Exploring the spatial relatedness network of the global system of international migration

Josef Novotný and Jiří Hasman

Department of Social Geography and Regional Development, Faculty of Science, Charles University, Albertov 6, Praha 2, 12843, Czech Republic

ABSTRACT
The global geography of international migration has a complex pattern that emerges from the distinct spatial distributions of many individual migrant groups. To visualise this structure, we use the spatial relatedness network of the global international migration system. Unlike traditional spatial networks based on real spatial interactions, spatial relatedness networks are based on possible relationships which are determined as a degree of agreement between the maps of the analysed phenomena. In this paper, the spatial relatedness network of the global migration system is conceptualised as an undirected network in which nodes represent individual migrant groups defined by the country of origin, and links between nodes correspond to their spatial relatedness in terms of similarity in their spatial distributions. We describe the construction of the network and explore some of its properties such as the clustering of migrant groups that share similar positions in the global migration system, signifying distinct spatial regimes of migrants’ destination choices.

1. Introduction
The global geography of international migration has a complex and inherently high-dimensional structure. It emerges as an assemblage of distinct spatial patterns that characterise distributions of different migrant groups defined by migrants’ country of origin. The global geography of international migration is thus difficult to depict in an aggregate way using traditional maps. In order to visualise the patterns of the global migration system we seek to apply the spatial relatedness approach, recently introduced to migration studies by Novotný and Hasman (2015). The spatial relatedness between a pair of migrant groups refers to the degree of agreement between world maps showing their concentrations in individual countries. We use the 2015–revised UN data on the stocks of international migrants disaggregated by individual destination countries and 233 source countries, and quantify the spatial relatedness matrix containing the spatial relatedness distances for all possible pairs of these 233 migrant groups (see Supplementary Data). Based on the matrix, we construct the spatial relatedness network of the global system of international migration and explore some of its properties.

2. Spatial relatedness networks
The network analysis has become a powerful tool that helps to understand various complex phenomena that are ubiquitous in both social and natural systems (e.g. Newman, 2010). While nodes are the basic building blocks of any network, links between nodes are essential components that determine the structural characteristics of networks. Usually, these links represent real interactions or relationships, as in a great majority of the social network analysis applications (e.g. networks based on kinship relations, friendships, collaboration ties, or other forms of social interactions). In some instances, these links, nodes, or both are explicitly spatial in character so the networks are labelled as spatial networks or spatially embedded networks (Barthélemy, 2011; Haggett and Chorley, 1969; O’Sullivan, 2014).

However, the links of a network can also represent potential interactions arising from a co-occurrence or another proximity measure, and therefore they can also indicate potential relationships. Some examples of the co-occurrence networks are various lexical networks based on the co-occurrences of certain syllables or words in texts (Dorogovtsev & Mendes, 2001; i Cancho & Solé, 2001; Jenssen, Lægreid, Komorowski, & Hovig, 2001), or various biological co-occurrence networks such as protein domain co-occurrence networks, genetic co-occurrence networks, microbial co-occurrence networks, or disease proximity networks (Faust et al., 2012; Hidalgo, Blumm, Barabási, & Christakis, 2009; Wuchty, 2001), among others. Unlike the analysis of real-interaction networks, the analysis of co-
occurrence networks primarily represents a data-mining or a dimension-reduction exercise that is helpful for exploring and visualising the underlying structure of high-dimensional systems in a low-dimensional space.

A subgroup of co-occurrence networks are networks in which the links correspond to a measure of co-occurrence within spatial units, or, in other words, to a measure of a spatial overlap between maps of analysed phenomena. A few examples of such spatial co-occurrence networks can be found in ecology, for instance, networks of spatial co-incidence of species (e.g. Araújo, Rozenfeld, Rahbek, & Marquet, 2011; Bell, King, Bohan, & Symondson, 2010). The spatial relatedness network of the global system of international migration presented in this paper represents just one example of such a spatial co-occurrence network. However, we use the term ‘spatial relatedness network’ in the present context to distinguish the measure of spatial relatedness described above from more traditional co-occurrence indices (e.g. Hayek, 1994). In addition, it also allows us to make a terminological distinction between spatial relatedness networks and traditional spatial networks where the links or nodes are explicitly spatially embedded.

3. Measuring spatial relatedness of migrant groups

To assess the spatial relatedness of two migrant groups we use the same approach as Novotný and Cheshire (2012) or Novotný and Hasman (2013), who applied an analogy to the methodology used by Hidalgo, Klinger, Barabási, and Hausmann (2007) for different purposes. The spatial relatedness of two migrant groups is conceptualised as the conditional probability of their joint spatial concentration where the concentration of a migrant group (i) in a country (r) is firstly quantified by the common localisation quotient as:

\[ LQ_{i,r} = \frac{F_{i,r}}{\sum_j F_{i,r} + \sum_i F_{i,r}} \]

where \( F_{i,r} \) refers to the population size of the group i in the country r.

The next step is a binarization of the continuous localisation quotient to assess whether group i established a concentration in particular countries by considering an intuitively appealing threshold of 1. This means that if \( LQ_{i,r} > 1 \), group i is concentrated in country r (we also tested various other thresholds and found the threshold of 1 to provide a good and intuitively appealing limiting value).

The size of the set of countries in which group i establishes its concentration can be formally written as \( \{r: LQ_{i,r} > 1\} \). Sets of counties in which groups i and j establish their concentrations, that is \( \{r: LQ_{i,r} > 1\} \) and \( \{r: LQ_{j,r} > 1\} \), are considered input data for the quantification of a pair-wise measure of spatial relatedness between these two migrant groups as:

\[ D_{ij} = \min(D^1_{ij}; D^2_{ij}), \]

where the term \( D^1_{ij} \) captures the probability that group i concentrates in country r conditional to the concentration of group j in this country:

\[ D^1_{ij} = P(LQ_{i,r} > 1 \mid LQ_{j,r} > 1) = \frac{|\{r: LQ_{i,r} > 1\} \cap \{r: LQ_{j,r} > 1\}|}{|\{r: LQ_{j,r} > 1\}|}, \]

and the term \( D^2_{ij} \) analogously corresponds to the probability that group j concentrates in country r conditional to the concentration of group i in this country:

\[ D^2_{ij} = P(LQ_{j,r} > 1 \mid LQ_{i,r} > 1) = \frac{|\{r: LQ_{j,r} > 1\} \cap \{r: LQ_{i,r} > 1\}|}{|\{r: LQ_{i,r} > 1\}|}. \]

The value of \( D_{ij} \) ranges between 0 (no joint concentrations) and 1 (both groups establish their concentrations exclusively in the same countries) and can be interpreted as the probability that one of the groups concentrates in a country where the other is concentrated.

As already noted, the spatial relatedness index \( D_{ij} \) can be considered a measure of agreement between two maps that depict the concentrations of a migrant group in individual countries. This approach can therefore be related to previous literature dealing with the measurement of (dis)agreement between maps (e.g. Foody, 2006; Pontius & Millones, 2011; Pontius, Peethambaram, & Castella, 2011; Pontius & Santacruz, 2014; Visser & De Nijs, 2006) that typically considers more complicated comparisons of nominal maps with multiple categories. One advantage of this measurement is that the initial calculation of localisation quotients for individual migrant groups and countries and the subsequent binarization makes the index less sensitive to extreme values. This is an important property of a spatial relatedness index especially when the focus is on the comparison of phenomena with highly right-skewed size distributions (Novotný & Nosek, 2009). Likewise, it is pertinent to the size distributions of immigrant populations across the countries studied in this paper.

4. Data

The results provided in this paper draw on the 2015-revised UN data on the stocks of international migrants disaggregated by 233 destination and source countries.
The data set summarises information from population censuses complemented by various other sources (population registries, representative surveys, and so on) and is freely available at: http://www.un.org/en/development/esa/population/migration/data/estimates2/estimates15.shtml. The data set containing original data as well as our results on the spatial relatedness $D_{i,j}$ observations between all possible pairs of 233 migrant groups (that is, 27,028 unique $D_{i,j}$ observations) can be accessed in the Supplementary Data file.

### 4.1 Constructing the spatial relatedness network of the global system of international migration

The spatial relatedness network of the global system of international migration appears in Main Map (parts S1 and S4). To construct this network, we used Cytoscape (Shannon et al., 2003) and applied an edge-weighted spring embedded algorithm on the matrix of spatial relatedness observations ($D_{i,j}$) between individual migrant groups. This algorithm considers weights proportional to the values of $D_{i,j}$ so that higher spatial relatedness observations are more consequential regarding the final network layout. By applying this technique, the network can be interpreted as an analogy to a physical system where nodes attract each other by forces proportional to their pair-wise relatedness, while the algorithm minimises the energy of the physical system and assigns the nodes to their positions accordingly. We uncovered that the size distribution of $D_{i,j}$ observations is highly right skewed (S3 in Main Map), while 42.6% of all 27,028 $D_{i,j}$ links correspond to zero, and many other observations have a value close to zero. Based on an inspection of the size distribution in S3, we considered the upper 28.5% of the $D_{i,j}$ observations with the value $D_{i,j} > 0.1$ for the construction of the network. In addition, for the purpose of simplicity, only significant links with the value of $D_{i,j} > 0.3$ are visible in S1 and S4. The node size in the main network (S1 and S4) is proportional to the square root of the population size of particular migrant groups. The colours of the nodes in S1 represent the respective world regions of individual source countries as depicted in the map of world macro-regions in S2 in Main Map.

In addition, Figure 1 shows the genesis of the spatial relatedness network of the global migration system (as in S1) in terms of different layers of the network determined by different thresholds of $D_{i,j}$. For example, based on the main network in S1, part A in Figure 1 shows solely 0.2% of the strongest links that satisfy $D_{i,j} > 0.7$ and their respective nodes (23.2% of the 233 analysed migrant groups). The other parts in Figure 1 apply lower $D_{i,j}$ thresholds, and therefore the visualisations converge consecutively to the main network in S1. In this way, we seek to uncover the positions of the most spatially related migrant groups and communities shown in the figures based on the higher $D_{i,j}$ thresholds. In addition, Table 1 depicts the top 15 pairs of the most tightly spatially related migrant groups. Evidently, these pairs contain mostly smaller countries with smaller but highly concentrated migrant populations. The set of all 27,028 spatial relatedness ($D_{i,j}$) observations appears in the Supplementary Data file.

### 5. Community structure of the global migration system

The spatial relatedness network in the part S1 of Main Map reveals a clear clustering of migrant groups according to their geographical and cultural similarity. The groups from the same world regions (as distinguished by the same colours shown in S2) mostly occupy similar parts of the network. Moreover, there are also some clearly recognisable patterns at both the lower and higher level of detail than captured by the macro-regional groupings. Regarding the network macro-structure, the cluster of mostly European migrant groups signifies the densest central core of the network surrounded by other clusters and communities more or less connected to the European core cluster. The pattern is intuitively straightforward: migrant groups from the Americas can be found to the bottom right of the central European cluster, those from Oceania and Asia to the bottom left, those from Africa in the upper left, and those from post-Soviet countries to the upper right of the European core. On a more granular level, some clear internal divisions are recognisable within particular macro-regional communities. This holds for the European core cluster, where the West and East European migrant groups form two clearly distinguishable parts: the African area of the network, where recognisable divisions exist between the West, East, and South African groups; and, for both the Asian and American parts of the network with similarly recognisable internal divisions.

The central position of the cluster of mostly (though not exclusively) European migrant groups can also be documented by their high degree of centrality within the network. A simple measure of centrality or positional embeddedness in the network can be expressed as the summation of all $D_{i,j}$ links of a given migrant group to all other groups in the network. Table 2 provides a list of the 15 most embedded migrant groups (the full set of results can be found in the Supplementary Data file), which confirms the central position of the core cluster consisting of mainly European groups and a few other source countries with a rich history of emigration (Israel, Iran).

Although already a visual inspection of the network in S1 uncovers interesting and meaningful patterns, we seek to delve deeper into the significance of clusters or
Figure 1. Genesis of the spatial relatedness network of the global system of international migration (as in S1). Notes: The figure shows different layers of the spatial relatedness network of the global system of international migration determined on the basis of different $D_{ij}$ thresholds. A – Nodes connected by links with $D_{ij} > 0.7$ (23.2% of all nodes, 0.2% of all links); B – Nodes connected by links with $D_{ij} > 0.6$ (44.2% of all nodes, 0.5% of all links); C – Nodes connected by links with $D_{ij} > 0.5$ (70.8% of all nodes, 1.3% of all links); D – Nodes connected by links with $D_{ij} > 0.4$ (89.7% of all nodes, 2.7% of all links); E – Nodes connected by links with $D_{ij} > 0.3$ (95.7% of all nodes, 5.6% of all links); F – Nodes connected by links with $D_{ij} > 0.2$ (99.6% of all nodes, 12.2% of all links).
Table 1. The 15 most tightly spatially related migrant group pairs.

| Migrant group 1         | Migrant group 2         | $D_{i,j}$ |
|-------------------------|-------------------------|-----------|
| Guadeloupe              | Martinique              | 0.900     |
| Kuwait                  | United Arab Emirates    | 0.895     |
| Costa Rica              | Panama                  | 0.864     |
| Barbados                | Saint Vincent and the Grenadines | 0.857 |
| Saint Vincent and the Grenadines | Trinidad and Tobago | 0.857     |
| Azerbaijan              | Belarus                 | 0.813     |
| Barbados                | Trinidad and Tobago     | 0.810     |
| Latvia                  | Lithuania               | 0.810     |
| Costa Rica              | Cuba                    | 0.792     |
| Jamaica                 | Saint Vincent and the Grenadines | 0.792 |
| Colombia                | Costa Rica              | 0.773     |
| Colombia                | Panama                  | 0.773     |
| Guyana                  | Jamaica                 | 0.769     |
| Jordan                  | Yemen                   | 0.769     |
| Saint Kitts and Nevis   | Saint Vincent and the Grenadines | 0.762 |

Table 2. The 15 most embedded migrant groups in the network.

| Sum of spatial relatedness links ($D_{i,j}$) to all other migrant groups |
|-----------------------------|
| Hungary                     | 36.2                      |
| Israel                      | 35.1                      |
| Finland                     | 35.1                      |
| Denmark                     | 33.2                      |
| Netherlands                 | 32.9                      |
| Greece                      | 32.8                      |
| Germany                     | 32.7                      |
| Belgium                     | 32.7                      |
| Austria                     | 31.9                      |
| Iran                        | 31.8                      |
| Poland                      | 31.7                      |
| Slovakia                    | 31.6                      |
| Norway                      | 31.2                      |
| Bulgaria                    | 31.0                      |
| Sweden                      | 30.7                      |

communities of migrant groups in the network. Accordingly, we used the affinity propagation technique, which is recommended for the purposes of community detection in networks (Bodenhofer, Kothmeier, & Hochreiter, 2011; Frey & Dueck, 2007). This technique classifies the nodes of a network into a set of mutually exclusive communities by identifying the most typical examples of particular parts of the network (referred to as exemplars) and the communities of nodes in their corresponding clusters. The communities obtained by affinity propagation applied to the set of $D_{i,j}$ observations examined for this paper have been distinguished by different colours in the network S4 and the geographical composition of these communities is depicted on the map in S5. The exemplars of particular communities have been marked with a red border of the respective nodes in the network S4 and appear in the captions for individual maps in S7. The S7 in Main Map indicates the spatial distribution of migrant groups in particular communities obtained by the affinity propagation. More specifically, these maps show the percentage share of migrant groups from all groups in a given community concentrated (i.e. with $LQ_{i,r} > 1$) in a given country.

The affinity propagation communities of migrant groups represent the subsystems of the global system of international migration that capture distinct spatial regimes of migrants’ destination choices. Each of the communities reveals a certain degree of its internal density in terms of the extent of similarity between migrant groups within a community. These communities also reveal some degree of their mutual relatedness. Table S6 in Main Map presents the results of our effort to quantify the internal density of the affinity propagation communities and their mutual external relatedness. The former has been expressed as the average value of $D_{i,j}$ within a given community (shown in the diagonal of S6), while the external relatedness was quantified as the average $D_{i,j}$ between these communities. The external relatedness figures are considerably lower than the values of internal density. This indicates the relevance of the community structure detected by affinity propagation clustering. Although all values of internal density depicted in the diagonal of S6 are considerably higher than the global average of $D_{i,j}$ calculated from all observations (0.079), there is a notable variation between particular affinity propagation communities. The highest internal density was revealed for the cluster of migrant groups from post-Soviet countries (A), while the lowest internal density was found for the communities ‘G’ and ‘I’, which can be due to the fact that they are rather residual communities composed of countries (or small groups of countries, e.g. those belonging to the Oceania group) which cannot be assigned to other communities in the initial iteration steps of affinity propagation.

The variation in external relatedness between particular communities mostly coincides with what can visually be inferred from the main network in S1 and S4. In addition, the last column of Table S6 shows the average external relatedness of individual affinity propagation communities. Interestingly, it indicates the degree of embeddedness of these particular communities in the spatial relatedness network. In other words, it provides information about how extensively these individual subsystems interact with the rest of the global migration system. The highest aggregate of external relatedness was revealed for the community of mostly (but not exclusively) European migrant groups (‘D’) that forms the main core of the network in S1 and S4. By contrast, the two lowest aggregate values of external relatedness were found for the communities of post-Soviet migrant groups (‘A’) and North-West African migrant groups (‘B’) which is mirrored by their positions in the upper parts of the network, largely disconnected from other communities except for the European core cluster (and a few links to some other African groups in the latter case).
6. Conclusions

The paper distinguished between spatial relatedness networks (as a specific type of spatial co-occurrence networks) based on possible relationships determined by a measure of spatial relatedness and conventional networks based on some real relationships, including conventional spatial networks with explicitly spatially embedded links or nodes. The construction of the spatial relatedness network of the global system of international migration based on the spatial relatedness measure $D_{ij}$ was described. Basically, this measure quantifies the extent of similarity between maps showing the distributions of spatial concentrations of two migrant groups.

We presented the spatial relatedness network of the global system of international migration which provides a unique, information-rich, and intuitively appealing representation of the patterns of the global migration system. The community structure of the network using affinity propagation clustering was explored. The migrant group communities identified represent distinct subsystems of the global migration system in terms of the distinct spatial regimes of migrants’ destination choices. The internal density and external relatedness of the affinity propagation communities of migrant groups were assessed. The former provides information about the extent of internal homogeneity of a given migration subsystem, while the latter reveals the degree of its connections to other communities, thus indicating its embeddedness with respect to the whole network, or, on the contrary, its separation from the rest of the global migration network.

Software

The matrix of spatial relatedness figures ($D_{ij}$) between individual migrant groups was calculated using EasyStat 1.0 (currently only available in Czech—http://web.natur.cuni.cz/~pepino/EasyStat_1_0_manual.pdf). The network visualisation in S1 and S4 was constructed in Cytoscape (Shannon et al., 2003) as described above. The maps in S2 and S5 were created in ArcGis 10.3.1. The Kernel density plot in S3 was created using SPSS and the affinity propagation clustering was obtained in APCluster library (Bodenhofer et al., 2011) for the R software (R Core Team, 2014). InDesign was used to organise the final layout of the Main Map.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Araújo, M. B., Rozenfeld, A., Rabbek, C., & Marquet, P. A. (2011). Using species co-occurrence networks to assess the impacts of climate change. Ecography, 34, 897–908. doi:10.1111/j.1600-0587.2011.06919.x
Barthélemy, M. (2011). Spatial networks. Physics Reports, 499, 1–101. doi:10.1016/j.physrep.2010.11.002
Bell, J. R., King, A. R., Bohan, D. A., & Symondson, W. O. (2010). Spatial co-occurrence networks predict the feeding histories of polyphagous arthropod predators at field scales. Ecography, 33, 64–72. doi:10.1111/j.1600-0587.2009.06046.x
Bodenhofer, U., Kothmeier, A., & Hochreiter, S. (2011). APCluster: An R package for affinity propagation clustering. Bioinformatics, 27, 2463–2464. doi:10.1093/bioinformatics/btr406
Dorogovtsev, S. N., & Mendes, J. F. F. (2001). Language as an evolving word web. Proceedings of the Royal Society of London B: Biological Sciences, 268, 2603–2606. doi:10.1098/rspb.2001.1824
Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., & Huttenhower, C. (2012). Microbial co-occurrence relationships in the human microbiome. PLoS Computational Biology, 8, e1002606. doi:10.1371/journal.pcbi.1002606
Foody, G. M. (2006). What is the difference between two maps? A remote senser’s view. Journal of Geographical Systems, 8, 119–130. doi:10.1007/s10109-006-0023-z
Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. Science, 315, 972–976. e136800. doi:10.1126/science.1136800
Haggett, P., & Chorley, R. J. (1969). Network analysis in geography. London: Edward Arnold.
Hayek, L. C. (1994). Analysis of amphibian biodiversity data. In W. R. Heyer, M. A. Donnelly, R. W. McDiardm, L. C. Hayek, & M. S. Foster (Eds.), Measuring and monitoring biological diversity: Standard methods for amphibians (pp. 207–269). Washington, DC: Smithsonian Press.
Hidalgo, C. A., Klíner, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. Science, 317, 482–487. doi:10.1126/science.114458
Hidalgo, C. A., Blumm, N., Barabási, A. L., & Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. PLoS Computational Biology, 5, e1000353. doi:10.1371/journal.pcbi.1000353
Cancho, R. F., & Solé, R. V. (2001). The small world of human language. Proceedings of the Royal Society of London B: Biological Sciences, 268, 2261–2265. doi:10.1098/rspb.2001.1800
Jensen, T. K., Lægreid, A., Komorowski, J., & Hovig, E. (2001). A literature network of human genes for high-throughput analysis of gene expression. Nature Genetics, 28, 21–28. doi:10.1038/ng0501-21
Newman, M. E. J. (2010). Networks: An introduction. New York: Oxford University Press.
Novotný, J., & Cheshire, J. A. (2012). The surname space of the Czech Republic: Examining population structure by network analysis of spatial co-occurrence of surnames. PLoS One, 7, e48568. doi:10.1371/journal.pone.0048568
Novotný, J., & Hasman, J. (2015). The emergence of regional immigrant concentrations in USA and Australia: A spatial relatedness approach. PLoS One, 10, e0126793. doi:10.1371/journal.pone.0126793
Novotný, J., & Nosek, V. (2009). Nomothetic geography revisited: Statistical distributions, their underlying principles, and inequality measures. Geografie, 114, 282–297.
O’Sullivan, D. (2014). Spatial network analysis. In M. M. Fischer & P. Nijkamp (Eds.), Handbook of Regional Science (pp. 1253–1273). Heidelberg: Springer.

Pontius, R. G., & Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, 32, 4407–4429. doi:10.1080/01431161.2011.552923

Pontius, R. G., Peethambaram, S., & Castella, J. C. (2011). Comparison of three maps at multiple resolutions: A case study of land change simulation in Cho Don district, Vietnam. Annals of the Association of American Geographers, 101, 45–62. doi:10.1080/00045608.2010.517742

Pontius, R. G., & Santacruz, A. (2014). Quantity, exchange, and shift components of difference in a square contingency table. International Journal of Remote Sensing, 35, 7543–7554. doi:10.1080/2150704X.2014.969814

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13, 2498–2504, e1239303. doi:10.1101/gr.1239303

Visser, H., & De Nijs, T. (2006). The map comparison kit. Environmental Modelling & Software, 21, 346–358. doi:10.1016/j.envsoft.2004.11.013

Wuchty, S. (2001). Scale-free behavior in protein domain networks. Molecular Biology and Evolution, 18, 1694–1702.