Environment and Planning B: Planning and Design 2015, volume 47, pages 1 – 21

doi: 10.1177/0265813515603868

Spatialization of user-generated content to uncover the multirelational world city network

Marco M Salvini, Sara I Fabrikant
Geography Department, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; e-mail: marco.salvini@bluewin.ch, sara.fabrikant@geo.uzh.ch
Received 8 September 2013; in revised form 6 October 2014

Abstract. In this paper we tackle a fundamental and long-time challenge in urban geography, to uncover a functionally differentiated global city network. To this day, the empirical investigation of a global multifunctional city network remains a challenge given the scarcity of appropriate relational and multifunctional data sources. To overcome this research gap, we present an interdisciplinary network modelling approach that integrates methods from geographic information science with social network analysis, including automated semantic analyses. We apply our modelling framework to a globally available, user-generated database (ie, Wikipedia), still underutilized in urban geography and planning research. The proposed visual analytical investigation of the multifunctional world city network also includes a systematic evaluation to assess the robustness of the proposed approach, and the adequacy of crowd-sourced databases for scientific uses. By example, we discuss economical and political relations of a latent multifunctional global city network which we uncovered with our data-driven approach. Our results not only empirically replicate previously well-established world city network theory, but also generate new research questions about multiple functions of cities, as hypothesized in world city network research. Furthermore, we showcase the potential of coupling text-based user-generated data analysis with geovisual analytics for scientific investigations in urban studies.

Keywords: world city network, world city, network analysis, network visualization

1. Introduction

On-going globalization processes alongside rapid technological developments shape the structure and organization of a networked information society. As Castells (1996) suggests, networks are the dominant organizational form in the contemporary information society. In defining cities as the spatial nodes of these global networks, Castells (1996) joins the long tradition of global cities and world cities research, shaped particularly by Friedmann (1986) and Sassen (1991).

Starting with economic change, as initiated by Hymer (1972), the integration of cities into an economic system is crucial for their integration in a world city network. The definition of global cities is closely related to the geography of the advanced producer services (APS) (Sassen, 1991; Taylor, 2004). Similarly, world cities play an important role in the international division of labor (Friedmann, 1986; Rozenblat, 2004). Globalization embraces not only the economy, but also cultural, political, and social processes (Taylor, 2005). The integration of cities in a world city network is driven by urban actors (Pumain, 2006). Urban actors are, for example, firms, people, or organizations which are located in a city, and they connect that city with other cities through their various activities (Pumain, 2006). Smith and Timberlake (1995a) classify urban actors into four interaction types: economic, political, cultural, and social, which in turn might be shaped by humans, materials, and information.

Castells (1996) follows the same line of argument, and suggests multiple networks into which world cities are integrated. The structure of each world city network and the position of cities in these networks depend on the considered linkage types. Alderson and Beckfield (2004, page 820) emphasize the multiplicity of this network integration process: "To explore and assess statements of the world city hypothesis, one would ideally construct a multirelational network, combining data on economic, political, social, and cultural linkages between cities."

The construction and the empirical investigation of this multirelational network is an open challenge for world cities research, and a main concern of this paper. As we will demonstrate in this paper, newly available user-generated databases, analyzed with an interdisciplinary set of methods, seem to be a worthwhile data source for the construction of a multirelational world city network. In the next section we first review established data sources for world city research, and then introduce our crowd-sourced data source, followed by the respective empirical data analysis and modelling steps. In section 3 we present an integrated modelling approach within the context of geographic information science (GIScience) to differentiate a multirelational world city network for further urban geographic research. We then discuss the potential risks and limitations of the chosen approach. Additionally, we include a systematic evaluation to test the robustness of the proposed methods, providing solid evidence of the validity of our results. In the discussion we highlight our main contributions to urban geography and conclude with recommendations to leverage the potential of user-generated data not only for answering urban geographic research questions, but also for the social sciences in general.

2. State of the art

While urban theoretical approaches and debates in world cities and global cities research are deep and comprehensive, data-driven, empirical validation approaches have been suggested to be a weak link in this research domain (Taylor, 1997). A series of contributions in the recent past have focused on the exploration of new data sources and modelling approaches to overcome the above mentioned weakness in global city network research. For example, Taylor et al., (2007a; 2011) and Derudder et al., (2012) provide an insightful overview of this relevant work. On the basis of the theoretical considerations by Sassen (1991) the Globalization and World Cities (GaWC) research network group defines a world city network by considering APS subsidiaries (Taylor, 2001). The primary assumption of this approach is that a relationship exists between APS subsidiaries that is proportional to the size of considered subsidiaries, and thus their importance within a firm. Second, an interlocking world city network can be obtained by aggregating intrafirm relationships based on the location of the APS subsidiaries. Already in one of their first-agenda setting contributions Taylor et al., (2002) consider 100 well-known global APS in over 300 cities around the world. The global extent of their considered set of cities, and the modelled intercity relationships are a first contribution to solving the data problems mentioned earlier. Ongoing work of GaWC specifies and enhances the approach initiated by Taylor (2001) by not only considering new modelling approaches (Hennemann and Derudder, 2014), and new visualization techniques (Hennemann, 2013), but also by proposing different urban network models (Derudder and Liu, 2013). Recent work also looks at alternative dimensions of world city network formation such as, transnational media (Hoyler and Watson, 2013), Nongovernmental organizations (NGOs), and energy corporations (Toly et al., 2012).

A line of research followed by Alderson and Beckfield (2004), Rozenblat (2004), and Wall and Van der Knaap (2011), considers a firm's geographic location for modelling a world city network. These authors embrace Friedmann's (1986) world city concept, and model a relationship network of multinational enterprises based on ownership data. Rozenblat and Pumain (2007) contend that the empirical basis for a network based on firm ownership relations is stronger than relying on APS information, especially when using headquarter and subsidiary locations appearing in the well-established Global 500 list of the *Fortune* magazine (Alderson and Beckfield, 2004).

A third, commonly employed, research direction considers infrastructure networks that build a world city network. For example, airline passenger movements between major city airports have been considered for modelling the postulated connectivity between cities (Derudder and Witlox, 2005; Smith and Timberlake, 1995b; 2001; Taylor et al., 2007b). Alternatively, telecommunication backbone networks (Malecki, 2002; Rutherford et al., 2004) have been investigated so as to overcome some of the biases directly related to airline passenger movements. More recent contributions including Derudder et al., (2014) looked at the integration of various infrastructure networks, such as transportation networks (ie, by air, on roads, and on train tracks), as well as digital computer networks. In the following sections we discuss the three briefly presented modelling approaches based on four aspects of the world city network: its relational nature, the duality between cities and urban actors, its global scope, and lastly, its multifunctionality.

One of the key elements of a world city network is its relational nature, and therefore data that are represented congruently in a relational structure seem especially attractive for modelling. Ownership relations (Alderson and Beckfield, 2004; Rozenblat, 2004; Wall and Van der Knaap, 2011), and infrastructural networks (Smith and Timberlake, 1995b; 2001) are explicitly relational datasets, and therefore particularly appropriate for modelling a network. However, as highlighted by Rozenblat and Pumain (2007), intrafirm relations used in the GaWC approach are based on assumptions that first need to be validated empirically.

The duality between cities (as spatial anchor points or nodes in a network), and their urban actors (as drivers behind intercity relationships) is another important aspect in the integration process of cities into a world city network. Interestingly, the only data that do not capture this duality are indeed the infrastructural network datasets. Other popular approaches consider the urban actors such as APS, transnational media, NGOs, energy corporations, and multinational enterprises.

While the globality of the investigated domain is well covered in all compared datasets — another key element of the world city network research — there are notable differences in the types and quantity of included urban actors. Taylor (2004) considers 7500 subsidiaries of 100 APS, and Rozenblat (2004) analyses 4000 subsidiaries of multinational enterprises.

The most significant aspect we wish to investigate further in the context of this paper is the multifunctionality of the modelled world city network. The two approaches based on APS and multinational enterprises, only consider economic aspects of the integration of the cities into a world city network. In other words, the modeled city networks are not multifunctional, as postulated by Alderson and Beckfield (2004) and Castells (1996). Political, social, and cultural processes are not taken into account in the two reviewed approaches. The approaches based on infrastructure networks could potentially consider a broader range of functions beyond just the economy. However, even if these data are relational, global, and potentially multifunctional, the different functions are not distinguishable. Hence, one of the most important limitations considering all prior attempts to investigate a multirelational network is the impossibility of distinguishing more than one of the various latent city functions. Early on, Taylor (2005) attempted to overcome this limitation by applying the GaWC interlocking approach to model four different functional networks. His economic city network considers APS and multinational enterprises. While the cultural city network is based on media conglomerates and architectural-engineering firms, the political network includes United Nations (UN) agencies, city government organizations, and embassies. Finally, NGOs and scientific publications constitute the social city network. As Taylor (2005) concludes, the study is only preliminary, and more effort is necessary to uncover multifunctional datasets for studying a multirelational world city network. Hoyler and Watson's (2013) investigations on transnational media, and the research by Toly et al., (2012) on NGOs and energy corporations are additional stepping stones to uncovering the multifunctionality of the world city network.

In terms of methods, social network analysis plays a considerable role in the investigation of the world city network (Smith and Timberlake, 1995a). This approach fits well with the relational trait of the postulated world city network. Nevertheless, Thierstein and Förster (2008) point to an important methodological improvement for this type of research: the systematic integration of advanced visual analytical techniques in the methodological frameworks, so as to facilitate data exploration and knowledge of construction from complex, multivariate relational datasets. The modelling approach presented in the next sections is an attempt to mine increasingly massive, multivariate datasets to disentangle the multifunctional world city network, and to transfer well-established state-of-the-art visualization techniques from GIScience and geovisualization to the domain of urban studies.

3. Methodological framework

We chose to harness the potential power of the globally generated and accessible, massive, crowd-sourced, multimedia database, such as the English version of the Wikipedia Encyclopedia, as our data source, and develop a methodological framework for the empirical investigation of the world city network. Wikipedia is a result of the fast growing, and crowd-sourced knowledge production on the Web 2.0. Crowd-sourcing is defined as “outsourcing [a function] to undefined (and generally large) network of people in the form of an open call” (Howe, 2006). Crowd-sourced data, such as Wikipedia, are increasingly harvested in academia for scientific purposes, and also for solving geographic research questions. Goodchild (2007) defines the voluntary and collective production of geographic data as volunteered geographic information (VGI). In Goodchild’s (2007) terms, Wikipedia can be considered as a massive VGI data source, and Wikipedia authors might be considered as human sensors through which we can see the world, similar to employing satellites as remote sensors to gain knowledge about the Earth’s surface.

First scientific attempts to harvest Wikipedia geographically were carried out by Hecht and colleagues (Hecht, 2007; Hecht and Moxley, 2009; Hecht and Raubal, 2008). In related work, Graham and Zook (2013) show how online query engines, such as GoogleMaps and Wikipedia, create online, digital representations of urban places through geocoded references to the economic, social, and political experiences of the city. In this paper we extend this line of research, and investigate the potential of user-generated content in general, and Wikipedia in particular, for the systematic modelling and analysis of a multirelational world city network. There are three reasons why Wikipedia seems an attractive data source for this endeavour. First, and most importantly, for the cities themselves and their urban actors (eg, people, firms), Wikipedia contains two key drivers, as identified in world city theory, to shape the urban process, and to determine the integration of cities in a world city network (eg, Pumain, 2006). Second, with its hyperlinked infrastructure, Wikipedia is relational and networked, and finally, being available globally and online, it is intrinsically rooted in the networked information society. Similarly, as Boulton et al., (2011) empirically exemplify, the value of web search engine data has become an important means of understanding cities as situated within, and constituted by, flows of digital information. They are able to demonstrate how, for example, the Google search engine can be used to specify a dynamic, informational classification of North American cities based on both the production and the consumption of web information.

As a first step, we downloaded a full copy of the English version of Wikipedia in XML format on January 13, 2010, preprocessed with the WikAPIdia API (Hecht, 2007), and stored the result in a MySQL database. The English Wikipedia copy used in this study consists of approximately 3.2 million articles, and 94.7 million links. About a half a million categories are included as metadata (ie, user-generated article tags). Categories are linked with one another, and organized in a thesaurus (Voss, 2006). The relationships in the category system can include holonymic (eg, a district is-a-part-of a city), or hyperonomic (eg, a pear is-a type of fruit) semantic relationships (Ponzetto and Strube, 2007). We can model the world city network through the article system by taking advantage of existing hyperlinks between articles about world cities, and we can distinguish the latent functions of the discovered intercity relations considering the user-generated category system. The modeling approach will be discussed in detail in the next two subsections.

3.1. Modeling the world city network

The entire workflow for modelling the world city network is described step by step in figure 1. All the data-handling steps (numbered from 1 to 6 in figure 1) are performed with queries to the above mentioned MySQL database. First, we selected cities listed with the Wikipedia entry 'List of towns and cities with 100 000 or more inhabitants', as defined by the Statistics

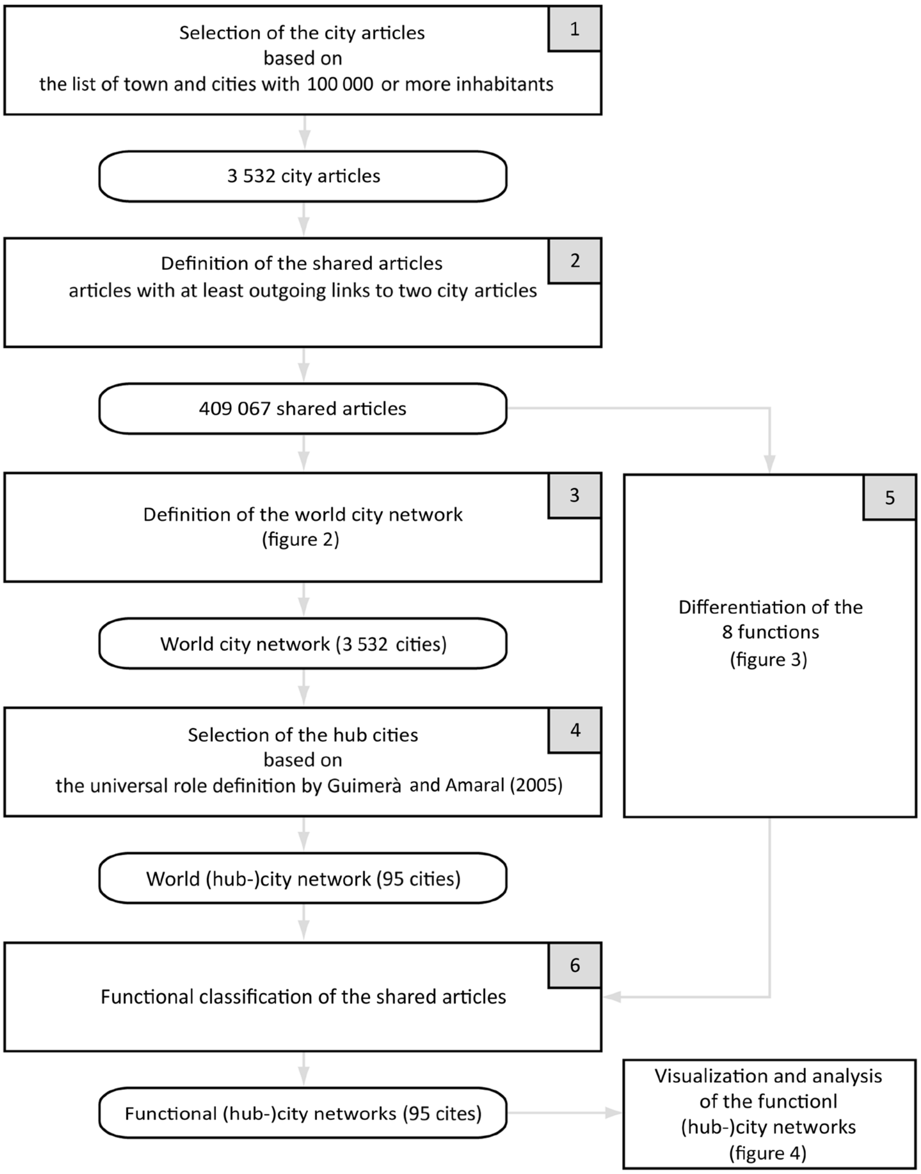

Figure 1. Workflow describing the transformation steps from the raw Wikipedia data into a world city network.

Division of the United Nations (step 1 in figure 1). As a result, 3532 cities are identified. Salvini (2012) demonstrated that the number of selected cities does not influence the intrinsic properties of the modelled network significantly. Therefore, we assume that these 3532 selected cities are an appropriate set for our network analysis.

A graph between the selected cities is built considering articles with at least one outgoing link to two cities (step 2 in figure 1). This is the minimal condition for an indirect relationship between articles about two cities. A relationship between two cities is given, when both articles describing these cities have incoming links from at least one shared article. This assumption can be illustrated by the example in figure 2. The articles about New York and Chicago

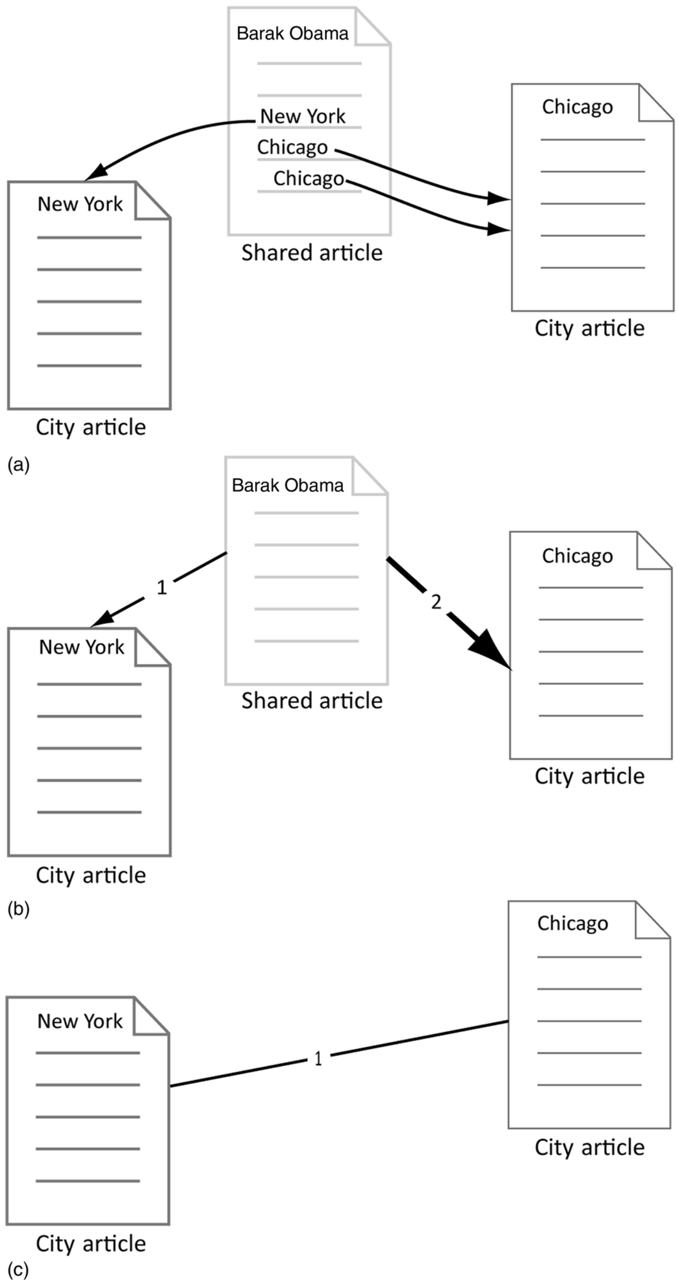

Figure 2. Definition of intercity relations and their weights in the world city networks: (a) article system; (b) count of the links; (c) max-flow–min-cut transformation.

have one or two incoming links from the shared article about Barack Obama. Therefore, New York and Chicago are indirectly related in the world city network through the article about Barack Obama. Using incoming links from shared articles has both conceptual and analytical advantages over other modelling solutions that employ direct links in the article system. We contend that the relationship between the shared articles and the city articles captures the interdependency between cities and urban actors in the integration process of cities in the world city network, as theorized by Pumain (2006). Specifically, the range of urban actors which determines the integration of cities in a world city network can be identified in the shared articles. This is because these articles include descriptions of activities of respective actors. Moreover, many of these activities are georeferenced with outgoing links to city articles. In other words, these three Wikipedia information items: shared articles, city articles, and outgoing links capture the theorized urban integration process, as formulated by Pumain (2006) including urban actors that are located in a city, and these urban actors connect this city with other cities through their various activities.

In our concrete example, the urban actor Barack Obama connected two world cities (ie, New York and Chicago) during his life course to a global city network in the real world. Similarly, the two Wikipedia city articles about New York and Chicago are connected through the encyclopaedic entry about the US President Barack Obama in the semantic world of Wikipedia. The world city network based on this approach considers 3532 city articles that are connected through approximately 400 000 shared articles.

We can now quantify the relationships between two cities using the number of shared articles, the number of links between city articles and shared articles, and the number of outgoing links from the shared articles. Figure 2(a) depicts the relationships between Chicago and New York in Wikipedia, again using the previous example with the urban actor Barack Obama. Two links connect the shared article about President Obama to the city article about Chicago, and one link connects the shared article about Barak Obama to the city article about New York. The shared article about Barak Obama has 351 links to other articles in Wikipedia. The first step in the definition of the link weights, as a representation of the relationship strengths between Chicago and New York, consists of a simple count of the links between the shared article Barack Obama, and the two connected city articles Chicago and New York [figure 2(b)]. Considering that this analysis focuses on the cities themselves rather than on the urban actors, in a next step, we transform the dual-mode urban-actor — city network link matrix to a single-mode city — city network [figure 2(c)]. The strength of the relationship between the two cities is determined considering the max-flow – min-cut theorem (Ford and Fulkerson, 1956), and this therefore corresponds to the minimal link weight along a path on the network between two city nodes. In the example shown in figures 2(b) and 2(c) , the link weight between articles Barak Obama and Chicago equals 2, and between the articles Barak Obama and New York the weight equals 1. Using the max-flow – min-cut approach the direct relation between the articles Chicago and New York is thus assigned the weight of 1. As articles can vary significantly in length, this can also determine the number of links an article might have. Short articles may have less than ten outgoing links, while longer articles may have hundreds of outgoing links. To reduce the potential influence of article length the weights of the city – city relationships are normalized following the normalization equation.

For $C \leq 5$

$$NW_{ij} = \frac{W_{ij}}{OL_{sa}}$$

and

For $C > 5$

$$NW_{ij} = \frac{W_{ij}}{C + (1 + \log_2 |OL_{sa} - C|)}$$

Where C is the number of outgoing links. In our example, the weight (= 1) is normalized on the basis of the total number of outgoing links of the shared articles (= 451). The number of outgoing links is normalized considering the overall structure of Wikipedia, previously investigated in studies by Hecht and Raubal (2008). If the shared article has fewer than five outgoing links, the normalization considers the total number of outgoing links from the shared article. Thus, the resulting relationship between Chicago and New York is the sum of all the weights, as computed for all the shared articles between these two city articles. The resulting relationship matrix represents a world city network considering 3532 cities, and 3577 545 relationships between these cities. The type of these intercity linkages is, however, not defined yet. Considering the massive amount of data to be modelled, we apply a functional data-reduction strategy to facilitate computation. Using Guimerà and Amaral's (2005) role definition approach (step 4 in figure 1) we differentiate hub-cities from non-hub-cities. Hub-cities show an above average intramodular strength compared with the other cities in the same group (Guimerà and Amaral, 2005). The resulting subset includes 95 hub-cities. The differentiation of the function (step 5 in figure 1), and the functional classification of the shared articles (step 6 in figure 1), are discussed in the next section.

3.2 Functional differentiation

As mentioned above, the category system includes user-generated tags which categorize the content of the articles, and the categories themselves are organized as a semistructured tree (Ponzetto and Strube, 2007). On the basis of the hyperonymic and hyponymic relationships, it is possible to quantify the similarity between the categories, and therefore between the respective articles. Using computed similarity relationships, the shared articles can be grouped into thematically coherent clusters. In essence, these clusters correspond to the functions in the world city network. Once each article is assigned a set of functions, the world city network can be differentiated, considering only the shared articles that belong to the same function.

Similarity assessment between the shared articles and delineation of functional clusters can be carried out following the generic principles of computational semantic analysis, for example, as with the topic modelling (TM) approach (Steyvers and Griffiths, 2007). TM starts with a text corpus in which the number of latent topics is a priori specified. Then, the semantic relationships between documents in the text corpus with latent topics are determined based on automatic similarity assessment using word co-occurrences. Following that, using the resulting document-topic similarity matrix, articles are grouped into clusters of similar documents. Figure 3 describes this workflow in more detail.

As shown in figure 3 our chosen text corpus for the semantic analysis contains articles that connect the ninety-five hub-city articles, including their category tags. Thus, this corpus contains 234 59 articles and 173,568 categories. An article together with its assigned categories is considered as one text document for our purpose (step 1 in figure 3). We selected twenty-two latent topics for our analysis, which is the same number of 'main categories' in Wikipedia. Main categories are defined as the subordinate categories of the top-level category 'main topic classification'. Main categories are assigned to one of the roots of the semistructured Wikipedia category tree, and thus all other categories are connected to it. To allow for a more detailed investigation of the article content, sixty-one subcategories connected to the main categories are also considered to be latent topics (step 2 in figure 3). We then determine semantic relationships between articles and the selected latent topics by assessing the path length along the branches of the semistructured category tree (step 3 in figure 3) that connect the eighty-three latent topics with the categories assigned to each article. The relationship between an article and a topic is strong if many categories are related to this topic, and if the paths along the branches of the category tree are short. Categories referring to time (ie, days, months, years, centuries, and millennia) and space (ie, cities,

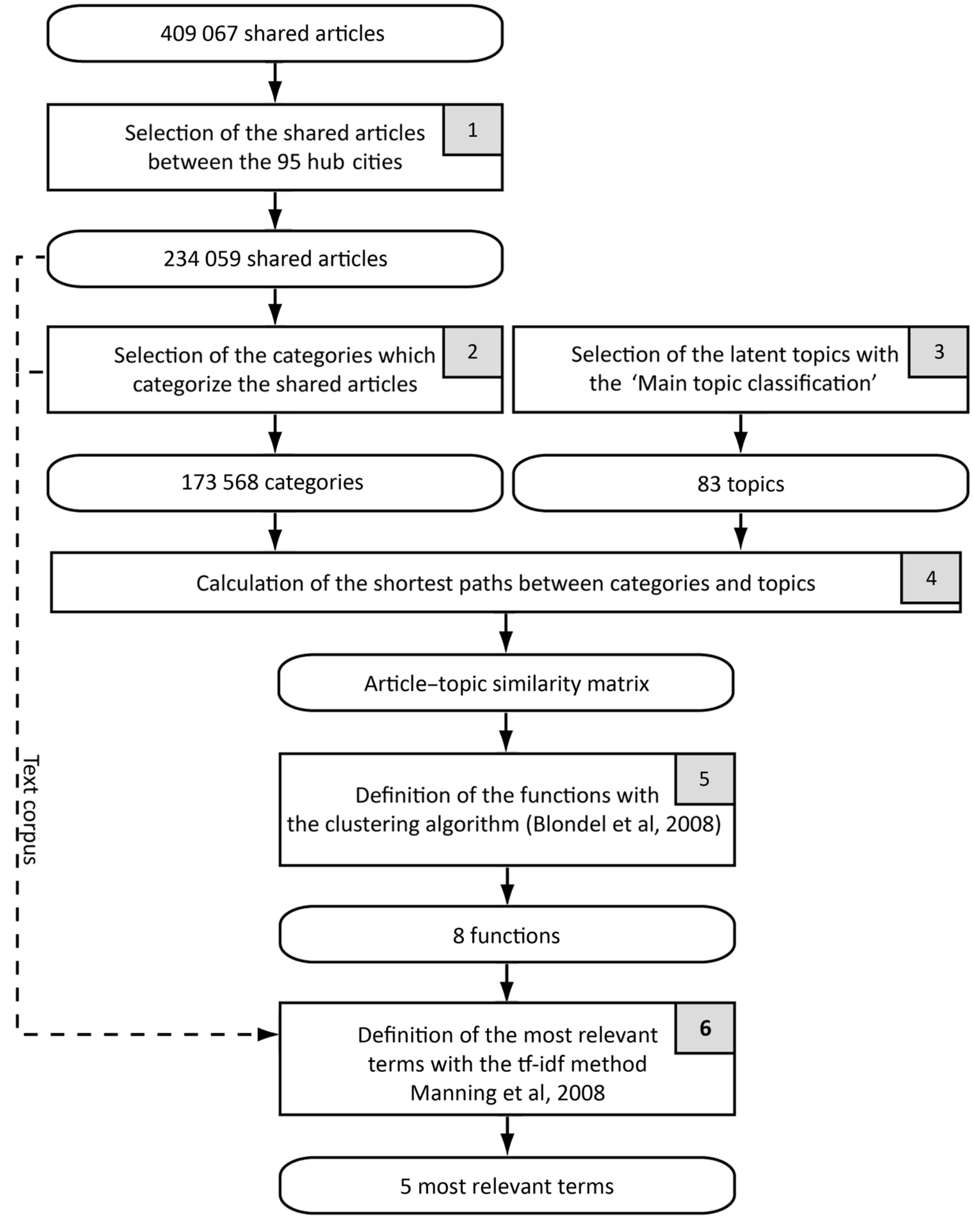

Figure 3. Workflow for the transformation of the raw Wikipedia data for a functional differentiation of the world city network.

regions, countries, and continents) are excluded from the analysis to emphasize the similarity assessment based on thematic aspects, rather than spatiotemporal orderings. The resulting article – topic matrix is normalized to reduce the influence of varying category numbers for each shared articles (step 4 in figure 3). The maximum possible similarity relationship between the shared articles and the latent topics scores equals 1 for each shared article. The normalized article – topic matrix is then clustered with the Blondel community detection algorithm (Blondel et al., 2008), which uncovers eight functional clusters (step 5 in figure 3). These functional clusters identify densely connected groups of nodes with many mutual relations (Blondel et al., 2008). The number of functional clusters is purely data driven, and not defined a priori, in contrast to the a priori definitions of Smith and Timberlake (1995a) and Taylor (2005). Similar to traditional cluster analyses, different solutions with a larger (more detailed) or smaller (less detailed) number of latent topics are possible. The content of each function cluster can be investigated further, for example, considering the most frequent or most relevant terms in the cluster. The text corpus for each functional cluster consists of the titles of the shared articles, and the titles of the categories which belong to the considered articles. The most relevant terms in the cluster are determined with the tf-idf method (Manning et al., 2008), commonly used in information retrieval (step 6 in figure 3). Relevant words occur more frequently in the considered functional cluster than in the rest of the text corpus. Given this functional information for all the shared articles, eight functionally differentiated world city networks can be modelled, each considering only the set of shared articles that belong to one of the eight uncovered functional clusters.

3.3. Visualization and analysis of the multirelational world city network

In general, the visualization of large and complex networks has become a significant scientific challenge in various disciplines. This is where GIScience and, in particular, cartography might offer attractive solutions. The following methodological considerations for visualizing the eight functionally differentiated world city networks are based on well established, long-standing, and more recently empirically validated visualization methods and principles from cartography and GIScience. Spatialization is the scientific research field within GIScience that deals with the cognitively inspired and perceptually salient visuospatial depiction

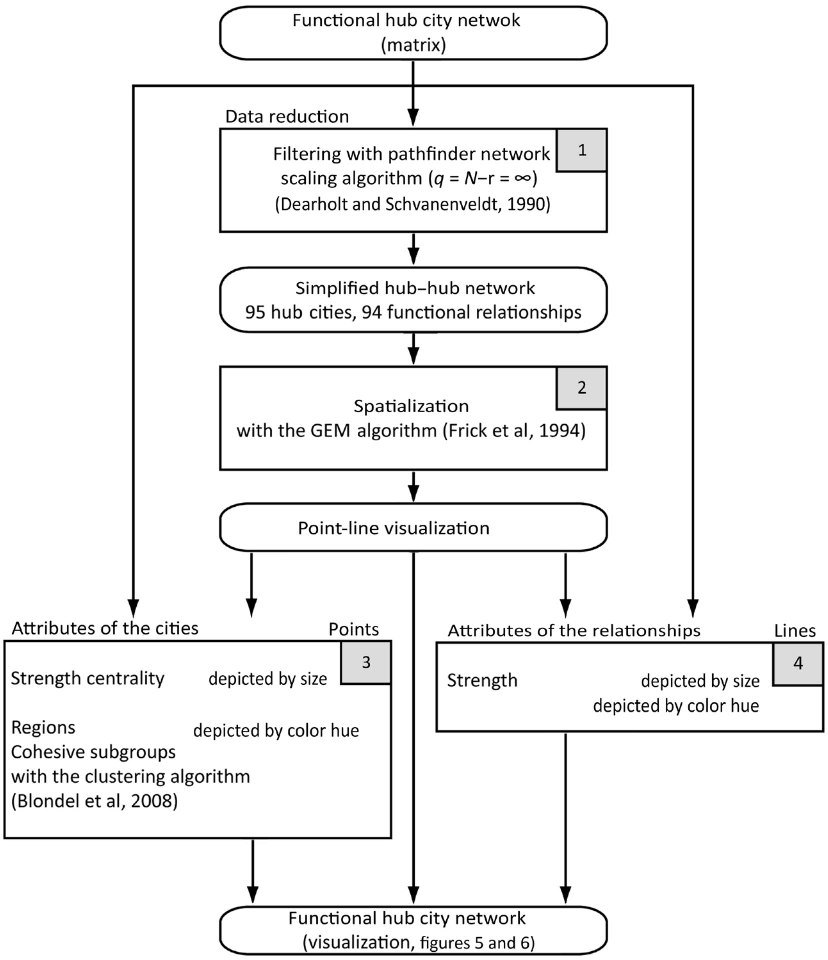

Figure 4. Workflow describing the data reduction and visualization of the multirelational world city network.

of inherently nonspatial data. The spatialization process is defined as the "systematic transformation of high-dimensional data sets into lower-dimensional, spatial representations for facilitating data exploration and knowledge construction" (Skupin and Fabrikant, 2007, p. 62) The spatialization of nonspatial data includes two steps: first, appropriate data filtering and data reduction must be applied to capture the essence of the meaning buried in large multivariate datasets. Secondy, the reduced data matrix needs to be visualized appropriately for human data exploration and sense making. Figure 4 summarizes the spatialization workflow to visualize two of the eight functionally differentiated world city networks, as depicted in figures 5 and 6. Further details about the spatialization of network data can be found in Fabrikant et al., (2004), Salvini (2012), and Fabrikant and Salvini (2011).

We start the spatialization process with the city – city matrix detailed in previous sections, which encapsulates the functional relationship between the ninety-five hub-cities, as shown in figure 4. As a first step, we filter the data. This is necessary not only to avoid cluttered and illegible visualizations, but also to carve out the structurally most salient relationships buried in the word city network. The pathfinder network scaling (PFNet) algorithm (Dearholt and Schvanenveldt, 1990) is often used in information visualization (Börner et al., 2003) to simplify complex networks by extracting the structurally fundamental edges (Chen, 1999). Using the PFNet algorithm, we extract the structurally most important relationships between the ninety-five hub-cities (step 1 in figure 4). The filtered matrix consists in ninety-four functional relationships among the

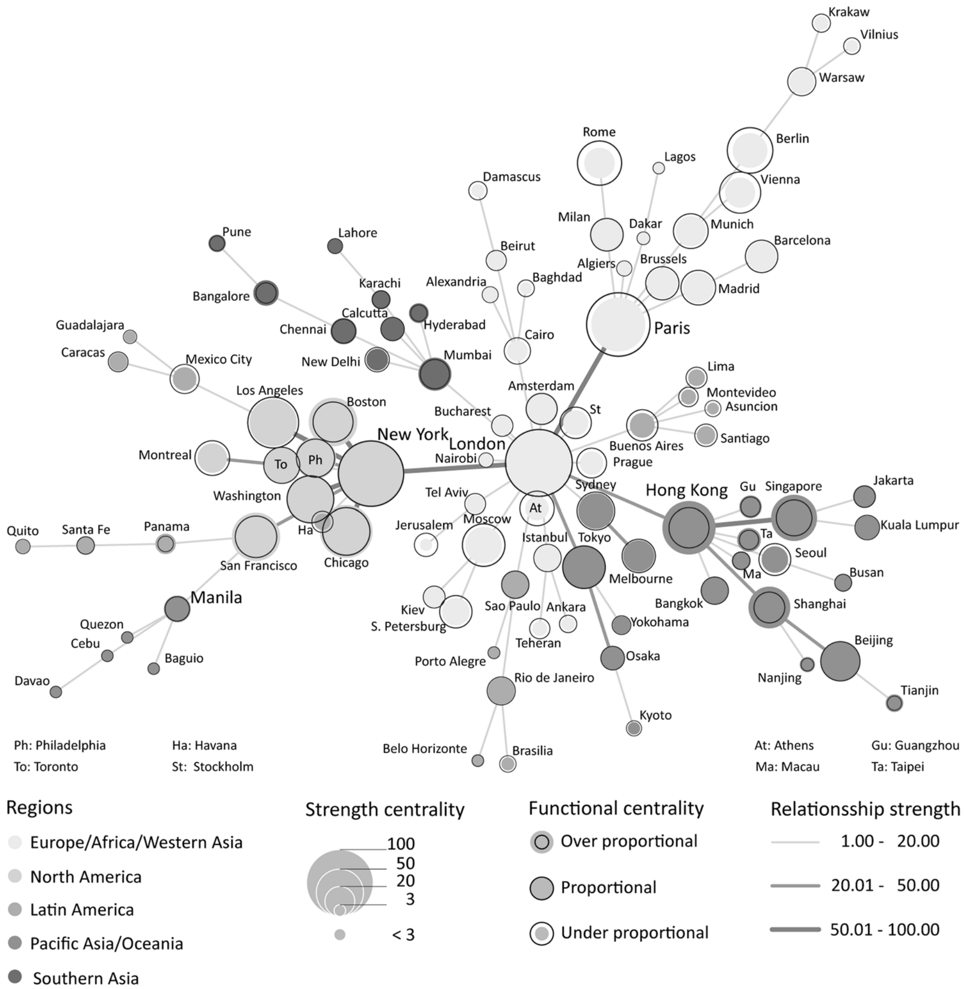

Figure 5. Economy/technology world city network with 95 hub cities.

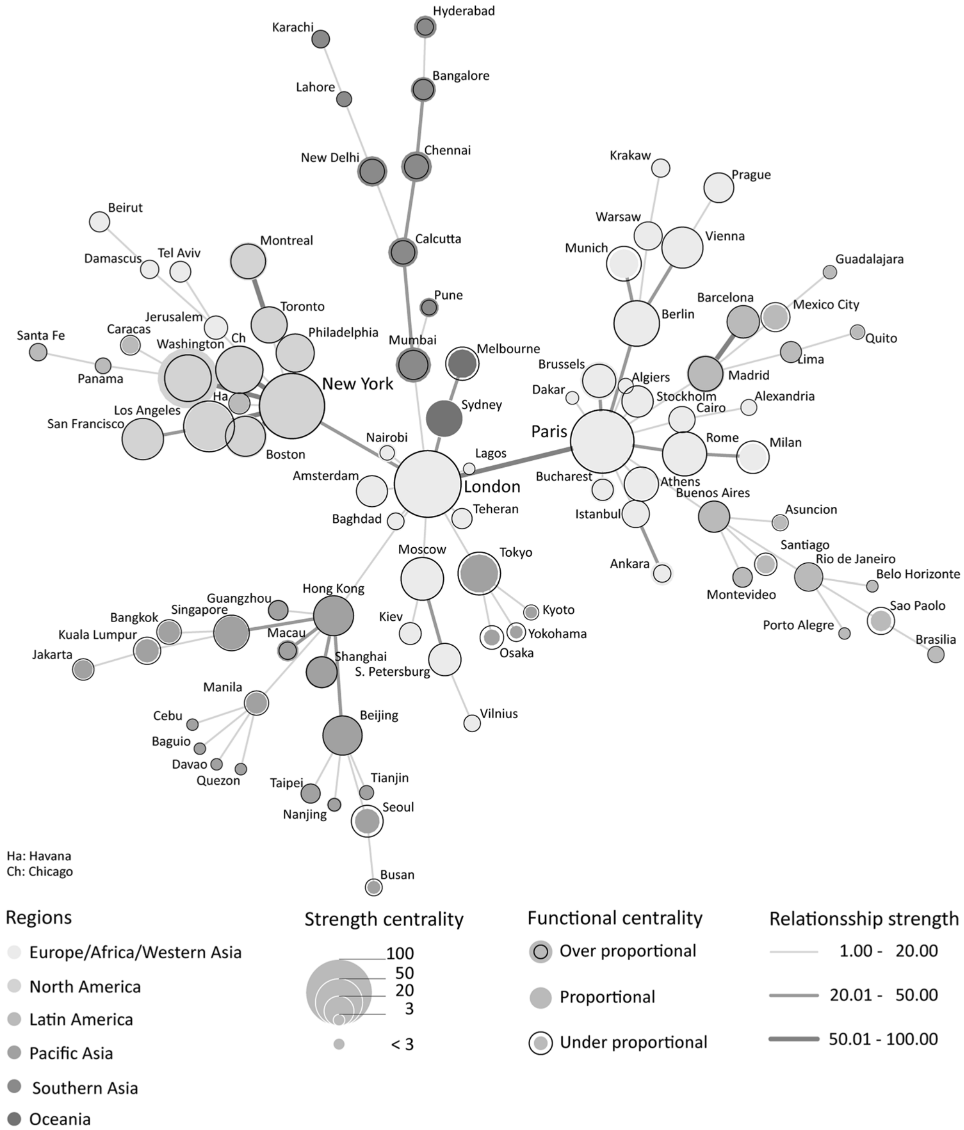

Figure 6. Politics world city network with 95 hub cities.

ninety-five hub-cities considered . The PFNet filtering approach allows us to uncover latent hierarchical structuring in the world city network, distinguishing, for example, smaller, less central cities from the few larger and highly central cities which drive the integration of the cities in the world city network, as can be seen in figures 5 and 6. To visualize the filtered network structure, we chose the GEM algorithm (Frick et al., 1994) (step 2 in figure 4). Cities (ie, nodes) which are strongly related to one another are placed closer together in the network than less strongly related cities (Tobler, 1970). The ninety-five cities are depicted as point symbols and their relations are visualized as edges in the network. The normalized centrality is calculated for each city in the network based on the full city – city matrix of ninety-five cities and 4465 intercity relations. These are depicted following well-established cartographic design principles and using the visual variable size (step 3 in figure 4). Hence, larger point symbols represent cities that are more central (eg, London), while cities that are more peripheral are depicted by smaller point symbols (eg, Lima). While various network metrics could be used to assess a city's role in a network, for our analysis we only consider strength centrality (Wasserman and Faust, 2008). This is defined as the sum of the weight of the direct relationships of one city with its neighbouring cities in a network. We can now depict and compare a city's centrality in a functionally differentiated network, or in a nondifferentiated, generic network, as shown in figures 5 and 6. The size of the gray shaded point symbols shows the normalized centrality of the cities in the functionally differentiated network; the size of the black outline depicts the normalized centrality in the generic, overall network. Cities that have a higher functional centrality compared with the centrality in the generic network can thus be depicted saliently (eg, Washington in figure 6 has a more central role in a political network, compared with its status in the generic network). On the one hand, centrality measures uncover the vertical structuring of the cities from the centre to the periphery of the network. On the other hand, cohesive subgroups identified by the Blondel community detection algorithm (see step 3 in figure 4) (Blondel et al., 2008), highlight the horizontal structuring in the network. In the context of a world city network, cohesive subgroups consist of cities with a high number of mutual relationships and, therefore, they can be considered as relation regions. The affiliation of a city to a relational region is depicted in figures 5 and 6 by different colour hues.

The uncovered intercity relationships are characterized by their function, and their linkage strength (step 4 in figure 4). As discussed earlier, a separate world city network is modelled for each function. The relationship strength and the location of the edges can vary within and across the modelled networks. The relationship strength is depicted by varying edge width, and color value. Larger and darker edges show stronger relationships between two cities (eg, the link Paris – London) compared with thinner and brighter relationships (eg, the link Buenos Aires – Lima).

4. Results

One of the main goals of this research is the bottom up modelling, visualization, and exploratory analysis of a functionally differentiated world city network. With the methods described earlier we uncovered eight functions (with the five most relevant terms that describe each) as follows:

(1) Sport: Olympics, summer, Olympic, players, and cup;
(2) People: actors, singers, film, politicians, and players;
(3) Art: actors, painters, films, film, and singers;
(4) Education: educational, institutions, actors, faculty, and academics;
(5) Environment: towns, settlements, cities, villages, and municipalities;
(6) Society/history/military: general, units, formations, conflicts, and battles;
(7) Economy/technology: companies, ships, built, class, and albums;
(8) Politics: politicians, party, bilateral, representatives, and ministers.

A comprehensive exploration of the content, and the motivation for the labelling of the functions is presented by Salvini (2012). In the remaining part of this section we explore further the uncovered structure for two selected functional world city networks, that is, economy/technology (figure 5), and politics (figure 6), based on ninety-five hub-cities.

As mentioned earlier, we derived the meaning of these two functional networks from the content of the articles, and their assigned categories. The function economy/technology captures, therefore, urban actors such as, service companies, heavy industry, technology companies, and IT firms, including music and fashion businesses. Politicians, party members, political representatives, and ambassadors are part of the uncovered function politics. In figures 5 and 6 we compare a city's status in functionally differentiated networks with the undifferentiated, generic world city network including all functions. Cities that have higher centrality in a functionally differentiated city network (ie, economy/technology or politics) compared with the generic world city network are assigned a graduated circle with a circular outline larger than its fill; cities that have lower centrality are assigned a smaller circular outline, and thus a larger fill, Cities whose centrality does not change between the functionally differentiated network and the generic network are depicted with graduated circles with perfectly matching fills and outlines.

The two world city networks, economy/technology, and politics, computed with PFNet, have three fundamental properties in common:

(1) New York, Paris, and specifically London are the most central cities. The other cities are directly or indirectly connected to them;
(2) The intercity relationships between New York, London, and Paris are the strongest compared with all cities in the network, and thus build the spine of the world city network;
(3) The relational regions that emerge with the Blondel community approach seem to reproduce the well-known world city regions in geographic space. The identified relational regions follow the borders between continents and subcontinents. Four regions are equivalent in the two compared networks: Europe/Africa/Western Asia, North America, Latin America, and Southern Asia. Pacific Asia and Oceania form a common region in the economy/technology network, while Pacific Asia and Oceania are two separate regions in the politics network.

Aside from these similarities, meaningful differences emerge between the economy/technology and the politics city network. These differences concern the cities themselves, as well as their relationships and their location in the relational regions:

(1) While New York, London, and Paris exhibit high centrality in the politics network, Hong Kong arises as the fourth most-central city in the economy/technology network.
(2) The Pacific Asia/Oceania region has a disproportionately large centrality in the economy/technology network. This is particularly apparent for Singapore and Shanghai. The continental European cities (eg, Paris, Rome, Berlin, and Vienna), the Western Asian cites (eg, Jerusalem and Teheran), and the Latin American cities (eg, Mexico City and Buenos Aires) have a disproportionately low centrality in the economy/technology network. In the politics network, national capitals (eg, Washington and Brussels) show a disproportionately high centrality compared with the generic network, as one would expect. Furthermore, the cities from the Southern Asian region have a disproportionately high centrality in the politics network.
(3) At a regional level, in addition to the split of the region Pacific Asia/Oceania the only difference in the politics network seems to be for Barcelona and Madrid. While the two Spanish cities are part of the European/African/Western Asian region in the economic/technology network, they are integrated in the Latin American region in the politics network. This is meaningful considering the shared language and common political history.
(4) Considering the relative locations of cities in the two networks, remarkable differences can be discovered. While the South American cities (eg, Sao Paolo, Buenos Aires, and Rio de Janeiro), the Eastern European cities (eg, Moscow, and Istanbul), and the Egyptian cities are strongly related to Paris in the politics network, they are strongly related to London in the economic/technology network. The relative location in the network of Jerusalem, Tel Aviv, Damascus, and Beirut is also different in the two compared networks: while they are strongly connected to New York in the politics network (perhaps due to the UN), they are related to London in the economy/technology city network, which might be related to their relative closeness to the European market.

5. Evaluation

Even if uncovered networks can replicate the relations between relevant world cities in a meaningful way, as postulated by prior theoretical work, the scientific use of user-generated information may include potential biases in the data and the applied methods, and thus pose possible risks related to the accuracy and generalizability of the results.

Of course this is an inherent problem for any currently employed dataset in global city network research, as discussed earlier. One of the biggest known potential risks that might affect the scientific use of Wikipedia is related to the inherent motivations and thematical expertise of the Wikipedia authors. This in turn might produce a potential bias in topic coverage and entry quality. For Wikipedia specifically, we have to deal with three main potential biases:

(1) Wikipedia English-version authors are mostly young, well-educated, English-speaking men without children (Glott et al., 2010). The world view represented in Wikipedia might, therefore, only represent a small portion of a globalized society.
(2) Wikipedia authors write mostly about topics which are close to them in space and time (Flanagin and Metzger, 2008). Therefore, one would expect the English version of Wikipedia to contain an over representation of topics related to English-speaking regions (ie, space-language bias) and that are of interest to young, well-educated, English-speaking men without children.
(3) As Neal (2012) points out, the structure of the world city network is highly determined by the structure of the dataset chosen for analysis. Therefore, the structure of the uncovered multifunctional world city network might be driven by the structure and internal governance of Wikipedia.

The listed biases are not only relevant for crowd-sourced databases such as Wikipedia, but are also important to consider for any dataset employed in world city network research (Neal, 2012). The question here is not if more or less bias might exist in crowd-sourced data compared with other data, but how one might deal with such kinds of biases, irrespective of the data source. The effects of known potential biases need to be and can be evaluated systematically. By comparing the world city network modelled using the English version of Wikipedia with a multilingual world city network as a reference, we offer one approach to handling potential space-language biases in similarly structured crowd-sourced datasets. The modelling of the reference world city network follows the workflow described in figure 1, but the input data also includes the French, German, and Italian versions of Wikipedia, together with the English version accessed earlier. Certainly, other languages can and should be used, but we chose these four reference languages as they are the four largest of the approximately 300 existing language versions of Wikipedia, and we wish to emphasize the method to deal with a potential language bias. To model this multilingual world city network we extracted 516 cities and their respective relations which are contained in all four selected language versions. The computed weights of the intercity relationships in these four world city networks are normalized between 1 and 1000, so that they can be aggregated to a single multilingual reference world city network.

In the following, we compare the English world city network quantitatively with the multilingual reference world city network to evaluate the potential space-language bias. Specifically, we assess commonly employed network properties as shown in table 1. The selected city centrality measures and the weights of the intercity relationships are compared using Pearson's correlation analysis. Furthermore, regions identified using the Blondel algorithm in the multilingual reference network are compared with those in the English world city network, using the Cohen's K coefficient. We hypothesize that the stronger the correlation between the English and the multilingual reference network, the smaller the effect of a potential space-language bias.

As shown in table 1, all correlations score above 0.80, and thus can be considered to be strong. The Cohen's K coefficient also shows a high correspondence between the relational regions in the two networks compared ($K = 0.89$). These scores confirm the statistically significant correlation between the two analyzed datasets. The strong correlation might be the first evidence against a potential space-language bias in our results.

Table 1. Association strengths between selected properties of the world city network derived from the English version of Wikipedia and that derived from the multilingual versions.

Property	**Association r (Pearson's correlation coefficient)**
Strength centrality of the cities	0.92
Intra-modular centrality of the cities	0.89
Degree centrality of the cities in the Pathfinder network	0.96
Strength of the relationships in the complete world city network	0.88
Strength of the relationships in the Pathfinder network	0.96

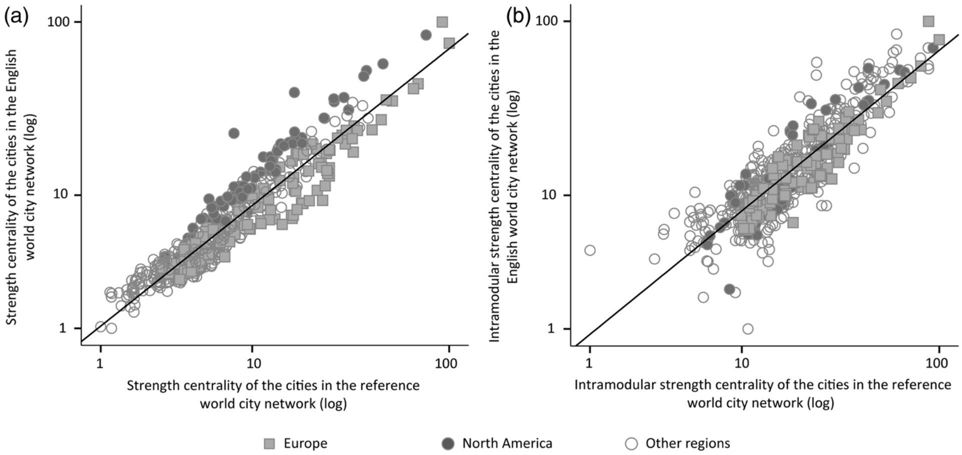

Figure 7. Scatter plots of two centrality measures of cities comparing the reference world city network with the English world city network: (a) strength centrality; (b) intramodular strength centrality.

Considering the first two centrality measures in table 1 (ie, strength and intramodular centrality) in more detail, further analysis of the potential effect of the space-language bias is possible. Figure 7 shows a scatter-plot matrix, plotting the two selected centrality measures against each other. Cities in the European region (light-gray squares) and the North American region (dark-grey dots) are highlighted in figure 7. The scatter plot in figure 7(a) compares the strength centrality of the two selected networks, and shows a pronounced pattern: the centrality of the North American cities (dots) is overestimated in the English world city network (above the 45° diagonal), while the centrality of the European cities (squares) is underestimated in the same network. This overestimation could serve as evidence for the initial assumption of a potential space-language bias on the results. The expected greater coverage of topics in the English speaking world regions does indeed affect the structural properties of the world city network. This pattern disappears in the scatter plot in figure 7(b), which compares the intramodular strength. The distribution of the North American and European cities appears to be distributed randomly above and below the 45° diagonal.

This more detailed comparison across the chosen properties in the English world city network compared with the multilingual reference world city network might suggests that only small space-language biases emerge in the modelled word city network. This additionally shows that with the appropriate methodological caution and a systematic evaluation of the results, a potential data bias can be identified, and systematically reduced.

6. Discussion and new insights for urban geography

Considering the results from a thematic point of view, the most fundamental new insight is the strong influence of a city's geographic location on the resulting relational structure of the world city network. The identified relational regions mostly follow the existing continental and regional borders. This congruence between functional relation and geographic region is also recognizable in the analysis of Derudder et al., (2003), and Rozenblat et al., (2010), but based on different datasets.

We are able to replicate prior work, with London and New York holding the leading positions in the world city network. Friedmann (1986; 1995), Sassen (1991), Alderson and Beckfield (2004), Wall and van der Knaap (2011), as well as Taylor (2004) previously identified the high centrality of these cities. Aside from New York and London, Paris also holds a central position in this network, but with a lower centrality compared with the English-speaking cities New York and London. This is especially visible in the economic network. The position of Paris seems more volatile. While in previous work based on multinational enterprises (eg, Alderson and Beckfield, 2004; Wall and van der Knaap, 2011) the centrality of Paris is comparable with that of London and New York, Taylor (2004), Friedmann (1986; 1995), and Sassen (1991) assign a less central role to Paris.

Replicating Taylor's (2004) findings, Hong Kong rises as a fourth centre in the economic world city network. In our work, Hong Kong is ranked even higher than Tokyo. From a political perspective, the centrality of Hong Kong seems medium-high. The leading position of Hong Kong in the Pacific Asia region appears in contrast to the leading position of Tokyo, which has been identified in prior related work (Alderson and Beckfield, 2004; Friedmann, 1986; 1995; Sassen, 1991; Wall and Van der Knaap, 2011). Furthermore, the centrality of the capitals Washington (United States of America) and Brussels (Belgium and The European Union) in the political world city network matches well with expectations related to the centrality of capital cities in a political network, and might be seen as yet another confirmation of the quality of the chosen approach.

In summary, while we could empirically confirm previously well-established world city network theory with our empirical approach, we are also able to generate new research questions about the latent structures of the world city network in the Pacific Asia region, specifically regarding the debated positions of Tokyo and Hong Kong in this network.

7. Conclusions

In this paper we presented our theoretically based world city network modelling approach. This approach integrates methods from GIScience with social network analysis and automated semantic analyses applied to the user-generated Wikipedia database. The overall goal of our approach is to uncover a multifunctional world network, as theorized by Castells (1996), Alderson and Beckfield (2004) and others. Our methodological contribution rests on a core theoretical concept of world city network formation, namely, the dual function of cities as spatial anchor points and urban actors. These functions are said to be the key drivers of intercity relationships (Pumain, 2006).

Our approach takes advantage of the emergence of massively crowd-sourced, potentially globally generated, and potentially worldwide accessible data such as Wikipedia to model, depict, and analyse a multirelational world city network and dual world city networks.

On applying this new methodological framework to world city research we are able to test long-standing multifunctional world city theory, generate new insights, and explore further research questions in urban geography. Furthermore, leveraging information extracted from about 400 00 shared Wikipedia articles, we are able to model inter-cities relationships at a significantly larger scale than the well-established approaches we reviewed.

Our most distinguishing contribution is the multifunctionality of the modelled world city network. Our network, based on Wikipedia content, includes eight user-generated data-driven functions. In addition to the economic emphasis discussed in prior related work, new city functions clearly emerge: politics, sport, education, environment, history/society/military, art, and people. In contrast to Taylor's (2005) work, these uncovered functions have emerged bottom-up, based on a shared understanding of various Wikipedia article editors, and thus is not determined a priori based on theoretical assumptions. However, the identified functions match well the taxonomy of Smith and Timberlake (1995a), and in addition complement it with two newly uncovered functions: sports and environment.

Furthermore, the integration of state-of-the-art visual analytical methods fills a methodological and analytical gap in the analysis of world city network theory, that was pointed out by Thierstein and Förster (2008).

Of course, method development should be complemented with careful validation and evaluation of the results produced. As detailed in the validation section above, aside from the discussed potentials of using user-generated content in a scientific context, these data might also bear some risks and challenges. The assessed potential influence on our results of the well-known space-language bias in Wikipedia indeed confirms that these risks do exist, and that they have to be considered systematically. However, the emerging bias in the analysis should be controlled and can be corrected with appropriate methods, as we have shown. In particular, the role definition (Guimerà and Amaral, 2005), and the PFnet approach can be employed to reduce the potential bias effect on the results. Therefore, a systematic evaluation of results obtained from user-generated content is always necessary, but this is also true for any other datasets used.

Our carefully evaluated results suggest the utility of crowd-sourced data sources such as Wikipedia for urban geographic purposes, and the validity of our proposed approach. Figures 5 and 6 depict a semantically and geometrically generalized world city network which emphasizes Friedmann's (1986) idea about hierarchical tendencies in world city formation. However, we are aware that the contemporary globalized information society is also shaped by many horizontal and direct relationships. Therefore, further work needs to focus on a broader range of urban geographic aspects and, in particular, on the structure of the world city network beyond hierarchy. Finally, the continuous and rapid growth of the Spanish, Russian, Chinese, and Japanese versions of Wikipedia will in the future allow for more comprehensive space-language bias analyses, and lead to more exhaustive validations of the framework presented here.

References

Alderson A S, Beckfield J, 2004, "Power and position in the world city system" *American Journal of Sociology* **109** 811–851

Blondel V, Guillaume J-L, Lambiotte R, Lefebre E, 2008, "Fast unfolding of communities in large networks" *Journal of Statistical Mechanics: Theory and Experiment* 2008 p10008

Börner K, Chen C, Boyack K, 2003, "Visualizing knowledge domains" *Annual Review of Information Science and Technology* **37** 179–255

Boulton A, Devriendt L, Brunn S D, Derudder B, Witlox F , 2011, "City networks in cyberspace and time: using Google hyperlinks to measure global economic and environmental crises", in I*CTs for Mobile and Ubiquitous Urban Infrastructures: Surveillance, Locative Media and Global Networks* eds JF Rodrigo, F Duarte' C Ultramari, (IGI Global, Hershey, PA) pp 67–87

Castells M, 1996 *The Rise of the Network Society* (Blackwell, Oxford)

Chen C, 1999, "Visualising semantic space and author co-citation networks in digital libraries" *Information Processing and Management* **35** 401–420

Dearholt D W, Schvanenveldt R W, 1990, "Properties of pathfinder networks", in *Pathfinder Associative Networks: Studies in Knowledge Organization* Ed. R W Schvanenveldt (Ablex, New York) pp 1–30

Derudder B, Liu X, 2013, "Analysing urban networks through the lens of corporate networks: a critical review" *Cities* **31** 430–437

Derudder B, Witlox F, 2005, "An appraisal of the use of airline data in assessing the world city network: a research note on data" *Urban Studies* **42** 2371–2388
Derudder B, Taylor P J, Witlox F, Catalano G, 2003, "Hierarchical tendencies and regional pattern in the world city network. A global urban analysis of 234 cities" *Regional Studies* **37** 875–886
Derudder B, Hoyler M, Taylor P J, Witlox F, 2012 *International Handbook of Globalization and World Cities* (Edward Elgar, Cheltenham, Glos)
Derudder B, Liu X, Kunaka C, Roberts M, 2014, "Theconnectivity of South Asian cities in infrastructure networks" *Journal of Maps* **10** 47–52
Fabrikant S I, Salvini M M, 2011, "Charting the ICA world of cartography 1999–2009", in P*roceedings of the 29th International Cartographic*)*, Paris,* http://www.geo.uzh.ch/˜sara/pubs/fabrikant_salvini_ica11.pdf
Fabrikant S I, Montello D, Ruocco M, Middleton R, 2004, "The distance-similarity metaphor in network display spatializations" *Cartography and Geographic Information Science* **31** 237–252
Flanagin A J, Metzger M J, 2008, "The credibility of vounteered geographic information" *GeoJournal* **72** 137–148
Ford L R, Fulkerson D R, 1956, "Maximal flow through a network" *Canadian Journal of Mathematics* **8** 399–404
Frick A, Ludwig A, Mehldau H, 1994, "A fast adaptive layout algorithm for undirected graphs", in *Proceedings of the DIMACS International Workshop on Graph Drawing Princeton, NJ* Eds R Tamassia, IG Tollis (Springer, Berlin) pp 388–483
Friedmann J, 1986, "The world city hypothesis" *Development and Change* **17** 69–84
Friedmann J, 1995, "Where we stand. A decade of world ctiy research", in *World Cities in a World System* Eds P L Knox, P J Taylor (Cambridge University Press, Cambridge) pp 21–47
Glott R, Schmidt P, Ghosh R, 2010, "Wikipedia survey - overview of results", http://wikipediasurvey.org/
Goodchild M F, 2007, "Citizens as sensors. The world of volunteered geography" *GeoJournal* **69** 211–221
Graham M, Zook M, 2013, "Augmented realities and uneven geographies: exploring the geo-linguistic contours of the web" *Environment and Planning A* **45** 77–99
Guimerà R, Amaral L A N, 2005, "Cartography of complex networks. Modules and universal roles" *Journal of Statistical Mechanics* 2005 1–12
Hecht B, 2007, "Utilizing Wikipedia as spatiotemporal knowledge repertory", master thesis, Department of Geography, University of California, Santa Barbara, CA
Hecht B, Moxley E, 2009, "Terabytes of Tobler. Evaluating the first law in a massive, domain-neutral representation of world knowldge", in *Proceedings of the Spatial Information Theory, International Conference* (*COSIT*), Aber Wrac'h, *France* Eds KG Hornsby, C Claramunt, M Denis (Springer, Berlin) pp 88–105
Hecht B, Raubal M, 2008, "GeoSR. Geographically explore semantic relations in world knowledge", in *Proceedings of the 11th International Conference on Geographic Information Science* (*AGILE*), *Girona* Eds L Barnard, A Friis-Christensen, H Pendt, I Compt (Springer, Berlin) pp 95–113
Hennemann S, 2013, "Information-rich visualisation of dense geographical networks" *Journal of Maps* **9** 68–75
Hennemann S, Derudder B, 2014, "An alternative approach to the calculation and analysis of the connectivity in the world city network" *Environment and Planning B: Planning and Design* **41** 392–412
Howe J, 2006, "The rise of crowdsourcing" *Wired* 14.06
Hoyler M, Watson A, 2013, "Global media cities in transnational media networks" *Tijdschrift voor Economische en Sociale Geografie* **104** 90–108
Hymer S, 1972, "The multinational corporation and the law of the uneven developement", in *Economics and World Order. From the 1970's to the 1990's* Ed. J N Bhagwati (Macmillan, London) pp 113–140
Malecki E J, 2002, "Geography of the Internet's infrastructure" *Economic Geography* **78** 399–424
Manning C D, Raghavan P, Schütze H, 2008 *Introduction to Information Retrieval. Classical and Web Information Retrieval Systems. Algorithms, Mathematical Foundations and Practical Issues* (Cambridge University Press, Cambridge)

Neal Z P, 2012, "Structural determinism in the interlocking world city network" *Geographical Analysis* **44** 162–170
Ponzetto S P, Strube M, 2007, "Deriving a large scale taconomy from Wikipedia" *Journal of Artificial Intelligence Research* **30** 181–212
Pumain D, 2006, "Systèmes de villes et niveaux d'organisation", in *Morphogenèse* Eds P Bourgine, A Lesne (Berlin Editions, Paris) pp 239–263
Rozenblat C, 2004, *Tissus de villes. Réseaux et systèmes urbains en Europe Habilitation à diriger des recherches* (Université Montpelier III, Montpellier)
Rozenblat C, Pumain D, 2007, "Firm linkages, innovation and the evolution of urban systems", in *Cities in Globalization: Practices,Policies and Theories* Ed P J Taylor (Routledge, London) pp 130–156
Rozenblat C, Mélançon G, Bourqui R, Auber D, 2010, "Comparing multilevel clustering methods on weighted graphs. the case of worldwide air passenger traffic 2000–2004", in *Multilevel Analysis and Visualization of Geographical Networks* Eds C Rozenblat, G Mélançon (Springer, Berlin) pp 141–154
Rutherford J, Gillespie A, Richardson R, 2004, "The territoriality of Pan-European telecommunications backbone networks" *Journal of Urban Technology* **11**(3) 1–34
Salvini M M, 2012, *Spatialization von Nutzergenerierten Inhalten für die Explorative Analyse des Globalen Städtenetzes* (Geographisches Institut, Universität Zürich, Zürich)
Sassen S, 1991, *The Global City. New York, London, Tokyo* (Princeton University Press, Princeton, NJ)
Skupin A, Fabrikant S I, 2007, "Spatialization", in *The Handbook of Geographical Information Science* Eds J P Wilson, A S Fotheringham (Blackwell, Reston, VA) pp 61–79
Smith D A, Timberlake M, 1995a, "Cities in global matrices. toward mapping the world-system's city system", in *World Cities in a World-system* Eds P L Knox, P J Taylor (Cambridge University Press, Cambridge) pp 79–97
Smith D A, Timberlake M, 1995b, "Conceptualising and mapping the structure of the world system's city system" *Urban Studies* **32** 287–302
Smith D A, Timberlake M, 2001, "World city networks and hierarchies, 1977 – 1997: an empirical anlysis of global air travel links" *American Behavioral Scientist* **44** 1656–1678
Steyvers M, Griffiths T, 2007, "Probabilistic topic models", in *Handbook of Latent Semantic Analysis* Eds T K Landauer, D S McNamara, S Dennis, W Kintsch (Lawrence Erlbaum, Hillsdale, NJ) pp 427–448
Taylor P J, 1997, "Hierarchical tendencies amongst world cities. A global research proposal" *Cities* **14** 323–332
Taylor P J, 2001, "Specification of the world city network" *Geographical Analysis* **33** 181–194
Taylor P J, 2004, *World City Network. A Global Urban Analysis* (Routledge, London)
Taylor P J, 2005, "Leading world cities. Empirical evaluations of urban nodes in multiple networks" *Urban Studies* **42** 1593–1608
Taylor P J, Catalano G, Walker D R F, 2002, "Measurement of the world city network" *Urban Studies* **39** 2367–2376
Taylor P J, Derudder B, Saey P, Witlox F (Eds), 2007a, *Cities in Globalization: Practices, Policies and Theories* (Routledge, London)
Taylor P J, Derudder B, Witlox F, 2007b, "Comparing airline passenger destinations with global service connectivities: a worldwide empirical study of 214 cities" *Urban Geography* **28** 232–248
Taylor P J, Ni P, Derudder B, Hoyler M, Huang J, Witlox F (Eds), 2011, *Global Urban Analysis: A Survey of Cities in Globalization* (Earthscan, London)
Thierstein A, Förster A, 2008, *The Image and the Region. Making Mega-city Regions Visible!* (Lars Müller, Baden)
Tobler W, 1970, "A computer movie simulating urban growth in the Detroit region" *Economic Geography* **46** 234–240
Toly N, Bouteligier S, Smith G, Gibson B, 2012, "New maps, new questions: Global cities beyond the advanced producer and financial services sector" *Globalizations* **9** 289–306

Voss J, 2006, "Collaborative thesaurus tagging the Wikipedia way" *ArXiv Computer Science*, http://arxiv.org/ftp/cs/papers/0604/0604036.pdf

Wall R, Van der Knaap B, 2011, "Sectoral differentiation and network structure within contemporary worldwide corporate networks" *Economic Geography* **87** 267–308

Wasserman S, Faust K, 2008, *Social Network Mnalysis. Methods and Applications* (Cambridge University Press, Cambridge)