biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

Even important connections are not always meaningful: on the use of a polarization measure in a typology of European cities in air transport networks

Van Nuffel, N., Derudder, B., Witlox, F.

In: *Tijdschrift voor Economische en Sociale Geografie*, 101 (3), 333-348, 2010.

doi: 10.1111/j.1467-9663.2009.00547.x

The definitive version is available at www.blackwell-synergy.com

To refer to or to cite this work, please use the citation to the published version:

Van Nuffel, N., Derudder, B., Witlox, F. (2010). Even important connections are not always meaningful: on the use of a polarization measure in a typology of European cities in air transport networks. *Tijdschrift voor Economische en Sociale Geografie* 101 (3), 333-348. doi: 10.1111/j.1467-9663.2009.00547.x

EVEN IMPORTANT CONNECTIONS ARE NOT ALWAYS MEANINGFUL: ON THE USE OF A POLARIZATION MEASURE IN A TYPOLOGY OF EUROPEAN CITIES IN AIR TRANSPORT NETWORKS

Nathalie Van Nuffel, Ben Derudder & Frank Witlox
Geography Department, Ghent University
Krijgslaan 281/S8, B9000 Gent, Belgium
E-mail: nathalie.vannuffel@skynet.be, ben.derudder@ugent.be,
frank.witlox@ugent.be

Abstract:

This paper aims to refine existing approaches for classifying nodes in spatial networks. Apart from generally used variables measuring (i) the total connectivity of a node and (ii) the spatial make-up of this total connectivity, we argue that a third indicator should be taken into account when describing the structure of a spatial network, i.e. (iii) the degree of ‘polarization’ in a node’s connectivity configuration. We measure ‘polarization’ by applying multiple linkage analysis, a technique that allows determining the number of meaningful connections for each node in the network. In this paper, we illustrate this approach by clustering 65 European cities based on their insertion in European air transport networks in 2005. The number of meaningful flows of a city is hereby considered alongside both traditional measures of connectivity. The results of the cluster analysis suggest that including a measure of polarization may indeed lead to a more satisfying typology.

Keywords:

Polarization, Europe, multiple linkage analysis, cluster analysis, air transport, AEA-data

INTRODUCTION

It seems fair to state that there has been a (renewed) interest in human geography in conceptual and empirical research on urban networks (e.g. Frenken & Hoekman 2006; Meijers 2007; Taylor et al. 2008). This paper aims to contribute to this literature by calling for the use of a polarization measure in the portrayal of a city’s connectivity profile. More specifically, we propose to determine the number of *meaningful* flows for each node in the network by applying multiple linkage analysis (MLA). MLA is a method that is specifically designed for this purpose (Van Nuffel 2007), but its usefulness has - to our knowledge - not been examined in the context of urban network research.

The overall objective of this paper can be cast in two different forms, one methodological and the other empirical. The methodological purpose is to discuss how the measurement of polarization through MLA may lead to a more satisfactory depiction of a city’s connectivity profile. The empirical purpose of this paper is to illustrate the practical implementation of this methodological assertion by broadly replicating the research presented in Burghouwt & Hakfoort (2002) and Taylor & Derudder (2004). More specifically, we classify 65 cities based on their position in the European air transport network in 2005¹. The degree of polarization in a city’s connectivity configuration is hereby considered alongside more traditional measures of connectivity.

This paper is organized as follows. In the next section, we discuss the recent rise of research on transnational urban networks in conceptual and empirical terms. This

review is used to call for adding a polarization indicator to the methodological toolbox, after which we introduce the MLA method and discuss our data and the transformations that were needed to make these suitable for our analysis. In the following section, we apply MLA to our air passenger data for 2005 and present the results of a cluster analysis of European cities in transport networks with *and* without this polarization measure. The paper is concluded with a discussion of our main findings.

CITIES IN TRANSNATIONAL NETWORKS

World city networks - The renewed interest in human geography in conceptual and empirical research on urban networks can loosely be traced back to the booming literature on the emergence of so-called 'world cities' or 'global cities'. Despite earlier uses of these terms, most contemporary research on globalized urbanization is based on two interrelated papers by Friedmann and Wolff (1982) and especially Friedmann (1986). Both texts framed the rise of world cities in the context of a major geographical transformation of the world economy. This restructuring, most commonly referred to as the 'New International Division of Labor', was basically premised on the internationalization of production and the ensuing complexity in the organizational structure of multinational enterprises (MNEs). This increased economic-geographical complexity, Friedmann (1986) argued, requires a limited number of control points in order to function, and world cities are deemed to be such points. Friedmann's observations quickly became the backbone for much research on the functional role of cities in an increasingly globalized economy. It provided the motivation for dozens of studies that sought to extend the theoretical framework (e.g. Sassen 1991); to create methods for measuring world city-formation (e.g. Beaverstock et al., 2000); to explicitly test some of the assumptions in the mainstream conceptualizations (e.g. Hill & Kim 2000); and to provide in-depth insights through case studies of selected cities (e.g. Wang 2003). Twenty years of research on world cities thus generated an enormous number of point of views (Brenner and Keil 2006), meta-theoretical narratives (Saey 2007) and methodological variations (Derudder 2006), which makes it altogether impossible to speak of a coherent research paradigm.

A key contribution in this literature has been the work of Saskia Sassen. In her path-breaking book *The Global City*, Sassen (1991) proposes to look afresh to the functional centrality of cities in the world economy, and she does so by focusing upon the attraction of advanced producer services firms (e.g. investment banks, accountancy, internationalized law,...) to major cities offering knowledge-rich and technology-enabled environments. In the 1980s and 1990s, many such service firms followed their global corporate clients to become important multinational enterprises in their own right, albeit that they tend to be more susceptible to the agglomeration economies offered in city locations. The transnational, city-centred spatial strategies of major advanced producer services (APS) firms have increasingly resulted in worldwide office networks covering major cities in most or all world regions, and it is the myriad of connections between these service complexes that gives, according to Sassen, way to the formation of a transnational urban *network* centered on global cities. Since the publication of Sassen's *The Global City* in 1991, it has become commonplace to recognize the networked character of transnational urban systems. This is, for instance, very clear in the work of an influential scholar such as Manuel Castells (2002) who argues that today cities can most fruitfully be understood when conceptualized in the context of 'the rise of the network society'.

Indicators of urban activity - Recent empirical studies of transnational urban networks have often taken on the form of classification exercises that draw on a combination of two sorts of *connectivity* indicators, i.e. (i) the total connectivity of

a city in transnational flows, and (ii) the spatial make-up of a city's connectivity in these transnational flows. The rationale for this dual approach is clearly set out in a paper by Burghouwt and Hakfoort (2002), in which the authors lament the fact that standard classifications of cities based on the connectivity of their major airports - such as those provided by the Federal Aviation Administration (2000) - nearly always draw rather simplistically on either the total number of passengers or the total seat capacity of an airport. Although this actual or potential capacity provides us with an even-handed measure of an airport's *total* connectivity, Burghouwt and Hakfoort (2002) rightly observe that it does not give any indication of the *actual spatial make-up* of this connectivity. As a consequence, potential differences in the configuration of a given level of connectivity are not taken into account in such classifications. However, an airport that is only directly connected to 1 or 2 other nodes clearly plays another role in the network than an airport that is directly connected to 10 or 20 other nodes, even if both airports have a similar capacity. In their classification of European airports, Burghouwt and Hakfoort (2002) therefore include - in addition to the total seat capacity of an airport - two related variables that gauge the spatial configuration of an airport's connectivity in the European air transport network: the number of destinations *per se* and the number of intercontinental destinations.

A similar approach to classifying nodes in a transnational network can be found in Taylor & Derudder (2004). In this paper, the authors develop an earlier classification of cities - presented in Derudder et al. (2003) and drawing on Sassen (1991) - in order to arrange European cities based on their position in the world city network created through the globalized location strategies of APS firms. In this empirical exercise, cities are clustered on the basis of the spatial arrangement of their inter-city relations, after which the various clusters are ordered through the total connectivity of their member cities. Despite the differences in units of analysis (cities versus airports) and the input data (inter-firm flows in the advanced producer services sector versus airline flows), both Taylor & Derudder (2004) and Burghouwt & Hakfoort (2002) take a similar dual approach to the classification of nodes in a spatial network. In this paper, we suggest that their combined use of (i) the total connectivity of a node and (ii) its spatial configuration may be enhanced by considering a third indicator, i.e. (iii) the degree of 'polarization' in a node's connectivity configuration. Polarization hereby refers to the extent to which a node's connections are (un)evenly distributed. In the context of this third measure, it is neither a city's total connectivity nor the spatial make-up of its connectivity that is important, but rather the *proportional magnitude of its different connections*.

Figures 1 and 2 illustrate the potential relevance of this extra indicator. The figures give an impression of the connectivity profile of four cities in the European air transport network, i.e. London, Amsterdam, Dublin, and Bucharest; the data used in these figures is discussed in some detail in the next section. Taking the dual approach set out in Burghouwt & Hakfoort (2002) and Taylor & Derudder (2004), one would assume that Amsterdam/London on the one hand and Bucharest/Dublin on the other hand take on largely similar roles in the European air transport network: (i) a comparable total connectivity (as measured by the total number of intra-European passengers in 2005: 28'182'754 and 17'989'712 for London and Amsterdam respectively, 2'200'234 and 2'206'690 for Dublin and Bucharest respectively) is in both cases matched by (ii) a comparable make-up of this total connectivity (as measured by the number of direct intra-European connections in 2005: 75 and 69 direct European connections for London and Amsterdam respectively, 22 and 19 for Dublin and Bucharest respectively). Figures 1 and 2 show an ordinal line graph of the overall passenger share of the European cities directly connected to London & Amsterdam and Dublin & Bucharest respectively. There are clearly quite different

patterns in these distributions: flows to/from Amsterdam and Dublin are dominated by a single city (in both cases London), while flows to/from London and Bucharest are far more evenly distributed (there is, for instance, no substantial difference between the importance of the flows to/from the most important and the second most important city). Thus despite having (i) similar levels of connectivity and (ii) this connectivity being distributed over a similar number of nodes, both Amsterdam/London and Dublin/Bucharest are quite different in their connectivity profile. Paraphrasing the observations made in Burghouwt & Hakfoort (2002), one might say that a city whose passengers are primarily directed towards 1 or 2 of the other nodes clearly plays another role in the network than a city whose passengers are rather evenly distributed over all the other nodes, even if both cities (i) have a similar seat capacity and (ii) are directly connected to a similar number of other cities. In the next section, we introduce multiple linkage analysis (MLA), a methodology that allows measuring the degree of polarization in a node's connectivity profile.

Figures 1-2 about here

DETERMINATION OF THE NUMBER OF SIGNIFICANT CONNECTIONS: MULTIPLE LINKAGE ANALYSIS

Method - MLA is a technique derived from mathematical graph theory. Concepts from graph theory were first introduced in geography by Nystuen & Dacey (1968), who used primary linkage analysis to reconstruct the urban hierarchy in Washington State on the basis of intercity telephone calls. In primary linkage analysis, only the dominant flow is taken into account when assessing a city's connectivity: "the 'dominant association' remains the critical concept in defining a nodal structure. The remaining, non-dominant associations are not used, even though the magnitudes of some of these associations may be relatively large" (Nystuen & Dacey 1968, p. 408; for a recent example, see Grubestic et al. 2008). Unsurprisingly, this neglect of subdominant flows is at the heart of the critique of this technique (e.g., Guldmann 2004). MLA is an extension of primary linkage analysis that has been explicitly developed to deal with subdominant flows that are important for the spatial structure (Puebla 1987). We begin by briefly setting out the mathematical basis of MLA, after which we will illustrate the technique with an example.

MLA attempts to determine the number of meaningful flows in a node's connectivity structure. To distinguish between significant and insignificant flows, all flows to/from a spatial unit are ranked from largest (w_1) to smallest (w_k). A set of cycles of expected flows $\{\hat{w}_i\}$, with $i \in I$ (the set of spatial units) is then generated as follows (Haggett *et al.* 1977):

$$\begin{aligned} 1^{\text{st}} \text{ cycle:} \quad & \hat{w}_1 = \sum_{i=1}^I w_i \\ & \hat{w}_2 = \hat{w}_3 = \dots = \hat{w}_k = 0 \end{aligned}$$

$$\begin{aligned} 2^{\text{nd}} \text{ cycle:} \quad & \hat{w}_1 = \hat{w}_2 = \frac{1}{2} \sum_{i=1}^I w_i \\ & \hat{w}_3 = \hat{w}_4 = \dots = \hat{w}_k = 0 \end{aligned}$$

$$j^{\text{th}} \text{ cycle } (j < k): \quad \hat{w}_1 = \hat{w}_2 = \dots = \hat{w}_j = \frac{1}{j} \sum_{i=1}^j w_i$$

$$\hat{w}_{j+1} = \hat{w}_{j+2} = \dots = \hat{w}_k = 0$$

$$k^{\text{th}} \text{ cycle:} \quad \hat{w}_1 = \hat{w}_2 = \dots = \hat{w}_k = \frac{1}{k} \sum_{i=1}^k w_i$$

The goodness of fit between the set of observed flows and each of the sets of expected flows is measured by the coefficient of determination (COD)ⁱⁱ. The number of significant flows corresponds to the j^{th} cycle with the highest COD value.

Example - The practical utilization of MLA can be shown through the example in Figure 3, in which X and Y are two spatial units connected to 8 other spatial units (a-h) by flows of different sizes. The spatial units can, for instance, be cities, and the corresponding flows outgoing commuters. Other possibilities include countries and the number of immigrants in a certain period, or - as is in our analysis - cities and their airline connections. The starting point of MLA, now, is that if we want to describe the structure of the network centered around X or Y, we do not need the full information available in Figure 3. Rather, MLA determines how many of the flows are really relevant for the specification of the spatial structure of the network around a spatial unit. Table 1 illustrates how this works in practice. In the table, the different connections of X and Y are ranked from largest to smallest.

Table 1 and Figure 3 about here

MLA compares the real configurations to ideal-typical configurations in which the flows are distributed over exactly 1, 2, 3,... destinations, in shares of equal magnitude. Put differently, in the first ideal-typical situation, X is only connected to a single spatial unit; in the second ideal-typical situation, both flows to/from X are distributed in equal terms over the two spatial units, etc. Next, the COD between the real distribution and each of the ideal-typical situations is calculated. These coefficients are shown in the last row. The number of destinations (flows) in the fictive distribution with this *largest* COD gives us the number of significant flows in the real distribution. From table 1, it can be seen that for spatial unit X, the number of significant flows is 3 (corresponding to a COD of 0.88), whereas for spatial unit Y the number of significant flows is only 1 (corresponding to a COD of 0.85).

Results - The COD of Dublin, Amsterdam, Bucharest and London for each consecutive cycle are given in Table 2. For Dublin and Amsterdam, the highest COD (0.93 and 0.75 respectively) is found between the observed flows and the first cycle of expected flows, which confirms the intuitive interpretation of Figures 1 and 2 that the connectivity profile of both cities is dominated by a single city (London). In the case of Bucharest, it is the 10th cycle of expected flows that has the highest COD, while for London it is the 18th cycle. Again, this conforms to the intuitive interpretation of the figures that the connectivity profile of both cities is more evenly spread. The MLA-values thus lead to the general conclusion that the polarization in the connectivity profile of Dublin and Amsterdam is high, whereas the polarization in the connectivity profile of Bucharest and London is relatively low. In the remainder of the paper, we systematically apply this technique to a database with information on European cities' air transport connections (discussed in the next section), after which we show that the classification of the urban network changes when this indicator is taken into account.

Table 2 about here

THE AEA-DATABASE

In recent years, air passenger data have become very popular, both for analysing transnational urban networks (see e.g. Shin & Timberlake 2000; Smith & Timberlake 2001; Matsumoto 2004; Derudder & Witlox 2005, 2008; Derudder *et al.* 2007a,b; Taylor *et al.* 2007) and for investigating the effects of deregulation and other changes in the airline industry (see e.g. Philips 1985; Goetz & Sutton 1997; Burghouwt & Hakfoort 2001; Reynolds-Feighan 2001; Burghouwt & Hakfoort 2002; Burghouwt *et al.* 2003; Frenken *et al.* 2004; Burghouwt & de Wit 2005; Zook & Brunn 2006; Alderighi *et al.* 2007; Grubestic & Zook 2007). Numerous authors have used air passenger data to develop classifications of both airports and cities (for an overview, see Burghouwt & Hakfoort 2001, 2002).

Description of the AEA-database - We make use of AEA-data. AEA stands for Association of European Airlines, a non-profit-making organisation that brings today together 34 major European airlinesⁱⁱⁱ. The AEA represents its member airlines in dialogue with all the relevant European and international organisations in the aviation value chain, thus ensuring the sustainable growth of the European airline industry in a global context. (<http://www.aea.be>)

Through the cooperation of an airline, we were able to obtain an AEA-database that contains, for each connection, information on the carrier, origin and destination (airport, city, country and region), number of passengers (subdivided into first class, business class and economy class), freight, mail, number of flights (subdivided into passenger flights and freight flights) and distance between origin and destination. The data is summarized on a monthly basis for the period January 2001 - April 2006, which allows a detailed analysis of recent data.

Obviously, the AEA-data is a very rich source of information. There are, however, two drawbacks that should be taken into account when interpreting the results. The first problem is the fact that the AEA is primarily made up of 'legacy carriers', which implies that there are no low-cost carrier members in the AEA. According to the European Low Fares Airline Association (www.elfaa.com), the low-cost carrier sector accounted for approximately 30% of all intra-European traffic in 2006. Contrary to full service carriers who experienced a shift towards hub-and-spoke configurations after the deregulation of the passenger aviation market, low-cost carriers tend to prefer a point-to-point organisation for their routes. Not including them will therefore have some influence on the results^{iv}. A second, and more important, disadvantage of the AEA-data, and a potential source of distortions and misinterpretations, is the lack of real origin-destination data: the database records the individual legs of a trip rather than the trip as a whole. For example, a flight from Oslo to Madrid via London will be recorded as two separate flights, one from Oslo to London and one from London to Madrid. Any possible stopovers are not registered as such, which implies that the connectivity of cities with an important hub function, like London and Paris, will be overestimated.

Transformation of the data - The AEA-database includes information on all flights carried out by its member airlines. For reasons of consistency, (i) we only retained those flights where both the origin city and destination city are European cities, while at the same time (ii) discarding intra-national flows. The first transformation

is necessary because the AEA flows are very partial for extra-European flights. For instance, a lot of non-AEA carriers are active on the EU-US market, which implies that intercontinental flows cannot be meaningfully compared with intra-European flights (where AEA carriers do provide meaningful measures as they dominate this market). The second transformation, on the other hand, is data-driven in that there are some homogeneity problems with the data for different years: between 2003 and 2004, the number of passengers shows a major increase due to a change in the registration procedure; from then onwards, domestic passengers are also recorded in the data. Omitting domestic flows thus results in consistent datasets across the period under investigation. Taken together, this implies that the connectivity profile (including the polarization measure) of cities that have sizable national or intercontinental connections may be rather different than expected intuitively (i.e. based on the overall connectivity of cities): the connectivity profiles presented here exclusively deal with cities' 'European flows', as neither national nor extra-European flows are considered.

We then converted the airport-to-airport database into a city-to-city database by summing the number of passengers over all the airports for a given city. Cities that do not figure as origin or destination in both 2002 and 2005 (the two years for which we conducted our analysis)^v were dropped. Finally, given that we do not know the home base of the travelers, we summed the passengers travelling from city A to city B with those travelling in the opposite direction, and grouped the same connections, resulting in a database of non-directional, international flows between cities. After these transformations, our database contains 116,015,489 passengers in 2002 and 128,809,672 passengers in 2005, distributed over 22 carriers^{vi}, 35 countries and 129 cities. We calculated the number of significant connections for each of the 129 cities in our database, but only the results for the 73 European so-called GaWC selection of world cities^{vii} will be discussed below. Because the analyses based on the 2002 and the 2005 data revealed largely parallel patterns, in the remainder of the paper we will only make use of the measurements based on the 2005 data.

POLARIZATION AND CLASSIFICATION OF EUROPEAN CITIES BASED ON THE AEA-DATA

Polarization - Table 3 shows the number of significant connections for each of the cities in the transformed AEA dataset for 2005. In general terms, the most important cities (in terms of total number of passengers) also have the largest number of significant connections. This suggests that - on average - the links of well-connected cities are more evenly distributed than those of less-connected cities. Thus major nodes such as London and Paris boast a large number of significant connections, while less important nodes such as Palermo and Cardiff have only one significant connection. However, Table 3 also illustrates that this relation is far from perfect. Amsterdam and Frankfurt, for instance, only have 1 and 2 significant connections respectively, despite being major nodes in the European air transport network. Meanwhile Riga and Bologna, which are obviously secondary nodes in the European air transport network, have a relatively even distribution. Taken together, the table suggests that polarization may not only be a valuable *conceptual* addition in the study of spatial networks, but also a helpful extra in *empirical* terms in that classification schemes may well be revised when this indicator is taken into account. To lend further evidence to this claim, we compare the results of a classification of cities in the European transport network with and without the results of such MLA analysis.

Table 3 about here

Classification rationale - Following Burghouwt & Hakfoort (2002) and Taylor & Derudder (2004), we classify our sample of European cities using an hierarchical cluster analysis^{viii}. Only cities that had at least three connections in 2005 were included in the analysis. This criterion reduces the number of cities from 73 to 65. Drawing on our general line of reasoning, we use the following variables in the cluster analysis: (i) the total number of international passengers within Europe as a measure of total connectivity, (ii) the total number of European connections as a measure of the make-up of this connectivity, and (iii) the number of significant European connections (as determined by MLA) as a measure of polarization. We first standardized the values of these variables by calculating z-scores for each of the cities. We carried out two analyses, i.e. (i) a classification drawing on the ‘standard’ approach which only makes use of the number of passenger and the number of destinations, and (ii) a classification that also incorporates the polarization indicator.

What are the *a priori* expectations for these analyses based on our general line of reasoning? Table 3 suggests that - on average - the relative lack of polarization is related to high connectivity levels, just as the number of destinations is related to high connectivity levels. However, Burghouwt & Hakfoort (2002) have convincingly shown that the latter correlation is far from perfect: including the number of destinations resulted in a different classification, which suggests that the new indicator contained additional information on a node’s position in the network. Here we take a similar approach: in a similar vein, it can be assumed that if (the lack of) polarization does provide further information on cities’ connectivity profiles - because polarization is not a mere replication of total connectivity and/or the total number of destinations - we will observe that the results will be different for both analyses.

Classification results - Different numbers of clusters can yield different salient results, and hence there is no firm theoretical basis for selecting the number of clusters for the classification analyses. Here, we will focus on the results for 5 clusters, which is a pragmatic choice after assessing several solutions of different classes. Tables 4 and 5 summarize the results of the cluster analyses. The detailed composition of the clusters is given in Appendix 1. Table 4 contains the results based on 2 indicators, Table 5 the results based on 3 indicators. For illustrative purposes, we have also added information on the polarization values for each of the clusters in Table 4, although this indicator was not used in the actual analysis.

For each cluster, the tables contain information on (i) two ‘representative’ member cities, (ii) the average value for each of the variables, and (iii) how this cluster average relates to the average value across the entire dataset. The latter information is summarized with signs indicating the degree to which the average value of the cluster differs from the average value across all cities. Because most passenger distributions are positively skewed (a lot of relatively small values and a few exceptionally large ones), we used different criteria for assessing deviations above and below the mean. Positive deviations from the mean are indicated with ‘+’ if they are smaller than one standard deviation, with ‘++’ if they are between one and two standard deviations, and with ‘+++’ if they are larger than two standard deviations. Negative deviations from the mean are indicated with ‘-’ if they are larger than the median, with ‘--’ if they are smaller than the median but larger than the 40th percentile, and with ‘---’ if they are smaller than the 40th percentile. For ease of interpretation, we renumbered the clusters according to decreasing average number of passengers.

Tables 4 and 5 about here

When comparing Tables 4 and 5, it becomes clear that adding information on polarization does indeed change the results of the classification. The most obvious change is that the addition of the polarization indicator brings about the break-up of cluster 1 in Table 4. Originally this cluster consisted of Europe's most-connected cities (i.e. London, Frankfurt, Amsterdam and Paris). However, because of the polarized profile of Amsterdam and Frankfurt, there are now two clusters of well-connected cities. The cluster of very well-connected cities with a polarized profile also features Munich, which was originally part of a cluster of secondary nodes. However, Munich's combination of fairly sizable number of passenger/destinations (above the average values of the initial centre of cluster 2) and a polarized structure implies that in practice this city is actually more like Amsterdam and Frankfurt, and this is now acknowledged in the cluster analysis.

In general terms, the original cluster 2 now constitutes the core of a new cluster that brings together cities that combine above-average passenger volumes and numbers of destinations with a fairly even distribution. This again explains why Munich is now in a different cluster, just as Bucharest is now 'upgraded' because it does not quite conform to the profile of a below-average connected city as set out in the original cluster 4: in contrast to most other cities in this original cluster, Bucharest has a very even distribution with 10 significant connections, which makes it more like Brussels and Rome. There are also some cities that are dropped from the three 'top' clusters. Manchester and Dusseldorf, for instance, are now put in cluster 4 because they have a very polarized profile, which suggests that their connectivity profile is actually more similar to that of Oslo and Lisbon. Looking at the meaning and the content of clusters 1-3 in both tables, it is clear that cities such as Munich and Bucharest now have secured a more satisfactory classification.

At first sight, clusters 4-5 seem to have retained their original meaning in that there is a clear-cut distinction between both clusters in terms of average number of passenger and destinations. However, this does not imply that nothing has changed: by incorporating the polarization measure in the classification, there now seems to be a more sophisticated classification in that the allocation of some 'in-between' cities is now more refined. Riga and Cologne provide us with comprehensible examples. When discarding the degree of polarization in their respective connectivity profiles, it can be said that both cities have largely similar connectivity profiles in that they both have below-average numbers of passengers and destinations. However, because Riga is slightly less-connected than Cologne, these cities ended up in cluster 5 and 4 respectively. When we also consider the degree of polarization in their connectivity profile (Cologne has a very polarized pattern, Riga is not polarized at all), both cities switch positions. Because Riga's passengers are spread out over the different destinations, it makes more sense to compare the city to the likes of Lisbon and Oslo, which combine below-average passenger/destination volumes with a more or less even distribution. A similar observation can be made with respect to Cologne, and the key point here is that including the polarization indicator results in a more meaningful classification.

CONCLUSION

In their cluster analysis of European airports, Burghouwt & Hakfoort (2002) make use of variables that measure both connectivity and its configuration based on air passenger flows. Although their analysis is a clear-cut improvement over former typologies that only focus on connectivity in the strict sense, in this paper, we have argued that their approach can be further refined by including an additional indicator of what we call 'polarization', i.e the extent to which one or only a few connections dominate a city's connectivity profile. The concept of polarization was

implemented through determining the number of significant connections, which was measured by multiple linkage analysis (MLA).

Our claim that considering polarization may be fruitful in this context has been based on both methodological and empirical grounds. In methodological terms, we have argued that cities with similar levels of connectivity may well be very different because of the dominance of one or more connections. In empirical terms, we have shown that although polarization is often related to overall connectivity, these indicators are different in empirical terms (e.g., a major node such as Amsterdam is very polarized, while a secondary node such as Riga is not polarized at all). This observation is in turn reinforced by the fact that a classification analysis of European cities seems to be more refined when the degree of polarization is taken into account. The main finding of our paper in general terms, therefore, is that even important connections (in terms of their size) are not always meaningful (in terms of their contribution to our understanding of the spatial structure of the network).

REFERENCES

- AEA (Association of European Airlines) (2003), *Yearbook 2003*. www.aea.be
- ALDERIGHI, M., A. CENTO, P. NIJKAMP & P. RIETVELD (2007), Assessment of New Hub-and-Spoke and Point-to-Point Airline Network Configurations. *Transport Reviews* 27, pp. 529-549.
- BEAVERSTOCK, J., R. SMITH, P. TAYLOR, D. WALKER & H. LORIMER. (2000), Globalization and World Cities: Some Measurement Methodologies. *Applied Geography* 20 (1), pp. 43-63.
- BRENNER, N. & R. KEIL (Eds.) (2006), *The Global Cities Reader*. London: Routledge.
- BURGHOUWT, G. & J. DE WIT (2005), Temporal Configurations of European Airline Networks. *Journal of Air Transport Management* 11, pp. 185-198.
- BURGHOUWT, G. & J. HAKFOORT (2001), The Evolution of the European Aviation Network, 1990-1998. *Journal of Air Transport Management* 7, pp. 311-318.
- BURGHOUWT, G. & J. HAKFOORT (2002), The Geography of Deregulation in the European Aviation Market. *Tijdschrift voor Economische en Sociale Geografie* 93, pp. 100-106.
- BURGHOUWT, G., J. HAKFOORT & J. RITSEMA VAN ECK (2003), The Spatial Configuration of Airline Networks in Europe. *Journal of Air Transport Management* 9, pp. 309-323.
- CASTELLS, M. (2002), Local and Global: Cities in the Network Society. *Tijdschrift voor Economische en Sociale Geografie* 93, pp. 548-558.
- DERUDDER, B. (2006), On conceptual confusion in empirical analyses of a transnational urban network. *Urban Studies* 43, pp. 2027-2046.
- DERUDDER, B., P. TAYLOR, F. WITLOX & G. CATALANO (2003), Hierarchical Tendencies and Regional Patterns in the World City Network: A Global Urban Analysis of 234 Cities. *Regional Studies* 37(9), pp. 875-886.
- DERUDDER, B. & P.J. TAYLOR (2005), The Cliquishness of World Cities. *Global Networks* 5, pp. 71-91.

DERUDDER, B. & F. WITLOX (2005), An Appraisal of the Use of Airline Data in Assessing the World City Network: A Research Note on Data. *Urban Studies* 42, pp. 2371-2388.

DERUDDER, B. & F. WITLOX (2008), Mapping Global City Networks through Airline Flows: Context, Relevance, and Problems. *Journal of Transport Geography* 16, pp. 305-312.

DERUDDER, B., L. DEVRIENDT & F. WITLOX (2007a), Flying where you don't want to go. An empirical analysis of hubs in the global airline network. *Tijdschrift voor Economische en Sociale Geografie* 98, pp. 307-324.

DERUDDER, B., F. WITLOX & P.J. TAYLOR (2007b), US Cities in the World City Network. Comparing their Positions using Global Origins and Destinations of Airline Passengers. *Urban Geography* 28, pp. 74-91.

FEDERAL AVIATION ADMINISTRATION (FAA) (2000), Airport Activity Statistics of Certificated Air Carriers: Summary Tables 2000. <http://www.bts.gov>. last accessed 15/12/2007.

FRENKEN, K., S. VAN TERWISGA, T. VERBURG & G. BURGHOUWT (2004), Airline Competition at European Airports. *Tijdschrift voor Economische en Sociale Geografie* 95, pp. 233-242.

FRENKEN, K. & J. HOEKMAN (2006), Convergence in an enlarged Europe: the Role of Network Cities. *Tijdschrift voor Economische en Sociale Geografie* 97, pp. 321-326.

FRIEDMANN, J. (1986), The world city hypothesis. *Development and Change* 17, pp. 69-83.

FRIEDMANN, J. & G. WOLFF (1982), World city formation: an agenda for research and action. *International journal of urban and regional research* 6, pp. 309-344.

GOETZ, A.R. & C.J. SUTTON (1997), The Geography of Deregulation in the U.S. Airline Industry. *Annals of the Association of American Geographers* 87, pp. 238-263.

GRUBESIC, T. & M. ZOOK (2007), A Ticket to Ride: Evolving Landscapes of Air Travel Accessibility in the United States. *Journal of Transport Geography* 15, pp. 417-430.

GRUBESIC, T., MATISZIW, T. & ZOOK M. (2008), Global airline networks and nodal regions. *GeoJournal* 71, pp. 53-66.

GULDMANN, J.-M. (2004), Spatial Interaction Models of International Telecommunication Flows. In: M.F. Goodchild & D.G. Jannelle, eds., *Best Practices in Spatially Integrated Social Science*, pp. 400-442. Oxford: Oxford University Press.

HAGGETT, P., A. CLIFF & A. FREY (1977), *Locational Analysis in Human Geography*. London: Edward Arnold.

HAIR, J.F., R.E. ANDERSON, R.L. TATHAM & W.C. BLACK (1998), *Multivariate Data Analysis*. 5th edition. London: Prentice Hall.

HILL R. & J. KIM (2000), Global cities and developmental states: New York, Tokyo, and Seoul. *Urban Studies* 37, pp. 2167-2195

HOAG, L.P. (1969), The Weaver Method: An Evaluation. *Professional Geographer* 21, pp. 244-246.

- JOHNSON, L.J. (1969), Further Comments on Weaver's Method. *Professional Geographer* 21, pp. 246-247.
- MATSUMOTO, H. (2004), International Urban Systems and Air Passenger and Cargo Flows: Some Calculations. *Journal of Air Transport Management* 10, pp. 241-249.
- MEIJERS, E. (2007), From Central Place to Network Model: Theory and Evidence of a Paradigm Change. *Tijdschrift voor Economische en Sociale Geografie* 98, pp. 245-259.
- NYSTUEN, J.D. & M.F. DACEY (1968), A Graph Theory Interpretation of Nodal Regions. In: B.J.L. Berry & D.F. Marble, eds., *Spatial Analysis. A Reader in Statistical Geography*, pp 407-418. New Jersey: Prentice Hall.
- PHILIPS, L.T. (1985), Structural Change in the Airline Industry: Carrier Concentration at Large Hub Airports and its Implications for Competitive Behavior. *Transportation Journal* 25, pp. 18-28.
- PUEBLA, J.G. (1987), Spatial Structures of Network Flows: a Graph Theoretical Approach. *Transportation Research B* 21, pp. 489-502.
- REYNOLDS-FEIGHAN, A. (2001), Traffic Distribution in Low-Cost and Full-Service Carrier Networks in the US Air Transportation Market. *Journal of Air Transport Management* 7, pp. 265-275.
- SAEY, P. (2007), How cities scientifically (do not) exist: Methodological appraisal of research on globalizing processes of intercity networking. In: P.J. Taylor, B. Derudder, P. Saey & F. Witlox, eds., *Cities in globalization: Practices, policies and theories*, pp. 298-313. Routledge: London and New York.
- SASSEN, S. (1991), *The global city: New York, London, Tokyo*. Princeton: Princeton University Press.
- SHIN, K.-H. & M. TIMBERLAKE (2000), World Cities in Asia: Cliques, Centrality and Connectedness. *Urban Studies* 37, pp. 2257-2285.
- SMITH, D.A. & M. TIMBERLAKE (2001), World City Networks and Hierarchies, 1977-1997: An Empirical Analysis of Global Air Travel Links. *American Behavioral Scientist* 44, pp. 1656-1678.
- TAYLOR, P.J. (2001), Specification of the World City Network. *Geographical Analysis* 33, pp. 181-194.
- TAYLOR, P.J. (2004), *World City Network: A Global Urban Analysis*. London and New York: Routledge.
- TAYLOR, P.J. & B. DERUDDER (2004), Porous Europe: European Cities in Global Urban Arenas. *Tijdschrift voor Economische en Sociale Geografie* 95, pp. 527-538.
- TAYLOR, P.J., G. CATALANO & D.R.F. WALKER (2002), Measurement of the World City Network. *Urban Studies* 39, pp. 2367-2376.
- TAYLOR, P.J., B. DERUDDER & F. WITLOX (2007), Comparing Airline Passenger Destinations with Global Service Connectivities: A Worldwide Empirical Study of 214 Cities. *Urban Geography* 28, pp. 232-248.

TAYLOR, P.J., M. HOYLER & R. VERBRUGGEN (2008), External Urban Relational Process: Introducing Central Flow Theory to Complement Central Place Theory. *GaWC Research Bulletin* 261. Accessed at: <<http://www.lboro.ac.uk/gawc/rb/rb261.html>>.

VAN NUFFEL, N. (2007), Determination of the Number of Significant Flows in Origin-Destination Specific Analysis. The Case of Commuting in Flanders. *Regional Studies* 41, pp. 509-526.

WANG, C. (2003) Taipei as a global city: A theoretical and empirical examination. *Urban Studies* 40, pp. 309-334.

WEAVER, J. C. (1954), Crop Combination Regions in the Middle West. *Geographical Review* 44, pp. 175-200.

WITLOX, F. & B. DERUDDER (2007), Airline Passenger Flows through Cities. Some New Evidence. In: P.J. Taylor, B. Derudder, P. Saey & F. Witlox, eds., *Cities in Globalization. Practices, Policies and Theories*, pp. 37-51. London and New York:: Routledge.

ZOOK, M.A. & S.D. BRUNN (2006), From Podes to Antipodes: Positionalities and Global Airline Geographies. *Annals of the Association of American Geographers* 96, pp. 471-490.

Appendix 1

Cluster analysis without polarization indicator	Cluster analysis with polarization indicator
<u>cluster 1</u>	<u>cluster 1</u>
Amsterdam	London
Frankfurt	Paris
London	
Paris	
<u>cluster 2</u>	<u>cluster 2</u>
Brussels	Amsterdam
Copenhagen	Frankfurt
Milan	Munich
Munich	
Vienna	
<u>cluster 3</u>	<u>cluster 3</u>
Athens	Athens
Barcelona	Barcelona
Budapest	Brussels
Dusseldorf	Bucharest
Geneva	Budapest
Helsinki	Copenhagen
Istanbul	Helsinki
Lisbon	Istanbul
Madrid	Madrid
Manchester	Milan
Oslo	Prague
Prague	Rome
Rome	Stockholm
Stockholm	Vienna
Zurich	
<u>cluster 4</u>	<u>cluster 4</u>
Ankara	Belgrade
Basle	Berlin
Belgrade	Birmingham
Berlin	Bologna
Birmingham	Düsseldorf
Bologna	Geneva
Bucharest	Hamburg
Cologne	Hanover
Dublin	Lisbon
Göteborg	Ljubljana
Hamburg	Lyons
Hanover	Manchester
Ljubljana	Marseilles
Lyons	Oslo
Stuttgart	Riga
Venice	Stuttgart
	Venice
	Warsaw
	Zurich
	<u>cluster 5</u>

Warsaw	
<u>cluster 5</u>	
Aberdeen	Aberdeen
Bergen	Ankara
Bilbao	Basle
Bordeaux	Bergen
Bristol	Bilbao
Edinburgh	Bordeaux
Glasgow	Bristol
Leeds	Cologne
Luxembourg City	Dublin
Marseilles	Edinburgh
Naples	Glasgow
Newcastle	Göteborg
Nottingham	Leeds
Nuremberg	Luxembourg City
Riga	Naples
Sarajevo	Newcastle
Seville	Nottingham
Strasbourg	Nuremberg
Tallinn	Sarajevo
Tirana	Seville
Turin	Strasbourg
Valencia	Tallinn
Vilnius	Tirana
Zagreb	Turin
	Valencia
	Vilnius
	Zagreb

Figure 1. *Distribution of the number of passengers for London and Amsterdam in 2005.*

Figure 2. *Distribution of the number of passengers for Dublin and Bucharest in 2005.*

Figure 3. *Fictive networks*

Table 1. *Determination of the number of significant flows in fictive networks*

Spatial Unit X										
real distribution				ideal-typical configurations						
connection	rank of connection	flow	% flow	fictive flow	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	
h	1	50	28,57	$\hat{w}_1$	100,00	50,00	33,33	25,00	20,00	
a	2	43	24,57	$\hat{w}_2$	0,00	50,00	33,33	25,00	20,00	
d	3	42	24,00	$\hat{w}_3$	0,00	0,00	33,33	25,00	20,00	
g	4	21	12,00	$\hat{w}_4$	0,00	0,00	0,00	25,00	20,00	
c	5	12	6,86	$\hat{w}_5$	0,00	0,00	0,00	0,00	20,00	
e	6	7	4,00	$\hat{w}_6$	0,00	0,00	0,00	0,00	0,00	
b	7	0	0,00	$\hat{w}_7$	0,00	0,00	0,00	0,00	0,00	
f	8	0	0,00	$\hat{w}_8$	0,00	0,00	0,00	0,00	0,00	
total		175	100,00		100,00	100,00	100,00	100,00	100,00	
				COD*	0,31	0,55	0,88	0,80	0,63	
Spatial Unit Y										
real distribution				ideal-typical configurations						
connection	rank of connection	flow	% flow	fictive flow	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	
f	1	85	48,57	$\hat{w}_1$	100,00	50,00	33,33	25,00	20,00	
d	2	32	18,29	$\hat{w}_2$	0,00	50,00	33,33	25,00	20,00	
h	3	20	11,43	$\hat{w}_3$	0,00	0,00	33,33	25,00	20,00	
g	4	19	10,86	$\hat{w}_4$	0,00	0,00	0,00	25,00	20,00	
a	5	12	6,86	$\hat{w}_5$	0,00	0,00	0,00	0,00	20,00	
c	6	5	2,86	$\hat{w}_6$	0,00	0,00	0,00	0,00	0,00	
b	7	2	1,14	$\hat{w}_7$	0,00	0,00	0,00	0,00	0,00	
e	8	0	0,00	$\hat{w}_8$	0,00	0,00	0,00	0,00	0,00	
total		175	100,00		100,00	100,00	100,00	100,00	100,00	
				COD*	0,85	0,67	0,51	0,44	0,34	

*: COD = coefficient of determination

Table 2. *Coefficients of determination between the number of observed flows and the number of expected flows for Dublin, Amsterdam, Bucharest and London in 2005 (highest value marked in bold).*

Number of expected flows	Coefficient of determination			
	Dublin	Amsterdam	Bucharest	London
1	0.93	0.75	0.48	0.47
2	0.82	0.73	0.67	0.63
3	0.74	0.72	0.72	0.70
4	0.70	0.72	0.73	0.73
5	0.65	0.71	0.75	0.75
6	...	...	0.79	0.76
7			0.80	0.78
8			0.82	0.79
9			0.83	0.79
10			0.84	0.80
11			0.83	0.81
12			...	0.81
13				0.82
14				0.82
15				0.82
16				0.82
17				0.82
18				0.83
19				0.82
...				...

Table 3. *Number of significant connections based on MLA.*

<i>Number of significant connections</i>	<i>Cities 2005</i>
1	Aberdeen, Amsterdam, Basle, Bordeaux, Cardiff, Cologne, Cracow, Dublin, Glasgow, Leeds, Luxembourg City, Newcastle, Palermo, Rotterdam, Sarajevo, Seville, Southampton, Tallinn, Trieste
2	Bergen, Bilbao, Bratislava, Bristol, Düsseldorf, Frankfurt, Genoa, Munich, Naples, Nottingham, Tirana, Valencia, Vilnius
3	Ankara, Birmingham, Edinburgh, Göteborg, Hamburg, Manchester, Nuremberg, Strasbourg, Turin, Zagreb
4	Berlin, Geneva, Oslo, Prague, Warsaw
5 or more	Athens, Barcelona, Belgrade, Bologna, Brussels, Bucharest, Budapest, Copenhagen, Hanover, Helsinki, Istanbul, Lisbon, Ljubljana, London, Lyons, Madrid, Marseilles, Milan, Paris, Riga, Rome, Stockholm, Stuttgart, Venice, Vienna, Zurich

Table 4. *Results cluster analysis without MLA, 2005. (Average number of significant connections has not been taken into account in the cluster analysis, and is given for illustrative purposes only.)*

<i>Cluster</i>	<i>Number of cities</i>	<i>Examples of member cities</i>	<i>Average number of passengers</i>	<i>Average number of destinations</i>	<i>Average number of significant connections</i>
1	4	Frankfurt, London	21,284,389 (+++)	72 (+++)	6 (+)
2	5	Brussels, Milan	8,942,270 (+)	63 (++)	6 (+)
3	17	Lisbon, Oslo	5,136,207 (+)	37 (+)	5 (+)
4	15	Basle, Berlin	1,470,869 (--)	20 (-)	4 (+)
5	24	Naples, Seville	409,790 (---)	7 (---)	2 (---)
all cities	65		3,718,950	26	4

+: less than one standard deviation above the mean, ++: between one and two standard deviations above the mean, +++: more than two standard deviations above the mean, -: smaller than the mean but larger than the median, --: smaller than the median but larger than the 40th percentile, ---: smaller than the 40th percentile.

Table 5. *Results cluster analysis with MLA, 2005.*

<i>Cluster</i>	<i>Number of cities</i>	<i>Examples of member cities</i>	<i>Average number of passengers</i>	<i>Average number of destinations</i>	<i>Average number of significant connections</i>
1	2	London, Paris	25,262,929 (+++)	76 (+++)	10 (+++)
2	3	Amsterdam, Frankfurt	14,940,826 (+++)	68 (+++)	2 (---)
3	14	Brussels, Rome	6,509,779 (+)	45 (+)	7 (++)
4	19	Lisbon, Oslo	2,148,892 (-)	24 (-)	5 (+)
5	27	Edinburgh, Seville	533,983 (---)	9 (---)	2 (---)
all cities	65		3,718,950	26	4

+: less than one standard deviation above the mean, ++: between one and two standard deviations above the mean, +++: more than two standard deviations above the mean, -: smaller than the mean but larger than the median, --: smaller than the median but larger than the 40th percentile, ---: smaller than the 40th percentile.

ⁱ Note that we cluster cities rather than airports. Both the differences in spatial units being clustered (cities versus airports), and the differences in data used (actual number of passengers versus seat capacity and airline data versus data on corporate networks) make a direct comparison of our results with those of Burghouwt & Hakfoort (2002) and Taylor & Derudder (2004) impossible.

ⁱⁱ In fact, MLA is an extension and improvement of Weaver's combination index (Weaver 1954; Haggett *et al.* 1977). The Weaver method uses the sum of the squared differences (instead of the coefficient of determination) between the fictive and the real percentages. The number of significant flows corresponds to the j^{th} cycle with the smallest sum of the squared differences. For a discussion of the value of the Weaver method, which seems to occasionally produce problematic results, see Hoag (1969) and Johnson (1969).

ⁱⁱⁱ Situation in January 2009.

^{iv} However, the distortion will probably not be very large. The results of a recent empirical analysis by Alderighi *et al.* (2007) indicate that the spatial network configuration of full service carriers and low-cost carriers is very similar. On the other hand, their temporal configuration is very different: by adopting a wave-system structure in the airline flight schedule, full service carriers show high temporal concentration, whereas low-cost carriers have almost a zero temporal concentration.

^v We preferred using data for 2002 instead of 2001 to avoid distortions due to the terrorist attacks of 11 September 2001. According to the AEA, the severe drop in air traffic after 11 September stabilized by the end of 2001, and by March 2002, traffic had almost returned to previous year's levels (AEA 2003).

^{vi} We only included those carriers that provide data for both 2002 and 2005. Moreover, some carriers only serve intercontinental flights or only freight flights. This explains why not all 35 AEA member airlines are present in our resulting database.

^{vii} http://www.lboro.ac.uk/gawc/datasets/da11_2.html (last accessed on 11 December 2008). The Globalization and World Cities (GaWC) Study Group takes a relational stance when analysing the world city network. Generally speaking, global inter-city matrices can be constructed in two different ways, which may be labelled the corporate organization approach and the infrastructure approach. The GaWC world cities we used in this paper are part of a list of 315 cities that is based on the corporate organization approach. This approach "is premised on the observation that inter-city relations are primarily created by firms that pursue global strategies and are thus prime world city agents." (Witlox & Derudder 2007, p. 39). For more information on the method, which makes use of data on the office networks of advanced producer service networks, see Taylor (2001, 2004), Taylor *et al.* (2002) and Derudder & Taylor (2005).

^{viii} We used Ward's method which minimizes the within-group variation by calculating at each step of the clustering procedure the sum of squares between each two clusters summed over all variables, and joining those two clusters that result in the smallest increase in this sum (Hair *et al.* 1998).