Returning the particular : understanding hierarchies in the Belgian logistics system

Reference:
Beckers Joris, Vanhoof Maarten, Verhetsel Ann.- Returning the particular : understanding hierarchies in the Belgian logistics system
Journal of transport geography - ISSN 0966-6923 - 76(2019), p. 315-324
Full text (Publisher's DOI): https://doi.org/10.1016/J.JTRANGEO.2017.09.015
To cite this reference: https://hdl.handle.net/10067/1465350151162165141

Returning the particular: understanding hierarchies in the Belgian logistics system

Joris Beckers[1*], *Maarten Vanhoof*[23], *Ann Verhetsel*[1]

1: joris.beckers@uantwerpen.be (*corresponding author)

Department of Transport and Regional Economics, University of Antwerp. Prinsstraat 13, 2000 Antwerpen, Belgium

2: m.vanhoof1@newcastle.ac.uk

Open Lab, Newcastle University. 89 Sandyford Rd, Newcastle upon Tyne NE1 8HW, UK

3 :

Departement SENSE, Orange Labs. 44 Avenue de la République, 92320 Châtillon, France

Keywords: Hierarchies, big data, logistics, community detection, damping values

Abstract: The recent proliferation of big data sources has given rise to a data deluge. Network theory has become the standard methodology to frame, develop and analyze such massive datasets. In line with the critique of Schwanen (2016), we argue in this paper that initiatives confronting network-based insights with (qualitative) location- and domain-specific insights are necessary in understanding, discussing and advancing the role network analysis can play in geography. By iterating a community detection algorithm to achieve different levels of communities and quantifying the borders between them through damping values (as proposed in Grauwin et al., 2017), we show how to derive the hierarchical structure within the logistics buyer-supplier network in Belgium. This allows for a richer geography, which has been missing in current big data studies.

1. Introduction

Over the last decade, a long list of big data sources have emerged that give rise to a true data deluge when it comes to capturing spatial economic systems. Recent advances in information technologies now allow for the continuous, positional tracing of a multitude of objects. GPS technologies can equip and trace almost any object, from watches to airplanes, revealing the true extent of worldwide connectedness. Online social networks and smartcard data capture the movement of large-scale populations, credit card transactions enable to follow up on low-level economic transactions within and between territories, and mobile phone data are capable of capturing large-scale communication patterns.

Obviously, the availability of such big data sources has revolutionized the study of movement and interaction in spatial economic systems. A common demeanor in studies integrating these new data sources is their reliance on network theory as both a methodological and conceptual framework. With applications ranging from the mere description of structural network properties to the simulation of complex dynamics, it is more than fair to state that network methodologies have widely advanced the empirical understanding of massive datasets describing spatial economic systems. Advancements, for instance, have been made in understanding the structure of air transportation networks and their potential role in global epidemics (Colizza et al., 2006; Guimerà et al., 2005); in detecting statistical properties of large-scale human mobility and elaborating their use in predictive models (Gonzalez, Hidalgo, & Barabasi, 2008; Simini, González, Maritan, & Barabási, 2012; Song, Koren, Wang, & Barabási, 2010); or in uncovering the spatial extent of interacting communities, hereby challenging existing delineations of space (Blondel et al., 2010; Ratti et al., 2010; Sobolevsky et al., 2013; Thiemann et al., 2010).

Despite the progress made, there has been a growing critique on the empiricist, positivist nature of big data and the related use of network analysis in geography (Graham & Shelton, 2013; Kitchin, 2013; Kwan, 2016). With respect to big data and network analysis in Transport Geography, Schwanen (2016) states that: "These methodological developments are to be welcomed because of … their generative character… . At the same time, caution needs to be exercised for epistemological reasons. There is a risk that generality – regularities, laws, basic principles, … - comes to trump particularity in the form of place and time specificity, uniqueness, singularity and also local knowledge again." (Schwanen, 2016, p2)

In this work we deem this critique to be valid. We discuss how current network methodologies are erasing the heterogeneity between observations and are pushing findings to be considered from a 'network internal perspective' only. Although we reckon that interdisciplinary discussion might be leading recent developments in network analysis to increasingly incorporate aspects of space, context and hierarchy, we argue that initiatives confronting network-based insights with (qualitative) location- and domain-specific insights, which Miller (2017) terms *mesogeography*, are necessary in understanding, discussing and advancing the role network analysis can play in (transport) geography.

Specifically, we attempt to demonstrate how fruitful the confrontation between local knowledge and network-based insights can be. Therefore we adhere to a recently developed network analysis tool as proposed by (Grauwin et al., 2017) that allows, for each node in a network and given a chosen community detection algorithm, the identification of its local, regional and national role and related connectivity. We show how the calculation of damping values in this method can be used to facilitate the confrontation with local insights and, by means of a case study on the Antwerp region, how relevant such knowledge can be for understanding, in this case, logistics systems.

Ultimately, we believe in the benefits of establishing clear touching points between network methodologies and the typical local, spatial and qualitative knowledge of (transport) geography. Our proposed methodology offers a first step in the development of one such touching point, while pushing the discussion on the applicability of community detection in (transport) geography. It is our argument that confrontations like this are beneficial for understanding, discussing, and advancing the role network analysis can play in transport geography.

2. The popularity of the network paradigm in geographical big data studies

The critique on the empirical and positivist nature of current big data practices, including network approaches, relates to a long history of reflection and debate on quantitative approaches within the geographical discipline (Graham & Shelton, 2013; Kitchin, 2013; Kwan, 2016; Schwanen, 2016), ranging from Haggett's reflection on his own pioneer book 'Locational Analysis in Human Geography' (Johnston & Sidaway, 2015), to numerous debates on the quantitative revolution, positivist approaches and, more recently, the role of GIS and the algorithmic treatment of big data (Goodchild, 1992; Graham & Shelton, 2013; Kwan, 2016). In other words, the rise of big data and the revival of network theory in transport geography do not necessarily represent entirely new 'threats'. Rather one can consider them as a next wave of quantitative analyses; this time being fueled by the emergence of new, different and bigger datasets, and the technical and methodological capabilities to treat them.

There is a lot to be said about the way this big data research and network analysis is integrating in current geographic scholarship, ranging from paradigm development to data-access inequality. However, one observation from Schwanen (2016) seems extremely relevant for this paper and, more in general, the field of transport geography. It is the observation that network science tends to erase heterogeneity and proposes differentiation solely based on characteristics of the network. Although explanations can be “imported from the outside” (for example by linking contextual information with connectivity measures), this approach heavily encourages explanations to be based on the network itself, leaving other perspectives sidelined. As such, the application of network science in transport geography is creating a reinforcing system that is “nudging geographers and transport researchers into adopting the internal network perspective” (Schwanen, 2016, p6).

This critique is easily illustrated by two widely used applications of network analysis in transportation geography: the analysis of degree distributions and the application of community detection algorithms, but is similarly relevant for other network applications and big data analytics that can be equally prone to the domination of an internal perspective

A first illustration of our critique is in the analysis of degree distributions. Many real-world networks, including economic, transport, mobility and communication networks, are found to have power-law like distributions. Their study has allowed the identification of scaling in networks, which in turn implies the presence of a hierarchical structure, i.e. with nodes of different importance (Barabasi & Albert, 1999; Jiang, 2013; Levy & Solomon, 1997; Pumain, 2006). Degrees themselves, however, are merely (aggregated) connectivity measures, or thus structural properties of the network. This means that they possess little to no information about local context, except for the relative one defined in the network. As such, a large part of the heterogeneity amongst nodes is simply not considered which, consequently, makes it hard to approach the prevalence of degrees from another perspective than the network one. Insights from local context and processes, geographical differences, or spatial relations are hence all sidelined in favor of a, both methodological and conceptual, internal network perspective like, for instance, the popular preferential attachment growth model (Barabasi & Albert, 1999).

Second, the use of community detection algorithms forms a second illustration of our critique. Community detection on spatial interaction networks (be it transport, economic, mobility or communication networks) has gained popularity ever since it was shown to allow for empirical regionalization. Although initially surprising, it has now been well established that in many spatial interaction networks communities tend to form spatially homogenous and contiguous groups, highlighting the spatial component of their linkages (Blondel et al., 2010; Nelson & Rae, 2016; Ratti et al., 2010; Thiemann et al., 2010). The implicit erasing of heterogeneity in community detection algorithms, however, lies in the assumption that all links in the networks (often expressed as the aggregated intensity of interaction between locations) are equal, no matter their context of occurrence in real life. Additionally, the primarily retrieved information from community detection exists of which node was classified in which community often revealing little to no information on which roles individual nodes have played in the detection process, which position they take in the entire network, or the attributes that distinguish between them, either in the network or in reality.

It is rather easy to remark that both these elements stimulate false interpretations of homogeneity within detected communities, and of comparability between whatever interactions make up the links in the network. Moreover, and with respect to the typical spatially homogenous communities found in interaction networks, both elements contribute to the imposition of an internal network perspective that promotes the intensity of interaction (or thus the structure of the network) to be the (only) driver of regionalization. This is a very restrictive, even false, perspective as intensity of interaction is simply not the main driver of regionalization, nor is it the reason for the existence of empirically observed communities or does it explain the complex organizations of such communities. Here again, the point is that alternative reasoning is easily excluded from both the methodological and conceptual setup of the network approach.

The question then becomes how, as a transport geography research field, to overcome the imposition of an internal network perspective when adhering to network analysis. We distinguish two possible ways, although more probably exist.

A first way is to stimulate the integration of geographical elements, like space, context and hierarchy in network methodologies. By now, such developments are well underway. To name a few, there is an expanding research field on spatial networks integrating spatial aspects like distance and position in network analysis (Barthelemy, 2010), multilayer, and especially multiplex networks, are allowing to differentiate between nodes and links on several levels instead of only one (Kivelä et al., 2014), and community detection algorithms have been well advanced to allow for overlapping communities and hierarchical structures (Fortunato & Hric, 2016). Somehow, however, it remains an open question whether these developments are influenced by interdisciplinary discussion between (transport) geography and network science, or rather from a next step in advancement of the latter. In other words, it remains to be seen to which degree the methodological developments in network science allow for enough depth to be applied to research questions and themes in transport geography.

A second way is for the field of transport geography to strive to confront, assess, reconcile and eventually appropriate elements of network analysis in its knowledge production. Appropriation done this way offers possibilities to critically asses network methodologies, challenge the hegemony of the internal network perspective and direct insights to themes that are of explicit interest in transport geography; all of which would advance the discussion on the role of network analysis in transport geography. To allow for confrontation in the first place, however, clear touching points between network methodologies and the typical local, spatial and qualitative knowledge base of transport geography need to be established. This requires, currently rare, case studies to be elaborated that embrace the inner workings of network analysis and show how their information can be coupled back to geographical knowledge.

Specifically, our interest in Grauwin et al.'s (2017) method lies in the possibility to extract information about the relation of all nodes to different hierarchical levels in the network. As such, information on hinterlands, the strength of borders, and the roles and relations with respect to the wider network can be retrieved for all individual locations. We consider such information a huge opportunity to be confronted with (qualitative) location and domain-specific knowledge and elaborate our point based on a case study of a logistics buyer-supplier network in Belgium, with a focus on Antwerp.

3. Methodology

Recently, Grauwin et al. (2017) proposed a network methodology to improve spatial interaction models, like the well-known gravity or radiation models. Their improvement exists in the replacement of continuous distance with a discrete distance denoting only the transgression of "hierarchical borders". The prediction of flows between locations hence depends on the number of transgressions that needs to be made between two locations with each transgression diminishing the predicted amount of interactions. Interestingly, in their framework the rate at which each transgression diminishes the predicted interaction, or thus the strength of each border, is a network-based measure called "damping value" that is unique for each location. In other words, Grauwin et al. (2017) show how, at least for communication networks in different countries, a spatial interaction model that is based on a discrete, yet location dependent description of distance, outperforms standard interaction models.

Relating to the description of two possible ways to overcome the internal network perspective in previous sections, Grauwin et al.'s (2017) work clearly classify under developments in network methodologies that incorporate elements of geography, in this case being the heterogeneous experience of distance by different locations. Although the construction of damping values and their proof of usability in spatial interaction models is a remarkable feat, in this paper we are not looking to replicate the methodology to another case. Rather, we are interested in the way damping values, and their network-based calculation of individual locations, can be confronted with local geographical knowledge in order to stimulate discussion on their interpretational value. As such, our work classifies under the second possible way described earlier; in which confrontation of insights is facilitated.

3.1 Calculating damping values

To calculate hierarchical borders, Grauwin et al. (2017) iterate a standard community detection algorithm on a given network. As each iteration groups tightly connected nodes in communities, the classification of a node in one community installs a border with all nodes classified in other communities. As such, in a first iteration, L_3-communities are defined based on the entire network. Next, a second iteration of the community detection algorithm is run, this time within each of the previously defined L_3-communities, resulting in a set of L_2-communities. Finally, a third iteration of the community detection algorithm is performed on the network of each of the created L_2-communities, resulting in a set of L_1-communities. In other words, the iterative community detection provides for a multi-level spatial delineation of the given network resulting in a set of L_1, L_2, L_3-community borders specific to each node in the network that we will continue to call *'hierarchical borders'*. Having defined the hierarchical borders of a network, we can start to assess the importance of each hierarchical border for each node. To do so, Grauwin et al. (2017) propose a measure called *damping value*:

$$q_i^h = \frac{T_i^{h+1}}{W_i^{h+1}} \frac{W_i^h}{T_i^h}$$

with h representing the set of nodes that are at a hierarchical distance h of node i. The hierarchical distance between two nodes h is 1 if both nodes are located in the same L_1 region, 2 if they are in different L_1 regions but the same L_2 region, 3 if they are in different L_2 regions but the same L_3 and 4 if they are located in different L_3 regions. T_i^h the total amount of linkages from node i to all zip codes at distance h and W_i^h the total amount of linkages from nodes within hierarchical distance h from node i (Grauwin et al., 2017). Interested readers may have noticed we flipped the terminology of the levels compared to the original authors. This mainly to ease understanding but also because it allows for more flexibility and comparability in case more than three levels will be identified in another study.

The interpretation of the calculation of the damping value q_i^h is rather straightforward. As each node is located in a specific L_1, L_2 and L_3 community, the importance of each hierarchical border (for each node) can be easily characterized by comparing the relative strength of a node in a certain hierarchy level, i.e. T_i^h/W_i^h, with its relative importance one level up the hierarchy, i.e. T_i^{h+1}/W_i^{h+1}. As a consequence, a low q_i^h value indicates a strong border between two hierarchical levels, meaning that a change from hierarchical level *h* to the next *h+1* is, for the investigated location *i*, a relevant change in its importance. Viewed from the hinterland perspective, low q_i^h values hence indicate that the investigated location is relatively more important in the hinterland at level *h* compared to the hinterland at level *h+1*. Consequently, a high q_i^h value implies the contrary indicating a gain in importance of the investigated location *i* when transgressing to the next hierarchical level. The different levels and damping values are summarized for an example zip code in Figure 2.

Remark that the definition of low and high damping values is a relative concept. Logically, damping values are expected to be below 1 as transgressing to a next hierarchical level implies a growing number of other localities (or in the hinterland metaphor, an extension of the hinterland) against which the locality under consideration is compared in order to determine its 'importance'. Currently, Grauwin et al. (2017) is the only research that has calculated damping values and this for large networks of mobile phone communication in different countries. The average damping values they calculated differ slightly from country to country but remain within the 0.10-0.25 range. Most remarkably, however, they found distributions of damping values to be similar for all different hierarchical borders (so for each hierarchical level *h = 1,2,3*) within each country.

3.2 K-means clustering

To understand the variation in damping values between nodes, we perform a k-means clustering on all nodes using their set of damping values for the three hierarchical levels as attributes. The chosen algorithm does not take the location of the nodes explicitly into account, which makes any resulting geographical image the outcome of the spatiality inherent to the original network. In our case, *k* is defined in an informal way by plotting the within-group sum of squares for a range of *k*'s choosing the partition that coincides with the bending point in the resulting graph, i.e. where an additional class has less impact on the within-group variation (Everitt & Hothorn, 2010). This approach is similar to the scree-plot used in factor analysis.

3.3 Community detection

Remark that in our case, we apply the 'Louvain method' community detection algorithm proposed by Blondel et al. (2008), contrary to Grauwin et al. (2017) who use the 'Combo' algorithm proposed in Sobolevsky et al. (2014). We recourse to the 'Louvain method' because of its widespread application in literature, its strong performance in a comparative analysis (Lancichinetti & Fortunato, 2009) and because of the familiarity due to earlier publication with the same dataset further used in this paper (Beckers et al., 2015). Comparison of the results between Louvain and Combo on our dataset showed little differences in obtained communities for all iterations and, as such, little differences in calculated damping values for both approaches.

Two known limitations of the Louvain method might be relevant for our case. Firstly, being a modularity-optimization approach, the detection of communities tend to operate at a coarse level preventing the formation of small communities (Fortunato & Barthelemy, 2006). An adapted Louvain algorithm to cope with this resolution limit exists, but demands prior knowledge on the network structure (Delvenne et al., 2013). In this paper however, we aim to detect various hierarchical levels without privileging one of them, finding smaller communities within larger ones, hereby partly bypassing the resolution problem. Secondly, the Louvain method has a dependency on the order of the input (network) data. To avoid this problem 100 random iterations were averaged into a final set of communities for the different levels with a membership value for each node (see Figure 1)

Ultimately, one can wonder why an iterative application of a standard community detection algorithm is preferred over hierarchical community detection algorithms and/or methods that allow for overlapping communities (Fortunato & Hric, 2016). The first reason is that our goal is to identify geographical hierarchy, which, as the results show, is inherently different and thus independent from the network hierarchy that emerges from the latter algorithms. The second reason is that neither application and interpretation of these methods is trivial, thus requiring a more methodological discussion which is outside the scope of this paper. It remains, however, future work to investigate the definition of, and differences between, damping values based on these approaches.

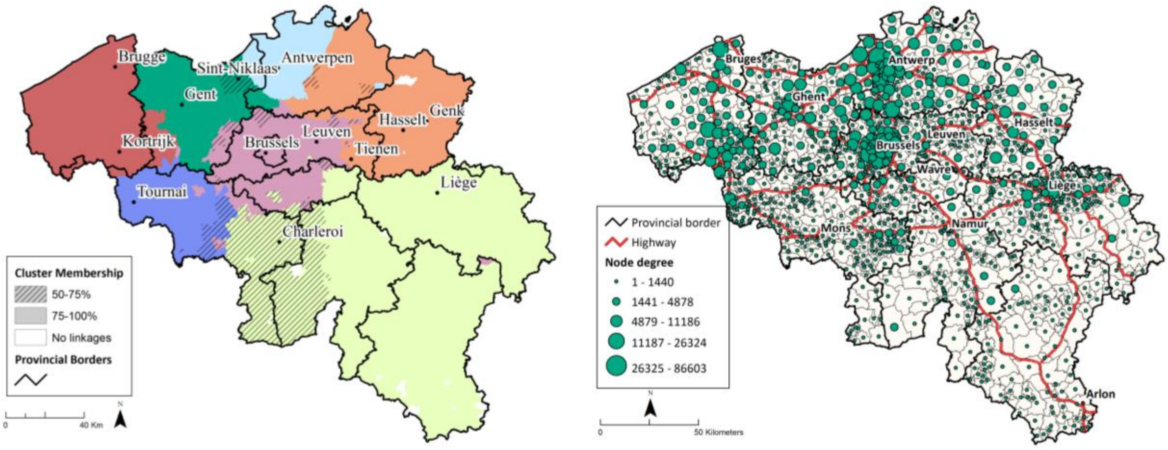
4. Case study: logistics network in Antwerp, Belgium

For the case study, we use a large dataset of micro-economic buyer-supplier linkages in Belgium as provided by the National Bank of Belgium. Each link captures the financial transaction between two companies of which at least one has to be involved in logistics activities. One example is the financial compensation for logistics services where the buyer may be a construction firm and the supplier a logistics carrier.

With over 800,000 observed linkages and the availability of nearly 170,000 unique firms (of which more than 83,000 logistic companies) the dataset represents a spatial expression of the Belgian logistic system in terms of service provision (value added, not transported goods) and enables for applied network analysis. The database spans the year 2011, and aggregates for privacy reasons the total number of buyer-supplier linkages between companies at zip code level. The final network consists of 1155 nodes (zip codes) and the total number of buyer-supplier linkages as weighted, undirected connections between them.

One shortcoming of the dataset is that it includes only linkages for which both the buyer's and supplier's headquarters are situated in Belgium. This probably diminishes the relative importance of some zip codes, especially in major economic centers and border regions. Secondly, the attribution of firms to zip codes is based on headquarter location, which might overemphasize the importance of the economic centers. However, with a strong presence of small and medium sized companies in the dataset, we are confident that the distortion between headquarter location and actual activities is limited and does not compromise the demands for our analysis.

In previous work, Beckers et al. (2015) use community detection analysis on this dataset to link the observation of co-location of logistics companies with higher local intensities of logistics buyer-supplier linkages. This in order to test the occurrence of the assumed Marshallian advantages in logistics clusters (Martin & Sunley, 2003; Porter, 1998). The application of a community detection algorithm on the buyer-suppliers network resulted in seven spatially homogeneous communities, displayed in figure 1A. The majority of these communities center around important industrial hot spots, like Antwerp, Brussels or Ghent and are interpreted as their hinterlands.



A B

Figure 1: Communities (left) and node degree (right) of the buyer-supplier linkages (Beckers et al., 2015)

Although the authors conduct a first attempt to find the drivers of the community delineation by calculating the within-module degrees and participation coefficients as proposed by Guimerà and Amaral (2005), the rich variety (both in size as in geography) of the node degrees observed in figure 1B is neglected in the community detection.

The comparison of figures 1A and 1B illustrates the limitations of two commonly used network approaches, as already mentioned in the introduction. Community detection, although rendering spatially homogenous communities, does not provide information on the local level, and as such, stimulates false interpretations of homogeneity within communities, as has been seen in other work. Second, the degree characteristics provide information on the variety of the nodes in the network, but ignore the spatial relations. The Antwerp community in blue for example has a strong internal heterogeneity with the port as the major economic hub but is surrounded by locations with different logistics importance. In addition, despite being classified in different communities, companies in the Antwerp community have many buyers and suppliers around the country. This information is important since our goal is to understand the buyer-supplier hierarchical structure within the country, but is addressed in neither of the two examples provided in Figure 1.

In what follows we apply Grauwin et al.'s (2017) methodology described in section 3 on the logistics buyer-supplier system in Belgium. First we display the results of the iterative runs of the community detection algorithm after which we focus on the communities and damping values of the Antwerp case. In the discussion we elaborate on the factors that contribute to the geographical and hierarchical variance of the damping value. The different levels and damping values explained above, are summarized for the example zip code in Figure 2 (city of Lier).

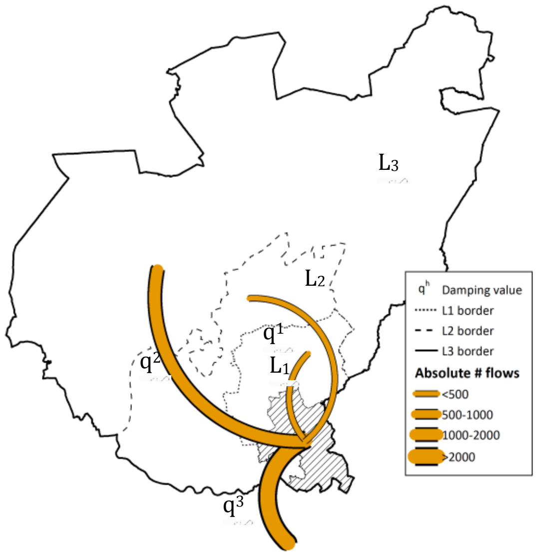


Figure 2: Hierarchical levels and corresponding damping values (q^h)

5. Results

Applying iterative community detection to the buyer-supplier network of Belgian logistics renders L_1 and L_2 communities as displayed in figure 3 (the L_3 communities are shown in figure 1A). A total of 1155 zip-codes were attributed to 7 L_3-communities, 40 L_2-communities and 81 L_1 communities, rendering three hierarchical borders for each individual zip code. Remark that we found almost all defined communities to be spatially contiguous, even though no explicit spatial criterion was introduced by the community detection algorithm. This interesting property aligns with findings in literature based on mobile phone and commuting data (Blondel et al., 2010; Grauwin et al., 2017; Ratti et al., 2010; Sobolevsky et al., 2013; Vanhoof, Smoreda, & Ratti, 2015) and suggests that geography matters, also in logistics buyer-supplier systems. From the point of view of each individual node iterative community detection on the network thus results in a spatial pattern where its closest hinterland is defined by its L_1-community, extending into a wider hinterland defined by its according L_2 community and a widest hinterland made up by its L_3-community before entering the scope of the entire spatial network. This can also be observed in Figure 2.

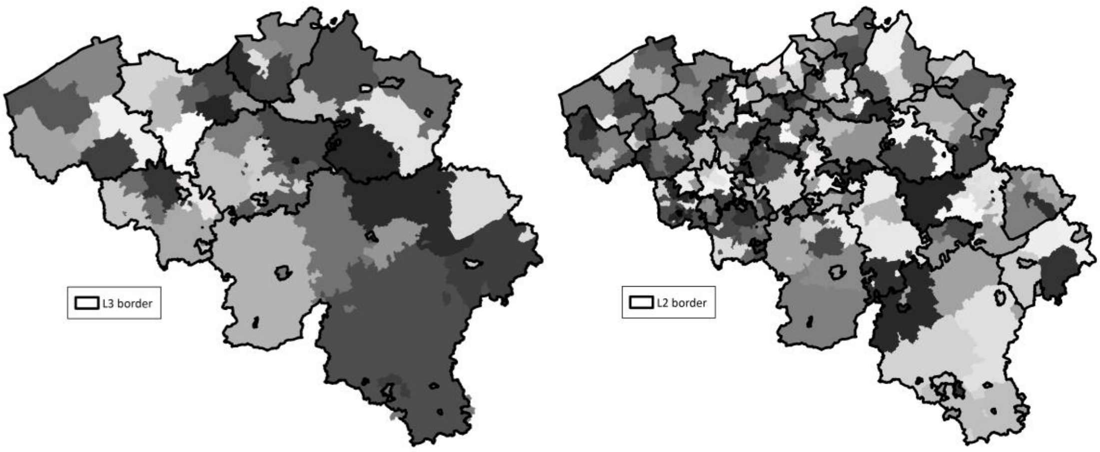


Figure 3: L2 (left) and L1 (right) communities in Belgium. Greyscale is used to differentiate the communities.

The distributions of the damping values for the hierarchical borders at level 1, 2 and 3 for all zip codes is given in figure 4. Distributions are more or less normally shaped but apparently hierarchical borders at different levels have different strengths in the Belgian logistics system with average damping values being 0.32 for the L_1 level, 0.55 for the L_2 level and 0.21 for the L_3 level. This contrasts with findings by Grauwin et al. (2017) where for mobile phone networks in different countries, damping values tended to have similar distributions for the different hierarchical levels.

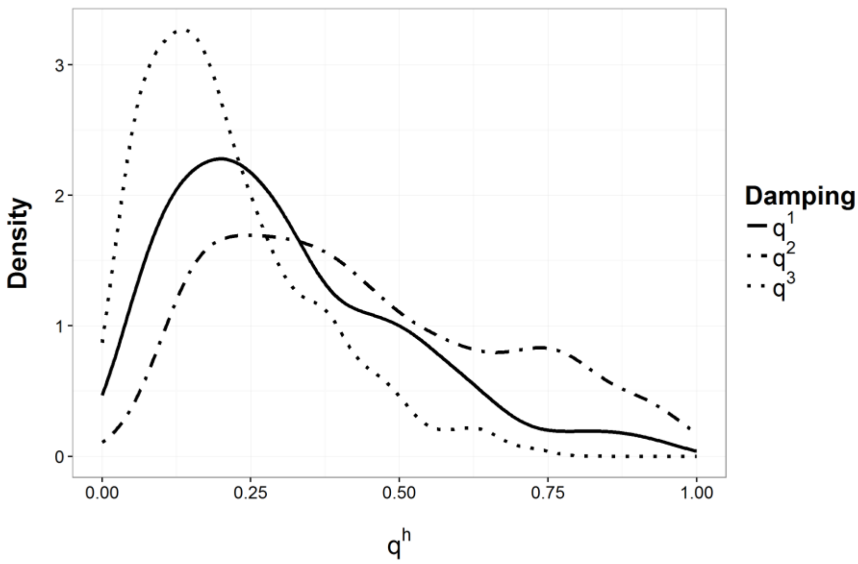


Figure 4: Distribution of different damping values

In order to understand how the definition of damping values at different hierarchical levels can lead to an increased interpretation of the particular context of individual zip codes we focus on the case of one L_3 community, the Antwerp community. The Antwerp community results from the first iteration of the community detection algorithm on the buyer-suppliers network. Figure 5 provides a zoom on the Antwerp L_3 community which corresponds to the blue area in figure 1A. The core of the community is the city of Antwerp of which the center is indicated with a star. The city is surrounded by a ring road and has a large port (purple) located north along both banks of the river Scheldt. Most economic activities outside the city are situated to the south, following the E19 and A12 highways towards Brussels, along the Albert canal towards the east and on the left bank in the west where one finds the most recent developed port areas. The northeastern part of the community is more suburban and even rural. A more detailed description of the location of logistics activities in Antwerp can be found in Verhetsel et al. (2015). In total, the community consists of 55 zip codes or thus 55 nodes in the analyzed logistics buyer-supplier network.

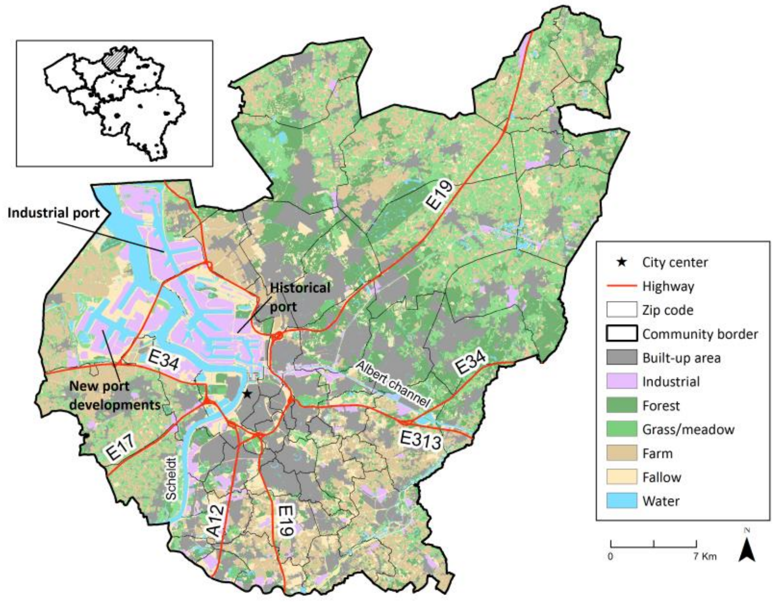


Figure 5: Situating the case of Antwerp

The L_2 and L_1 communities relating to the L_3 Antwerp community are mapped in figure 6. At the L_2 level one can recognize the central city with the old port to its north. The industrial port with its chemical cluster northwest of the city at the fringe of the L_3 community, and the newest port developments on the left bank show up as individual L_2 communities as well. Further our community detection iteration identifies the southern metropolitan area and the more residential/rural northeast. While all detected L_2 communities contain at least two different zip codes, some zip codes form a community of their own at the L_1 level. This is due to strong internal loops in the concerned zip codes, meaning that buyer-supplier linkages within the own zip code are prominent.

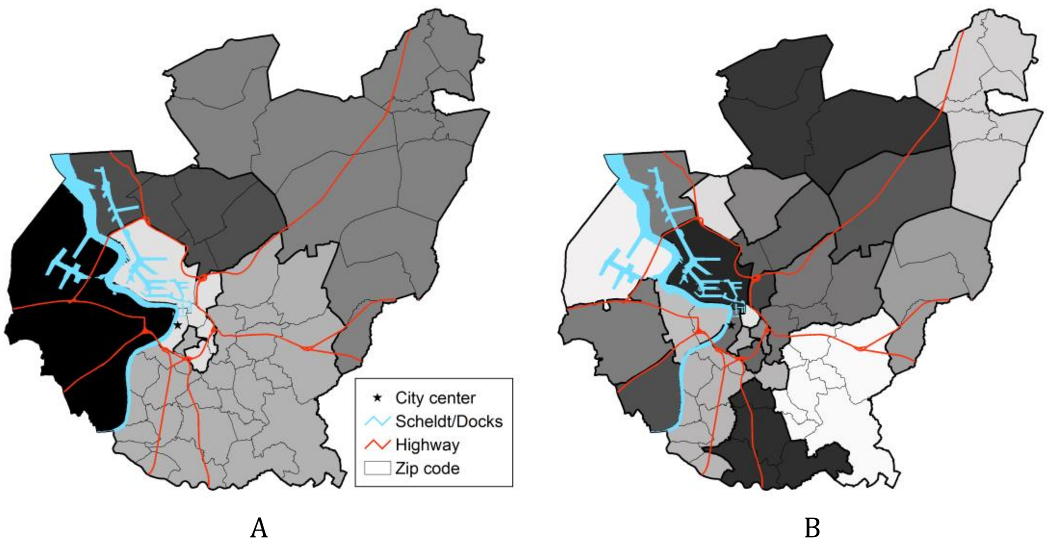
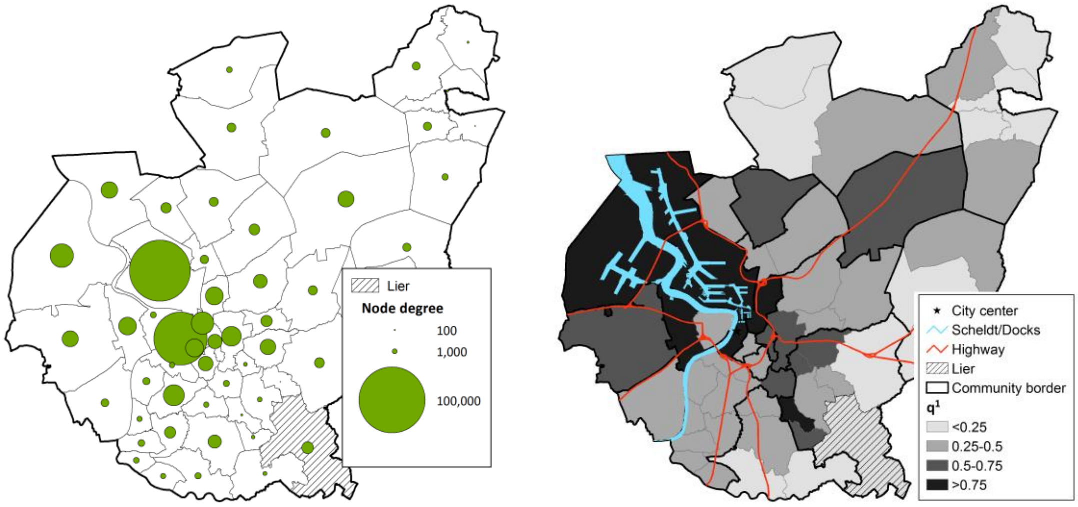


Figure 6: L_2 (A) and L_1 (B) communities for Antwerp L_3 Community (figure 1). Greyscale is used to differentiate the communities

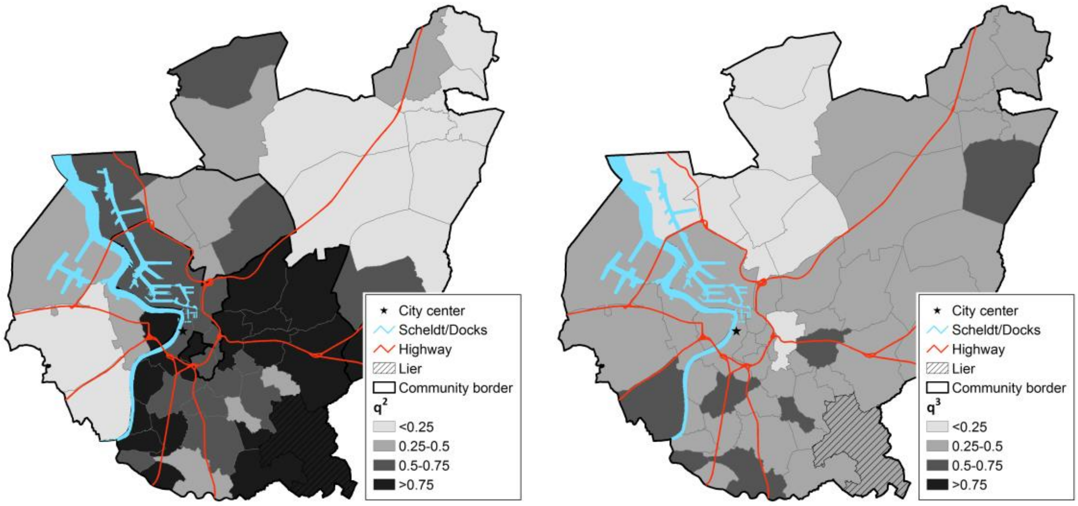
Apart from observing spatially contiguous communities, the question arises to what extent an hierarchical structure is present within these communities. In other words, what exactly are the damping values related to these hierarchical borders, what kind of spatial patterns appears and what do they tell us about the role of individual zip codes?

Figure 7 plots the q^1, q^2, q^3 values for all zip codes in the Antwerp community. These values represent the strength of the L_1, L_2, and L_3 borders respectively. Overall the damping values show strong geographical variance at each level as well as over the different levels. Although the damping values are normally distributed around an average, figure 7 clearly shows that summarizing this information in one damping value for all levels over the entire study area as proposed by Grauwin et al (2017) in his search for a universal model, comes with a neglect of their spatial variance.

The variance of damping values points out the unique positions that each zip code holds in the overall network, depicting different relations to different borders at different levels. For example, the high damping values for the L_1 level obtained for the Antwerp port area (see figure 7B) indicate a smaller importance of this level for those zip-codes. Likewise, in figure 7C a group of zip codes in the southeast show significantly higher damping values at the L_2 border compared to the L_1 and L_3 ones, suggesting the relative insignificance of their L_2 level. Remark that, given their definition, the construction of damping values for a zip code is independent from its absolute number of buyer-supplier linkages as can be observed by comparing the spatial pattern of node degrees in figure 7A with the spatial patterns of the damping values in figure 7B,C,D.



A B



C D

Figure 7:A: Node degrees, B: damping values L_1, C: damping values L_2 D: damping values L_3 values for Antwerp community

Clearly, the analysis of damping values counters the suggestion of similarity between zip codes that is implied by the discrete, spatially homogenous result of community detection. Indeed, each zip code has its own characteristics when it comes to positioning in the proposed L_1, L_2 and L_3 communities and resultantly, a richer geography is detected. Still, some spatial grouping seems to emerge from figure 7, with adjoining zip codes depicting similar damping values over different hierarchy levels.

To more formally investigate such similarities and the related spatial pattern, we applied a k-means clustering of the damping values. Figure 8 indicates three clusters as the ideal solution, which are mapped in figure 9 and, they too, are spatially quite contiguous but differ rather strongly from the presented L_2 or L_1 communities in figure 6, meaning that a different kind of characterization took place compared to standard community detection. In other words, compared to the network hierarchy resulting from the iterative community detection, we now observe the geographical hierarchy within the region. The applied k-means clustering yields three different classes of zip codes. Their average damping values for the different hierarchical borders are shown in figure 9A, together with the average damping values for all zip codes in Belgium.

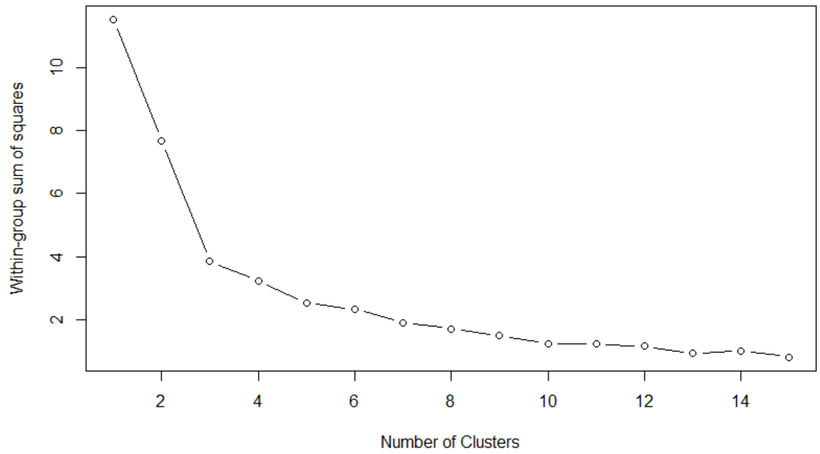


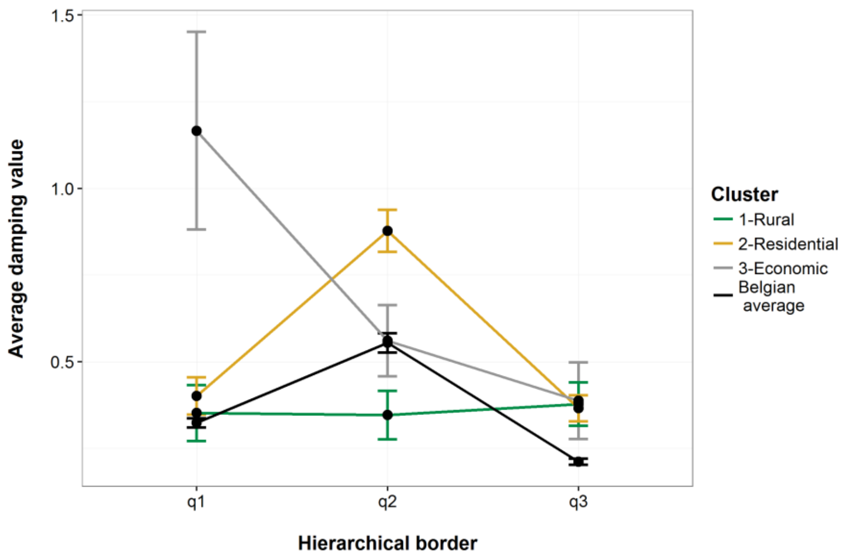
Figure 8: Within-cluster sum of squares for different number of clusters.

Several insights can be derived from the properties of the k-means clustering based on damping values as shown in Figure 7.

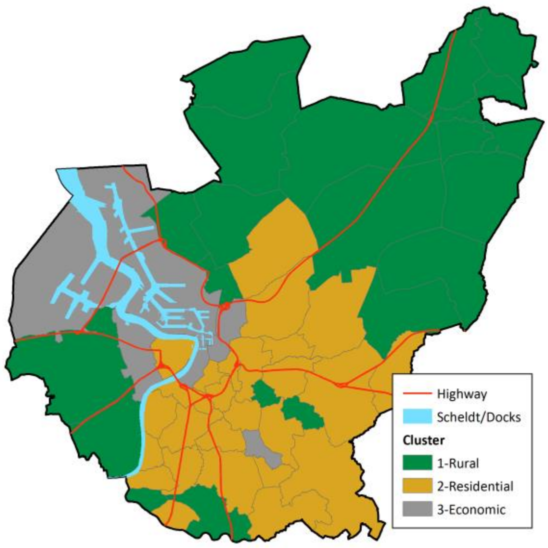
Firstly the damping values obtained for the L_3 community borders are not differentiating the different clusters. Interestingly, compared to the Belgian average, q^3 damping values of the zip codes in the Antwerp community are significantly higher, demonstrating the importance of the nodes in the Antwerp community for the entire Belgian logistic system.

Secondly, it shows that zip codes attributed to cluster 1 and 2 (Figure 9) do not differentiate amongst themselves with relation to damping values for the q^1 and q^3 borders. Rather, it is the q^2 border that separates their profiles from each other. A rather high damping value at the L_2 border for zip codes in cluster 2 suggests the difference between L_2 and L_3 communities to be less relevant compared to the other zip codes in the Antwerp community. The zip codes attributed to cluster 1 show extremely low damping values, indicating their peripheral role in the wider logistics network. Crosschecking figures 9 and 5 reveals the division between clusters 1 and 2 results from the distinct geography in both regions. Cluster 1 comprises the most rural zip codes while cluster 2 consists mostly of small cities and residential areas within Antwerp. In the remainder of this paper we will speak of the rural and residential clusters respectively.

Thirdly, it is clear that the lowest hierarchical border, q^1, is differentiating zip codes attributed to cluster 3 in the k-means algorithm. The zip codes in cluster 3 include the economic areas in the port and city center and will be called economic cluster from here on. Despite their large amount of internal linkages (some of them form an L_1 community on their own), their relative importance is significantly higher at their L_2 community level compared to their L_1 level. Interestingly, for these zip codes, the distinction between L_2 and L_3 communities is similar to the average Belgian zip code, indicating the relevance of this second hierarchical level (figure 6A), even for these important nodes. Since these second level communities correspond to the different port areas, the strong q^2 indicates the relative important difference between the left bank, the old port and the industrial port. While in previous work they were classified in one Antwerp logistics cluster, we now observe the existence of the sub clusters, proving the presence of an important geographical hierarchy in the network.



A



B

Figure 9: A: Damping profile of each k-means cluster. B: Geographical pattern of k-means clusters.

6. Discussion

In this work we introduced a geographical perspective on Grauwin et al.'s (2017) methodology to delineate hierarchical levels and define the strength of their borders in large networks. Our approach provides an empirical way to include the role of each individual location while assessing their relationships, a combination which has been overlooked in recent network applications in (transport) geography.

When applying our approach to a network of logistics buyer-supplier linkages in Belgium, the results clearly show that nodes located in the same community can exhibit strongly differing relations with different hierarchical levels. This is due to a strong geographical and hierarchical variation of the distributions of links in space for different nodes.

In our methodology, we showed how one can capture such geographical and hierarchical variation for each node individually by calculating its damping values. One of our main findings is a high diversity among damping values calculated for buyer-supplier linkages, both between nodes and between hierarchical levels (figure 4). This in contrast to the findings for mobile phone communication networks where damping values show similar distributions at different hierarchical levels (Grauwin et al., 2017). This variation indicates the need for better insights both on the local context of individual nodes as on the existence of hierarchies when interpreting or even modelling linkages.

A second important finding is that despite their high variety, damping values for logistics buyer-supplier linkages show spatial patterns at the regional level (figure 7). Performing a simple k-means cluster algorithm based on the calculated damping values unveils the underlying geography of the Antwerp community (figure 9). The fact the regions obtained by clustering damping values differ from the communities produced by (iterative) community detection points to the presence of an important hierarchical structure and results in a renewed insight in regional differences now based on the relation of individual nodes over hierarchical levels. These clusters can then be interpret by combining expert, i.e. geographical, knowledge and a detailed study of the damping value profiles.

The analysis of the different cluster profiles in figure 9 allows us to understand how hierarchical levels are experienced differently for each cluster. Given its rather complex definition, the disadvantage of damping values in this perspective is that their interpretation is not always trivial.

Concerning the low q^1 damping values for the rural and residential clusters, for example, it is worthwhile to understand that we observe the main driver to be the expansion of the set of considered zip codes. Indeed, when climbing an hierarchical level and given that the added zip codes show similar network characteristics, the W in the definition of damping values will increase, triggering a decrease of the damping values.

To demonstrate this we can take the example of a zip code attributed to the residential cluster (city of Lier, zip code 2500, see figure 7). The low damping value for the q^1 border results out of $T^1_{2500} \approx T^{(2)}_{2500}$ while $W^2_{2500} \gg W^1_{2500}$ (table 1). The diminishing importance of this zip code at the L_2 level is due to the L_2 level being a merge of several zip codes with similar node degree (size of logistics activities) and linkages distribution (diversity of its network). In the L_1 community each zip code has a share amongst 4 to 10 similar zip codes and in the L_2 community they are joined by more or less 20 other similar zip codes, decreasing the relative importance of a single zip code, like Lier, which translates in an important damping between the two levels.

Table 1:Absolute flows and damping value over each hierarchy for one zip code per cluster

	Lier (zip 2500)		
Level	T^h	W^h	q^h
L_1	460	19253	0,24
L_2	437	75959	0,97
L_3	1715	306934	0,41
Belgium	3088	1332829	

If, when scaling up, added zip codes do not show similar characteristics, the observed damping values become the result of a complex interplay of relative gains in importance and connectivity at the higher hierarchical level, despite the enlargement of the set of considered zip codes. This is the case for the residential cluster where transferring from level 2 to level 3 means adding zip codes that are either very important (port areas) or less significant (northeastern rural areas), resulting in a more complex construction of damping values and thus a less trivial interpretation. The q^2 for the city of Lier in the residential cluster is a good example. Adding the port area at the L_3 level results in a higher $W_{2500}^{(3)}$ due to the many buyer-supplier linkages at the port. In the meantime the port is an important buyer and supplier for the city as well, increasing $T_{2500}^{(3)}$. The extend of this connectivity, enforced by the lower connectivity of the zip codes from the northeastern rural areas that were also added at the L_3 level results in a rather high damping value for the city, even though $W_{2500}^{(3)}$ went significantly up by adding highly connected zip codes like the port area to the equation.

In summary, the profiles of the rural and residential clusters demonstrate the mechanics that are behind the calculation of damping values and that make their interpretation non-trivial. The main premise, however, stands: the absolute damping value depends on how important a node is within its L_h community compared to the joining communities at L_{h+1}. The more important it is (in terms of flows and connectivity) the lower your damping value will be.

An understanding of these mechanics allows for an interpretation of the different obtained clusters. Clearly, the rural and residential clusters are locally well embedded with their zip codes having a higher importance at lower scale levels. For the former, this loss of importance is constant when transgressing all hierarchical levels meaning that increasing hierarchy implicates decreasing importance for the involved zip codes. This is not surprising given the rural nature of the involved zip codes.

The residential cluster, on the other hand, does not display constant damping values for all hierarchical levels and has a remarkable high damping value for q^2. The clear gain of importance in the L_3 level is due to two factors. First, the L_3 level introduces a strong relation between the central zip codes of the residential cluster and the port area. Secondly, at the L_1 level, the central zip codes contrast with the less important zip codes from the rural cluster, both of which add to the importance of the residential one at the level of the Antwerp community. All of this yields an interpretation of the residential cluster being a set of zip codes that have similar characteristics and are rather locally focused, but as a group play an important role at the level of the Antwerp community.

Finally, the economic cluster is the most interesting case. The extremely high q^1 value indicates the high within-community links at the L_2 hierarchical level (which are the three different parts of the port) compared to the L_1 level (which are the different zip codes of the port individually). In addition, the higher than Belgian average value for q^2 implicates the zip codes in the economic cluster are extremely well linked within the L_3 level too. This pattern is indicative for the zip codes in the economic cluster being the glue of the Antwerp community. They can be perceived as economic hubs at the center of their region, which is not difficult to imagine when one evaluates their location within the central business district and the port area.

Even without assessing the absolute number of buyer-supplier linkages it is easy to understand that, in traditional community detection algorithms (like deployed to arrive at the L_3 level) it are the zip codes of the economic cluster that are most important in delineating the Antwerp community. However, as the zip codes in the two other clusters constitute the largest share of the overall network, it becomes clear how typical applications of community detection do not tap into a significant part of the available information that emergent large datasets provide.

7. Conclusions and future research

In this paper we apply three iterations of the Louvain community detection algorithm to uncover the hierarchical structure in the logistics buyer-supplier network in Belgium. By applying the methodology proposed by Grauwin et al. (2017), we receive second and third level communities within the ones previously defined for the whole Belgium scale (Beckers et al., 2015). Similarly to the high level communities and community delineations in other studies (Blondel et al., 2010; Kung et al., 2014; Nelson & Rae, 2016; Ratti et al., 2010; Vanhoof et al., 2015), the second and first level communities exhibit a strong spatial contiguity. Next, like Grauwin and his colleagues, we calculate three damping values for each node in the network, providing information on the strength of the borders between each hierarchical level. This strength represents the in- or decrease of the relative importance of a node amongst its peers when transitioning a level. To better understand the spatial patterns of damping values in our case study, we apply a simple k-means clustering, which yields three distinct regions and allows for a better understanding of the forces behind the community delineations.

To our knowledge, creating spatial patterns from hierarchical characteristics in networks has not yet been performed in previous work. Our approach therefore could enhance studies like Ratti et al. (2010) or Nelson and Rae (2016) that try to redraw the map of the UK or recreate the US economic geography by one-level community detection algorithms only. The problem our approach solves is that, as we discuss for the Antwerp community, high-level communities delineated by using one iteration of community detection are mostly based on the delineation of the strongest nodes hereby concealing the influence and relations of less important nodes. Given that less important nodes often constitute a considerable share of the overall network, many of the community detection applications overlook a significant part of their data hereby grasping only a small portion of the overwhelming potential the new big datasets provide. A critique that has also been posed by Schwanen (2016) and Kwan (2016), who notice the lack of attention for the local context in current big data analysis.

Furthermore, we find that damping values calculated on the buyer-supplier linkages display a strong geographical and hierarchical variance, highlighting the difference among individual nodes within the dataset. Especially the large hierarchical variance of damping values is in contrast with previous work where findings of constant damping led to a generalising predictive model of mobile phone communication (Grauwin et al., 2017). We can conclude that for logistics such a model will probably not hold meaning that insights on the local context and existence of hierarchies are necessary to understand the complex relations of buyer-supplier linkages in logistics, rather than generalizing principles.

In our case study, the k-means clustering of nodes based on their damping values serves well in relating these values to a local context and facilitated the detection of a geographical hierarchy that contrasts the network hierarchy. While we argue that our methodology directs attention to the specific nodes in the network, we acknowledge that our attempt is merely a first step on one of plausibly multiple pathways that can reconcile network-based insights and more qualitative, localized insights. It has, in other words, not been in the scope of this paper to fully integrate localized qualitative insights with damping values. Rather, we showed the potential is there. Yet, it would be an overestimation of our experiences to propose a clear path how to do so. What we can state, however, is that if we were to fully exploit the potential of network-based big data analysis for transport geography, the combination of qualitative insights and quantitative analysis should eventually lead to theorization or relevant input for decision making.

Potential ideas, and thus suggestions for future research, are to link the resulting geographical hierarchy to consecutive steps in a hub-and-spoke structured supply chain, in which case different regions demand different zonal planning, although additional input from logistics stakeholders would be necessary. Or, to elaborate a comparison of the k-means clusters over the country to allow the identification of over- and underperformers from a logistics perspective. Understanding the drivers of the observed differences and the comparison with previously identified key factors influencing the location decision of logistics companies (e.g. Verhetsel et al., 2015) could potentially allow for new insights in the latter. Similarly, the application of the used methodology on public transport data may help to identify missing links between some regions and overcapacity on other connections when comparing resulting hierarchical communities with, for example, delineations of daily urban systems. Although in these examples our proposed methodology can lead to more advanced geographical insights, hence aiding in reaching the level of mesogeography Miller (2017) advocates, they all demand that extra step of integrative qualitative analysis that we believe now forms the next step to address.

Acknowledgments

This research was supported by the Flemish Government through the Urban Logistics and Mobility project (ULM) financed by IWT-SBO. We are grateful to the National Bank of Belgium for providing data.

Bibliography

Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–12. http://doi.org/10.1126/science.286.5439.509

Barthelemy, M. (2010). Spatial Networks. *Physics Reports*, 1–86. http://doi.org/10.1016/j.physrep.2010.11.002

Beckers, J., Vanoutrive, T., Thomas, I., & Verhetsel, A. (2015). Geographical Clustering of Buyer-Supplier Linkages in the Logistics Sector in Belgium. In R. Soora & H. Timmermans (Eds.), *BIVEC/GIBET Transport Research Day 2015* (pp. 44–57). Eindhoven.

Blondel, V., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of community hierarchies in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10*, 1–6. http://doi.org/10.1088/1742-5468/2008/10/P10008

Blondel, V., Krings, G., & Thomas, I. (2010). Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. *Brussels Studies*, *32*(42), 0–12.

Colizza, V., Barrat, A., Barthelemy, M., & Vespignani, A. (2006). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, *103*(7), 2015–2020. http://doi.org/10.1073/pnas.0510525103

Delvenne, J.-C., Schaub, M. T., Yaliraki, S. N., & Barahona, M. (2013). The stability of a graph partition: A dynamics-based framework for community detection. http://doi.org/10.1007/978-1-4614-6729-8_11

Everitt, B. S., & Hothorn, T. (2010). *A handbook of Statistical Analyses Using R* (2nd ed.). CRC Pres.

Fortunato, S., & Barthelemy, M. (2006). Resolution limit in community detection, *104*(1). http://doi.org/10.1073/pnas.0605965104

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide, 1–43. http://doi.org/10.1016/j.physrep.2016.09.002

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individial human mobility patterns. *Nature*, *453*(7196), 779–782.

Goodchild, M. F. (1992). Geographical information science. *International Journal of Geographical Information Systems*, *6*(1), 31–45.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, *3*(3), 255–261. http://doi.org/10.1177/2043820613513121

Grauwin, S., Szell, M., Sobolevsky, S., Hövel, P., Simini, F., Vanhoof, M., … Ratti, C. (2017). Identifying the structural discontinuities of human interactions. *Scientific Reports*, *7*(46677). http://doi.org/10.1038/srep46677

Guimerà, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature*, *433*(7028), 895–900. http://doi.org/10.1038/nature03288

Guimerà, R., Mossa, S., Turtschi, A., & Nunes Amaral, L. a. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(22), 7794–7799. http://doi.org/10.1073/pnas.0407994102

Jiang, B. (2013). Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution. *The Professional Geographer*, *65*(3), 482–494. http://doi.org/10.1080/00330124.2012.700499

Johnston, R., & Sidaway, J. D. (2015). *Geography and Geographers: Angelo-American Human Geography Since 1945*. London: Routledge.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, *3*(3), 262–267. http://doi.org/10.1177/2043820613513388

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, *2*(3), 203–271. http://doi.org/10.1093/comnet/cnu016

Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS ONE*, *9*(6). http://doi.org/10.1371/journal.pone.0096180

Kwan, M. (2016). Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge. *Annals of the American Association of Geographers*, *106*(2), 274–282. http://doi.org/10.1080/00045608.2015.1117937

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical Review E*, 12. Physics and Society; Computational Physics. http://doi.org/10.1103/PhysRevE.80.056117

Levy, M., & Solomon, S. (1997). New evidence for the power-law distribution of wealth Moshe. *Physica A*, *242*, 90–94.

Martin, R., & Sunley, P. (2003). Deconstructing clusters: chaotic concept or policy panacea? *Journal of Economic Geography*, *3*(1), 5–35. http://doi.org/10.1093/jeg/3.1.5

Miller, H. J. (2017). Geographic information science II: Mesogeography. *Progress in Human Geography*, 30913251771215. http://doi.org/10.1177/0309132517712154

Nelson, G. D., & Rae, A. (2016). An Economic Geography of the United States: From Commutes to Megaregions. *Plos One*, *11*(11). http://doi.org/10.1371/journal.pone.0166083

Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review*, *76*(6), 77–90.

Pumain, D. (Ed.). (2006). *Hierarchy in Natural and Social Sciences*. Springer.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., … Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PLoS ONE*, *5*(12). http://doi.org/10.1371/journal.pone.0014248

Schwanen, T. (2016). Geographies of Transport II : Reconciling the General and the Particular. *Progress in Human Geography*. http://doi.org/10.1177/0309132516628259

Simini, F., González, M. C., Maritan, A., & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, *484*(7392), 96–100. http://doi.org/10.1038/nature10856

Sobolevsky, S., Campari, R., Belyi, A., & Ratti, C. (2014). General optimization technique for high-quality community detection in complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *90*(1), 1–8. http://doi.org/10.1103/PhysRevE.90.012811

Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., & Ratti, C. (2013). Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLoS ONE*, *8*(12). http://doi.org/10.1371/journal.pone.0081707

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, *6*(10), 818–823. http://doi.org/10.1038/nphys1760

Thiemann, C., Theis, F., Grady, D., Brune, R., & Brockmann, D. (2010). The structure of borders in a small world. *PLoS ONE*, *5*(11), 1–7. http://doi.org/10.1371/journal.pone.0015422

Vanhoof, M., Smoreda, Z., & Ratti, C. (2015). Drawing Boundaries of Social Interacties. *Uses and Value*, *52*, 2–5.

Verhetsel, A., Kessels, R., Goos, P., Zijlstra, T., Blomme, N., & Cant, J. (2015). Location of logistics companies: a stated preference study to disentangle the impact of accessibility. *Journal of Transport Geography*, *42*, 110–121. http://doi.org/10.1016/j.jtrangeo.2014.12.002