Baerveldt et al. 2014) and many more, within the system.

A key motivation for this paper is that a better understanding of proximity contributions to bicycle ownership would also contribute to efforts to encourage greater ownership and usage. The geographical layout of the countries already presents a corresponding topology and since we were interested in studying their associative patterns, a network or graph-based approach is advantageous. Wall et al. (2011) used a network model to investigate relationships based on "geographical embeddedness" and corporate activity between nations. Zhukov and Stewart (2013) also took advantage of geographic topology to formulate their network model for political diffusion.

Mathematically, a graph is set of nodes and edges. An edge is a line connecting a pair of nodes in a graph. The degree of a node is number of edges incident from it, and this can be immediately seen as a measure of the node's connectedness in the graph. A pair of nodes is said to be adjacent if the nodes are connected by one or more edges, i.e. one node is at the tail, while the other is at the head of the connecting edge. For a given node, the set of its adjacent nodes are its neighbors.

Notation Given a graph $G(V, E)$, with nodes $u, v \in V$ and edges $e_{uv} \in E$. The tail node of an edge e_{uv} is u and its head node is v. The set of neighboring nodes is denoted by $N(u)$ for a node u. A brief summary of the relevant graph notation is given in Table 3.

3.2 Mixing

Various properties can be of interest when a system is modeled as a network. Some of these include patterns of growth and resilience, which are relevant for population networks and infrastructural systems, respectively. Measures of influence of each node can also be important. This is especially true for social networks (Anagnostopoulos et al. 2008). If the nodes already belong to certain groups, another property of interest is mixing, which is concerned with nature of the connectedness of nodes by their groupings. Mixing patterns can provide insight into the associational behaviors of entities in a network (Newman 2002; Hens et al. 2009; Piraveenan et al. 2012).

To quantify the likelihood of nodes in various groups to be connected or cross-connected, a mixing matrix $\boldsymbol{E}$ is defined such that each element E_{ij} is given by

$$E_{ij} = \sum_{u \in I} \sum_{v \in J} A_{uv} \tag{3.1}$$

where A_{uv} is the edge-incidence matrix of the network, and i, j are indicial bijections onto the groups $I \in \boldsymbol{I}$, $J \in \boldsymbol{J}$, i.e. the index i in the matrix $\boldsymbol{E}$ represents the group I, and so forth. The matrix $\boldsymbol{A}$ is constructed such that $A_{uv} = 1$ if there is an edge connecting nodes u and v, and 0 otherwise. The normalized mixing matrix $\boldsymbol{e}$ is then defined as[5]

$$e_{ij} = \frac{E_{ij}}{\sum_{i=1}^{|\boldsymbol{I}|} \sum_{j=1}^{|\boldsymbol{J}|} E_{ij}} \tag{3.2}$$

The mixing coefficient (Newman 2003), which measures the level of connectedness of nodes in the same group, is defined by

$$r_m = \frac{\sum_i e_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2} \tag{3.3}$$

where $a_i = \sum_j e_{ij}$. The error of r_m is evaluated by its standard deviation, which, as Newman has noted (Newman 2003) is similar to that of intraclass correlation measurements (Fleiss et al. 1969) and given by

$$\sigma(r_m) = \frac{1}{M} \frac{\sum_i a_i^2 + \left(\sum_i a_i\right)^2 - 2\sum_i a_i^3}{1 - \sum_i a_i^2} \tag{3.4}$$

where M is the number of connections (edges) in the network. Our criteria for statistical significance is given by[6]

$$\frac{r_m}{\sigma(r_m)} \geq 5 \tag{3.5}$$

To measure the likelihood of cross-group mixing, we define the conditional probability P as

$$P(k|i) = \frac{e_{ik}}{\sum_j e_{ij}}. \tag{3.6}$$

Thus, given that a node is in group i, P is the likelihood that its neighbor is in group k.

3.3 Degree assortativity

Assortativity is a more general term that defines nodal associations based on a shared property in the network. For the purposes of this paper, we consider the term "mixing" as specific to group association, while "assortativity" is a generalization to other properties. The assortativity of a network based on the degree of each node is an established metric, first formally defined by Newman (2002, 2003). When nodes of high degree have more connections with nodes of high degree, the network is said to be assortative. If more connections appear between nodes of high degree to those of low degree, then the network is said to be disassortative.

Various measures of assortativity have been defined. The Pearson correlation coefficient, however, is typically used to measure the assortative strength in a network (Newman 2003; Foster et al. 2010). Two further definitions are needed before we describe the assortativity coefficient. First, e_{bc}^d is the probability that an edge chosen at random in the network connects

[5] Subsequent summations over i or j are over the same ranges indicated in Eq. 3.2.

[6] This threshold corresponds to the Z-score, for which 3σ or 2σ are commonly employed as the minimum value for significance. Here, we use the more stringent 5σ threshold (Colquhoun 2014; Foster et al. 2010; Newman 2002).

nodes of degrees b and c. Second, the excess degrees of two connected nodes are the degrees of each respective node without counting the edge connecting them. The excess degree distribution is given by the probability q_c for each degree c in the network. The probability q_c is therefore the probability that for a node chosen at random, the degree of all the other attached edges except for the edge being considered is c:

$$q_c = \sum_{b \in D} e^d_{bc} \tag{3.7}$$

The degree assortativity coefficient (Newman 2003) for associations based on node degree is thus[7]

$$r_d = \frac{\sum_{b,c \in D} bc(e^d_{bc} - q_b q_c)}{\sigma_q^2} \tag{3.8}$$

where σ_q is the variance of the excess degree distribution q_c given by Newman (2002):

$$\sigma_q = \sum_c c^2 q_c - \left(\sum_c k q_c \right)^2. \tag{3.9}$$

The statistical significance of r_d can be evaluated by its size relative to its standard deviation, which is calculated according to the following Newman (2003):

$$\sigma(r_d) = \frac{1}{M} \frac{\sum_c q_c^2 + \left(\sum_c q_c\right)^2 - 2\sum_c q_c^3}{1 - \sum_c q_c^2} \tag{3.10}$$

where M is the number of edges in the network. For the purposes of this discussion, we will consider weighted node degree assortativity as significant if the following condition is satisfied (see footnote 6):

$$\frac{r_d}{\sigma(r_d)} \geq 5 \tag{3.11}$$

Assortativity profiles can also be plotted to investigate the property of interest in a network (Piraveenan et al. 2010). They are useful for observing individual element contributions to the assortativity or disassortativity of the network.

3.4 A simple network model for bicycle ownership

We consider the 150 countries in our dataset as nodes in an undirected graph whose edges represent the land borders between adjacent countries. Data on node adjacency or borders can easily be read from a map, but we obtained a compiled list from Visualign (2013). From this, we obtained the degree of each node, which represents the number of unique land borders a country has. These values are shown in Table 6.

A key component of defining a network is defining the criteria for connections between nodes (and their strength, in the case of a weighted network) (Zhukov and Stewart 2013). Since we are investigating a network structure with regard to bicycle ownership, we want the edges to be weighted as a measure of the separation in bicycle ownership between adjacent countries. In this model, we simplify this calculation by simply using the difference in group rank, with a maximum possible weight of 4 (closest) and a minimum of 1 (furthest apart). We therefore define the edge weight $w(e)$ or w_{uv} as

$$w_{uv} = 4 - |\mathrm{Gr}(u) - \mathrm{Gr}(v)|, \tag{3.12}$$

[7] Subsequent summations over b or c are on the same set D, except where otherwise noted.

where $\text{Gr}(\cdot) \in \{1, 2, 3, 4\}$ is the group rank. Thus, an edge connecting two nodes in the same group will have a weight of 4, while the weight of the edge joining two nodes of maximum group separation will have a weight of $4 - 3 = 1$.

The degree of each node can therefore be weighted by the difference in group rank between the node and each of its neighbors. Since there are no loops or multiple edges in this network, we can define the weighted degree as follows.

Definition 1 (*Weighted degree*) Given a node u, the weighted degree is given by d^w as

$$d^w(u) = \sum_{v \in N(u)} w_{uv}. \tag{3.13}$$

Following Eq. 3.1, the mixing matrix elements for this network are given by

$$E_{ij} = \sum_{\text{Gr}(u)=i} \sum_{\text{Gr}(v)=j} A_{uv} \tag{3.14}$$

We will use the normalized mixing matrix and the corresponding conditional probabilities, as defined in Eqs. 3.2 and 3.6, to measure the group-mixing tendencies in the network. The group mixing coefficient for this network (Eq. 3.3) will indicate the overall strength of same-group mixing in the network.

In Sect. 3.3, we described assortativity with respect to the degree distributions of the nodes. For the purposes of this model, we will now consider the assortativity by weighted degree, as given by Eq. 3.13. We motivate this as follows: the degree of each node in the network without considering bicycle ownership groupings is simply a measure of the number of land borders the represented country has. This value remains the same except political factors intervene in the creation or destruction of one or more countries. By considering the weighted degree, the number of neighbors a country has can then be valued by their group separation, which is the difference between the group rank of the country and that of each of its neighbors. (We recall from the definition in Eq. 3.12 that the weight increases with decreasing separation by group). For instance, given two countries A and B with the same number of neighbors: if the weighted degree of A is 16 and that of B is 8, then A is clearly more associated with countries whose groups are closer to that of A, while the neighbors of B are further in separation by group. Based on Eq. 3.8, we now define the weighted degree assortativity coefficient, WDAC (Leung and Chau 2007; Noldus and Van Mieghem 2015).

Definition 2 (*Weighted degree assortativity coefficient, WDAC*) For a network with weighted degrees $b, c \in D^w$, where D^w is the set of all the weighted degree values that exist in the network, the weighted degree assortativity coefficient is given by

$$r_{d^w} = \frac{\sum_{bc \in D^w} bc(e_{bc}^{d^w} - q_b q_c)}{\sigma_q^2} \tag{3.15}$$

The WDAC will indicate the strength of assortativity or disassortativity by the weighted degrees of each node. We will also consider the value of r_{d^w} by the contribution of each group to investigate underlying behavior.

Finally, we will use assortativity profiles to enable us to visualize the nature of the contribution of each node to the overall assortativeness of the network. When assortativity is considered with respect to node degree mixing, the average nearest neighbor degree (ANND) can be used to evaluate the assortativity by observing the trend of ANND as a function of the degree of each node (Serrano et al. 2007). As we are dealing with the weighted degree in this case and definition of "neighbor" is equivalent to the "nearest neighbor" for each

Table 4 Mixing matrix of the clusters in the bicycle ownership network

Group	1	2	3	4
1	14	12	8	2
2	12	50	36	14
3	8	36	52	49
4	2	14	49	110

node, we now define the average neighbor weighted degree (ANWD), which is computed by taking the mean of the weighted degrees of the neighbors of a given node. Averaging the weighted degrees in the neighbor set in our case also has the added benefit of allowing for some adjustment for the number of borders a country has. This will be our function of interest for the assortativity profiles we use here. (See Barrat et al. 2004 for further background on this quantity.)

Definition 3 (*Average neighbor weighted degree, ANWD*) For given node u with a neighbor set $N(u)$, the average neighbor weighted degree is given by

$$d_N^w(u) = \frac{1}{|N(u)|} \sum_{v \in N(u)} w_{uv}. \tag{3.16}$$

The weighted degree and average neighbor weighted degree values for each country in the dataset are given in Table 6.
We note here that we did not consider countries in our dataset with degree 0, i.e. those that do not share a physical land border with any other country. These were Australia, Japan, New Zealand (Group 2) and the Phillipines (Group 3). The countries with no ownership data were also excluded from this model. Thus, the United Arab Emirates (Group 2), South Korea (Group 3) and Tunisia (Group 4) were not included in the analyses, as they were bordered by countries for which data were not available. The 7 exclusions represent only about 4.7% of the entire dataset.

4 Results

4.1 Mixing by group

The number of group pairings of neighboring countries is shown in Table 4. The number of connections increases by decreasing group rank, as expected. The table represents the symmetric mixing matrix $\boldsymbol{E}$. The sum of the elements in the upper triangle of $\boldsymbol{E}$ is 347, which is the number of unique land borders of the countries in the network.

The normalized mixing matrix $\boldsymbol{e}$ is derived from $\boldsymbol{E}$ (according to Eq. 3.2) and evaluates to

$$\boldsymbol{e} = \begin{bmatrix} 0.027 & 0.023 & 0.017 & 0.004 \\ & 0.097 & 0.071 & 0.027 \\ & & 0.116 & 0.112 \\ & & & 0.251 \end{bmatrix} \tag{4.1}$$

The group mixing coefficient r_m is 0.204 and $\sigma(r_m) = 0.0139$, a 15σ value. This indicates that, overall, associations between members in the same group are more likely.

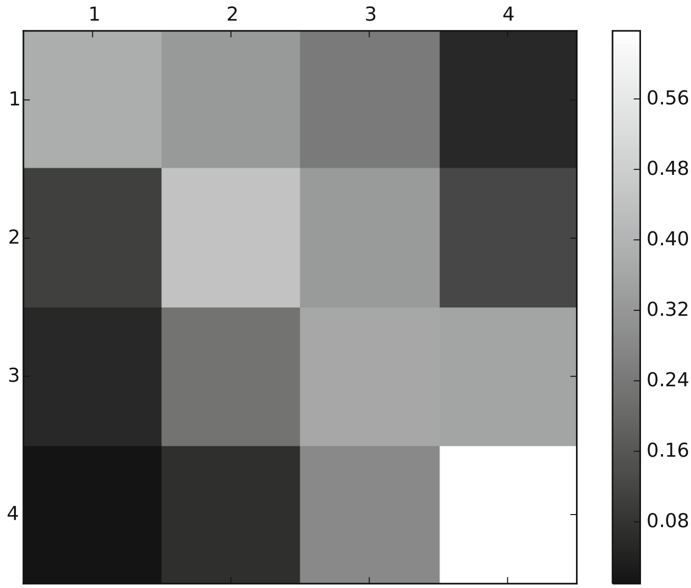


Fig. 2 Grayscale map of conditional probabilities of group mixing matrix given in Eq. 4.2. The colors get *darker* as the probabilities approach zero. The map shows that mixing is stronger between members of the same group (particularly so for Group 4) and weakest for those that are furthest apart in bicycle ownership

Finally, we compute the matrix of conditional probabilities that the neighbor of a given node is in Group j given that the node is in Group i.

$$\boldsymbol{P} = \begin{bmatrix} 0.378 & 0.324 & 0.243 & 0.054 \\ 0.106 & 0.442 & 0.327 & 0.124 \\ 0.055 & 0.226 & 0.366 & 0.354 \\ 0.010 & 0.069 & 0.284 & 0.637 \end{bmatrix} \tag{4.2}$$

This matrix indicates that Group 1 countries are more likely to be neighbors to other Group 1 countries. However, the likelihood that Group 2 members are neighbors to those Group 1 is nearly just as likely, as $P_{12} = 0.324$ and $P_{11} = 0.374$. A similar trend follows for each group in that their neighbors are most likely to be members of the same group, while the next most likely neighbors occupy the nearest group. However, the spread between these likelihoods is greatest for Group 4, indicating that its members are most closely linked.

To better illustrate these transitions, we plot a grayscale map of $\boldsymbol{P}$ (Fig. 2).

4.2 Weighted degree assortativity

We compute the weighted degree assortativity coefficient for the network, and also for each of the Groups 1 through 4. The results are shown in Table 5. For the entire network, the value $r_{d^w} = 0.05$ ($\approx 4\sigma(r_{d^w})$) indicates a statistically insignificant (by Eq. 3.11) weighted degree assortativity. However, when the contributions are considered by group relative to the network, further details emerge. For Group 1, the strength of disassortativity is quite small, but this is statistically significant. Group 4, however, is strongly disassortative with respect to the entire network ($r_{d^w} = -0.519 \approx -41\sigma(r_{d^w})$). Thus, both Group 1 and Group 4

Table 5 Weighted degree assortativity coefficients, with standard deviations and relative values

Group	r_{d^w}	$\sigma(r_d^w)$	$\lvert r_{d^w}/\sigma(r_d^w)\rvert$
1	−0.092	0.0157	5.9
2	0.383	0.0128	29.9
3	0.463	0.0129	35.9
4	−0.519	0.0126	41.1
All	0.050	0.0115	4.3

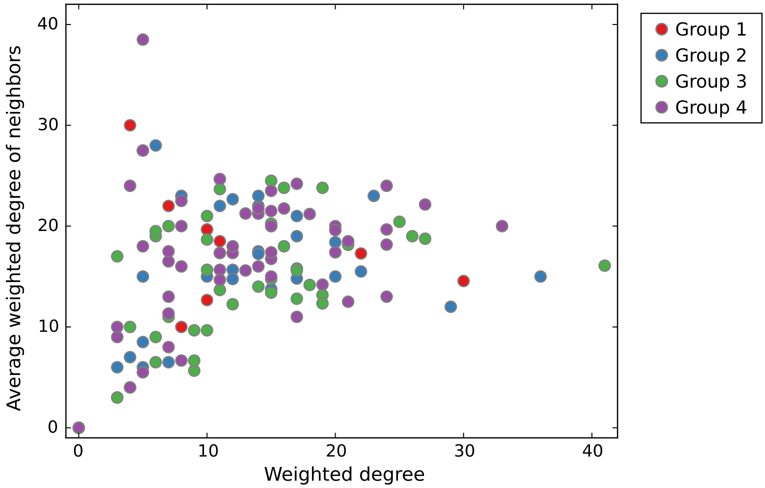


Fig. 3 Assortativity profile for the network (by average weighted neighbor degree) showing all groups. (Color figure online)

member countries of high weighted degree significantly tend to be connected to countries of low weighted degree (and vice versa).

Groups 2 and 3 are assortative with respect to the entire network, a tendency that is stronger for Group 3. The members of these two groups are therefore more likely to be connected with nodes of similar weighted degree.

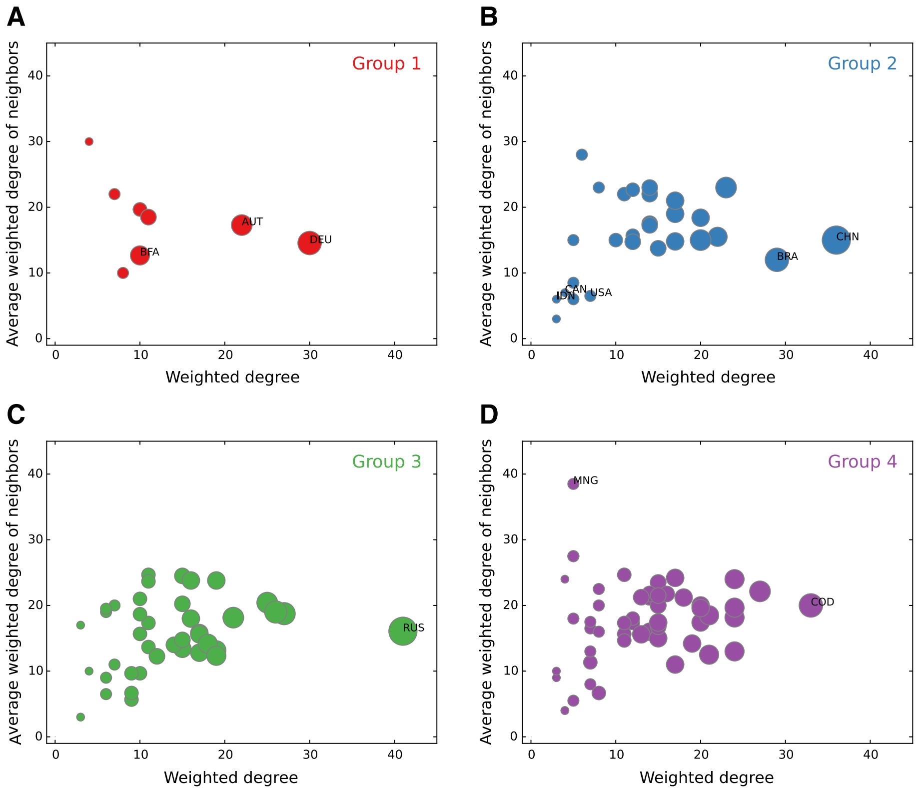
4.3 Assortativity profiles based on average weighted degree of neighbors

We plot the assortativity profiles by average weighted neighbor degree for the entire network (Fig. 3) and also by the contribution of each group (Fig. 4). The results support the coefficient values computed in Sect. 4.2.

While the profile of Group 1 does indicate a negative trend, the sparsity of points explains why its disassortativity is weak. Group 3 shows the stronger assortative profile compared to Group 2 (Fig. 4b, c). The profile for Group 4 (Fig. 4d) shows an accumulation of low weighted degree points with much higher weighted degree neighbor averages. Outlying points in the group profiles are labeled and discussed in Sect. 5.

5 Discussion

The mixing matrix and its associated conditional matrix P indicate that countries in the same group are more likely to be found near each other than to those from other groups. This effect



Country abbreviation reference:

AUT	Austria	CHN	China	MNG	Mongolia
BFA	Burkina Faso	COD	Cote D'Ivoire	RUS	Russia
BRA	Brazil	DEU	Germany	USA	United States
CAN	Canada	IDN	Indonesia		

Fig. 4 Assortativity profiles for Groups 1 through 4 based on the average weighted neighbor degree. The size of the markers indicates the actual degree of the node

is most pronounced for Group 4, indicating that many of them are geographically proximal. Furthermore, countries in Groups 1 and 3 are almost as likely to be found next to those in Groups 2 and 4. Thus, for Groups 1 and 3, there is most likely to be a cluster separation of 0 or 1 in relation to their neighbors. For all the groups, member countries are least likely to be proximally associated with those of widening separation by ownership group.

We consider assortativity with regard to weighted degree. Overall, the network is mildly assortative, based on the value of the computed assortativity coefficient. Thus, for all nodes in the four groups, there is a slight tendency for associations between nodes of similar weighted degree. This can be interpreted by saying countries with increased connectivity (in terms of both the number of neighbors and the closeness of group) are found adjacent to countries with similar characteristics. At the group level, this is not always the case. For Groups 1 and 4, this behavior is opposite (i.e. disassortative) and more strongly so for Group 4. As the connectivity of member countries increases, the average weighted neighbor degree reduces, which is a result of the reduced likelihood that those neighbors are in the same group. The highest point in Fig. 4d represents Mongolia which, while having only 2 neighbors, is connected to countries, namely Russia and China, with considerably more neighbors and higher bicycle ownership. The furthermost point along the x-axis represents DR Congo, which has the largest number of land borders in Africa.

Group 1 in particular shows the opposite trend, but its seeming disassortativity is insignificant, as can be inferred from the fact that there are far fewer countries in Group 1 compared to the other groups. In Fig. 4a, Burkina Faso is at the lowest point in the profile, indicating that it is surrounded by countries with a greater degree of cluster separation. The furthest points, i.e. those with the highest weighted degree (Austria and Germany) have an even greater number of neighbors. That they are higher up than Burkina Faso indicates that they are surrounded by neighbors closer in ownership.

For Groups 2 and 3, there is a strong assortative trend. In Group 2, two countries are found well beyond a weighted degree of 23: Brazil and China (Fig. 4b). While China has the largest degree (number of neighbors) in the set, its neighbors are further apart (in terms of ownership level), hence its position in the profile. Brazil also has a large number of neighbors but they are more distantly separated by group, which is why it is not further up in the profile. The points closest to the origin of the plot represent countries with only 2, 1 or no neighbors (e.g. USA, Canada, Ireland, Indonesia), indicating their limited influence in spite of their high ownership levels. Group 3 countries are more strongly assortative. As seen in Fig. 4c, Russia lies further up the x-axis than any other member of Group 3, as it has just as many land borders as China (13) but its weighted degree is larger—a potential indicator of its influence.

6 Conclusion and outlook

We have presented a simple network model that enables us to further investigate how the level of ownership in the countries considered is related to their neighbors. To do this, we defined a measure of connectivity for each pair of countries based on the presence of a shared land border and the separation between the ownership group occupied by each. We then applied principles of mixing and assortativity to describe patterns in associations between the bicycle ownership groups defined in an earlier study (Oke et al. 2015). Mixing was considered strictly on the basis of group, while assortativity was considered by the measure of connectivity defined—the weighted degree.

Our results show that countries with high ownership are generally likely to be neighbors to those with similar ownership or just one step lower. The countries with the lowest ownership, however, are the most isolated from other countries in terms of ownership separation. Furthermore, our weighted degree assortativity analyses enable us to better understand how ownership and proximity are related between countries and their neighbors in the four different groups. We observe that on the whole, weighted degree assortativity within the ownership network is statistically insignificant. However, for the countries in Groups 2 and 3, assortativity is observed, indicating that high weighted degree countries are more connected with those of equally high weighted degree. The opposite is the case for Groups 1 and 4, the latter exhibiting stronger weighted degree disassortativity. The group assortativity profiles by average neighbor weighted degree provide greater insight into individual country behavior and may serve as starting points for further investigations into these countries.

This work is an initial step toward applying network analysis tools in an effort to not only understand the factors influencing bicycle ownership, but also how the ownership levels in the countries themselves indicate those of neighboring countries. While the nodes in this network are largely static with regard to position (i.e. the degrees will largely remain the same), the levels of ownership can certainly change. For policymakers trying to weigh options to improve the quality of life in their countries, regions or continents by encouraging more ownership of bicycles, the analyses present a few approaches. For instance, if a regional coalition is interested improving the level of bicycle ownership in their member countries, initiatives might be better piloted in the most connected country.

A dynamic model of how the nature of the network might evolve in time would be useful for policymakers trying to weigh options to improve the quality of life in their countries, regions or continents by encouraging more ownership of bicycles. In such an instance, knowledge of how countries have moved into higher-ownership groups and their respective weighted connectivities can inform the selection of a pilot country in which to launch targeted initiatives, whose benefits would hopefully influence its neighbors in time. With more bicycle ownership data and future clustering analyses, perhaps more groups might also emerge. We would also like to better define the nature of country-to-country influence by accounting for other factors, such as climate, mobility, income, history (political and cultural) and topography. In our next round of efforts, we plan to utilize readily available data on some on these features to further investigate the nature of these relationships to whatever degree they exist. Further, with a more accurate and translatable network representation, we can hopefully provide more insights for stakeholders and policymakers in sustainable transit.

Acknowledgements We thank the participants of the 2014 INFORMS Workshop on Data Mining and Analytics participants for their valuable comments in the earlier stages of our work. The Gordon Croft Fellowship awarded by the Energy, Environment, Sustainability and Health Institute (E^2SHI) at The Johns Hopkins University, Baltimore, Maryland, funded this research in part. This quality of this paper was greatly improved as a result of the work of the two anonymous reviewers to whom the authors are grateful.

Appendix: Country parameters

Table 6 lists the degree parameters for each country. The code (written in Python) and supporting data for this work are available at http://modl.jhu.edu/resources/spatial-bicycle-ownership/.

Table 6 Degree, weighted degree, average neighbor weighted degree and group membership of the countries in the dataset

ISO	Country	Degree	Group	Weighted degree	Average neighbor weighted degree
AFG	Afghanistan	5	3	16	18
AGO	Angola	4	4	15	23.5
ALB	Albania	3	3	11	13.7
ARE	United Arab Emirates	0	2	0	0
ARG	Argentina	5	2	17	14.8
ARM	Armenia	3	4	11	15.7
AUS	Australia	0	2	0	0
AUT	Austria	7	1	22	17.3
AZE	Azerbaijan	4	4	14	21.3
BDI	Burundi	3	4	11	24.7
BEL	Belgium	4	2	14	17.5
BEN	Benin	4	3	12	12.3
BFA	Burkina Faso	6	1	10	12.7

Table 6 continued

ISO	Country	Degree	Group	Weighted degree	Average neighbor weighted degree
BGD	Bangladesh	2	4	5	18
BGR	Bulgaria	5	4	14	16
BIH	Bosnia and Herzegovina	3	3	11	17.3
BLR	Belarus	5	3	19	23.8
BLZ	Belize	2	2	5	8.5
BOL	Bolivia	5	3	17	15.8
BRA	Brazil	9	2	29	12
BTN	Bhutan	2	4	5	27.5
BWA	Botswana	4	4	15	20
CAF	Central African Republic	6	4	24	24
CAN	Canada	1	2	4	7
CHE	Switzerland	4	2	14	22
CHL	Chile	3	3	10	15.7
CHN	China	13	2	36	15
CIV	Cote d’Ivoire	5	3	15	13.4
CMR	Cameroon	5	4	20	17.4
COD	Congo DRC	9	4	33	20
COG	Congo	5	4	20	20
COL	Colombia	5	3	17	12.8
COM	Comoros	0	4	0	0
CRI	Costa Rica	2	2	5	6
CZE	Czech Republic	4	2	14	23
DEU	Germany	9	1	30	14.6
DJI	Djibouti	3	4	12	17.3
DNK	Denmark	1	1	4	30
DOM	Dominican Republic	1	4	4	4
ECU	Ecuador	2	2	5	15
EGY	Egypt	2	4	7	16.5
ERI	Eritrea	4	4	16	21.8
ESP	Spain	3	3	10	9.7
EST	Estonia	2	2	6	28
ETH	Ethiopia	6	4	24	18.2
FIN	Finland	3	1	10	19.7
FRA	France	6	2	22	15.5
GAB	Gabon	2	4	8	20
GBR	United Kingdom	1	3	3	3
GEO	Georgia	4	4	14	21.3
GHA	Ghana	3	4	7	11.3
GIN	Guinea	6	4	21	12.5
GMB	Gambia	1	3	3	17
GNB	Guinea-Bissau	2	3	6	19

Table 6 continued

ISO	Country	Degree	Group	Weighted degree	Average neighbor weighted degree
GRC	Greece	4	3	14	14
GTM	Guatemala	3	4	8	6.7
GUY	Guyana	3	2	10	15
HND	Honduras	2	3	6	6.5
HRV	Croatia	5	3	17	15.6
HTI	Haiti	1	4	4	4
HUN	Hungary	7	3	21	18.1
IDN	Indonesia	1	2	3	6
IND	India	6	3	19	13.2
IRL	Ireland	1	2	3	3
IRQ	Iraq	2	4	7	13
ISR	Israel	3	3	9	5.7
ITA	Italy	4	2	14	17.3
JOR	Jordan	2	4	7	8
JPN	Japan	0	2	0	0
KAZ	Kazakhstan	5	4	17	24.2
KEN	Kenya	5	4	18	21.2
KGZ	Kyrgyzstan	4	4	14	21.3
KHM	Cambodia	3	2	12	15.7
KOR	Republic of Korea	0	3	0	0
LAO	Laos	5	2	20	18.4
LBN	Lebanon	1	4	3	9
LBR	Liberia	3	4	11	14.7
LKA	Sri Lanka	0	3	0	0
LSO	Lesotho	1	4	4	24
LTU	Lithuania	4	3	15	24.5
LUX	Luxembourg	3	2	11	22
LVA	Latvia	4	3	15	20.3
MAR	Morocco	1	4	3	10
MDA	Republic of Moldova	2	3	7	20
MDG	Madagascar	0	4	0	0
MDV	Maldives	0	3	0	0
MEX	Mexico	3	3	9	6.7
MKD	Macedonia	4	2	12	14.8
MLI	Mali	6	3	18	14.2
MLT	Malta	0	3	0	0
MMR	Myanmar	5	2	17	19
MNE	Montenegro	4	3	15	14.8
MNG	Mongolia	2	4	5	38.5
MOZ	Mozambique	6	4	21	18.5
MRT	Mauritania	2	4	7	17.5

Table 6 continued

ISO	Country	Degree	Group	Weighted degree	Average neighbor weighted degree
MUS	Mauritius	0	2	0	0
MWI	Malawi	3	3	11	24.7
MYS	Malaysia	2	3	6	9
NAM	Namibia	4	4	15	20
NER	Niger	5	4	15	15
NGA	Nigeria	4	4	15	16.8
NIC	Nicaragua	2	4	5	5.5
NLD	Netherlands	2	1	7	22
NOR	Norway	3	1	10	19.7
NPL	Nepal	2	4	5	27.5
NZL	New Zealand	0	2	0	0
PAK	Pakistan	3	3	11	23.7
PAN	Panama	2	3	7	11
PER	Peru	5	4	13	15.6
PHL	Philippines	0	4	0	0
POL	Poland	7	2	23	23
PRT	Portugal	1	3	4	10
PRY	Paraguay	3	3	10	21
ROM	Romania	5	4	15	17.4
RUS	Russia	13	3	41	16.1
RWA	Rwanda	4	4	14	21.8
SDN	Sudan	6	4	24	19.7
SEN	Senegal	5	4	17	11
SLE	Sierra Leone	2	4	8	16
SOM	Somalia	3	4	12	18
SSD	South Sudan	7	4	27	22.1
STP	Sao Tome and Principe	0	4	0	0
SUR	Suriname	2	3	6	19.5
SVK	Slovakia	5	2	17	21
SVN	Slovenia	4	1	11	18.5
SWE	Sweden	2	1	8	10
SWZ	Swaziland	2	4	8	22.5
TCD	Chad	5	4	20	19.6
TGO	Togo	3	3	9	9.7
THA	Thailand	4	2	15	13.8
TJK	Tajikistan	4	4	13	21.3
TKM	Turkmenistan	3	4	11	17.3
TLS	Timor-Leste	0	4	0	0
TTO	Trinidad and Tobago	0	2	0	0
TUN	Tunisia	0	3	0	0
TUR	Turkey	6	3	19	12.3

Table 6 continued

ISO	Country	Degree	Group	Weighted degree	Average neighbor weighted degree
TZA	Tanzania	8	3	27	18.8
UGA	Uganda	5	3	16	23.8
UKR	Ukraine	7	3	25	20.4
URY	Uruguay	2	2	8	23
USA	United States	2	2	7	6.5
UZB	Uzbekistan	5	4	19	14.2
VEN	Venezuela	3	3	10	18.7
VNM	Vietnam	3	2	12	22.7
VUT	Vanuatu	0	4	0	0
YEM	Yemen	0	4	0	0
ZAF	South Africa	6	4	24	13
ZMB	Zambia	8	3	26	19
ZWE	Zimbabwe	4	4	15	21.5
ZZZX	Serbia	7	2	20	15

References

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'08)* (pp. 7–15). New York, NY, USA: ACM. doi:10.1145/1401890.1401897.

Baerveldt, C., Bunt, G., & Vermande, M. M. (2014). Selection patterns, gender and friendship aim in classroom networks. *Zeitschrift für Erziehungswissenschaft*, *17*, 171–188. doi:10.1007/s11618-014-0546-x.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747–3752. doi:10.1073/pnas.0400087101.

Bellman, R., & Kalaba, R. (1959). A mathematical theory of adaptive control processes. *Proceedings of the National Academy of Sciences of the United States of America*, *4*(2), 1–9.

Capocci, A., Servedio, V. D., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., et al. (2006). Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E*, *74*(3), 036,116.

Caprarelli, G., & Fletcher, S. (2014). A brief review of spatial analysis concepts and tools used for mapping, containment and risk modelling of infectious diseases and other illnesses. *Parasitology*, *141*, 581–601.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Open Science*, *1*, 140216.

Dales, J., & Jones, P. (2014). *International cycling infrastructure best practice study*. London: Report for transport for London, Urban Movement, Phil Jones Associates.

Darmofal, D. (2015). *Spatial analysis for the social sciences. Analytical methods for social research*. New York, NY: Cambridge University Press.

Dijkstra, A., Hak, E., & Janssen, F. (2013). A systematic review of the application of spatial analysis in pharmacoepidemiologic research. *Annals of Epidemiology*, *23*(8), 504–514.

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, *72*(5), 323–327.

Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, *107*(24), 10,815–10,820. doi:10.1073/pnas.0912671107.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: The DTW package. *Journal of Statistical Software*, *31*(7), 1–24.

Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Van Damme, P., & Beutels, P. (2009). Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infectious Diseases*, *9*(1), 1–18. doi:10.1186/1471-2334-9-5.

Jacobsen, P. L. (2013). Safety in numbers: More walkers and bicyclists, safer walking and bicycling. *Injury Prevention*, *9*(3), 205–209.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264–323.

Leung, C., & Chau, H. (2007). Weighted assortative and disassortative networks model. *Physica A*, *378*(2), 591–602.

Li, H., Xu, S. H., & Kuo, W. (2014). Asymptotic analysis of simultaneous damages in spatial boolean models. *Annals of Operations Research*, *212*(1), 139–154.

Ma, H., Fang, C., Pang, B., & Li, G. (2014). The effect of geographical proximity on scientific cooperation among Chinese cities from 1990 to 2010. *PLoS ONE*, *9*(11), e111705. doi:10.1371/journal.pone.0111705.

Mérigot, B., Durbec, J. P., & Gaertner, J. C. (2010). On goodness-of-fit measure for dendrogram-based analyses. *Ecology*, *91*(6), 1850–1859.

Midgley, P. (2011). *Bicycle-sharing schemes: Enhancing sustainable mobility in urban areas*. Background Paper 8, Global Transport Knowledge Partnership, International Road Federation, New York.

Milgram, S. (1967). The small-world problem. *Psychology Today*, *1*(1), 61–67.

Newman, M. E. J. (2002). Assortative patterns in networks. *Physical Review Letters*, *89*(20).

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, *67*(2).

Newman, M. E. J. (2004). Detecting community structure in networks. *The European Physical Journal B*, *38*, 321–330.

Noldus, R., & Van Mieghem, P. (2015). Assortativity in complex networks. *Journal of Complex Networks*, *3*(4), 507–542. doi:10.1093/comnet/cnv005.

Oke, O., Bhalla, K., Love, D. C., & Siddiqui, S. (2015). Tracking global bicycle ownership patterns. *Journal of Transport & Health*, *2*(4), 490–501.

Páez, A., & Scott, D. M. (2005). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, *61*(1), 53–67. doi:10.1007/s10708-005-0877-5.

Piraveenan, M., Prokopenko, M., & Zomaya, A. Y. (2010). Classifying complex networks using unbiased local assortativity. In *Proceedings of the Alife XII conference* (pp. 329–336). Odense, Denmark.

Piraveenan, M., Prokopenko, M., & Zomaya, A. (2012). Assortative mixing in directed biological networks. *IEEE-ACM Transactions on Computational Biology and Bioinformatics*, *9*(1), 66–78.

Roseland, M. (2012). *Toward sustainable communities: Solutions for citizens and their governments* (4th ed.). Gabriola Island, BC: New Society Publishers.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on ASSP*, *26*(1), 43–49.

Serrano, M. A., Boguná, M., Pastor-Satorras, R., & Vespignani, A. (2007). *Large scale structure and dynamics of complex networks: From information technology to finance and natural sciences*. World Scientific, chap Correlations in complex networks (pp. 35–66).

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269.

Steglich, C., Snijders, T. A. B., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, *40*(1), 329–393.

Szmyt, J. (2014). Spatial statistics in ecological analysis: From indices to functions. *Silva Fennica*. doi:10.14214/sf.1008.

Teichmann, S. A., & Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature Genetics*, *36*(5), 492–496.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, *63*(2), 411–423.

Tsui, K. L., Chen, V. C. P., Jiang, W., & Aslandogan, Y. A. (2006). *Springer handbook of engineering statistics*, Book chapter data mining methods and applications (pp. 651–669). London: Springer.

Vandenbulcke-Plasschaert, G. (2011). *Spatial analysis of bicycle use and accident risks for cyclists*. Thèses de l'Université catholique de Louvain, Presses universitaires de Louvain, Louvain-la-Neuve, Belgium.

Vieira, I. T., Cheng, R. C. H., Harper, P. R., & de Senna, V. (2010). Small world network models of the dynamics of HIV infection. *Annals of Operations Research*, *178*(1), 173–200.

Visualign. (2013). Number of neighbors for world countries. http://docs.google.com/spreadsheets/d/1agWAs620UntEQGY0yXUo5_C6zZesLEjBj1TVd3tTEDc/edit#gid=0. Accessed at http://visualign.wordpress.com/2011/10/06/number-of-neighbors-for-world-countries/.

Wall, R. S., Burger, M. J., & van der Knaap, G. A. B. (2011). The geography of global corporate networks: The poor, the rich, and the happy few countries. *Environment and Planning A*, *43*, 904–927.
Zhukov, Y. M., & Stewart, B. M. (2013). Choosing your neighbors: Networks of diffusion in international relations. *International Studies Quarterly*, *57*, 271–287.