Researching commuting to work using the methods of complex network analysis

Pálóczi, Gábor

University of Debrecen

10 October 2016

STUDIES

Researching commuting to work using the methods of complex network analysis

Gábor Pálóczi
University of Debrecen, Hungary
E-mail: paloczig@gmail.com

In the current paper the possible utilization of complex network analysis in spatial researches was investigated. The organizational and developmental regularities of networks were demonstrated from the aspect of regional development planning. The reviewed regularities provide a new approach of the regional developments.

The dependencies of settlements were analysed with the application of disparity method on the basis of the commuting matrix of the census from 2011. The disparity of out-commuting exceeded the level of in-commuting in all population categories, producing a more significant dependency relation in case of out-commuting. In general, the value of disparity increases with decreasing population number in settlements and dependency grows. This can be related with decrease in the level of degree and commuting distance. According to detailed results, the method of disparity might be effectively used in additional spatial analyses as well.

The community detection procedures of the complex network analysis were also applied for spatial division. Modularity optimization with the Louvain method was successfully used in the delimitation of larger territorial units. Smaller units can be created by the increase of the resolution but modularity stability deteriorates. At the same time the composition of the units changes. In the light of the results, it could be stated that regions formed by commuting relations (according to the process of regionalism) did not match the Hungarian NUTS2 statistical regions, but natural borders and NUTS-3 level administrative boundaries could be detected in more cases. The differences between the results and NUTS-3 boundaries are not unique distortions caused by the methodology but these reflect real commuting relations (the local labour system units were discussed in a previous study). The methodology might be appropriate to detect the hierarchical order of the local labour system's units. The method is adaptable for additional analysis of spatial interactions.

Keywords:
network analysis, commuting, disparity, dependency, regionalization, community detection

Introduction

Spatial research, apart from studying the characteristics and dynamics of spatial units, is focused on processes (innovation, migration and labour migration) between spatial units (settlements and public administration units) and resistance (accessibility via public roads). More detailed knowledge regarding flow and the increase of intensity of connections may help in reducing regional unevenness.

Regarding flow between settlements, only a few databases are available with data that were collected using united methods and suitable for settlement level analyses. Although the data series of commuting to work obtained in the course of the censuses meet the requirements, they are not used frequently (Kiss–Szalkai 2014). For further analyses, a new methodological approach was explored in international literature. For analysing commuting data, complex network analysis (Russo et al. 2007; De Montis et al. 2010; Caschili–De Montis 2013) is used mostly apart from spatial interaction models (Patuelli et al. 2007; LeSage–Fischer 2010). The primary aim of the present research is to analyse commuting relations using complex network analysis, presenting the possible applications of the method, its limits and further potentials.

Description of network analysis

The toolbar of complex network analysis is used frequently in international literature. Its popularity is based on its approach towards problems as differing from the widely used analytical and contextual data network analysis, which uses relational data. Analytical data are information related to a settlement that cannot be separated, such as unemployment ratio. Contextual data compare a parameter of a given settlement to a similar parameter of its surroundings, so that it can be decided whether, for example, the unemployment ratio in the settlement is better than in the region. Relational data, however, indicate a relationship between two settlements and sometimes the strength of this relationship (e.g. commuting to work and traffic relations; Letenyei 2005). This method is based on graph theory. Scientists simplify reality to nodes and edges, connecting them during the application of the method. Nodes of the graphs could be the given settlements, while flow between them could be the edges of the graphs. Several ways are available to plot real processes. Networks can be directed so that edges are not lines but arrows pointing from one node (source node) towards the other (destination node). Basically, the presence or lack of connection between the given nodes is recorded using binary coding (value of 0 or 1 – dichotomous network). Edges can be weighted based on the volume of material or information flow, the capacity of the infrastructural element or the length of the edge. The geographical space modelled by nodes and edges can be analysed using the methods of graph theory, linear algebra (Dusek 2013) and network analysis.

Network analysis and spatial research

Applying network analysis in spatial research can be regarded as a renaissance. Papers analysing spatial interactions on graph theory basis were published as early as the 1960s and 1970s (Dusek 2013, Klapka et al. 2013). The methodological background for the present work was taken mostly from studies in physics and biology. Analogies and methods are frequently taken from physics in spatial studies, such as gravitational models (Dusek 2003) and the spring-block model (Máté et al. 2013).

Network analysis was developed in numerous disciplines – simultaneously in many cases – in recent decades according to the challenges in the given scientific fields. It has become increasingly widespread in the Hungarian social geographical and regional geographical literature, besides being used physics, medical sciences, biology, ecology (Szabó et al. 2012) and sociology (Letenyei 2005). Its application in spatial research is justified as the intensity, volume and balance of spatial relations can be considered as development indicators: developed, dynamic regions have complex and intensive inner and outer relations, while in peripheral regions, relations are weak and one-sided, indicating the presence of problems (Nemes Nagy 2009). Companies prefer settlements with a developed network of connections (e.g. good accessibility; Kozma 1998).

Space categories and terms of spatial science can be interpreted in network research as well (Vida 2013), and this method can be applied to – beside others – the delimitation of peripheral areas (regarding methods of identification, see Pénzes 2014) and areas of attraction. According to Hungarian literature, network analysis has been applied to studying scientific cooperation (Vida 2012), tourism (Madarász–Papp 2013), interbank accounting transactions (Pál 2014), creation of polycentric urban regions (Fleischer 2009), traffic (Géber 2007, Pálóczi–Pénzes 2011, Szabó et al. 2013), social networks (Lengyel et al. 2015) and subregional commuting (Letenyei 2000). In the future, the methods could be applied successfully to studying online spaces and notoriety networks (Jakobi–Lengyel 2014) or global value chains (Molnár 2012).

Nodes of networks represent geographical locations in spatial research. Considering the national level, they would be settlements, while at the settlement level, they would be road junctions or parts of settlements. Edges would represent elements of the linear infrastructure (e.g. roads and railways) or material and information flow along them (Dusek 2013).

Applied data – adjacency matrix

This study is based on the commuting to work data of the census in 2011. The database involves data on daily commuting between the place of residence and workplace in order to carry out work. Settlements represent the nodes, whereas commuting between settlements represents the edges of a network. Data were sorted into an adjacency matrix and the following modifications were carried out:

- data of those commuting to abroad and variable destinations were removed;
- connections of the districts of Budapest were aggregated; that is, Budapest is adjacent to every settlement in our matrix with which any of the districts were in connection regarding employment, (weighing factors were of course summed);
- applied parameters cannot handle the connection of the settlements with themselves (local employees); therefore, in the diagonal of the matrix, the value is zero and
- the symmetric version of the directed network matrix was prepared as well; that is, values according to the direction of commuting between the settlement pairs were summed, and the obtained value was used for both in- and out-commuting.

The modifications are in harmony with the data processing described in the literature (Pénzes et al. 2014; De Montis et al. 2007).

Nodes of the network – in the case of directed networks – can be characterised best by the number of edges running in (in-degree) and out (out-degree) of them. There are 128,610 edges among the 3,154 nodes in the network. The highest number of edges belongs to Budapest (2,162), while Gagyapáti is an isolated inclusion and connected to no settlements. The weight of an average edge is 18; that is, 18 people are commuting in an average relation, whereas more than 13,000 people are commuting between Budapest and Budaörs. A strong, positive connection was detected between the number of residents and grade (r=0.73).

Characteristics of the commuting system as a network

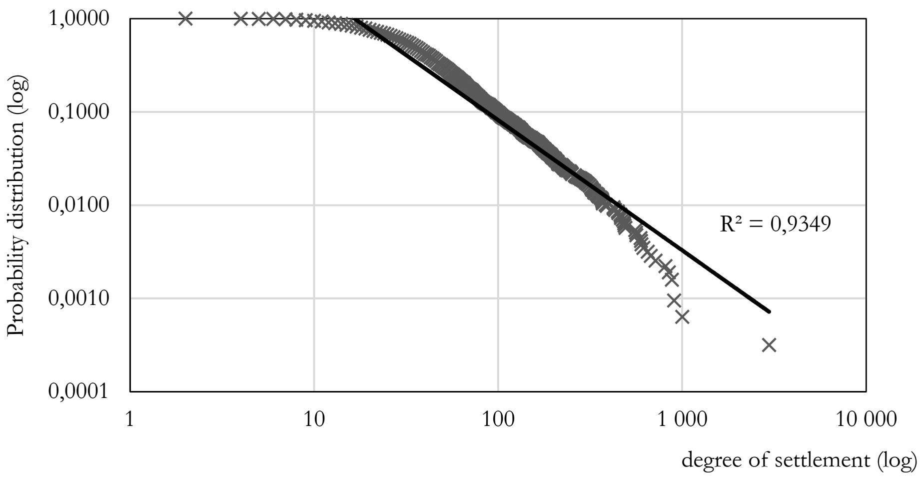
Numerous universal regularities have been identified when studying the complex networks, which are the result of the structure of networks and are independent of the subject of the study. However, the behaviour of networks can be understood better on the basis of their investigation. Fundamentally random and scale-free networks can be identified. Most networks present in nature and in society are scale-free and have the following characteristics (Barabási 2006):

- Degree distribution corresponding to the power law;
- Short-average (network) distance;
- Preferential attachment;
- High level of clusters.

The degree distribution of scale-free networks corresponds to the power law, that is, such networks are composed of a high number of nodes having a small number of connections and a small group of nodes with a very high degree. (A special case of the power law, the Pareto distribution, is well known in social sciences. In this distribution, 20% of the households have 80% of the income, and the rank-size rule of settlements is similar; Czaller 2012.) Upon depicting the cumulative probability distribution of degree on a logarithmic scale, it is clear that the majority of the values fit well with the power trend line. Accuracy of this fit is $R^2=0.9349$; that is, the degree distribution of the settlements of the network can be described by the power function with an accuracy of 93% (Figure 1). Most of the settlements have labour market connections with less than 50 other settlements.

Figure 1

Cumulative probability distribution of the degree (log-log plot), 2011



Source: Author's own construction based on data of the Hungarian Central Statistical Office (HCSO).

Short-average distance indicates that any two nodes of the network – despite the high element number – can reach each other with a few steps via a series of edges; that is, information (e.g. innovation and labour market effects) can flow very rapidly between any two nodes. The idea of a contact network among people appeared first in a novel by Frigyes Karinthy entitled *Láncszemek* (Chain-links). Milgram was the first to try to prove it with experiments. According to his results, any person in the United States is only five to six handshakes away from each other ('six degrees of separation', Milgram 1967). Upon selection of the participants, the experiment was proved to be false (Kleinfeld 2002), but the presence of a so-called 'small-world' in numerous networks remained – social networks among them (e.g. Backstrom et al. 2011). The number of steps increase with the logarithm of the number of elements at the most (Barabási 2006).

In the Hungarian commuting network, the average shortest way among the settlements is 2.37. The number of settlements in Hungary is 3154, the logarithm of which is 3.49. The magnitude of the two values is the same; therefore, the commuting network in Hungary in 2011 can be regarded as a small world. The characteristic presence of small-worlds in commuting networks was proved (De Montis et al. 2007). In the case of commuting, small-world characteristics are limited in space due to the cost of commuting (travel time and financial costs).

The model of scale-free networks sets the rules of the formation of new nodes and edges. According to this, new nodes and new edges of old nodes connect to nodes with higher degree with more probability (preferential attachment). Based on this statement, it can be assumed that job seekers probably search for and find jobs in areas to which more people commute. Distribution corresponding to the power law and preferential attachment together contribute to the phenomenon of 'the rich become richer'; that is, nodes with a special role in the topology of the network retain and even increase their position and have a hand in the development of the network (Barabási 2006). Apparently, the analytic criteria of networks influence their development; that is, the network topology does not determine the development of particular settlements. Different conditions of the nodes can be fit into the models, describing the development of scale-free networks by introducing the variable of 'fitting' (Bianconi–Barabási 2001a). Based on the results of the studies, the competition of nodes attributed with the fitting characteristic is comparable to the movement of particles of the Bose gas in quantum mechanics (Bianconi–Barabási 2001b). No studies using this result for modelling social conditions were found in the literature, but this could be another interesting chapter in the social scientific analogies of features in physics.

High clustering indicates that nodes were close to each other regarding the topology in the groups of nodes (cliques), and cliques were connected via some nodes characterised by very high degrees, that is, based on a degree in which some type of a hierarchy of the nodes could be determined (Watts–Strogatz 1998). This characteristic is associated with the regularity exposed in the course of the diffusion of innovation, that is, innovation with a hierarchy character across adjacent connections (Nemes Nagy 2009).

Certain characteristics of scale-free networks described above may affect several social-economic fields (political relations, company cooperation, flow of innovation, commuting, migration and certain fields of traffic), as a result of which the population of settlements may show a distribution characteristic for scale-free networks – in harmony with Zipf's law – in the course of their development.

Knowing the nature of networks, regional development measures could be evaluated. For example, the effects of the major road network belonging to the group of random networks – the development of which could improve the development of a balanced settlement network (Fleischer 2009) – were reduced by the development of an extremely Budapest-centric highway network (Tóth 2005) that contributed to unbalanced regional development with its quasi-star topology.

Dependency of the commuting relations of settlements

Regarding commuting relations, the key question is how dependent are the inhabitants of the given settlements on the employment conditions of a single settlement or whether they live in a stable environment as part of a polycentric settlement network. In order to answer the question, the strength of the nodes was calculated first using the following formula:

$$s(i) = \sum_{j \in V(i)} w_{ij},$$

where w_{ij} is the weight of the edge between nodes i and j.

The parameter indicates the number of those commuting in and out of a node in the case of undirected networks. In the commuting network in 2011, the minimum and the maximum values were two and 273,421 (Budapest), respectively, while the average was 753 people.

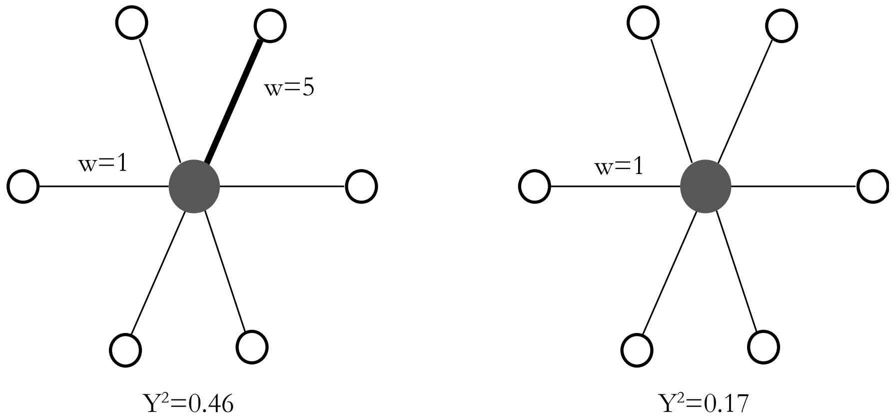
In order to study dependency conditions, the disparity (Y2) parameter (Barthélemy et al. 2005) was calculated. It was developed to measure the heterogeneity of weighted relations as follows:

$$Y_2(i) = \sum_{j \in V(i)} \left(\frac{w_{ij}}{s_i} \right)^2$$

The Y2 parameter shows the distribution of the commuting people of a given settlement among all of the connections. If the number of commuters is the same for all settlements, the value of the parameter will equal the root of its grade (Y2 ≈ 1/k). If one settlement dominates, Y2 will equal to 1 (Figure 2).

Figure 2

Illustration of the disparity indicator in two different cases



Source: De Montis et al. 2007, page 6.

Differing from the literature (De Montis et al. 2007), disparity was calculated in relation to not only undirected but also directed graphs; that is, dependency of in-commuting and out-commuting people were also studied for each settlement (Table 1).

Based on the data, in the case of any population category in 2011, the disparity value of out-commuting is greater than that of in-commuting; that is, the dependency is more characteristic in the case of out-commuting (Table 1). One of the reasons for the significant – twofold or threefold – difference can be the great gap between the values of in-degree and out-degree. Apparently (assuming similar settlement density), in- and out-commuting people have to travel further in case of settlements with higher grade[1]; that is, the weight of the edges decreases. Many edges with small intensity result in a decreasing disparity value as well.

Disparity of in-commuting is basically balanced regarding the population categories. Dependency is more characteristic of the labour market of settlements smaller than 1,000 people (Y_2>0.3); one reason for this could be that in-commuting had become occasional. Effects of employment in small and tiny villages affect only a few nearby settlements, and due to the small number of in-commuting people, certain commuting relations become dominant with a higher chance.

Regarding the disparity of out-commuting, the value of out-degree is decisive as well. The grade of dependency is significantly smaller in case of settlements larger than 50,000 inhabitants, while it is more-or-less similar (Y_2>~0.3) in case of settlements smaller than 50,000 inhabitants.

Table 1

Disparity of settlements by categories of the population number, 2011

Category of population	In-commuting		Out-commuting		Disparity of undirected graph
	Disparity	In-degree	Disparity	Out-degree	
Budapest	0.012	2088.00	0.044	876.00	0.014
100,000 – 999,999	0.029	534.88	0.069	258.75	0.029
50,000 – 99,999	0.043	380.30	0.108	179.60	0.068
20,000 – 49,999	0.089	211.52	0.306	118.62	0.164
10,000 – 19,999	0.105	122.35	0.290	83.49	0.194
5,000 – 9,999	0.126	64.76	0.324	58.32	0.229
2,000 – 4,999	0.176	34.07	0.360	38.26	0.293
1,000 – 1,999	0.198	21.46	0.352	27.34	
1 – 999	0.307	9.73	0.343	14.04	
Total	*0.246*	*26.87*	*0.344*	*26.87*	

Source: Own calculation on the basis of census data, HCSO.

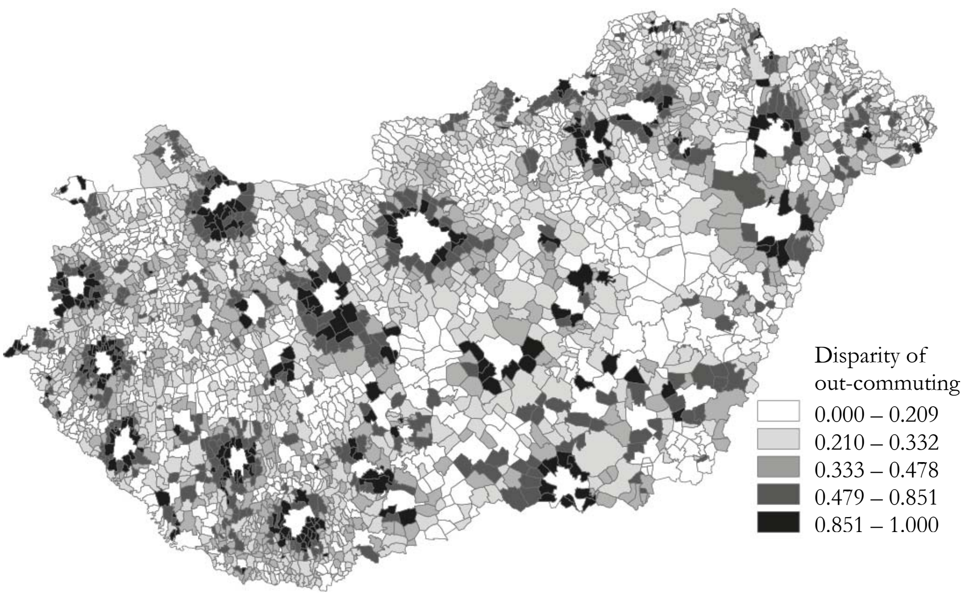
[1] From settlements smaller than 5,000 inhabitants, the commuting people make a beeline towards a 15–20 km area on average. This value in the case of settlements with 5,500 and more than 50,000 inhabitants is 20–30 km and 40–50 km, respectively. In contrast, from Budapest, people commute up to only 30 km (Pálóczi et al. 2014).

The data of the undirected graph shows that aggregation of the data covers the characteristics among groups; that is, modification of the method in the literature (De Montis et al. 2007) contributed to a more detailed analysis. In general, with the decreasing number of inhabitants, one-sided dependency increases and the direction of out- and in-commuting becomes increasingly concentrated. It is worth noting that dependency is not disadvantageous in itself and it can be interpreted while considering other factors as well (e.g. economic activity, unemployment and commuting intensity).

By studying the spatial distribution of the disparity of commuting, two statements can be made (Figure 3). The grade of dependency of out-commuting is high in the regions of the employment centres. The greatest continuous regions can be found around Győr, Székesfehérvár and Budapest.

Figure 3

Disparity of out-commuting, 2011



Source: Own construction based on the census data, HCSO.

The extent of dependent areas is influenced by the drawing effect of employment centres besides the settlement network factors. A fine example of it can be observed in the limited extent of the region in the Sajó Valley (Kazincbarcika) depending on Miskolc or the mangled appearance of the southwestern part of the agglomeration around Budapest due to the effects of economic suburbanisation (Budaörs, Törökbálint). Cross commuting in these regions may have a positive effect on the grade of utilisation of the traffic network, and the profitability of mass transport as opposite traffic may become significant in peak time periods.

This method may help in exposing the reasons behind the regional wage curve (wage difference in case of employees with similar parameters is formed in accordance with unemployment in the vicinity of their workplace; for more information, see Kertesi–Köllő 1998) and in planning measures. Disparity regions characterised by high ratio of unemployment could be classified into settlements from which commuting is possible only to the closest town due to travel costs and insufficient traffic relations.

Disparity could also be applied to classify peripheries of commuting where settlements with poor commuting intensity and low disparity are found (small-scale, branching commuting). The results could be used for studying the spatial pattern of social marginalisation (Nagy et al. 2015).

Spatial division procedures

Identifying spatial units in which real spatial relations and processes operate presents a problem for several scientific fields (e.g. public administration, spatial science, and geography). Spatial units cannot be regarded as natural entities as they are designated via classification based on objective and subjective criteria (Nemes Nagy 1998; Hajdú 2005); therefore, there are no universal methods for designating spatial units. A fine example is the debate regarding the agglomeration around Budapest (Tóth–Schuchmann 2010). Criteria are selected based on their own considerations and purpose of the research. Delimitations using commuting datasets belong to the attraction district-based type (Barancsuk et al. 2014; for details on labour market district delimitations see Pénzes et al. 2014).

In the present study, a social-physical method of the complex network analysis was applied instead of the more widespread – associated with commuting – spatial division mechanisms. In the course of the network analysis, techniques suitable for spatial division are called locally dense partial graph search (Tibély 2011) or simply community detection (Kovács et al. 2012). Henceforward, partial graph search and community detection are used as synonyms.

Although there is no precise, generally accepted definition of network tangles (in other words, groups, clusters, modules and communities; Tibély 2011), they could be described as 'dense' partial graphs of the network within which nodes connect to each other more frequently and – in the case of weighted networks – more intensively than in the other parts of the network (Derényi et al. 2006).

Within group searching procedures, three categories – based on local, global, and nodal similarities – can be identified (Fortunato 2010). In the *local* approach, tangles of sub-graphs and their adjacency conditions are in focus while other parts of the graph are ignored. The *global* definition treats sub-graphs as structural units of the graph; that is, identifiable characters of the sub-graph become recognisable when the sub-graph is compared with the whole graph. Definitions based on *nodal* similarity select sub-graphs as a group of nodes similar to each other. The criterion of similarity is the presence or lack of an edge between node pairs (De Montis et al. 2013).

Results of partial graph search performed on non-weighted graphs can identify the modules based on topology. In contrast, when weighted networks are analysed, clusters are determined by not only topology but also the weight of the given edges. As a result, commuting was studied as a weighted network in our study.

For searching groups of nodes, there are numerous methods and algorithms available in the literature. Fortunato 2010 identifies three major classes:

- divisive algorithms,
- optimisation methods and
- spectral analysis.

Methods that cannot be categorised in the above three are also known, such as Q-state, Potts model, clique percolation (Derényi et al. 2006), random walk, Markov cluster algorithm, maximum likelihood and the L-shell method (Fortunato 2010).

Although there is extensive literature on the procedures developed for different disciplines, intensive, detailed and careful evaluation and comparison of the given indexes and procedures besides the extensive method development are missing (Tibély 2011). In the present paper, the results of two methods are mentioned and one is presented in detail using the example of a Hungarian dataset.

Structural equivalence analysis

In a previous study, the possibilities of one of the methods being based on node similarities – CONCOR (CONvergence of iterated CORrelations) analysis – were discussed (Pálóczi–Pénzes 2015). Using this method, structurally equivalent members can be identified. Structurally equivalent members belong to a group that is connected to the same settlements with similar intensities (for details, see Kürtösi 2005). According to our hypothesis, this method is suitable for identifying settlement groups that connect to similar employment centres with similar intensities; that is, it can be assumed that their employment conditions and number of available jobs via commuting would change in the same way in cases of both recession and economic boom. (Of course, the qualification of employees and their industrial branch would cause differences.) Based on the results, this method cannot be applied successfully to complex, branching connection systems such as a commuting dataset. Although the analysis based on the method detected the intensive attraction districts of the major employment centres and exposed their inner conditions, classification of the majority of the settlements was not possible. The number of groups depended on the subjective decision of the scientists, and this influenced the final results significantly (Pálóczi–Pénzes 2015). Furthermore, in the case of networks with higher number of elements, the method requires significant computer aid.

Clique percolation algorithm

The clique percolation algorithm identifies groups among whom overlapping is possible. Using this method, groups will comprise of the densest elements of the network, called cliques. Within a clique, every node is connected to every other node and the groups are continuous chains of such cliques (for description of the method, see Derényi et al. 2006).

In the case of real networks, there seem to be overlaps among the groups. For example, in contact networks, one person can be a member of several groups that know each other well (e.g. family, colleagues and friends), and overlapping is possible among groups (e.g. classmate and friend). If overlap-free groups are searched (see the results of Louvain optimisation), then members who do not 'know' each other may be included into designated groups (false positive) and several nodes may be included into different clusters (false negative) despite their apparent interdependence. In theory, the clique percolation algorithm is suitable for avoiding such mistakes. This is why the commuting dataset was analysed using this method; however, the analysis was not successful as it identified 5,000 groups. Although it reduced the number of groups in weighted networks, we did not find it suitable for spatial research; therefore, the results are not discussed in detail here.

Modularity optimisation using the Louvain method

The Louvain algorithm for optimisation (Blondel et al. 2008) was chosen because it was used successfully in commuting research (De Montis et al. 2013). The method is based on maximisation of the objective function called modularity (Q_w; Newman–Girvan 2004). The value of the function shows the number of edges within a given community – comparing it with the number of edges outside the community. The value makes it possible to compare how a partial graph suits the criteria in relation to every other possible variation. This value may vary between -1 and +1. It can also be zero if no more partial graphs can be formed within the given sub-group. A negative value indicates that there is no point in further dividing the network: the best division is the given group. The function can be defined in the case of weighted networks as follows:

$$Q_w = \frac{1}{2W} * \sum_{ij} \left(w_{ij} - \frac{S_i S_j}{2W} \right) * \delta(c_i, c_j)$$

where w_{ij} is the weight of edges connected to nodes 'i' and 'j', $S_i = \sum_j w_{ij}$ (node strength): sum of the weight of edges connected to node '', $W = \frac{1}{2} \sum_{ij} w_{ij}$ sum of the weight of all edges, and $\delta c_i, c_j$ is a function that equals one if nodes 'I' and 'j' belong to the same community and are not connected to any other at all.

The greatest advantage of the Louvain method is that the number of partial graphs comes from the algorithm. Without an ordered number at the beginning, it avoids the subjectivity of scientists in contrast to other methods. The algorithm is based on the following iterative steps (Table 2):

Table 2

Iterative process of the modularity optimisation

Step	Task
1	Each node is assigned to a unique single community.
2	Neighbour nodes of each target node are preferentially included in the same community if the variation of the modularity (ΔQ_W) is positive.
3	This aggregation process proceeds until the modularity function Q_W reaches a maximum.
4	A new network is then built in which nodes correspond to the communities obtained in step 3; each link connecting a pair of communities is featured by a weight equal to the sum of the weights of the external links originally between them. The internal links are represented by a self-loop, whose weight is equal to the sum of their internal weights.
5	Step 1 applies to the last network.

Source: Blonde et al. 2008: Table 1, referring to: De Montis et al. 2013.

The function ΔQ – applied in step 2 – measures the volume of change caused by the separation of the given node from 'C' community and the connection of the given node to 'C' community.

The study was performed on a symmetric, undirected weighted graph (web link: 1) using the software Pajek. Besides the original algorithm, the software enables the control of the level of community resolution using the resolution parameter (for description of the procedure, see Arenas et al. 2008). In case the value of the resolution is 1, the algorithm runs in an unchanged version. If the value is increased, the software carries the iteration and produces several smaller modules or, if resolution is decreased, fewer but greater modules are produced.

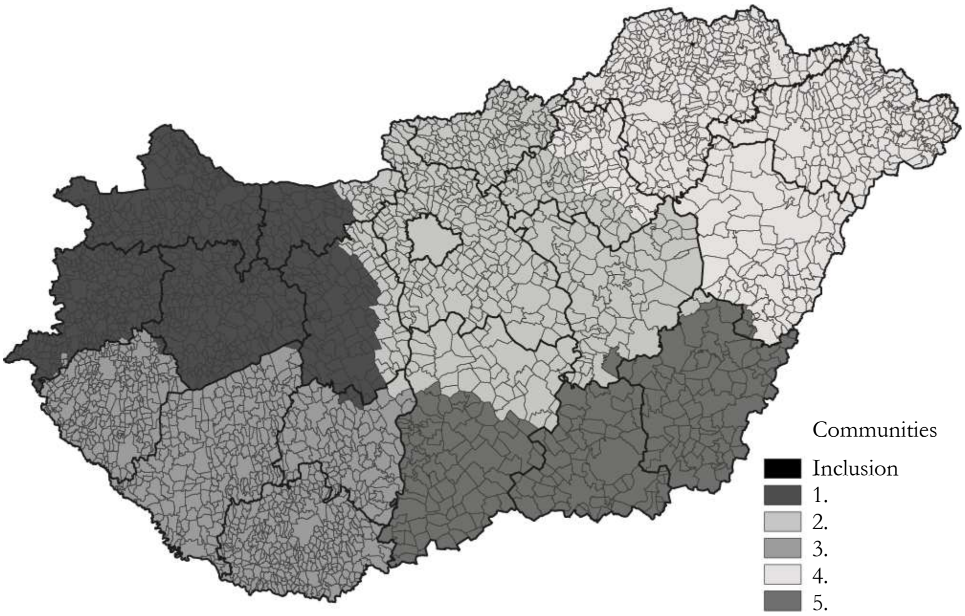
Based on experiences, running the algorithm once frequently results in an unstable module structure; therefore, significant differences may occur among the modules at similar parameters. Consequently, the algorithm was run several times at the same resolution value. Spatial divisions belonging to the same resolution values were tested using the recommended correlation analysis (web link: 1). In conclusion, the stability of the modules belonging to resolution values of 0.2 and 1 exceeded the Cramer's V value of 0.99.

Based on the results, the method produced spatially continuous settlement groups that can be interpreted professionally. Thus, the method is suitable for regionalising commuting connections (Figures 4 and 5).

Results revealed that labour market settlement groups are not in harmony with the NUTS2 level regions of Hungary (Figure 4). The most significant modification to the regions is caused by the complex and intensive connection system of Budapest, determining central Hungary. The spatial extent of module 2 (Figure 4) is much less towards the west due to the marked employment role of Northwestern Hungary. The borderline of the Komárom-Esztergom and Fejér counties is the same as the borderline in Figure 5. This stable borderline indicates the balance of the attraction of employment centres. County-sized areas are connected to settlement group 2. At this resolution (0.2), county borders and natural barriers (Balaton and certain sections of the Danube) play important role. Regarding settlement groups, number 5 seems to be most similar to the NUTS2 region (except for northern Bács-Kiskun).

Figure 4

Network communities based on the Louvain modularity (resolution=0.2), 2011



Source: Own construction based on census data of HCSO.

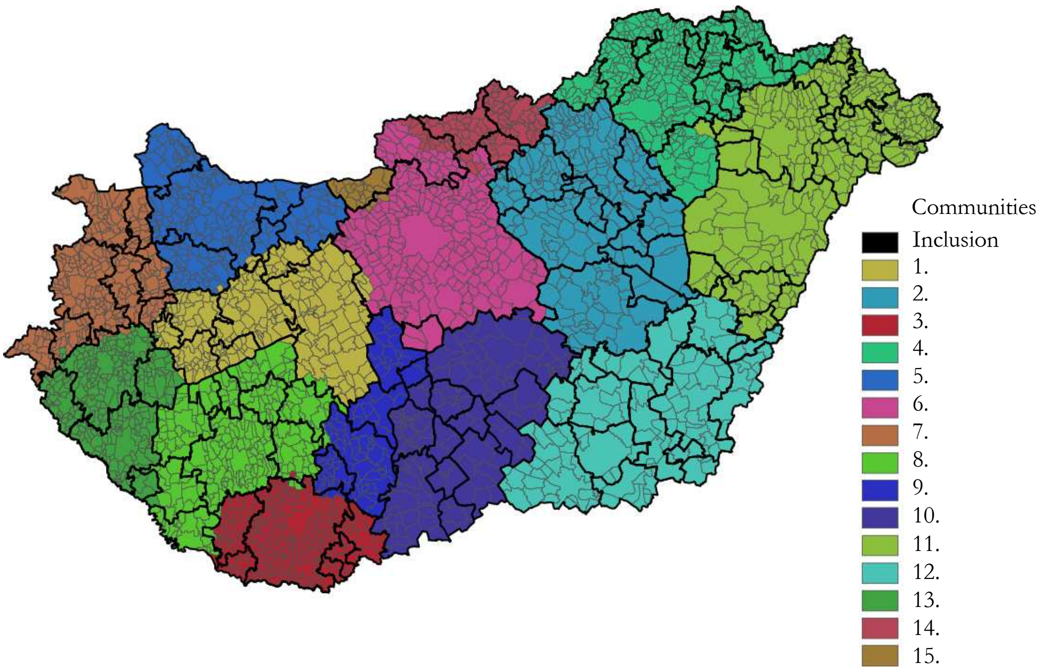
Results were interpreted together with a former local labour market attraction district delimitation (Local Labour System [LLS]); thus, the advantages and disadvantages of the applied method became more visible. The other spatial delimitation was made using the same dataset, however, with a completely different methodical approach (Pénzes et al. 2014). In general, the LLS yielded more detailed results that could be interpreted easily, whereas with the network analysis system, level relationships could be recognised more easily, and the relationships of LLS districts with each other could be understood as well.

Unfortunately, comparison of the two studies is limited due to differences in magnitude between the numbers of spatial units (15 and 123); however, comparison of the modules and LLS districts yielded information (Figure 5). In numerous cases, the border of modules and LLS districts is the same and modules are composed of several LLS districts, so that these modules represent higher hierarchy levels. In order to overview the results, they are compared with the well-known county borders (Figure 6).

In cases in which the attraction of an LLS centre crosses its county border, the adjustment of the module border to the LLS district is expected (LLS districts of Mezőkövesd, Tiszaújváros, Szeghalom, Szarvas, Kecskemét, Dunaújváros, Dombóvár, Tamási, Nagykanizsa, and Keszthely-Hévíz). In contrast, certain LLS districts are connected to the module of the neighbouring county (LLS districts of Kunszentmárton, Solt, Csurgó, Sopron and Kapuvár).

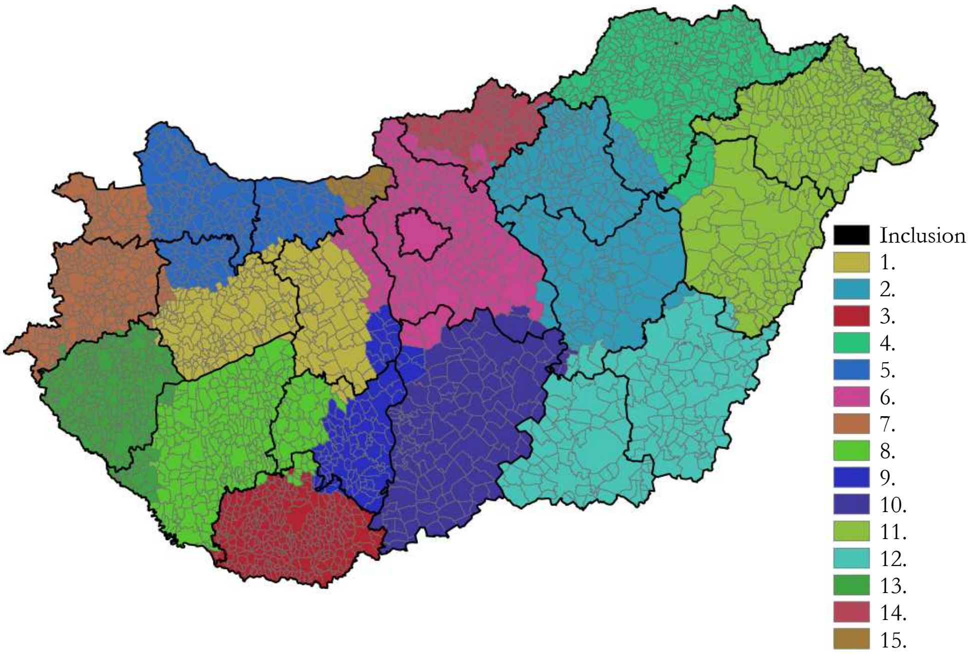
Figure 5

Network communities based on the Louvain modularity (resolution=1), 2011



Source: Own construction based on census data of HCSO.

Figure 6
Network communities based on the Louvain, modularity (resolution=1), 2011



Source: Own construction based on census data of HCSO.

Module 5 (Győr) is composed of more than half a dozen LLS districts, and module 9 is located in the eastern parts of Fejér and Tolna counties and on the left side of the Danube, thanks to the bridges at Dunaföldvár and Pentele in M8 motorway, which are conglomerates that differ from the rest of the small-border modifications due to their size. In their cases, the settlement groups differ significantly from their counties.

Conclusions

In the present study, the application possibilities of graph-theory-based network analysis in spatial research were studied. The most important characteristics of random and scale-free networks were presented based on national and international literature. When characteristics are presented, their role in studying features in spatial research is also discussed. Conclusions from the theoretical description may contribute to better understanding of the regularities of flow of resources between settlements and ways to strengthen or reduce these regularities according to the purpose of regional development.

Dependency conditions of out- and in-commuting people in settlements were analysed based on the commuting data of the census in 2011. Out-commuting generates stronger dependency relationships in the case of all settlement categories than in-commuting. In general, the decreasing population of settlements results in stronger dependency relationships. According to the results, the disparity index could be used successfully in other spatial research as well.

Attempts were made to apply the group forming procedures of network analysis to spatial division. Using previous research, the evaluation of three methodological applications was performed. By applying the CONCOR analysis, delineation of structurally equivalent elements can be made. This procedure is less suitable for analysing a complex, branching relation system such as commuting. It can be used successfully to identify groups in the attraction districts of only the most significant employment centres. Objectivity of the method is reduced if the researcher decides the number of groups to be formed.

The other group identifying procedure described was clique percolation. By applying this method, detecting the overlaps in the attraction districts of commuting becomes possible. However, the procedure delineated too many groups at the national level, even when the parameters were modified. This method could be suitable for carrying out district- or county-level analyses.

Modularity-optimisation conducted using the Louvain method proved to be suitable for delineating greater spatial units as well. Based on the results, settlement groups (regions) formed according to employment relations – that is according to regionalism – adjusted less to NUTS2 regions and resembled the county borders more. Differences from the county system are not single differences caused by specialities of the method; they reflect real commuting relations proven with the LLS districts taken from our previous study. The method could be used to determine the hierarchy of local labour-market attraction districts and to study further spatial interactions.

REFERENCES

ARENAS, A.–FERNANDEZ, A.–GOMEZ, S. (2008): Analysis of the structure of complex networks at different resolution levels *New Journal of Physics* 10 No. 053039.

BACKSTROM, L.–BOLDI, P.–ROSA, M.–UGANDER, J.–VIGNA, S. (2011): *Four Degrees of Separation* http://arxiv.org/pdf/1111.4570.pdf (downloaded: January 2016.)

BARABÁSI, A. L.–ALBERT, R. (1999): Emergence of Scaling in Random Networks *Science* 286: 509–512.

BARABÁSI, A. L. (2006): A hálózatok tudománya: a társadalomtól a webig *Magyar Tudomány* 167 (11): 1298–1308.

BARANCSUK, Á.–GYAPAY, B.–SZALKAI, G. (2013): Az alsó középszintű térfelosztás elméleti és gyakorlati lehetőségei *Területi Statisztika* 53 (2): 107–129.

BARTHÉLEMY, M.–BARRAT, A.–PASTOR-SATORRAS, R.–VESPIGNANI, A. (2005): Characterization and modeling of weighted networks *Physica A: Statistical mechanics and its applications* 346: 34–43.

BIANCONI, G.–BARABÁSI, A. L. (2001a): Competition and multiscaling in evolving networks *Europhysics Letters* 54 (4): 436–442.

BIANCONI, G.–BARABÁSI, A. L. (2001b): Bose-Einstein condensation in complex networks *Physical Review Letters* 86 (24): 5632–5635.

BLONDEL, V. D.–GUILLAUME, J. L.–LAMBIOTTE, R.–LEFEBVRE, E. (2008): Fast unfolding of communities in large networks *Journal of Statistical Mechanics: Theory and Experiment* P10008.

CASCHILI, S.–DE MONTIS, A. (2013): Accessibility and Complex Network Analysis of the U.S. commuting system *Cities, Special Section: Analysis and Planning of Urban Settlements: The Role of Accessibility* 30: 4–17.

CZALLER, L. (2012): A Zipf-törvény érvényesülése a világ országaiban *Területi Statisztika* 52 (5): 461–478.

DE MONTIS, A.–BARTHÉLEMY, M.–CHESSA, A.–VESPIGNANI, A. (2007): The structure of interurban traffic: A weighted network analysis *Environment and Planning B: Planning and Design* 34: 905–924.

DE MONTIS, A.–CASCHILI, S.–CHESSA, A. (2013): Commuter networks and community detection: a method for planning sub regional areas *The European Physical Journal Special Topics* 215: 75–91.

DE MONTIS, A.–CHESSA, A.–CAMPAGNA, M.–CASCHILI, S.–DEPLANO, G. (2010): Modeling commuting systems through a complex network analysis: A study of the Italian islands of Sardinia and Sicily *Journal of Transport and Land Use* 2 (3/4): 39–55.

DERÉNYI, I.–FARKAS, I.–PALLA, G.–VICSEK, T. (2006): Csoportosulások szociológiai, technológiai és biológiai hálózatokban *Magyar Tudomány* 167 (11): 1319–1324.

DROBNE, S.–KONJAR, M.–LISEC, A.–PICHLER MILANOVIĆ, N.–ZAVODNIK LAMOVŠEK, A. (2010): Functional Regions Defined by Urban Centres of (Inter)National Importance – The Case of Slovenia In: SCHRENK, M.–POPOVICH, V. V.–ZEILE, P. (ed.): *Cities for Everyone. Liveable, Healthy, Prosperous. Promising Vision or Unrealistic Fantasy?* pp.: 295–304., Real Corp 2010 15th International Conference on Urban Planning, Regional Planning and Information Society. CORP – Competence Center of Urban and Regional Planning, Wien.

DUSEK, T. (2003): A gravitációs modell és a gravitációs törvény összehasonlítása *Tér és Társadalom* 17 (1): 41–58.

DUSEK, T. (2013): *A területi statisztika egyes térparamétereket használó elemzési eszközei* Habilitációs értekezés, Kézirat, Győr.

ERDŐS, P.–RÉNYI, A. (1959): On Random Graphs I. *Publicationes Mathematicae* 6: 290–297.

ERDŐS, P.–RÉNYI, A. (1960): On The Evolution of Random Graphs *Magyar Tudományos Akadémiai Matematikai Kutató Intézet Közleményei* 5: 17–61.

FLEISCHER, T. (2009): A közlekedés szerepe a policentrikus városhálózat fejlesztésében *Tér és Társadalom* 23 (1): 19–42.

FORTUNATO, S. (2010): Community detection in graphs *Physics Reports* 486: 75–174.

GÉBER, J. (2007): Határmenti hálózatok A kapcsolatháló elemzés alkalmazási területei a határmenti térségekben In: SÜLI-ZAKAR, I. (ed.) *A határok és a határon átnyúló (CBC) kapcsolatok szerepe a kibővült Európai Unió keleti perifériáján* pp. 99–106., Kossuth Egyetemi Kiadó, Debrecen.

HAJDÚ, Z. (2005): *Magyarország közigazgatási földrajza* Dialóg Campus Kiadó, Budapest.

JAKOBI, Á.–LENGYEL, B. (2014): Egy online közösségi háló offline földrajza, avagy a távolság és a méret szerepének magyar empíriái *Tér és Társadalom* 28 (1): 40–61.

KERTESI, G.–KÖLLŐ, J. (1998): Regionális munkanélküliség és bérek az átmenet éveiben. A bérszerkezet alakulása Magyarországon II. rész. *Közgazdasági szemle* 45 (7–8): 621–652.

KISS, J. P.–SZALKAI, G. (2014): A foglalkoztatás területi koncentrációjának változásai Magyarországon *Területi Statisztika* 54 (5): 415–447.

KLAPKA, P.–HALÁS, M.–TONEV, P.–BEDNÁR, M. (2013): Functional regions of the Czech Republic: comparison of simpler and more advanced methods of regional taxonomy *Acta Universitatis Palackianae Olomucensis – Geographica* 44 (1): 45–57.

KLEINFELD, S. J. (2002): The small world problem *Society* 39 (2): 61–66.

KOVÁCS, L.–OROSZ, K.–POLLNER, P. (2012) Magyar szóasszociációk hálózata *Magyar Tudomány* 173 (6): 699–705.

KOZMA, G. (1998): A gazdasági élet szereplőinek térbeli preferenciái *Falu Város Régió* 9: 7–14.

KÜRTÖSI, ZS. (2005): A társadalmi kapcsolatháló elemzés módszertani alapjai In: LETENYEI, L. (ed.) *Településkutatás szöveggyűjtemény* pp. 663–685., Ráció Kiadó, Budapest.

LAMBIOTTE, R.–DELVENNE, J. C.–BARAHONA, M.. (2015): Laplacian Dynamics and Multiscale Modular Structure in Networks *IEEE Transactions on Network Science and Engineering* 1 (2): 76–90.

LESAGE, J. P.–FISCHER, M. M. (2010): Spatial Econometric Methods for Modeling Origin-Destination Flows In: FISCHER, M. M.–GETIS, A. (ed.) *Handbook of Applied Spatial Analysis* pp. 409–433., Springer Berlin-Heidelberg.

LETENYEI, L. (2000): Regionális társadalmi hálózatok *Falu Város Régió* 7: 21–26.

LETENYEI, L. (2005): Kapcsolatháló-elemzés In: LETENYEI, L. (ed.) *Településkutatás* pp. 187–248., Rátió Kiadó, Budapest.

MADARÁSZ, E.–PAPP, ZS. (2013): Delimiting the "Balaton Riviera" tourist destination by using network analysis *Hungarian Geographical Bulletin* 62 (3): 289–312.

MÁTÉ, G.–KOVÁCS, A.–NÉDA, Z. (2013): Hierarchical Settlement Networks *Regional Statistics* 3 (1): 30–41.

MILGRAM, S. (1967): The Small-World Problem *Psychology Today* 1 (1): 61–67.

MOLNÁR, E. (2012): Kelet-Közép-Európa az autóipar nemzetközi munkamegosztásában *Tér és Társadalom* 26 (1): 123–137.

NAGY, E.–TIMÁR, J.–NAGY, G.–VELKEY, G. (2015): A társadalmi-térbeli marginalizáció folyamatai a leszakadó vidéki térségekben *Tér és Társadalom* 29 (1): 35–52.

NEMES NAGY, J. (1998): *A tér a társadalomtudományban* Hilscher Rezső Szociálpolitikai Egyesület, Budapest.

NEMES NAGY, J. (2009): *Terek, helyek, régiók* Akadémiai Kiadó, Budapest.

NEWMAN, M.–GIRVAN, M. (2004): Finding and evaluating community structure in networks *Physical Review E* 69 (2): 026113.

OECD (2012): *Defining "Urban". A New Way to Measure Metropolitan Areas* Organization for Economic Co-operation and Development, Paris.

PÁL, ZS. (2014): A magyarországi települések közötti bankközi pénzforgalom földrajzi és globális hálózati jellemzőinek vizsgálata *Területi Statisztika* 17 (2): 134–151.

PÁLÓCZI, G.–LENGYEL, I. M.–MOLNÁR, E. (2014): *Changing commuting pattern in Hungary after the turn of the millennium* Oradea. (kézirat)

PÁLÓCZI, G.–PÉNZES, J. (2011): A közösségi közlekedési rendszer térinformatikai vizsgálatának módszerei Hajdú-Bihar megye példáján In: LÓKI, J. (ed.) *Az elmélet és a gyakorlat találkozása a térinformatikában* pp. 443–449., Debreceni Egyetem, Debrecen.

PÁLÓCZI, G.–PÉNZES, J. (2015): A kapcsolatháló-elemzés alkalmazási lehetősége a munkaerő-piaci vonzáskörzetek lehatárolásában In: BODA, J. (ed.) *Az elmélet és a gyakorlat találkozása a térinformatikában* pp. 331–338., Debreceni Egyetemi Kiadó, Debrecen.

PATUELLI, R.–REGGIANI, A.–GORMAN, S. P.–NIJKAMP, P.–BADE, F. J. (2007): Network Analysis of Commuting Flows: A Comparative Static Approach to German Data *Networks & Spatial Economics* 7 (4): 315–331.

PÉNZES, J.–MOLNÁR, E.–PÁLÓCZI, G. (2014): Helyi munkaerő-piaci vonzáskörzetek az ezredforduló utáni Magyarországon *Területi Statisztika* 54 (5): 474–490.

PÉNZES, J. (2014): *Periférikus térségek lehatárolása – dilemmák és lehetőségek* Didakt Kiadó, Debrecen.

RADVÁNSZKI, Á.–SÜTŐ, A. (2007): Hol a határ? Helyi munkaerőpiaci rendszerek Magyarországon – Egy közép-európai transznacionális projekt újdonságai a hazai településpolitika számára *Falu Város Régió* 14 (3): 45–54.

RUSSO, G.–REGGIANI, A.–NIJKAMP, P. (2007): Spatial activity and labour market patterns: A connectivity analysis of commuting flows in Germany *Annals of Regional Science* 41 (4): 789–811.

SZABÓ, L.–HORVÁTH, B.–HORVÁTH, R.–GAÁL, B. (2013): A Győri agglomeráció közforgalmú közlekedési rendszerének vizsgálata *Területi Statisztika* 53 (6): 578–592.

SZABÓ, SZ.–NOVÁK, T.–ELEK, Z. (2012): Distance models in ecological network management: A case study of patch connectivity in a grassland network *Journal for Nature Conservation* 20 (5): 293–300.

TIBÉLY, G. (2011): *Mesoscopic structure of complex networks* Doktori disszertáció, Budapest.

TÓTH, G.–SCHUCHMANN, P. (2010): A budapesti agglomeráció területi kiterjedésének vizsgálata *Területi Statisztika* 50 (5): 510–529.

TÓTH G. (2005): *Az autópályák szerepe a regionális folyamatokban* Központi Statisztikai Hivatal, Budapest.

TÓTH, G. (2013): *Az elérhetőség és alkalmazása a regionális vizsgálatokban* Központi Statisztikai Hivatal, Budapest.

VIDA, ZS. V. (2012): Regionális Tudományi Tanulmányok hivatkozási hálózatai In: NEMES NAGY, J. (ed.) *Térfolyamatok, térkategóriák, térelemzés* pp. 23–29., ELTE Regionális Tudományi Tanszék, Budapest.

VIDA, ZS. V. (2013): A térkapcsolatok értelmezési lehetőségei hálózatokon In: BOTTLIK, ZS. (ed.) *Önálló lépések a tudomány területén* pp. 115–129., ELTE TTK Földtudományi Doktori Iskola, Budapest.

WATTS, D. J.–STROGATZ, S. H. (1998): Collective dynamics of 'small-world' networks *Nature* 393: 440–442.

WEB LINKS

http://mrvar.fdv.uni-lj.si/pajek/community/CommunityDrawExample.htm