

Research Article

The Settlement Structure Is Reflected in Personal Investments: Distance-Dependent Network Modularity-Based Measurement of Regional Attractiveness

Laszlo Gadar,^{1,2} Zsolt T. Kosztyan,^{1,3} and Janos Abonyi ⁴

¹MTA-PE Budapest Ranking Research Group, University of Pannonia, Veszprém, Hungary

²Innopod Solutions Kft., Budapest, Hungary

³Department of Quantitative Methods, University of Pannonia, Veszprém, Hungary

⁴MTA-PE “Lendulet” Complex Systems Monitoring Research Group, University of Pannonia, Veszprém, Hungary

Correspondence should be addressed to Janos Abonyi; janos@abonyilab.com

Received 10 August 2018; Revised 24 October 2018; Accepted 8 November 2018; Published 5 December 2018

Academic Editor: Pasquale De Meo

Copyright © 2018 Laszlo Gadar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How are ownership relationships distributed in the geographical space? Is physical proximity a significant factor in investment decisions? What is the impact of the capital city? How can the structure of investment patterns characterize the attractiveness and development of economic regions? To explore these issues, we analyze the network of company ownership in Hungary and determine how connections are distributed in geographical space. Based on the calculation of the internal and external linking probabilities, we propose several measures to evaluate the attractiveness of towns and geographic regions. Community detection based on several null models indicates that modules of the network coincide with administrative regions, in which Budapest is the absolute centre, and where county centres function as hubs. Gravity model-based modularity analysis highlights that, besides the strong attraction of Budapest, geographical distance has a significant influence over the frequency of connections and the target nodes play the most significant role in link formation, which confirms that the analysis of the directed company-ownership network gives a good indication of regional attractiveness.

1. Introduction

Mining valuable information from social networks is a hard problem due to its dynamic nature [1, 2], complex structure [3, 4], and multidimensionality [5]. This paper deals with the structural issues as it tries to evaluate regional attractiveness based on a set of goal-oriented null models identified to describe the geographical distributions of company-ownership relations.

Complex multivariate socioeconomic data is widely used to monitor regional policy [6, 7]. As the usage of a different set of variables results in various rankings, the definition and selection of socioeconomic variables are the key issue in these applications. The drawback of these indicator-based approaches is that although economic behavior is socially constructed and embedded in networks of interpersonal relations [8] and strong related to location

[9], the network structure of the economy is neglected.

This paper adds a viewpoint to regional studies based on the analysis of how the network of personal investments and the founding of companies relate to the settlement hierarchy. We assume that the socially embedded economy must have a network-based imprint in the company-ownership network which is a good indication of regional attractiveness.

Attractiveness is meaningful in preferential attachment networks, where the likelihood of a new connection is proportional to degree [10] and fitness [11] of the node. These models were generalized to handle initial attractiveness [12] and latecomer nodes with a higher degree of fitness [11, 13]. It is important to note that these models generate power-law (degree) distributions that are similar to the distribution of socioeconomic variables of settlements indicating that preferential attachment is a process that can be used to

describe city grow [14–18]. In the case of geographically distributed networks, the likelihood of link formation is dependent on distance due to the cost of establishing connections and spatial constraints [19]. Connection costs also favor the formation of cliques and thus increase the clustering coefficient [20]. Space is important in social networks as most individuals connect with their spatial neighbors [20] to minimize their effort and maintain social ties [21]; e.g., the majority of our friends are in our spatial neighborhood [22]. The probability $P(d)$ that distance d separates two connected individuals is found to behave as $P(d) \sim d^{-2}$ in terms of Belgian mobile phone data [23], or generally $P(d) \sim d^{-\delta}$, as has been shown in the case of the social network of more than one million bloggers in the USA [24], in friendship network of Facebook users, and in email communication networks [25, 26].

The attractiveness of airports [27], countries for foreign investments [28], and touristic destinations [29] is evaluated based on socioeconomic variables. As many origins and destinations are present in these applications, the theory of bilateral trade flows accounts for the relative attractiveness of origin-destination pairs. The gravity model is one of the most successful empirical models in economics developed to describe such interactions across the space [30]. Almost 40 years ago, before the emergence of network science, Anderson suggested that as a force between two mass points, the number of trips from location i to location j , follows the (economic version) of the ‘‘Gravity’’ law, $F(d) \sim P(d) \sim I_1^\alpha I_2^\alpha d^{-\delta}$ [31]. Nowadays, many complex networks embedded in space and spatial constraints may have an effect on their connectivity patterns such as trade markets [32], migration [33], traffic flow [34], and mobile communication [23] that can be successfully modeled by a gravity model, which was also successfully applied in link prediction [35].

We assume that regions that heavily rely on local resources consist of more internal connections that form modules in networks, so the modularity of the networks which reflect socioeconomic relationships can be used to measure regional attractiveness. The goal of modularity analysis is to separate the network into groups of vertices that have fewer connections between them than inside the communities [36]. In social network analysis, community detection is a basic step in understanding the structure, function, and semantics of networks [4]. Community analysis is performed in two separate phases: first, detection of meaningful community structure from a network, and second, evaluation of the appropriateness of the detected community structure [37]. Systematic deviations from a random configuration allow us to define a quantity called modularity, that is a measure of the quality of partitions. Newman-Girvan modularity considers only the degree of nodes as a null model which is equivalent to rewiring the network whilst preserving the degree sequence [38, 39]. This random model overlooks the spatial nature of the network; thus, modules are blind to spatial anomalies and fails to uncover modules determined by factors other than mere physical proximity [19], which is the reason why several distance-dependent null models have been proposed recently [19, 37, 40, 41].

Our goal is to use the tools of network community detection to evaluate the attractiveness of the elements of settlement hierarchies (towns, statistical subregions, counties, and regions) based on their modularities as well as internal and external connection densities. We study the internal connections of the ownership network through the point of view of Newman-Girvan, spatial and gravity-based null models. As the modularity is based on the difference between the actual and evaluated values of weight of edges, the real spatial network more accurately describes the null model, and the total modularity tends to be zero, so the modules highlight the hidden structural similarities. We developed a visualization technique to analyze these unknown effects on community structure which can explain the attractiveness of a settlement/region. Besides measuring the attractiveness, we utilize the Louvain community detection algorithm [42, 43] to identify closely related regions. We examine the complete investment network of Hungarian companies to explore how the ownership connections are geographically distributed, what is the structure of the network, and what are the common connection directions, as well as how the extracted information is correlated to the settlement hierarchy. The studied database contains information about the owners and addresses of the companies. The results highlight the fact that distance dependence of the investment connections is more significant than was found in online social networks [22, 26, 44]. The analysis shows that the network is hierarchical and modular as well as shaped according to the settlement hierarchy, in which Budapest is the absolute center, and the centers of counties function as hubs.

The outline of this paper is as follows: Section 2.1 presents the company-ownership network. The metrics related to attractiveness are given in the Appendix. Section 2.2 describes the null models designed by us to measure modularity as well as handling physical proximity and presents how closely related regions can be explored based on the modularity-related merging of towns and subregions. The results and discussion are provided in Section 3.

2. Problem Formulation: Settlement Hierarchy and Community Structure in Personal Investment Patterns

2.1. Network Representation of Personal Investment Patterns. The proposed methodology is based on the analysis of a directed investment network represented by an asymmetric biadjacency matrix $\mathbf{A}^{[p,co]}$, whose elements are defined as

$$a_{i,j}^{[p,co]} = \begin{cases} 1 & \text{if the } i\text{-th person owns the } j\text{-th company} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

As the addresses of the owners and their companies are known, connections between companies and their owners define ties between geographic locations.

According to the levels of the settlement hierarchy, a four-level study can be defined to describe how towns, regions,

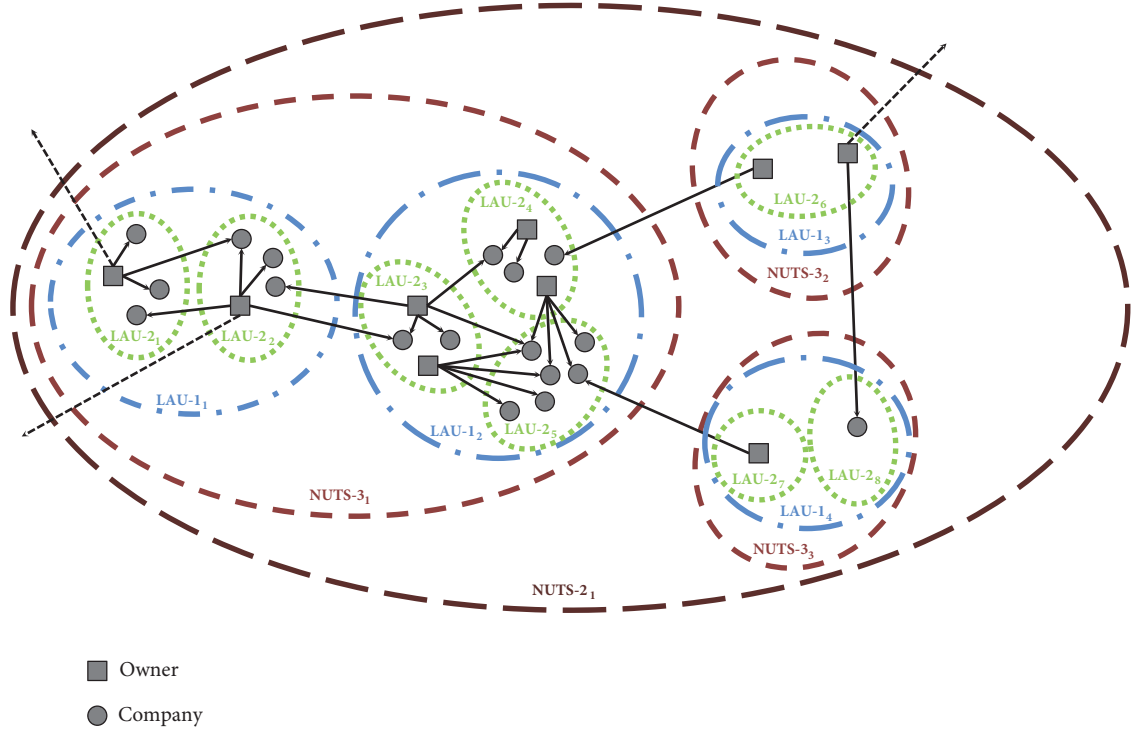


FIGURE 1: Company-ownership relations connect the elements of the settlement hierarchy (Settlement (LAU 2), statistical subregion (LAU 1), small-region (NUTS 3), and region (NUTS 2)).

or counties are connected through company ownerships (see Figure 1). Although companies also own shares in other companies, as we intended to study the attractiveness of economic regions based on personal investment decisions, we examined only companies that belong to individuals.

The levels of the settlement hierarchy [l] are defined based on the nomenclature of territorial units for statistics classification (NUTS) and the two levels of local administrative units (LAUs):

$$l = \begin{cases} 1 & \text{town/settlement - LAU 2, formally NUTS 5 level} \\ 2 & \text{statistical sub-region - LAU 1, formally NUTS 4 level} \\ 3 & \text{small regions/counties, NUTS 3 level} \\ 4 & \text{regions of regional policies, NUTS 2 level} \end{cases} \quad (2)$$

(Please note that, for simplicity, the term “town” is used for all cities and villages.)

People and their companies are assigned to geographic regions by the $\mathbf{A}^{[co,l]}$ and $\mathbf{A}^{[p,l]}$ incidence matrices, whose elements are defined as follows:

- (i) $a_{i,j}^{[co,l]}$ with element one if the headquarter of the i -th company is situated in the j -th geographic region at the level l of the settlement hierarchy,
- (ii) $a_{i,j}^{[p,l]}$ with element one if the i -th person is situated in the j -th geographic region at the level l of the settlement hierarchy,

so the directed weighted network that defines the number of investment connections between the regions can be defined as

$$\mathbf{A}^{[l]} = (\mathbf{A}^{[p,l]})^T \times \mathbf{A}^{[p,co]} \times \mathbf{A}^{[co,l]}. \quad (3)$$

Although companies may have many local divisions, the links between the towns are defined only by connecting the permanent addresses of the owners and the location of the headquarter. This arrangement results in a transparent and easily interpretable network as people and companies are assigned to only one location. The resultant network describes how investments unite the locations; e.g., the adjacency matrix $\mathbf{A}^{[1]}$ defines the number of links between the towns, and the degrees of the nodes represent the number of incoming and outgoing investments to the j -th and from the i -th town, respectively:

$$k_j^{[l,in]} = \sum_i a_{i,j}^{[l]} \quad (4)$$

$$k_i^{[l,out]} = \sum_j a_{i,j}^{[l]}. \quad (5)$$

The total number of ownership relationships is equal to the sum of the edge weights of the networks:

$$L = \sum_i \sum_j a_{i,j}^{[l]}, \quad \forall l, \quad (6)$$

where i and j represent the indices of the geographic regions at the level l of the settlement hierarchy.

It should be noted that as L represents the total number of connections, its value is independent of at which hierarchy level the edge weights are summarised.

Similarly, the total number of companies and investors can be calculated by summing the number of companies and people at any hierarchy level, respectively:

$$\begin{aligned} N^{[co]} &= \sum_{j=1} n_j^{[l,co]}, \\ N^{[p]} &= \sum_{j=1} n_j^{[l,p]}, \end{aligned} \quad (7)$$

$\forall l,$

where j represents the index of the geographic regions at the level l of the settlement hierarchy.

As people and companies are assigned only to one geographical region with the $\mathbf{A}^{[co,l]}$ and $\mathbf{A}^{[p,l]}$ incidence matrices, the number of people and companies at the j -th region of the l -th level of the settlement hierarchy can be calculated as

$$n_j^{[l,co]} = \sum_i a_{i,j}^{[co,l]} \quad (8)$$

$$n_j^{[l,p]} = \sum_i a_{i,j}^{[p,l]}. \quad (9)$$

The number of internal and external links of the network and the analysis of the local densities can be used to measure the attractiveness of the regions (see the Appendix). The following main body of the paper focuses on models that can be used to explore the communities in the network.

2.2. Evaluation of the Community Structure in the Settlement Hierarchy. The key idea of the methodology is that geographical regions can be interpreted as nonoverlapping communities of investors and companies as they belong to exactly one region among the set of these regions on the l -th level of the hierarchy, $C^{[l]} = \{C_1^{[l]}, C_2^{[l]}, \dots, C_l^{[l]}, \dots, C_{n_{c,ik}}^{[l]}\}$.

From the view of a community, the external degree is the number of links that connect the i -th community to the rest of the network, while the internal degree is the number of links between companies and owners in the same community, in other words, at the same location at the l -th level of the hierarchy (for more details see Appendix A). Recently, a wide variety of $f(C)$ metrics have been proposed to evaluate the quality of communities on the basis of the connectivity of their nodes [37]. The following subsections will demonstrate how these metrics can be interpreted to evaluate the attractiveness of geographical regions.

2.2.1. Modularity of a Region and Level of a Settlement Hierarchy. Classical modularity optimization-based community detection methods utilize $f(C)$ metrics that are based on the difference between the internal number of edges and their expected number [39, 45]:

$$\begin{aligned} f(C) &= (\text{fraction of edges within communities}) \\ &\quad - (\text{expected fraction of such edges}). \end{aligned} \quad (10)$$

In the case of the proposed directed network, this difference can be formulated as

$$f(C^{[l]}) = \frac{1}{L} \sum_{i,j} (a_{i,j}^{[1]} - p_{i,j}^{[1]}) \delta(C_i^{[l]}, C_j^{[l]}), \quad (11)$$

where $p_{i,j}^{[1]}$ represents the number of estimated investments proceeding from the i -th to the j -th town and $\delta(C_i^{[l]}, C_j^{[l]})$ is the Kronecker delta function that is equal to one, if the i -th and j -th towns are assigned to the same region on the l -th level of the hierarchy (e.g., $\delta(C_A^{[2]}, C_B^{[2]}) = 1$ when towns A and B are situated in the same statistical subregion).

The modularity of the partition $C^{[l]}$ can be calculated as the sum of the modularities of the $C_c^{[l]}, c = 1, \dots, n_c^{[l]}$ communities:

$$M_c^{[l]} = \frac{1}{L} \sum_{(i,j) \in C_c^{[l]}} (a_{i,j}^{[1]} - p_{i,j}^{[1]}). \quad (12)$$

The value of the modularity $M_c^{[l]}$ of a cluster/region $C_c^{[l]}$ can be positive, negative, or zero. Should it be equal to zero, the community has as many links as the null model predicts. When the modularity is positive, then the $C_c^{[l]}$ subgraph tends to be a community that exhibits a stronger degree of internal cohesion than the model predicts.

Using the proposed matrix representation, the calculation of the internal links at a given level of the hierarchy is straightforward, so the modularity can be easily calculated based on the diagonal elements of the adjacency matrices of the network and its null model:

$$f(C^{[l]}) = \sum_{c=1}^{n_c^{[l]}} M_c^{[l]} = \frac{1}{L} \sum_c a_{c,c}^{[l]} - \frac{1}{L} \sum_c p_{c,c}^{[l]}, \quad (13)$$

where $a_{c,c}^{[l]}$ represents the number of internal links in the c -th community/region on the l -th hierarchy level while $p_{c,c}^{[l]}$ is the expected number of these internal links calculated by the null model.

2.2.2. Null Models for Representing Regional Attractiveness. The critical element of the methodology is how the $p_{i,j}^{[1]}$ connection probabilities of the towns are calculated. The most widely applied *null model* is the random configuration model which calculates the edge probabilities assuming a random graph conditioned to preserve the degree sequence of the original network:

$$p_{i,j}^{[1]} = \frac{k_i^{[1,out]} k_j^{[1,in]}}{L}. \quad (14)$$

This randomized null model is inaccurate in most real-world networks [41].

As we measure the attractiveness of the regions based on the probability of link formation, it is beneficial to utilize attractiveness-related variables in the model as well as taking the distance-dependent link structure into account. Firstly,

we generalize the model by defining the node importance measures I_i^{out} and I_j^{in} :

$$p_{i,j}^{[1]} = \gamma I_i^{out} I_j^{in}. \quad (15)$$

As is expected from the null model, to fulfill the following equality,

$$\sum_{i,j} p_{i,j}^{[1]} = \sum_{i,j} a_{i,j}^{[1]} = L, \quad (16)$$

the importance measures are normalized as $\sum_i I_i^{out} = 1$ and $\sum_j I_j^{in} = 1$:

$$I_i^{out} = \frac{x_i^\alpha}{\sum_j x_j^\alpha}, \quad (17)$$

$$I_j^{in} = \frac{x_j^\beta}{\sum_i x_i^\beta},$$

where the parameters $\alpha, \beta > 0$ reflect the importance of the x_i and x_j variables used to express the probability of forming an edge from the i -th to the j -th node. Please note that when $\alpha = 1$ and $\beta = 1$, $x_i = k_i^{[1,out]}$, $x_j = k_j^{[1,in]}$, and $\gamma = L$, the model is identical to the random configuration model of a weighted directed graph.

To model the probability of distance-dependent link formation, the model defined by (15) is extended by a deterrence function $f(d_{i,j})$ which describes the effect of space [20]:

$$p_{i,j}^{[1]} = \gamma I_i^{out} I_j^{in} f(d_{i,j}). \quad (18)$$

The function $f(d_{i,j})$ can be directly measured from the data by a binning procedure similar to that used in [19]:

$$f(d) = \frac{\sum_{i,j|d_{i,j}=d} a_{i,j}^{[1]}}{\sum_{i,j|d_{i,j}=d} I_i^{out} I_j^{in}} \quad (19)$$

whose function is proportional to the weighted average of probability $(1/\gamma) a_{i,j}^{[1]}/(I_i^{out} I_j^{in})$ of a link existing at distance d .

When the distance dependence of the connection probability is handled by an explicit function, various modifications of the gravity law-based configuration model can be defined: $f(d) = 1/d_{i,j}^\delta$ [34, 46], $f(d) = \exp(-d_{i,j}/\delta)$ [47], or $f(d) = d_{i,j}^{-\delta} \exp(-d_{i,j}/\kappa)$ [48].

To ensure that the sum of the expected number of links is equal to L (see (16)), in this distance-dependent model γ should be normalized as

$$\gamma = \frac{L}{\sum_{i,j} I_i^{out} I_j^{in} f(d_{i,j})}. \quad (20)$$

Several models can be defined based on what kind of indicators are selected in the model. When the nodes are considered to be equally important, in other words, $I_i = I_j = 1$, only the distance determine the link formation probability,

$f(d_{i,j})$. The importance of the nodes can be interpreted as the number of investors and companies, so $I_i = (n_i^{[l,p]})^\alpha$ and $I_j = (n_j^{[l,co]})^\beta$. The null model can be defined based on the random configuration model, which results in the selection of the variables as $I_i = (k_i^{[l,out]})^\alpha$ and $I_j = (k_j^{[l,in]})^\beta$. Finally, socioeconomic indicators, like the number of inhabitants, or their complex combinations can be utilized.

When $f(d) = 1/d_{i,j}^\delta$, the parameters α, β, δ can be estimated as a regression problem. The identified parameters indicate the sensitivity, i.e., importance, of the variables that can be sorted by their importance as suggested in classical gravity law-based studies, like in [20].

2.2.3. Economic Relations of the Regions. Connections that interlink communities indicate their relationships and possibilities to merge modules/regions that are strongly connected. We combine regions and determine the gain of the merged modularity in a similar way to the Louvain community detection algorithm [42]. The $\Delta M_{i,j}$ modularity change obtained by merging the i -th and j -th communities can be calculated as the difference between the actual and predicted number of interlinking nodes:

$$\Delta M_{i,j}^{[l]} = \frac{1}{L} (a_{i,j}^{[l]} - p_{i,j}^{[l]}) + \frac{1}{L} (a_{j,i}^{[l]} - p_{j,i}^{[l]}). \quad (21)$$

The resultant symmetric modularity gain matrix can be calculated as

$$\Delta \mathbf{M}^{[l]} = (\mathbf{B}^{[l]})^T + \mathbf{B}^{[l]}, \quad (22)$$

where $\mathbf{B}^{[l]} = \mathbf{A}^{[l]} - \mathbf{P}^{[l]}$ is the so-called modularity matrix [38].

The Louvain algorithm moves a node i in the community for which the gain in modularity is the largest. If no positive gain occurs, i remains in its original community. After merging the nodes/regions, a new network is constructed whose nodes are in the communities identified earlier. This method can be used to explore regions (modules) formed by the elements of the l -th settlement hierarchy with different null models. Although model-based communities can be identified by this approach and compared to regions of a larger hierarchy level as modules of ground truth, the main goal of the analysis of $\mathbf{M}^{[l]}$ is to measure the strength of relationships between the regions.

The following section demonstrates the applicability of the previously presented toolset in the analysis of the network of Hungarian companies.

3. Results and Discussion

3.1. Description of the Studied Dataset. The studied dataset represents $L = 1,077,090$ ownership relations between $N^{[pl]} = 531,249$ people and $N^{[co]} = 868,591$ Hungarian companies in 2013. It should be noted that only less than 10% of the ownership connections are defined based on how companies possess shares in other companies, so, although only personal investments are studied, the results reflect the attractiveness

TABLE 1: Number of edges inside the settlement hierarchies.

	Town-level	sub-Region-level	County-level	Region-level
Number of nodes, N	3,111	175	20	7
Number of internal ties	797,492	846,309	893,559	969,995
Number of external ties	279,598	230,781	183,531	107,095

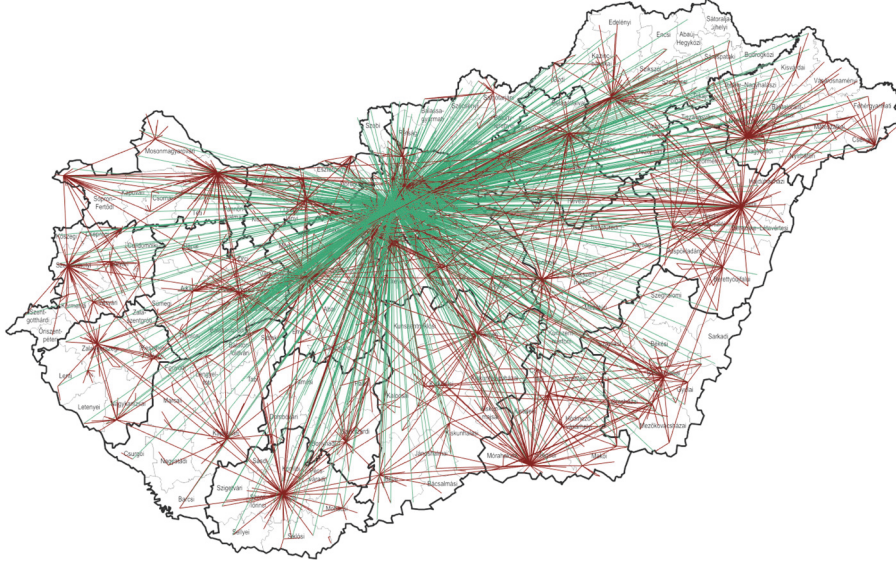


FIGURE 2: Map of the town-level company-ownership network. Edges with more than 10 ownership connections are shown. Edges connected to the capital (Budapest) are denoted by green lines.

of the towns and regions as the generated network covers more than 90% of the investment-type connections.

The owners and companies were assigned to settlements, and the related settlement hierarchy covers 3,155 towns (level LAU 2, formally level NUTS 5), 175 statistical subregions (level LAU 1, formally level NUTS 4), 20 small regions/counties in level NUTS 3, and 7 regions in level NUTS 2.

74% of the connections remain within the borders of the towns, which also reflects the high degree of modularity of the network (for more details, see Table 1). 302,781 connections are within Budapest and 45,559 connections point out of the city, while 89,944 connections point into the capital. The map of the regional connections between the people and companies can be generated using the obtained connectivity matrix and the latitudes and longitudes of the towns (see Figure 2). It can be seen that the network reveals a hierarchical and modular structure reflecting that the Hungarian economy is concentrated around the capitals of the counties and Budapest, the capital of the country. The majority of the companies are situated in these locations; consequently, the network follows the structure of online social networks [44]; in other words, it is also structured according to the settlement hierarchy, in which Budapest is the absolute center of the network and the centers of counties also function as hubs.

3.2. Measuring Attractiveness. The densities inside towns and regions can highlight the modular structure of the

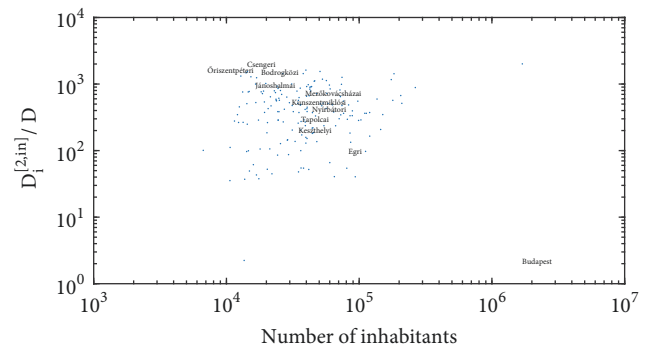


FIGURE 3: Network density as a function of the number of inhabitants on the level LAU 1.

company-ownership network. As shown in Figure 3, these densities are significantly higher in most subregions and a negative correlation exists between the size of the regions and the number of their inner connections ($r = 0.298$, $p < 10^{-4}$). As illustrated by the results, smaller locations are much more isolated than larger ones, like Budapest. The same result is obtained by the analysis of the external density-based openness measure which we consider as a main measure of attractiveness (see Appendix A for more details). As shown in Figure 4, bigger regions exhibit larger openness values reflecting their higher degree of attractiveness ($r = 0.94$, $p < 10^{-10}$).

TABLE 2: Performances of distance-dependent null models.

Nodes/Null models	p^{spa}	$p^{\alpha,\beta}$	p^{grav}
$I_i^{out} = I_j^{in} = 1$	0.28100	0.28113	0.28093
$I_i^{out} = N_i^{[p]}, I_j^{in} = N_j^{[co]}$	0.08915	0.01359	0.00651
$I_i^{out} = k_i^{[1,out]}, I_j^{in} = k_j^{[1,in]}$	0.05759	0.01389	0.00642
$I_i^{out} = \text{Inhabitants}_i, I_j^{in} = \text{Inhabitants}_j$	0.12106	0.01456	0.00650
$I_i^{out} = \text{TDI}_i, I_j^{in} = \text{TDI}_j$	0.07142	0.01482	0.00644

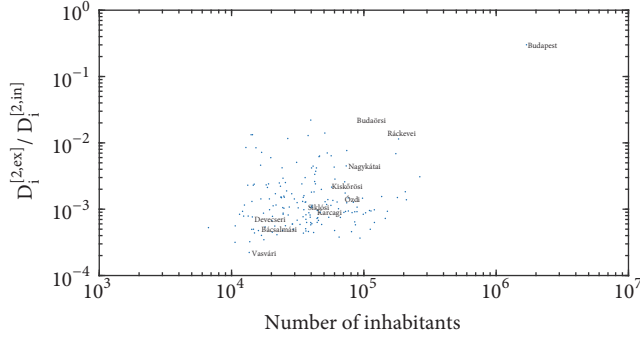
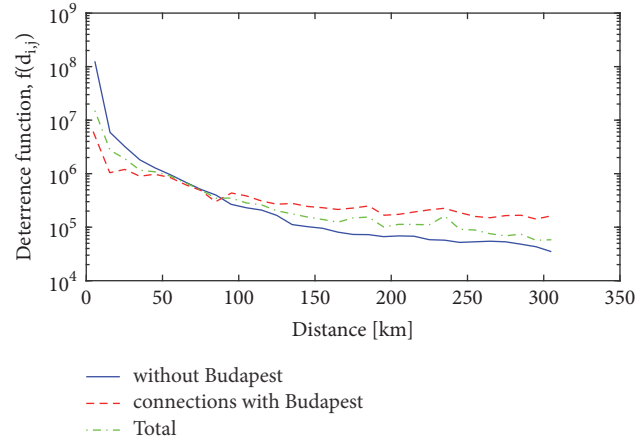


FIGURE 4: Openness of small regions (LAU 1 level) as a function of the number of their inhabitants.

FIGURE 5: Empirically derived deterrence function determined by (19), where $I_i^{[in]} = n_i^{[1,p]}$, $I_j^{[in]} = n_j^{[1,co]}$.

3.3. *The Effect of Geographical Distance.* To address the effect of distance decay on link formation, the observed ties between the towns were compared with their expected number calculated from a probabilistic model.

A resolution of 10 km was used for binning the distance distribution (see Figure 5). The exponent of distance decay according to our data is -1.1057. It should be noted that the effect of the capital city is so high, the probability of forming connections with Budapest is slightly less distance-dependent, and the exponent of distance decay with regard to these connections is only -0.6385.

The distance-dependent link formation probability can be explained by the notion that the costs of establishing and

maintaining the connections are also distance-dependent. This assumption can be confirmed by the fact that the distance has a much stronger effect on investment ties than on online social networks in Hungary (where the exponent of distance decay is -0.6) [44], probably since the cost of keeping connections is less dependent on distance than the management of a company far from the permanent address of the owner.

3.4. *Comparison of the Null Models.* Based on the utilized distance function, three different types of models can be defined. When $f(d)$ is a deterrence function defined by (19), the models are denoted as $p^{spa} = \gamma I_i^{out} I_j^{in} f(d)$. $p^{\alpha,\beta} = \gamma (I_i^{out})^\alpha (I_j^{in})^\beta f(d)$ represents the parametric version of this model, when the exponents α and β are optimized to achieve a more accurate approximation of connections between towns. $p_{i,j}^{grav} = \gamma (I_i^{out})^\alpha (I_j^{in})^\beta / d^\delta$ represents the gravity-type models.

Five sets of I_i^{out}, I_j^{in} variables were defined, including simple metrics like the numbers of nodes and edges [1] in addition to socioeconomic variables, like the number of inhabitants and Total Domestic Income (total income received by all sectors of the economy including the sum of all wages, profits, and taxes, minus subsidies). Based on the combination of different variables and distance functions, 15 different models were identified:

$$\min_{\alpha,\beta,\gamma} E_m(\alpha, \beta, \gamma) = \frac{1}{L} \|\mathbf{A}^{[1]} - \mathbf{P}^{[1]}\|_2. \quad (23)$$

As summarized in Table 2, by taking the distance into account, the accuracy of the model is significantly improved. Among distance-dependent models, the gravity models perform best (in comparison, the accuracy of the distance independent random configuration model is 0.16494).

The Total Domestic Income (TDI) is one of the best indicators. The identified α, β , and δ parameters reflect the importance of the I_i^{out}, I_j^{in} , and d variables in the models (e.g., in the case where $I_j^{in} = \text{TDI}_j$ and $I_i^{out} = \text{TDI}_i$, the resultant nonlinear regression model is $p_{i,j} = 0.12 \cdot ((I_i^{out})^{0.37} \cdot (I_j^{in})^{0.81}) / d^{1.58}$ (see Table 3)), which can be interpreted as the notion that the number of connections between location i and location j is increased by 0.37% as a result of 1.0% growth of TDI in location i . Similarly, the number of connections between location i and location j is increased by 0.81% as a result of 1.0% growth of TDI in location j . According to the gravity-type models, the importance of the target/destination

TABLE 3: Coefficients of the parametric models that reflect the importance of the variables.

Nodes/Parameters	$p^{\alpha,\beta} = \gamma (I_i^{out})^\alpha (I_j^{in})^\beta f(d)$		$p_{i,j}^{grav} = \gamma (I_i^{out})^\alpha (I_j^{in})^\beta / d^\delta$		
	α	β	α	β	δ
$I_i^{out} = N_i^{[p]}, I_j^{in} = N_j^{[co]}$	1.08373	0.91787	0.34984	0.67191	1.63711
$I_i^{out} = k_i^{[1,out]}, I_j^{in} = k_j^{[1,in]}$	1.05439	0.94455	0.35652	0.69045	1.59439
$I_i^{out} = \text{Inhabitants}_i, I_j^{in} = \text{Inhabitants}_j$	0.99347	1.15642	0.40654	0.88313	1.52391
$I_i^{out} = TDI_i, I_j^{in} = TDI_j$	0.98571	1.03669	0.37367	0.81425	1.58060

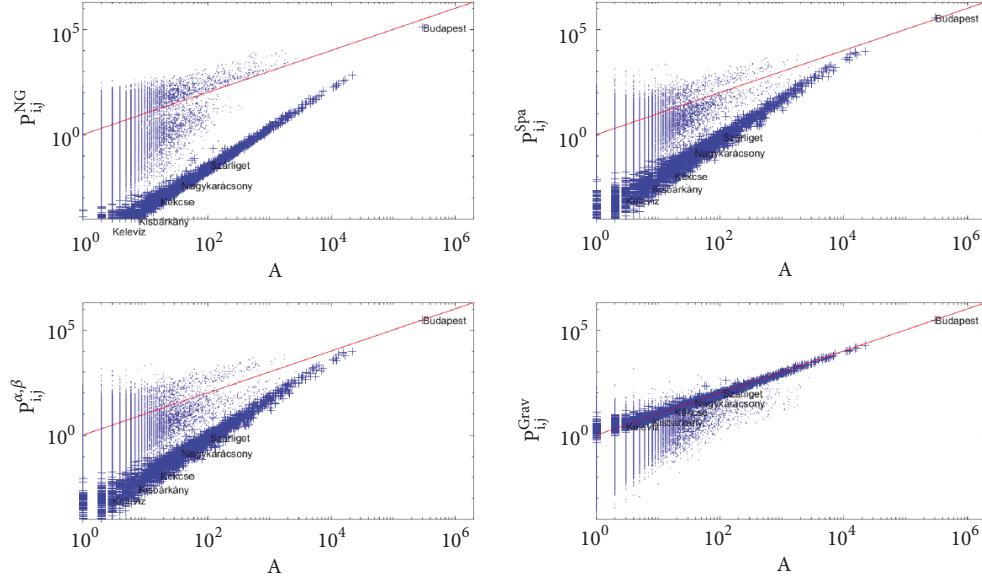


FIGURE 6: Comparison between the number of the edge weights $a_{i,j}^{[1]}$ and their estimated values $p_{i,j}^{[1]}$ generated by different null models on the town level (LAU 2) settlement hierarchy when $I_i^{out} = k_i^{[1,out]}$ and $I_j^{in} = k_j^{[1,in]}$. The + symbols represent the inner connections that form a separate cluster. This plot directly reflects the goodness of fit as the model estimates the connections of the towns.

locations (β) is greater than the importance of the sources (α) regardless of how the strengths of the nodes are interpreted.

3.5. Evaluation of the Modularities. As modularity-based community detection evaluates the set of $a_{i,j}^{[1]} > p_{i,j}^{[1]}$ edges (and the related nodes) whose weights are underestimated by the null model (see (11)), we designed a plot that compares $a_{i,j}^{[1]}$ with $p_{i,j}^{[1]}$ to highlight the set of potential edges that can be used to form communities.

Four null models based on the $I_1 := k_i^{[1,out]}$ and $I_2 := k_j^{[1,in]}$ Newman and Girvan model are compared in Figure 6. In all models, the inner connections (represented by +) form a separate cluster which confirms that 74% of the connections remain within the borders of the towns. The first model (p^{NG}) shows that more inner connections exist than would be expected based on the random configuration network. The spatial models p^{Spat} and $p^{\alpha,\beta}$ handle the dependence on distance of the connections, so a slightly smaller difference is shown in the number of the experienced and expected inner connections. It is reflected in Figure 7 that during the aggregation procedure the qualitative behavior of the models does not change.

The difference between the expected number of interconnections is higher in the case of smaller settlements which indicates that small regions are not as attractive as would be expected from their number of nodes. The gravity model p^{Grav} well estimates the inner connections thanks to the exponents $\alpha = 0.35652$ and $\beta = 0.69045$ whose parameters effectively represent that the increase in the number of connections affects the attractiveness in a nonlinear fashion. This phenomenon is much more interesting when the utilized variables can be interpreted as economic potentials. When TDI is applied in the gravity model, $\alpha = 0.37367$ and $\beta = 0.81425$. These values and Figure 8 confirm that gravity-based models behave similarly and, therefore, reflect the same mechanism of attractiveness.

3.6. Forming Communities. Connections that interlink communities are indicative of their relationships. The effect of these interlinks can be studied by the change in modularity (see (21)) expressed as $\Delta \mathbf{M}^{[l]} = (\mathbf{B}^{[l]})^T + \mathbf{B}^{[l]}$.

To determine the community structure, the MATLAB implementation [49] of the greedy Louvain algorithm [50] was used. Towns and subregions were used as an initial

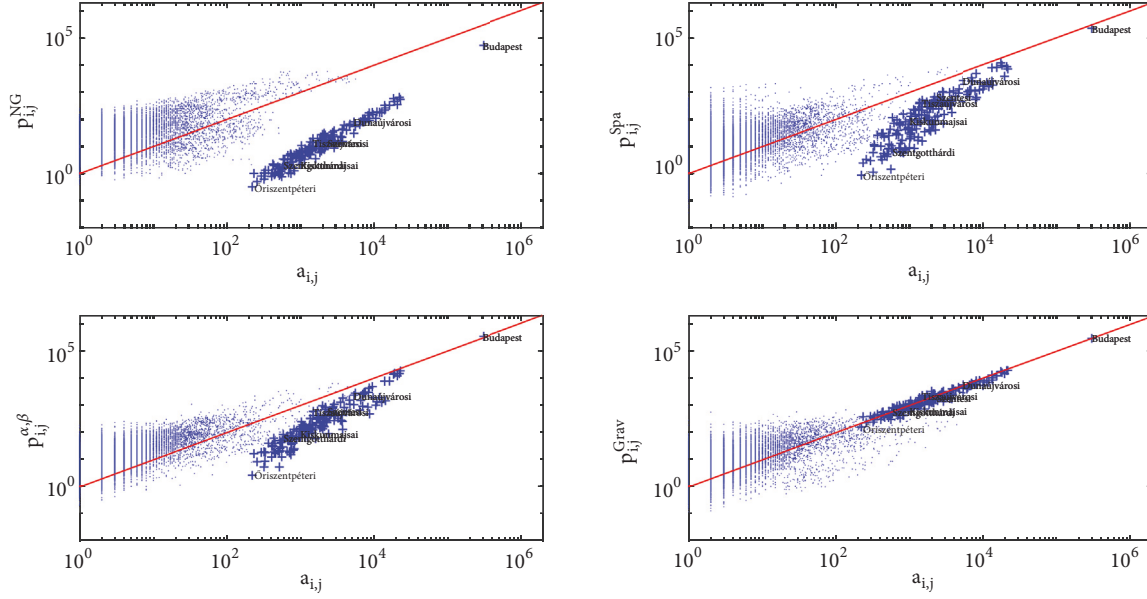


FIGURE 7: Comparison between the number of the edge weights $a_{i,j}^{[2]}$ and their estimated values $p_{i,j}^{[2]}$ generated by different null models at level LAU 1 of the settlement hierarchy when $I_i^{out} = k_i^{[1,out]}$ and $I_j^{in} = k_j^{[1,in]}$. The + symbols represent the inner connections that form a separate cluster. This plot reflects that, during the aggregation procedure, the qualitative behavior of the models does not change; furthermore, the same phenomena can be observed as in Figure 6.

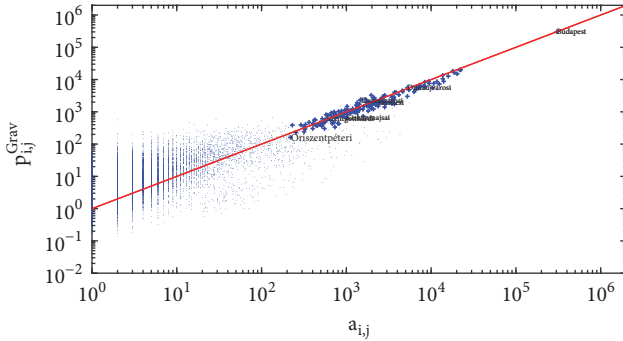


FIGURE 8: Comparison between the number of the edge weights $a_{i,j}^{[2]}$ and their estimated values $p_{i,j}^{[2]}$ generated by the gravity null model at level LAU 1 of the settlement hierarchy when $I_i^{out} = TDI_i$ and $I_j^{in} = TDI_j$. The + symbols represent the inner connections that form a separate cluster.

partition $\mathbf{B}^{[1]}$. As shown in Figure 9, the community structure formed based on the null model p^{NG} almost perfectly reconstructs the counties confirming that the settlement structure is reflected in terms of the personal investments.

Different null models provide different viewpoints with regard to community detection. The NG null model does not handle the distance dependence of the connections so the matrix $\mathbf{B}^{[1]} = \mathbf{A}^{[1]} - \mathbf{P}^{[1]}$ of the modeling errors reflects the distance dependence of the connections. Therefore, the resulting communities form spatial clusters. On the contrary, communities formed by the gravitational models

reflect distance-dependent differences less. According to the resultant maps, the attractiveness of Budapest is highlighted as only small since closed regions were not assigned to the module of the capital (see Figure 10(a)). It is interesting to note that all the centers of counties were assigned to the community of Budapest in gravitational model which also confirms the hierarchical structure of the network. To highlight the hierarchical structure and increase the sensitivity of the model, a resolution parameter was introduced into the model (see Appendix B) that can be adapted to detect similar region-pairs as shown in Figure 10(b).

Communities formed with the NG null model (see Figure 9) and the TDI-based gravity models (see Figure 10) significantly differ. The interpretation of the communities and these differences should rely on the understanding of the concept of the modularity. The utilised modularity detection algorithm generates partitions in which the links are more abundant within communities than would be expected from the employed model.

As the NG null model only uses the basic structural information encoded in the adjacency matrix, when the probabilities of the connections are dependent on distance, the resulting communities will represent closer geographical regions. As Table 1 and Figures 6 and 7 show, most of the connections remain within the county borders, so it is natural that the resultant 30 communities are almost identical to the counties.

Since the Hungarian road network reflects the administrative regions, it can be shown that the distance strongly affects the probability of the connections. This distance dependence of the connection probability can be incorporated into the null model by the proposed gravity model.

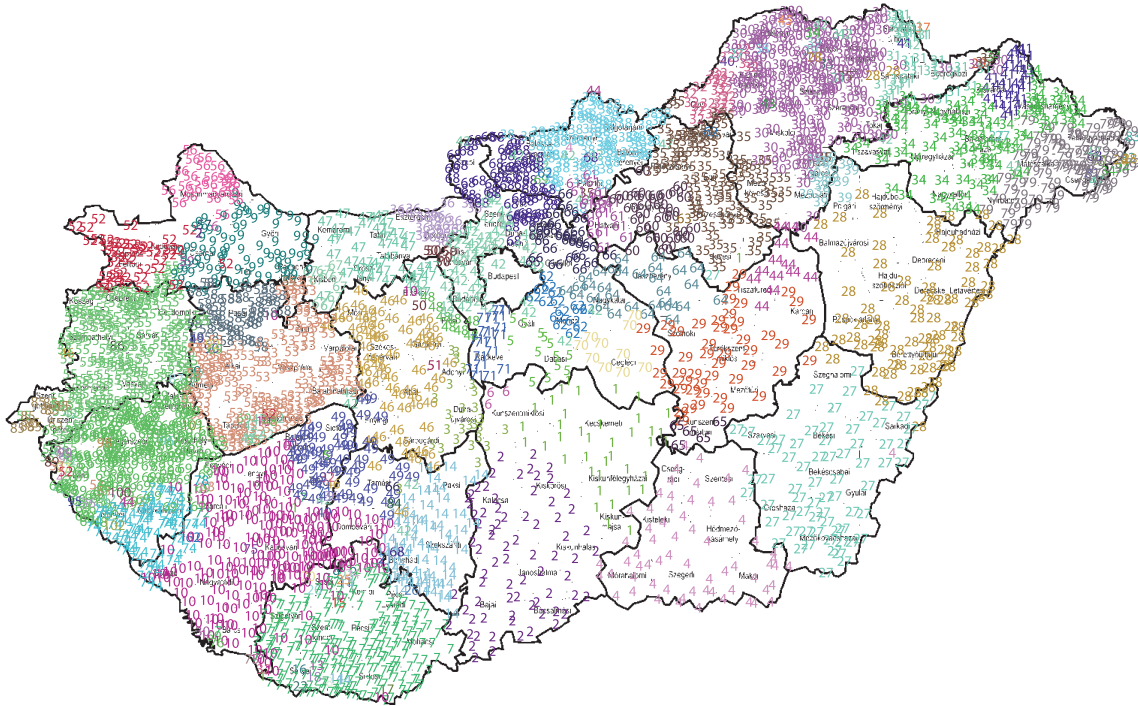
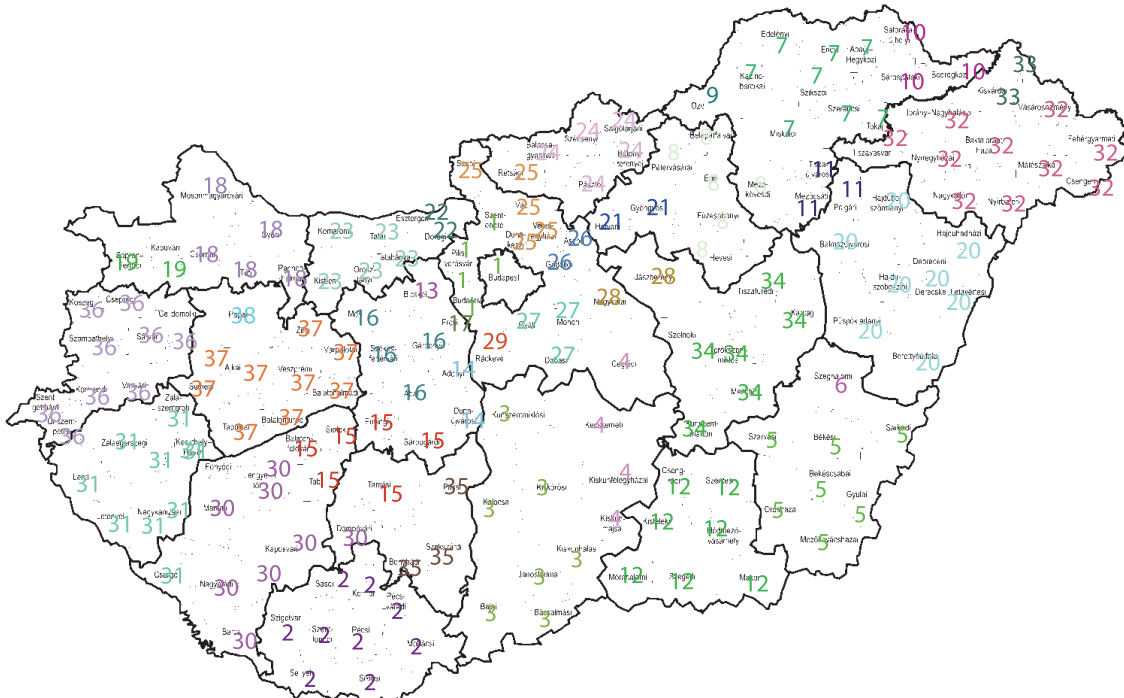
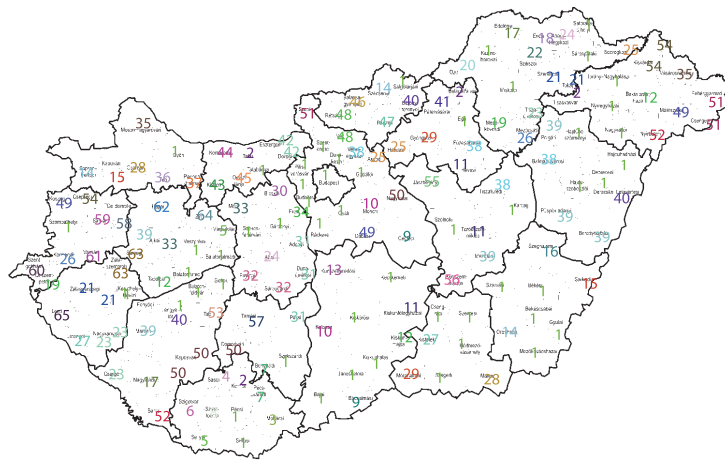
(a) Initial nodes are towns ($l = 1$)(b) Initial nodes are subregions ($l = 2$)

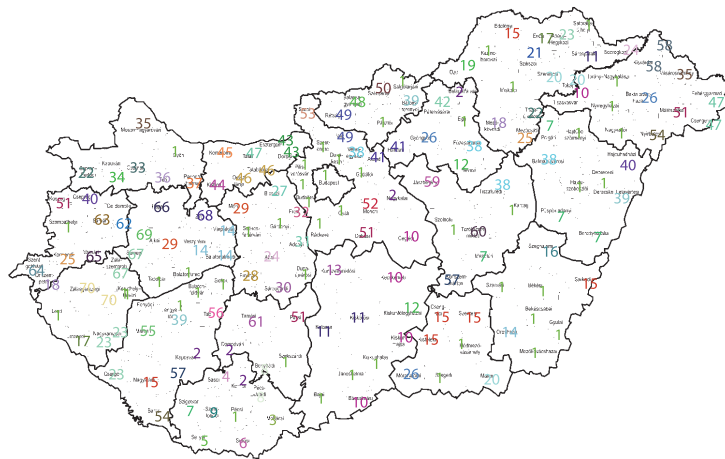
FIGURE 9: Communities formed by the Louvain method and Newman-Girvan (NG) null model ($I_i = k_i^{out}$ and $J_j = k_j^{in}$) reflect the settlement hierarchy as the resultant communities are almost identical to the counties.

In this case, the resultant communities will reflect another unmodelled surplus in the number of connections. When the attractiveness and the distances are considered in the null model, the communities will reflect the additional economic attractiveness/similarity of the regions.

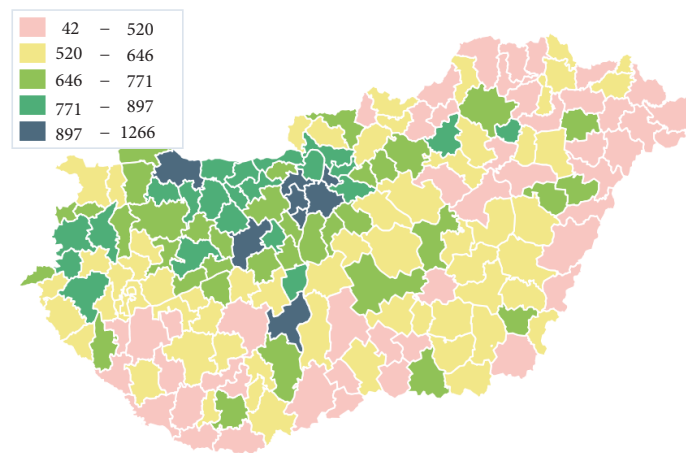
As Figure 10 shows, the algorithm generates a huge cluster of a well developed regions with Budapest, the larger cities and county seats with high TDIs, and several small communities related to isolated and less developed subregions.



(a) TDI-based gravitational model: Initial nodes are subregions ($l = 2$)



(b) The same TDI-based gravitational model at higher resolution $\gamma_r = 1.1$



(c) Spatial distribution of the TDI per capita (in 1000 HUF)

FIGURE 10: Communities formed by the Louvain method and gravitational null models reflect the attractiveness of Budapest as only less developed closed regions were not assigned to the module of the capital.

4. Conclusions

Regional policy-making and monitoring are firm-centered, incentive-based, and state-driven. Personal investments define ties between geographical locations. We analyzed the structure of this ownership network and proposed a methodology to characterize regional attractiveness based on a set of null models identified to approximate the probabilities of link formation. According to the levels of the settlement hierarchy, a four-level study was conducted.

Based on the calculation of the internal and external network densities, several measures were proposed to evaluate the attractiveness and development of towns and geographical regions. The results indicate that small and less competitive regions have less internal connections, while larger cities are much more open.

To provide a more in-depth insight into the network, the dependence of link formation on distance was studied. The probability of connections between owners and their companies shows a much more rapid degree of distance decay than experienced in social networks. The attractiveness of the capital is so high that its connections are much less dependent on distance than other cities.

Based on the combination of three deterrence models and five sets of indicators, 15 different null models were identified besides the classical Newman-Girvan random configuration model. Communities statistically have more significant edge weights that would be wired according to the null model. As it was highlighted that underestimated link probabilities are the sources of modularity, a scatter plot was designed to visualize how the null model approximates the real structure of the network.

The identification of gravity-type models highlighted that link formation is nonlinearly dependent on the studied variables. Furthermore, the target nodes are much more important when determining the probability of link formation than the source nodes which also confirms why the structural analysis of company-ownership networks can be used to measure regional attractiveness.

We applied the Louvain community detection algorithm to form clusters of cities and subregions and compared the resultant communities to administrative regions. When the null model more closely approximates the real structure of the network, then the modularity is expected to be lower. As community detection forms modules whose internal link densities are significantly higher than what would be expected from the applied null models, spatial clusters that were highlighted by the distance independent random configuration model are almost identical to the counties. Communities generated based on the gravitational models, which correctly estimate the number of internal nodes and the dependence of link formation on distance, exploited the attractiveness of the capital, as they form a massive cluster that includes most of the centers of each county, bigger cities, and the competitive touristic regions, while the remaining small clusters reflect isolated regions that are less developed and less attractive.

Appendix

A. Internal and External Connection-Based Evaluation

Finding community structure means the assignment of the nodes into groups, where within the nodes are highly connected and across the nodes of the communities they are much loosely connected to each other [51].

The density of the whole network can be calculated as

$$D = \frac{L}{N^{[p]} N^{[co]}}. \quad (\text{A.1})$$

while the internal density of the region is calculated as

$$D_i^{[l, in]} = \frac{a_{i,i}^{[l]}}{n_i^{[l,p]} n_i^{[l,co]}}. \quad (\text{A.2})$$

$D_i^{[l, in]}/D$ compares internal complexity of the regions to the whole network.

The probability of an external tie, in other words, the external density, can be calculated in a similar fashion:

$$D_i^{[l, ex]} = \frac{\sum_{i \neq j} a_{i,j}^{[l]}}{N^{[l,p]} (N^{[l,co]} - n_i^{[l,co]})}, \quad (\text{A.3})$$

where $N^{[l,co]} - n_i^{[l,co]}$ represents the number of companies that are outside of the i -th region at the $[l]$ -th level of the settlement hierarchy.

To evaluate the openness as a measure of the attractiveness of the region, the ratio of the external to internal probabilities can be defined as

$$O_i^{[l]} = \frac{D_i^{[l, ex]}}{D_i^{[l, in]}}. \quad (\text{A.4})$$

Apart from taking into account internal and external links, the direction of the connections can be considered. Expansion computes the number of edges pointing outside the community [37]:

$$E_i^{[l]} = \frac{\sum_i a_{i,j}^{[l]} - \sum_i a_{i,i}^{[l]}}{n_i^{[l,p]}}. \quad (\text{A.5})$$

Similarly, the ability of a community to collect links can be determined by the normalized number of links that point inside the community:

$$LCA_i^{[l]} = \frac{\sum_j a_{i,j}^{[l]} - a_{i,i}^{[l]}}{n_i^{[l,co]}}. \quad (\text{A.6})$$

Cut ratio is similar to the internal density as it computes the fraction of edges pointing out and the number of possible edges that are pointing outside the community:

$$CR_i^{[l]} = \frac{\sum_j a_{i,j}^{[l]} - a_{i,i}^{[l]}}{n_i^{[l,p]} (N^{[l,co]} - n_i^{[l,co]})}. \quad (\text{A.7})$$

TABLE 4: Parameters of the power-law distributions fitted to networks at different settlement hierarchy levels.

Distribution	k_{sat}	k_{cut}	γ
$k_j^{[1,in]}$ (LAU 2)	120	15061	1.85
$k_j^{[1,out]}$ (LAU 2)	138	19709	1.87
$k_j^{[2,in]}$ (LAU 1)	1974	392724	2.04
$k_j^{[2,out]}$ (LAU 1)	2070	348339	2.04
$k_j^{[3,in]}$ (NUTS 3)	19693	392724	2.54
$k_j^{[3,out]}$ (NUTS 3)	20401	348339	2.49
$k_j^{[4,in]}$ (NUTS 2)	74161	557112	3.31
$k_j^{[4,out]}$ (NUTS 2)	77042	519967	3.35

B. Improvement of the Resolution

The modularity always increases when small communities are assigned to one group [52]. Modularity optimization with the null model p^{NG} has a resolution threshold which means it fails to identify small communities in large networks and communities consisting of less than $(\sqrt{L}/2-1)$ internal links [53]. Reichardt and Bornholdt (RB) generalized the modularity function by introducing an adjustable γ_r parameter [54, 55] to handle this problem, which for our directed and weighted networks is

$$M_{RB}^{dir} = \frac{1}{L} \sum_i \sum_j \left(a_{ij} - \gamma_r \frac{k_i^{out} k_j^{in}}{L} \right) \delta(C_i, C_j). \quad (B.1)$$

Arenas, Fernandez, and Gomez (AFG) also proposed a multiresolution method by adding r self-loops to each node [56]. This algorithm increases the strength of a node without altering the topological characteristics of the original network, as $\mathbf{A}_r = \mathbf{A} + r \mathbf{I}$, where \mathbf{I} denotes the identity matrix and r the weight of the self-loops of each node:

$$M_{AFG} = \frac{1}{L'} \sum_i \sum_j \left(a'_{i,j} - \frac{k_i^{out'} k_j^{in'}}{L'} \right) \delta(C_i, C_j), \quad (B.2)$$

where $L' = L + Nr$, $L = \sum_{i,j} a_{i,j}$, $k_i^{out'} = k_i^{out} + r$, $k_j^{in'} = k_j^{in} + r$, and

$$a'_{i,j} = \begin{cases} a_{i,j}, & \text{if } i \neq j, \\ a_{i,j} + r, & \text{if } i = j. \end{cases} \quad (B.3)$$

These methods still have the intrinsic limitation, so large communities may have been split before small communities became visible. The theoretical results indicated that this limitation depends on the degree of interconnectedness of small communities and the difference between the sizes of the communities, while being independent of the size of the whole network [52].

It should be noted that the modularity decreases when $p_{i,j}$ more closely approximates the real $a_{i,j}$ values which is equivalent to finding the null model that most closely fits.

C. Network Topology Analysis

The degree distribution was determined in all levels of the settlement hierarchy by following the methodology presented

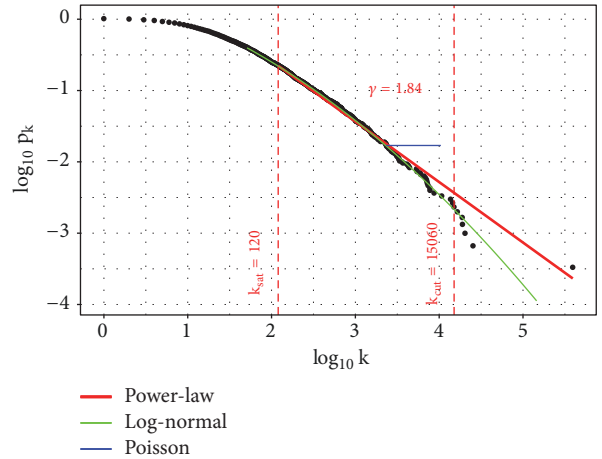


FIGURE 11: Distribution of the $k_j^{[1,in]}$ edges at the LAU 2 settlement hierarchy level.

in [13]. Figure 11 shows that the distribution shows small-degree saturation and high-degree cutoff. Several distribution functions were fitted. The two-sided Young's test statistic [57] showed that exponential and Poisson distributions which reflect the randomness of connections could be rejected. According to this test, the power-law distribution cannot be rejected. The estimated parameters are shown in Table 4. The power-law distribution of the incoming and outgoing connections reflects the preferential attachment-type structure of the network.

In hierarchical networks, nodes with high degree tend to connect to nodes that are less connected to others [58]. Therefore, the hierarchical structure of the network is reflected by the dependence of the local clustering coefficient $C(k)$ on the degree of the nodes. As Figure 12 shows, $C(k)$ decreases with increasing k with $C(k) \approx k^{-0.3}$ which indicates the hierarchical structure of the network [58, 59].

D. Notations

p: Person/investor who is equivalent to the owner of a company

co: Company

[l]: Level of the settlement hierarchy (see (2))

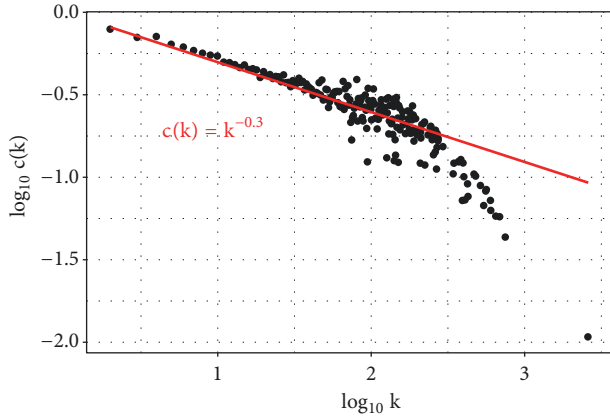


FIGURE 12: Local clustering coefficient as a function of the $k_j^{[1, in]}$ node degrees.

- $entity^{[l]}$: Aggregation of an *entity* at level l of the settlement hierarchy
- $\mathbf{A}^{[p, co]}$: Biadjacency matrix of person-company ownership network
- $a_{i, j}^{[p, co]}$: An element (edge weight) of the $\mathbf{A}^{[p, co]}$ biadjacency matrix of person-company ownership network
- $\mathbf{A}^{[p, l]}, \mathbf{A}^{[co, l]}$: Incidence matrices of person-location and company-location bipartite networks at the level l of the settlement hierarchy
- $\mathbf{A}^{[l]}$: Simpler notation of an adjacency matrix of location network at l level of settlement hierarchy (see (3))
- $k_j^{[l, in]}$: In-degree of the j -th node (geographic region) at level l of the settlement hierarchy
- $k_i^{[l, out]}$: Out-degree of the i -th node (geographic region) at level l of the settlement hierarchy
- $n_j^{[l, co]}, n_j^{[l, p]}$: Numbers of companies and people in the j -th region at level l of the settlement hierarchy
- $N^{[co]}, N^{[p]}$: Number of companies and people/owners/investors in the network
- L : Number of links in the network
- C : Set of communities (each node is a member of exactly one community)
- $C^{[l]}$: Set of communities at level l of the settlement hierarchy (C^1 denotes the set of towns)
- $n_c^{[l]}$: Number of communities at level l of the settlement hierarchy
- $f(C)$: Generally a metric as a function of community structure that indicates the goodness-of-fit of the community on the basis of the connectivity of nodes in it

- $f(C^{[l]})$: Metric of the goodness-of-fit of the community structure which is the level l of the settlement hierarchy
- M : A special $f(C)$ defined by (11) called modularity of network
- M_c : Modularity of community c (sum of the modularity of each community yields the modularity M of the network)
- $D_i^{[l, in]}, D_i^{[l, ex]}$: Internal and external densities of the i -th community at level l of the settlement hierarchy, defined by (A.2) and (A.3)
- $O_i^{[l]}$: Openness of the i -th community at level l of the settlement hierarchy, defined by (A.4)
- $E_i^{[l]}$: Expansion of the i -th community at level l of the settlement hierarchy, defined by (A.5)
- $LCA_i^{[l]}$: Link-collection ability of i -th community at level l of the settlement hierarchy, defined by (A.6)
- $CR_i^{[l]}$: Cut ratio of the i -th community at level l of the settlement hierarchy, defined by (A.7).

Data Availability

The data used to support the findings of this study are available from the website of the corresponding author (<https://www.abonyilab.com/network-science/structural-analysis>).

Disclosure

Parts of the research have been presented at the 16th Annual Meeting of the Hungarian Regional Science Association (18 October 2018, Kecskemet, Hungary) in an oral presentation entitled “Measurement of Regional Attractiveness Based on Company-Ownership Networks.”

Conflicts of Interest

The authors declare that no conflicts of interest exist with regard to the publication of this paper.

Acknowledgments

This research was supported by the National Research, Development, and Innovation Office (NKFIH), through the project OTKA-116674 (Process Mining and Deep Learning in the Natural Sciences and Process Development), and the European Union, as well as Hungary and cofinanced by the European Social Fund through the project EFOP-3.6.2-16-2017-00017, titled “Sustainable, Intelligent, and Inclusive Regional and City Models.”

References

- [1] L. Kendrick, K. Musial, and B. Gabrys, “Change point detection in social networks—critical review with experiments,” *Computer Science Review*, vol. 29, pp. 1–13, 2018.

- [2] K. Musial, M. Budka, and K. Juszczyszyn, "Creation and growth of online social network: How do social networks evolve?" *World Wide Web*, vol. 16, no. 4, pp. 421–447, 2013.
- [3] P. Bródka, K. Musial, and P. Kazienko, "A Method for Group Extraction in Complex Social Networks," in *Knowledge Management, Information Systems, E-Learning, and Sustainability Research*, vol. 111 of *Communications in Computer and Information Science*, pp. 238–247, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [4] M. Qin, D. Jin, D. He, B. Gabrys, and K. Musial, "Adaptive community detection incorporating topology and content in social networks," in *Proceedings of the 9th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2017*, pp. 675–682, Australia, August 2017.
- [5] P. Kazienko, K. Musial, E. Kukla, T. Kajdanowicz, and P. Bródka, "Multidimensional Social Network: Model and Analysis," in *Computational Collective Intelligence. Technologies and Applications*, vol. 6922 of *Lecture Notes in Computer Science*, pp. 378–387, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [6] S. T. Cavusgil, T. Kiyak, and S. Yeniyurt, "Complementary approaches to preliminary foreign market opportunity assessment: Country clustering and country ranking," *Industrial Marketing Management*, vol. 33, no. 7, pp. 607–617, 2004.
- [7] C. del Campo, C. M. F. Monteiro, and J. O. Soares, "The European regional policy and the socio-economic diversity of European regions: A multivariate analysis," *European Journal of Operational Research*, vol. 187, no. 2, pp. 600–612, 2008.
- [8] A. Amin, "An Institutionalist Perspective on Regional Economic Development," *International Journal of Urban and Regional Research*, vol. 23, no. 2, pp. 365–378, 1999.
- [9] M. Wang, "Location Is (Still) Everything: The Surprising Influence of the Real World on How We Search, Shop, and Sell in the Virtual One by David R. Bell," *Southeastern Geographer*, vol. 56, no. 4, pp. 476–477, 2016.
- [10] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [11] G. Bianconi and A.-L. Barabási, "Competition and multiscaling in evolving networks," *EPL (Europhysics Letters)*, vol. 54, no. 4, pp. 436–442, 2001.
- [12] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Structure of growing networks with preferential linking," *Physical Review Letters*, vol. 85, no. 21, pp. 4633–4636, 2000.
- [13] A.-L. Barabási, *Network science book*, Center for Complex Network, Northeastern University, Boston, 2014, <http://barabasi.com/networksciencebook>.
- [14] A. Blank and S. Solomon, "Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components)," *Physica A: Statistical Mechanics and its Applications*, vol. 287, no. 1-2, pp. 279–288, 2000.
- [15] G. Duranton and D. Puga, "The Growth of Cities," *Handbook of Economic Growth*, vol. 2, pp. 781–853, 2014.
- [16] X. Gabaix, "Zipf's law for cities: an explanation," *The Quarterly Journal of Economics*, vol. 114, no. 3, pp. 739–767, 1999.
- [17] M. Cristelli, M. Batty, and L. Pietronero, "There is more than a power law in Zipf," *Scientific Reports*, vol. 2, article no. 812, 2012.
- [18] M. Reba, F. Reitsma, and K. C. Seto, "Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000," *Scientific Data*, vol. 3, 2016.
- [19] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, "Uncovering space-independent communities in spatial networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 19, pp. 7663–7668, 2011.
- [20] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1–3, pp. 1–101, 2011.
- [21] G. K. Zipf, *Human behavior and the principle of least effort*, Addison-Wesley Press, 1949.
- [22] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora, "Distance matters: Geo-social metrics for online social networks," in *Proceedings of the 3rd Conference on Online Social Networks, WOSN'10*, USENIX Association, 2010.
- [23] R. Lambiotte, V. D. Blondel, C. de Kerchove et al., "Geographical dispersal of mobile communication networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 21, pp. 5317–5325, 2008.
- [24] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins, "Geographic routing in social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 33, pp. 11623–11628, 2005.
- [25] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: improving geographical prediction with social and spatial proximity," in *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, pp. 61–70, ACM, Raleigh, NC, USA, April 2010.
- [26] J. Goldenberg and M. Levy, *Distance is not dead: Social interaction and geographical distance in the internet era*, 2009, <https://arxiv.org/abs/0906.3202>.
- [27] A. Reynolds-Feighan and P. McLay, "Accessibility and attractiveness of European airports: A simple small community perspective," *Journal of Air Transport Management*, vol. 12, no. 6, pp. 313–323, 2006.
- [28] A. P. Groh and M. Wich, "A Composite Measure to Determine a Host Country's Attractiveness for Foreign Direct Investment," *SSRN Electronic Journal*.
- [29] C. E. Gearing, W. W. Swart, and T. Var, "Establishing a Measure of Touristic Attractiveness," *Journal of Travel Research*, vol. 12, no. 4, pp. 1–8, 1974.
- [30] J. E. Anderson, "The gravity model," *Annual Review of Economics*, vol. 3, pp. 133–160, 2011.
- [31] J. E. Anderson, "A theoretical foundation for the gravity equation," *American Economic Review*, vol. 69, no. 1, pp. 106–116, 1979.
- [32] K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna, "The international trade network: Weighted network analysis and modelling," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 2, 2008.
- [33] M. Levy, "Scale-free human migration and the geography of social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 21, pp. 4913–4917, 2010.
- [34] W.-S. Jung, F. Wang, and H. E. Stanley, "Gravity model in the Korean highway," *EPL (Europhysics Letters)*, vol. 81, no. 4, Article ID 48005, 6 pages, 2008.
- [35] A. Wahid -Ul- Ashraf, M. Budka, and K. Musial-Gabrys, "Newton's Gravitational Law for Link Prediction in Social Networks," in *Complex Networks & Their Applications VI*, vol. 689 of *Studies in Computational Intelligence*, pp. 93–104, Springer International Publishing, Cham, 2018.
- [36] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.
- [37] T. Chakraborty, A. Dalmia, A. Mukherjee, and N. Ganguly, "Metrics for Community Analysis," *ACM Computing Surveys*, vol. 50, no. 4, pp. 1–37, 2017.

- [38] E. A. Leicht and M. E. J. Newman, "Community structure in directed networks," *Physical Review Letters*, vol. 100, no. 11, Article ID 118703, 2008.
- [39] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 3, Article ID 036104, 19 pages, 2006.
- [40] R. Cazabet, P. Borgnat, and P. Jensen, "Enhancing Space-Aware Community Detection Using Degree Constrained Spatial Null Model," in *Complex Networks VIII*, Springer Proceedings in Complexity, pp. 47–55, Springer International Publishing, Cham, 2017.
- [41] X. Liu, T. Murata, and K. Wakita, *Extending modularity by incorporating distance functions in the null model*, 2012, CoRR, abs/1210.4007.
- [42] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.
- [43] P. Schuetz and A. Caflisch, "Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 4, Article ID 046112, 2008.
- [44] B. Lengyel, A. Varga, B. Sagvari, A. Jakobi, and J. Kertesz, "Geographies of an online social network," *PLoS ONE*, vol. 10, no. 9, 2015.
- [45] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowledge and Information Systems*, vol. 42, no. 1, pp. 181–213, 2015.
- [46] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, "Urban gravity: A model for inter-city telecommunication flows," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 7, 2009.
- [47] D. Balcan, V. Colizza, B. Gonalves, H. Hud, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 51, pp. 21484–21489, 2009.
- [48] P. Kaluza, A. Kolzsch, M. T. Gastner, and B. Blasius, "The complex network of global cargo ship movements," *Journal of the Royal Society Interface*, vol. 7, no. 48, pp. 1093–1103, 2010.
- [49] I. S. Jutla, L. G. Jeub, and P. J. Mucha, *A generalized Louvain method for community detection implemented in MATLAB*, 2011, <http://netwiki.amath.unc.edu/GenLouvain/GenLouvain>.
- [50] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, 2010.
- [51] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [52] J. Xiang and K. Hu, "Limitation of multi-resolution methods in community detection," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 20, pp. 4995–5003, 2012.
- [53] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 1, pp. 36–41, 2006.
- [54] J. Reichardt and S. Bornholdt, "Detecting fuzzy community structures in complex networks with a potts model," *Physical Review Letters*, vol. 93, no. 21, 2004.
- [55] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 1, Article ID 016110, 2006.
- [56] A. Arenas, A. Fernandez, and S. Gomez, "Analysis of the structure of complex networks at different resolution levels," *New Journal of Physics*, vol. 10, Article ID 053039, 2008.
- [57] Q. H. Vuong, "Likelihood ratio tests for model selection and nonnested hypotheses," *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989.
- [58] E. Ravasz and A. Barabasi, "Hierarchical organization in complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 67, no. 2, Article ID 026112, 2003.
- [59] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Pseudo-fractal scale-free web," *Physical Review E*, vol. 65, no. 6, pp. 66–122, 2002.



Hindawi

Submit your manuscripts at
www.hindawi.com

