Applied Network Science

RESEARCH Open Access

Analysis of ownership network of European companies using gravity models

Zsolt Tibor Kosztyán[1*], Ferenc Király[1] and Marcell T. Kurbucz[2,3]

*Correspondence:
kzst@gtk.uni-pannon.hu

[1] Department of Quantitative Methods, University of Pannonia, Egyetem Street 10, Veszprém 8200, Hungary
[2] Department of Computational Sciences, Wigner Research Centre for Physics, 29-33 Konkoly Thege Miklós Street, Budapest 1121, Hungary
[3] Institute of Data Analytics and Information Systems, Corvinus University of Budapest, 8 Fővám Square, Budapest 1093, Hungary

Abstract

Social network analysis is increasingly applied to modeling regional relationships. However, in this scenario, we cannot ignore the geographical economic and technological nature of the relationships. In this study, the tools of social network analysis and the gravity model are combined. Our study is based on the Amadeus database of European organizations, which includes 24 million companies. The ownership of parent subsidiaries was modeled using economic, technological, and geographic factors. Ownership was aggregated to the NUTS 3 regional level, to which average corporate profitability indicators, the GDP per capita characterizing the economic environment, and the number of patents, which is a proxy of the technological environment, were assigned to NUTS 3 regions. The formation of the ownership network between 2010 and 2018 was characterized using this dataset. As the proposed model accurately describes the formation of ownership relationships marked with edges, it is possible to estimate network properties, such as modularity and centrality.

Keywords: Gravity models, Regional analysis, Temporal and spatial networks, Distance-based modules

Introduction

Social network analysis (SNA) is a highly visual technique that can be used to describe the overall structure of various networks, as well as the relationship between their nodes, by transforming the spatial system into quantitative relational data (Ye et al. 2022). Over the past 10–15 years, this approach has become an important methodology in regional science and economic geography (Hui et al. 2020). Among the wide range of applications in these fields, SNA has been successfully employed to study the spatial structure of urban and economic agglomeration (Van Meeteren et al. 2016; Liu et al. 2018; Searle et al. 2018) and to analyze innovation, knowledge (Morrison 2008; Sebestyén and Varga 2013; Dahesh et al. 2020; Abonyi et al. 2020; Czvetkó et al. 2021; Weidenfeld et al. 2021), trade (Bhattacharya et al. 2008; Mao and Cheng 2019), and various tourism networks (Liu et al. 2012a; D'Agata et al. 2013; Asero et al. 2016; Mou et al. 2020; Seok et al. 2021). In contrast to these disciplines, the systematic description, modeling, and analysis of network relationships in international business studies remains in its infancy (Kurt and Kurt 2020).

While power-related connections between corporations play an important role in understanding our global corporate system (Vitali et al. 2011), few papers have investigated such networks. For example, Nakamoto et al. (2019) employed the so-called Orbis database (Dijk 2018) to identify and analyze high-risk intermediate companies used for international profit shifting. Using the same database, Khalife et al. (2021) modeled the ownership network to establish a methodology to extract and analyze meaningful patterns of capitalistic influence from the graph structure. Mizuno et al. (2020) applied a new model on this network to measure a shareholder's power to control corporations. Based on their findings, the landscape of global corporate control appears different if we adequately evaluate indirect influence via dispersed ownership. Finally, Takes et al. (2018) investigated the essential building blocks (multiplex motifs) of this graph to provide a better understanding of multiplex corporate networks.

This paper analyzes the European subset of the Orbis database called Amadeus[1] to provide further insights into the structure of European companies' ownership networks. To this end, we combined the tools of SNA with a gravity model containing different economic, technological, and geographic indicators. Ownership was aggregated to the NUTS 3 regional level, to which average corporate profitability indicators, the GDP per capita characterizing the economic environment, and the number of patents and industrial designs characterizing the technological environment were assigned to NUTS regions. The formation of the ownership network between 2010 and 2018 was then characterized using this dataset. As the proposed model accurately describes the formation of ownership relationships marked with edges, it is possible to estimate network properties, such as modularity and centrality.

The aims of the study are twofold, namely, proposing a new, economic null model and identify and analyze economic-investment communities (EICs). The stated aims (As) of this study are grouped as follows:

I. Propose a new gravity-based economic null model (GEN):

A_1 To improve link prediction with GEN.
A_2 To predict derivative network characteristics such as centralities and modularity values.

II. Identify and analyze companies' ownership structure:

A_3 To identify EICs.
A_4 To analyze the stability of EICs over time.

Incorporating gravity models into null models offers us to predict links more accurately (see A_1). In this way, the formation of the companies' ownership network, as well as its

[1] Source: https://www.bvdinfo.com/en-gb/our-products/data/international/amadeus?gclid=Cj0KCQiA-qGNBhD3ARIsAO_o7ylaUtImuqqfTflrsCmF2pzLuP0VqEYFG6EIyWvWZIbCgn3yjkNALNMaAhy8EALw_wcB, retrieved: 5 May 2022.

properties, such as centralities and modularity values, can also be predicted more accurately (see A_2).

The modularity value is a measure of the structure of networks, which measures the strength of a network's communities organized into modules. The proposed GEN-based modules specify economic-investment communities on the companies' ownership network (see A_3). Within these communities, property relationships—that can be treated as investments—are denser than those between two different communities. In addition, yearly data between 2010 and 2018 provide the opportunity to analyze the stability of EICs (see A_4).

The rest of this paper is organized as follows. "Data and methods" section introduces the data utilized in this study and the applied methodology. "Results" section presents the results of the analysis. "Discussion" section discusses the results, followed by "Summary and conclusion" section, which provides a summary and conclusions. Finally, "Limitations and future works" section highlights the limitations and proposes further research directions.

Data and methods

The study combines data-driven and model-driven approaches. The employed data-driven methods came from network sciences, such as calculating centralities to identify the key regions in investment and community-based modularity detection to identify communities of regions. On the other hand, the applied gravity model is a frequently used economic model, where the rate of flows, such as migration, trade exchanges—or in this case, the number of investments—has to be modeled.

The employed data-driven approach, in contrast to the traditional model-driven approaches, could not be based on a preliminary research model and the associated research hypotheses. However, clearly defined research purposes and the associated research questions are formulated. In addition, the combination of data-driven and model-driven approaches allows scholars to formulate more specific research questions (RQs) as follows:

I. Methodological research questions:

RQ_1 Is it possible to improve link prediction via the proposed GEN null model?
RQ_2 Is it possible to improve the derivative network coefficients, such as centralities and modularity values, by the proposed GEN link prediction?

II. Applications of null models:

RQ_3 Do administrative (such as country) borders affect investments?
RQ_4 How do investments change if distance does not play a role?
RQ_5 What kind of EICs can be identified with the GEN null model? Are they stable in time and space?

RQ_1 and RQ_2 are derived from A_1 and A_2. One of the main goals of this study is to propose a better null model that better predicts the links (i.e., the number of owners) between regions. In a spatiotemporal network, which is the company ownership network, not only the distance but also the economic and technological environment, as well as the financial status of companies, can influence the links between nodes. Therefore, not only the links (see A_1–RQ_1) but also the derived network characteristics, such as centrality and modularity values, can be predicted more accurately (see A_2–RQ_2).

Without underestimating the importance of methodological questions, the interesting questions can be centered around the interpretation of the results (see A_3–A_4 and RQ_3– RQ_5). The establishment of a new subsidiary can be considered an investment in which technology and knowledge transfer also take place.

The configuration model of Newman (2010) shows that if links between nodes are concentrated around geographical locations, then the distance between nodes should be considered in null models (Expert et al. 2011). At the same time, distance dependence alone does not explain why administrative boundaries are returned during a module search (see RQ_3). In that case, we can rightly assume that other economic, technological, or corporate characteristics also influence the decision of investments. Indeed, while in the European Union, the federalist and sovereignist positions fight with each other in almost all areas of decision-making (Saurugger 2018; Heidbreder 2022), an important question may be whether the administrative borders (here primarily the country borders) play a role.

Nevertheless, using the distance-dependent null model in community detection gives us the opportunity to ask questions about what kind of relationships would develop if distance did not play a role (see RQ_4). In addition, by using GEN-based null models, the following question can be answered: how does an EIC change in time and space (see RQ_5). Since EICs show regions where the connections between regions are denser than the gravity model predicts if administrative borders are returned as modules, then this indicates that the financial, economic, and technological differences are still decisive in the unified economic area.

Since the gravity model is a classical economic model that follows a model-driven approach, research hypotheses (RHs) can also be stated.

RH_1 The links within companies' ownership networks can be modeled by the distance between regions, the economic and technological properties of the regions, and the financial status of the companies.

RH_2 The administrative borders play an important role in the formation of EICs, which are stable in space and time.

The RH_1 determines the groups of indicators involved in the gravity models. The applied GEN-model-based community detection specifies EICs, which mainly reflect the country's borders. We also assume that these EICs are stable in space and time (see RH_2).

Data employed

The database was collected and aggregated by employing the freely available database of Eurostat,[2] as well as the databases of Amadeus,[3] and PATSTAT.[4] Although the latter two are commercial databases, they collect freely available data from European companies and patents. Amadeus consists of over 24 million company data from all over Europe. In addition to the headquarters of companies, it contains balance sheet and income statement data, as well as ownership relations among companies. The PATSTAT database consists of filed and accepted patents, industrial designs, and trademarks from around the world, as well as the addresses of the inventor and the exploiter. Most European headquarters are assigned to a NUTS 3 region. In addition, we collected per capita GDP adjusted for purchasing power parity (PPP) and population data for NUTS 3 regions. Note that GDP (PPP) is often used as an indicator that is suitable for international comparisons (see, e.g., Abrham and Vosta 2010); however, within the European Union, there is a relatively moderate difference between nominal GDP and GDP PPP.[5] Two distinct data tables are specified. The first table provides the data of nodes (i.e., data of NUTS 3 regions). All indicators are attached and aggregated to a NUTS 3 region. Data from Amadeus (indicators $m_2 - m_{15}$) are also aggregated to NUTS 3 regions. The mean values of company data are calculated for indicators $m_2 - m_{14}$. The edge dataset connects two regions (i and j), and the distances ($d_{i,j}$) between two regions are collected from the official site of Eurostat.[6]

Table 1 shows the list of applied indicators and the data sources.

Note that in the Eurostat database, there is no information about the NUTS 3 GDP data for Iceland (2 regions), Liechtenstein (1 region), Switzerland (25 regions), and the United Kingdom (179 regions); therefore, we used the GDP per capita values for all countries.[7]

To test RH_1, data on companies' financial status ($m_1 - m_{14}$), as well as regional economic (m_{15}) and technological (m_{16}) indicators and interregional distances ($d_{i,j}$)are collected. These variables were treated as independent variables, while the dependent variable was the number of owners of i region companies in the j region ($a_{i,j}$).

Methods employed

Network representation of ownership

The parent-daughter relationship between firms was characterized by means of a binary adjacency matrix $\mathbf{A}$, whose elements are defined as:

[2] Source: https://ec.europa.eu/eurostat/data/database, retrieved: 5 May 2022.

[3] Source: https://www.bvdinfo.com/en-gb/our-products/data/international/amadeus?gclid=Cj0KCQiA-qGNBhD3ARIsAO_o7ylaUtImuqqfTflrsCmF2pzLuP0VqEYFG6EIyWvWZIbCgn3yjkNALNMaAhy8EALw_wcB, retrieved: 5 May 2022.

[4] Source: https://www.epo.org/searching-for-patents/business/patstat.html, retrieved: 5 May 2022.

[5] Source: https://statisticstimes.com/economy/gdp-nominal-vs-gdp-ppp.php, retrieved: 5 July 2022. Note that no significant changes in coefficients were experienced when we replaced GDP (PPP) with GDP in the gravity model. A similar conclusion can be found, e.g., in Paas et al. (2008), which examined international trade within the European Union using a gravity model.

[6] Source: https://ec.europa.eu/Eurostat, retrieved: 5 May 2022.

[7] Source: https://ec.europa.eu/eurostat/databrowser/view/NAMA_10_PC__custom_1874928/default/table?lang=en, retrieved: 5 May 2022.

Table 1 Applied indicators

v	Indicators*	Description	Data source
Node dataset (NUTS 3 regional data)			
m_1	TA	Total assets	Amadeus
m_2	SR	Solvency ratio (Asset based) (%)	Amadeus
m_3	SH	Shareholders' funds	Amadeus
m_4	RB	ROE using P/L before tax (%)	Amadeus
m_5	RCB	ROCE using P/L before tax (%)	Amadeus
m_6	PM	Profit margin (%)	Amadeus
m_7	PLF	P/L for period	Amadeus
m_8	PLB	P/L before tax	Amadeus
m_9	OR	Operating revenue	Amadeus
m_{10}	FA	Fixed assets	Amadeus
m_{11}	EN	Number of employees	Amadeus
m_{12}	CR	Current ratio	Amadeus
m_{13}	CF	Cash flow	Amadeus
m_{14}	CO	Number of companies	Amadeus
m_{15}	GDP	GDP/ capita in purchasing power priority	Eurostat
m_{16}	PI	Patents	PATSTAT
Edges			
i	FROM	The NUTS 3 ID of parent companies	Amadeus
j	TO	The NUTS 3 ID of daughter companies	Amadeus
$d_{i,j}$	Dist	Distance between regions	Eurostat
$a_{i,j}$	OWN	Number of ownerships	Amadeus

*Remark: The definition of the indicators can be found in the "Appendix"

$$a_{i,j} = \begin{cases} 1 \text{ if the } i\text{-th company owns the } j\text{-th company.} \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

Because of the difficulties of interpreting aggregations, the rate of ownership is not considered. The adjacency matrix $\mathbf{A}$ is further called the *company ownership matrix (COM)*. The database contains the exact geographical location of each company. Moreover, since all our economic and technological indicators were provided at the NUTS 3 level and we also wanted to preserve the anonymity of the companies, we aggregated the data to the NUTS 3 level. However, these data are stored separately to use in link (i.e., the number of ownerships) prediction between NUTS 3 regions. Each settlement was assigned to a NUTS 3 region (county). Companies are assigned to geographic regions by the $\mathbf{A}^{[\text{mo,NUTS 3}]}$ and $\mathbf{A}^{[\text{da,NUTS 3}]}$ incidence matrices, whose elements are defined as:

- $a_{i,j}^{[\text{mo,NUTS 3}]}$ with element one if the headquarters of the i-th mother company is situated in the j-th NUTS 3 geographic region,
- $a_{i,j}^{[\text{da,NUTS 3}]}$ with element one if the i-th daughter is situated in the j-th NUTS 3 geographic region,

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

$$\mathbf{A}^{[\text{NUTS 3}]} = \left(\mathbf{A}^{[\text{da,NUTS 3}]}\right)^T \times \mathbf{A} \times \mathbf{A}^{[\text{mo,NUTS 3}]}, \tag{2}$$

where $\mathbf{A}^{[\text{NUTS 3}]}$ is the *(aggregated) company ownership matrix (ACOM)*. If both the subsidiary (daughter) and the parent company are stated in the same NUTS 3 region, then a self-loop is formated in a NUTS 3 level. $a_{i,j} \in \mathbf{A}^{[\text{NUTS 3}]}$ represents the number of owners between NUTS 3 region i and j.

The advantage of this *company-to-local transformation* is to create the opportunity to analyze connections between regions via yearly cross-sectional analysis.

If we examine several periods, we obtain three-dimensional arrays instead of adjacency matrices, where the third dimension is time (i.e., year). As we address intercounty relations throughout, the NUTS 3 notation is neglected. An adjacent matrix in year t is denoted as $\mathbf{A}_t = \mathbf{A}_t^{[\text{NUTS 3}]}$.

Applied null models

Null models predict connections between nodes. The most widely applied *null model* is the random configuration model specified by Newman and Girvan (2004), which calculates the prediction assuming a random graph conditioned to preserve the degree sequence of the original network:

$$a_{i,j} \sim p_{i,j}^{[NG]} = \frac{k_i^{[out]} k_j^{[in]}}{L}, \tag{3}$$

where $k_i^{[out]} = \sum_j a_{i,j}$, $k_j^{[in]} = \sum_i a_{i,j}$, and $L = \sum_i \sum_j a_{i,j}$. Note that self-loops created during the regional aggregation of the ownership network have to be treated. To this end, Arenas et al. (2008) proposed a multiresolution method called AFG (after the authors, Arenas, Fernandez, and Gomez) by adding r self-loops to each node. This algorithm increases the strength of a node without altering the topological characteristics of the original network, as follows: $\mathbf{A}_r = \mathbf{A} + r\,\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix and r the weight of the self-loops of each node. We used this correction in the case of finding modules; however, this compensation underestimates the self-loops.

The so-called *randomized null model* presented by Eq. (3) is inaccurate in most real-world networks Liu et al. (2012b). Nevertheless, several community-based detection methods, such as modularity detection, are based on this random configuration model (Newman 2010).

One of the main disadvantages of the randomized null model is that it neglects the distance dependency between nodes (i.e., regions). The following null model can be specified by considering distance dependency and the use of the attractiveness or importance of nodes instead of the sum of incoming or outgoing edges (Barthélemy 2011; Expert et al. 2011):

$$a_{i,j} \sim p_{i,j}^{[spat]} = \gamma \left(I_i^{[out]}\right)^{\alpha} \left(I_j^{[in]}\right)^{\beta} f(d_{i,j}), \tag{4}$$

where $I_i^{[out]}$ ($I_j^{[in]}$) denotes the importance (or attractiveness) of nodes. α, β are fitting parameters. Since $\sum_i \sum_j p_{i,j} = \sum_i \sum_j a_{i,j}$, $\gamma = \frac{L}{\sum_i \sum_j \left(I_i^{[out]}\right)^{\alpha} \left(I_j^{[in]} f(d_{i,j})\right)^{\beta}}$. The function

$f(d_{i,j})$ can be directly measured from the data by means of a binning procedure, where the prediction error should be minimized, similar to that used in Expert et al. (2011):

$$f(d) = \frac{\sum_{i,j|d_{i,j}=d} a_{i,j}}{\sum_{i,j|d_{i,j}=d} I_i^{out} I_j^{in}}. \tag{5}$$

Note that Eq. (4) is the generalized version of Eq. (3). Additionally, these null models are identical if $\alpha = \beta = f(d) = 1, \gamma = 1/L$. AFG correction can also be used in distance-dependent predictions; however, if $f(d_{ij}) \neq \infty, \forall d_{i,j} = 0$ then all self-loops can be predicted by Eq. (4). It is important to note that Eq. (4) is already a hybrid null model. Since Eq. (3) predicts links solely by network characteristics, such as incoming and outgoing edges, excluding any other influence indicator, which determines the weights of edges between nodes, Eq. (4) already includes the distance dependency in the model. In addition, regression parameters also allow distinguishing the importance of incoming and outgoing edges, which has already different meanings.

Only one step remains to estimate the probability of connections with a gravity model, where $f(d_{i,j}) = d^{\delta}$. Following the notation of gravity models, instead of I, we denote m as the characteristics of nodes (i.e., regions) (Gadár et al. 2018), such as GDP per capita and population. The null model is generalized as follows:

$$a_{i,j} \sim p_{i,j}^{[grav]} = \gamma d_{i,j}^{\delta} \prod_{\nu=1}^{N} m_{i_\nu}^{\alpha_\nu} m_{j_\nu}^{\beta_\nu}, \tag{6}$$

where N is the number of indicators belonging to the nodes. $\alpha, \beta, \gamma, \delta$ are regression coefficients. Eq. (6) further called this the gravity-based economic null model. If $d_{i,j} \neq 0$ regression parameters can be estimated via the logarithmic version of GEN (see Eq. (6)):

$$\log a_{i,j} \sim \log p_{i,j}^{[grav]} = \log \gamma + \delta \log d_{i,j} + \sum_{\nu=1}^{N} \alpha_\nu \log m_{i_\nu} + \sum_{\nu=1}^{N} \beta_\nu \log m_{j_\nu}. \tag{7}$$

In this study, $\forall m_i > 0$, however, because of the self-loops, $d_{i,i} = 0$. If there is no exact knowledge about the distances, there are two ways to handle self-loops. One way is to add 1 km to every distance. In this way, $\log(d_{i,i} + 1) = 0$, and Eq. (7) can be solved. Nevertheless, Burger et al. (2009) showed that this correction can distort the estimation; therefore, they suggested solving Eq. (6) by Poisson regression instead of solving Eq. (7) directly. At the same time, the geocoded location of the company exists; therefore, in the aggregation, the average distance is used in NUTS 3 regional self-loops instead of using only one correction value. Note that this average distance can be calculated for every pair of regions; however, this correction had no significant effect and was therefore only used in self-loops.

Since Eq. (7) provides a linear regression model, and all assumptions of regression models, such as normality, homogeneity and independence (i.e., there is no multicollinearity), must be satisfied. To test for multicollinearity, we used the variance inflation factor (VIF).

$$VIF_i = \frac{1}{1 - R_i^2} \tag{8}$$

where $VIF_i \in [1, \infty[$ is the variance inflation factor for variable i. R_i^2 is the coefficient of determination of the regression equation $X_i = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_{i-1} X_{i-1} + \alpha_{i+1} X_{i+1} + \cdots + \alpha_n X_n + \epsilon$.

To reduce the multicollinearity, the greatest VIF should be less than 2.5 ($\max_i VIF_i < 2.5$) Johnston et al. (2018).

The proposed gravity-based economic null model (GEN, see Eq. (6)) and its logarithmic version (see Eq. (7)) are purely economic models, and there is no network property involved. At the same time, we assume that GEN provides better link predictions than other null models (see A_1, RQ_1). In addition, via better link prediction, an estimated network can also be predicted where the network properties, such as centralities and modularity values, can be calculated. Better link prediction also provides lower prediction error in derived properties (see A_2, RQ_2). At the same time, it must not be forgotten that the GEN is purely an economic model, which thus models not only the formation of edges but also the formation of centralities and modules.

The goodness of fit of null models is determined by how well edges are estimated. Therefore, if there are variable parameters, the absolute differences between the real and predicted edge values must be minimized. Formally:

$$\min \longleftarrow \epsilon = ||\mathbf{A} - \mathbf{P}||. \tag{9}$$

This section introduces three kinds of null models. Newman and Girvan (2004)'s model considers only network properties during link prediction. While Expert et al. (2011)'s distant dependent null model already considers the spatial dependencies between nodes, it can be considered a hybrid model because spatial and network properties are involved simultaneously. Several other null models can be found in Barthélemy (2011), but to the best of our knowledge, the proposed GEN model is the first, which predicts links based on purely spatial, economic, technological, and corporate financial data but does not employ network-property data.

Note that in the case of GEN, the minimization problem is very similar to the gravity-based economic models. Since in the regression model the square estimation error, while in the case of optimizing null models the absolute difference between the original and the predicted links should be minimized. This similarity provides for the employment of gravity models for link prediction and via link prediction the prediction of the company ownership network.

Communities

One of the main applications of null models is to detect communities. Classical modularity optimization-based community detection methods utilize $f(C)$ metrics based on the difference between the internal number of edges and their link prediction (Newman and Girvan 2004; Yang and Leskovec 2015).

$$f(C) = (\text{fraction of edges within communities}) - (\text{expected fraction of such edges}). \tag{10}$$

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

$$f(C) = \frac{1}{L}\sum_i \sum_j \left(a_{i,j} - p_{i,j}\right)\delta\left(C_i, C_j\right), \tag{11}$$

where $p_{i,j}$ represents the number of estimated ownership relationships from region i to region j and $\delta\left(C_i, C_j\right)$ is the Kronecker delta function, which is equal to one if the i-th and j-th regions are assigned.

The modularity of the partition C can be calculated as the sum of the modularities of the C_c , $c = 1, \ldots, n_c$ communities:

$$\max \longleftarrow M_c = \frac{1}{L}\sum_{(i,j)\in C_c} \left(a_{i,j} - p_{i,j}\right). \tag{12}$$

The value of the modularity M_c of a cluster C_c 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, the C_c subgraph tends to be a community that exhibits a stronger degree of internal cohesion than the model predicts. When specifying modules, Eq. (12) must be maximized. When using randomized null models, the modules specify communities where connections are stronger between members within a community than between members of two distinct communities (Newman 2010). A number of links between nodes are dependent on the distances on a spatial network (Expert et al. 2011); therefore, modules give a set of nodes that are close together in geographical terms; however, if they give larger regional units, such as countries, then other formation forces can also be guessed. Therefore, it is a question to be answered whether modules provide larger regions (see RQ_3).

The distance-dependent modules already compensate for the effect of spatial distances between regions. Therefore, the modules can be treated as a module without regional distances. In other words, we can analyze what happens if there are no spatial distances between regions (see RQ_4).

In the case of gravity models, modules specify the *area of investments* (Gadár et al. 2018); we further called *economic-investment communities (EICs)*. EICs specify a set of regions where the strength of investments (modeled by a number of ownerships) are denser than the economic, financial, and technological opportunities, as well as the geographical distances predict. If EICs also give back the administrative boundaries, it indicates that the administrative boundaries are the main formation force in investments (see RH_2), which should be considered at the European Union level.

This study proposes a generalization of gravity null models (GEN). This model also highlights which economic and technological indicators influence the formation of investment areas of regions (see RH_1).

Eq. (12) is typically solved via Louvain's algorithm (Blondel et al. 2008); however, to increase the stability of the results, the recent Leiden's algorithm is applied in this study (Traag et al. 2019).

Since a company ownership network (CON) can represent a static network of ownership, if it is important to analyze CON in time, the one way is to specify a multilayer network, where every layer represents a year. Yearly null models deal only with one layer

at once; therefore, all predictions can be performed simultaneously. The other way is to use the dynamic network, where edges between nodes are specified within a time frame. However, this model is better in the case of continuous time intervals. While the multilayer network represents a set of yearly static networks, the existing null models can be extended in an easier way to a multilayer network.

Both algorithms can be generalized to multilayer networks, where layers represent a time slice. Thus, the proposed Gravity null models can be used to predict the links in a multilayer network, and modules can specify the *yearly EICs*.

The whole network formation can be modeled via link prediction; in this way, several network properties, such as centralities, can be estimated. In addition, the formation of these coefficients can be explained, and their changes over time can be predicted.

Multilayer network as a discrete model of a spatial-temporal network

A multilayer network is a pair $\mathcal{M} = (\mathcal{G}, \mathcal{C})$, where $\mathcal{G} = \{G_\alpha = (V_\alpha, E_\alpha, W_\alpha), \alpha \in \{1, .., m\}\}$ is a family of (directed or undirected, weighted graphs (called layers of $\mathcal{M}$), where V_α is the set of vertices (set of nodes), $E_\alpha \subseteq V_\alpha \times V_\alpha$ is the set of edges (links, or arcs), and $W_\alpha : V_\alpha \times V_\alpha \to \mathbb{R}_0^+$ is the weight matrix of edges of graph G_α in layer α and

$$\mathcal{C} = \{E_{\alpha,\beta} \subseteq V_\alpha \times V_\beta, W_{\alpha,\beta} : V_\alpha \times V_\beta \to \mathbb{R}_0^+, \alpha, \beta \in \{1, .., m\}, \alpha \neq \beta\} \tag{13}$$

is the set of interconnections between nodes of different layers $G_\alpha, G_\beta \in \mathcal{M}$ with $\alpha \neq \beta$.

In this study, the set of interconnections is not specified; therefore, it is assumed that $\mathcal{C} = \emptyset$. Note that in the case of spatial and temporal networks, a layer can represent a time slice (i.e., a year), $\alpha = t$. In addition, the regions are time invariant; therefore, $V_t = V, \forall t$. Only the weights of edges may change over time. Thus, the connections between regions can be estimated separately (see Eq. 14) using a yearly gravity model:

$$\log a_{i,j,t} \sim \log p_{i,j,t}^{[grav]} = \log \gamma_t + \delta_t \log d_{i,j} + \sum_{\nu=1}^{N} \alpha_{\nu_t} \log m_{i_{t,\nu}} + \sum_{\nu=1}^{N} \beta_{\nu_t} \log m_{j_{t,\nu}}. \tag{14}$$

Shifts in regression parameters indicate changes in the role of geographical, economic and technological indicators. The analysis of embeddedness using the multilayer version of centralities indicates shifts in role-player regions.

Finally, analyzing the shifts in modules in time and space indicates the changes in EICs, while calculating modules in a multilayer structure provides time-invariant economic-investment communities.

Centralities

Centralities are traditionally used as descriptive network properties in network science to identify key nodes (roleplayer) in a network. However, if not only links but also links, the whole network can be predicted, and the centralities can be calculated for the predicted network. In other words, in this way, the centralities are predicted. This prediction offers scholars to analyze which indicators influence a region to become a roleplayer. For this analysis, centralities should be modeled as much as possible (see A_2).

Since a directed graph is employed to distinguish the mother-daughter relationships of companies, Only the directed versions and generalized versions of centralities are used,

such as in-degree, out-degree, betweenness, in-closeness, out-closeness, authorities, host, and PageRank centralities.

Degree centrality is defined as the number of links incident upon a node (i.e., the number of ties that a node has). In the case of a directed network (where ties have direction), we usually define two separate measures of degree centrality, namely, in-degree and out-degree. In-degree is a count of the number of ties directed to the node, and out-degree is the number of ties that the node directs to others. The degree centrality of a vertex v is defined as:

$$C_D v = \text{deg}(v), \tag{15}$$

$$C_D^+ v = \text{in-deg}(v), \tag{16}$$

$$C_D^- v = \text{out-deg}(v). \tag{17}$$

In a connected graph, the normalized closeness centrality (or closeness) of a node is the average length of the shortest path between the node and all other nodes in the graph. Thus, the more central a node is, the closer it is to all other nodes.

Closeness is defined by Bavelas (1950) as the reciprocal of farness, that is:

$$C_c = \frac{1}{\sum_w d(v,w)}, \tag{18}$$

where $d(v, w)$ is the graph distance between vertices v and w. Distances from or to all other nodes are irrelevant in undirected graphs, whereas they can produce totally different results in directed graphs.

Betweenness centrality (C_B) quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. Vertices that have a high probability of occurring on a randomly chosen shortest path between two randomly chosen vertices have a high betweenness.

PageRank satisfies the following equation:

$$v_i = \alpha \sum_j a_{ji} \frac{v_j}{L(j)} + \frac{1-\alpha}{N}, \tag{19}$$

where

$$L(j) = \sum_i a_{ji} \tag{20}$$

is the number of neighbors of node j. $\alpha \in [0, 1]$, where N is the number of nodes.

Hub centrality (C_H) and authority centrality (C_A) are calculated to obtain the ranking results. The hub value is the centrality of a node in its ability to make a relation with other nodes, while the authority value is the centrality value of a node based on the number of relations to the node.

The Newman and Girvan (2004), Expert et al. (2011) and proposed GEN models predict links and networks; thus, the centralities can be calculated for both the original and predicted networks. The absolute error of the centralities can be calculated as follows:

$$\epsilon_C = \frac{1}{N}\sum_{v} |C(v) - \widehat{C}(v)|, \tag{21}$$

where $C(v)$ is the centrality measure for vertex v, N is the number of nodes, C is the original, and $\widehat{C}$ is the predicted centrality measure.

Do not forget that in the case of low ϵ_C, the GEN-based prediction, which uses purely economic, corporate financial, and technological indicators in the prediction, models centralities indirectly. This model shows which kind of mixture of spatial-economic-financial-technological indicators can increase the role of a region.

Results

Descriptive statistics

During the analysis, we investigated the ownership network of European companies between 2010 and 2018. The Amadeus database[8] contains data from 23,381,325 companies. From these data, we identified 1,872,272 companies as mother companies or subsidiaries within the examined time period. After data cleaning, we obtained 1,620,340 different parent companies and subsidiaries. The investigated subsidiaries and parent companies are related to 1,435 NUTS 3 regions, which form the nodes of our temporal network. The 87,708 identified ownership relations between these companies over the studied time period are indicated by the edges of the network. Note that connections within the same NUTS 3 region resulted in self-loops.

The 2 table contains descriptive statistics of the main economic and technological data for the examined time period.

The profit and loss (P/L) statement shows the mean value in thousand € for years considering all the companies, not aggregated for NUTS 3 regions. P/L before taxes is the mean value of companies by year in thousand €. In the cash flow line, we can see the mean value of cash flows for all companies by year in thousand €. The number of employees shows the mean values of the number of the companies' employees by year.

Except for the last 3 years of patents, all values are increasing over time. The patent information comes from the PATSTAT database; we have access to the Spring 2021 release of PATSTAT. Because the database itself contains only information about applications that have already been published and because the standard publishing time is usually more than 18 months, the number of PATSTAT entries is much lower in 2017 and 2018.

Null models as link prediction

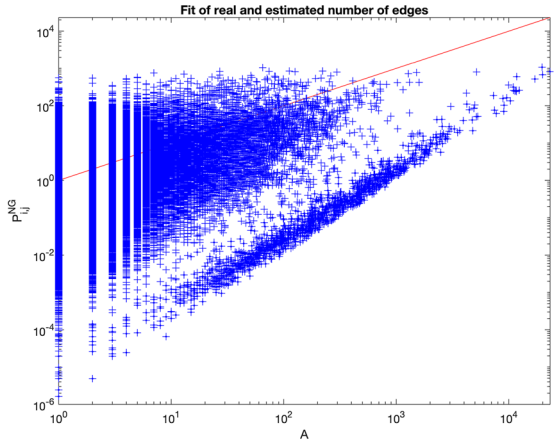
Null models predict links. At the same time, via link prediction, a predicted network is proposed. Figure 1 shows the fits of null models, where **A** contains the adjacency matrix of the original company ownership network (CON); **P** represents the adjacency matrix of predicted networks.

The Newman and Girvan (2004)'a model assumes a random network (see Fig. 1a); however, this model cannot explain loops, and the probability of links between spatial

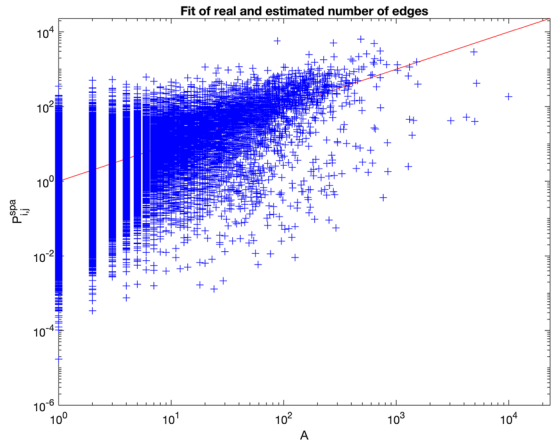
[8] https://www.bvdinfo.com/en-gb/our-products/data/international/amadeus?gclid=Cj0KCQiA-qGNBhD3ARIsAO_o7ylaUtImuqqfTflrsCmF2pzLuP0VqEYFG6EIyWvWZIbCgn3yjkNALNMaAhy8EALw_wcB, retrieved: 5 May 2022.

Table 2 Descriptive statistics of (absolute) indicators

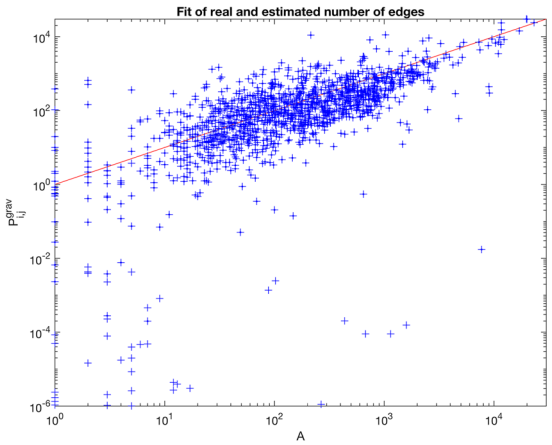
Description	2010	2011	2012	2013	2014	2015	2016	2017	2018	All
Mean value of P/L for period	1083	1122	1167	1262	1353	1473	1648	1874	2008	1.443
Mean value of P/L before taxes	1.353	1.411	1.458	1.493	1.628	1.757	1.919	2.193	2.341	1.728
Mean value of cash flow	1.426	1.477	1.568	1.622	1.725	1.909	2.045	2.154	2.231	1.795
Total number of employees	51.791	56.005	59.190	63.453	66.819	72.696	77.410	80.243	81.375	67.695
Total Number of patents	26.109	27.088	27.868	28.275	28.877	29.065	24.313	9.753	1.385	202.733



(a) Newman and Girvan (2004)'s model: $||\mathbf{A} - \mathbf{P}^{\mathrm{NG}}|| = \epsilon^{NG} = 0.0191$



(b) Expert et al. (2011)'s model: $||\mathbf{A} - \mathbf{P}^{\mathrm{spa}}|| = \epsilon^{spa} = 0.0112$



(c) GEN model: $||\mathbf{A} - \mathbf{P}^{\mathrm{grav}}|| = \epsilon^{grav} = 0.0080$

Fig. 1 Fits of null models (2018)

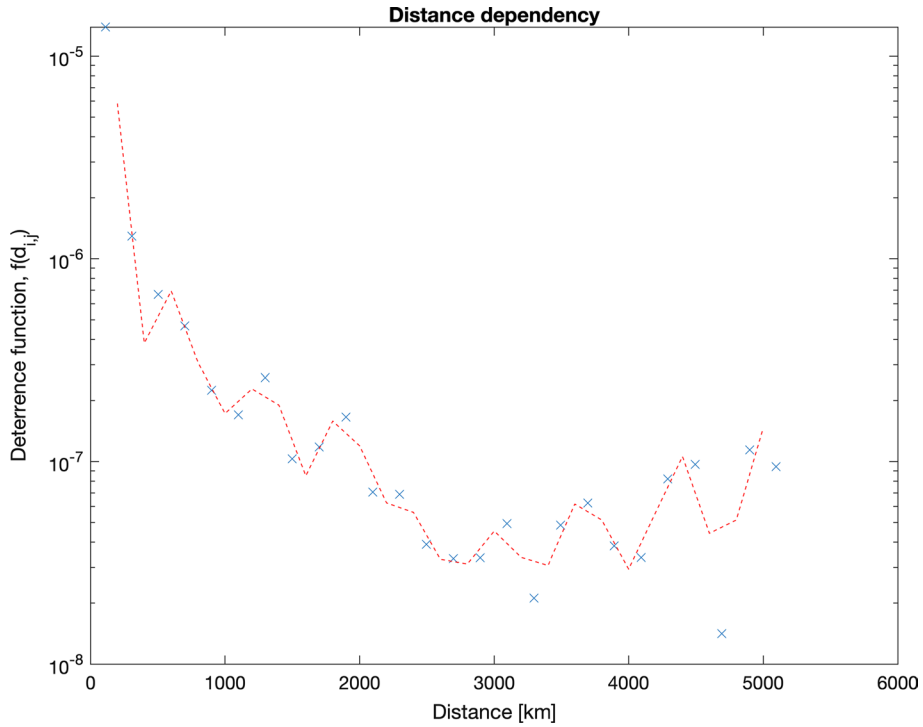


Fig. 2 Distance deterrence function (2018)

nodes (i.e., NUTS 3 regions) is dependent on distance (see Fig. 2). The Expert et al. (2011) formula already considers the nonlinear distance dependency between nodes (see Fig. 1b), and in this way, the group loops disappear. The ($f(d)$) distance dependency is compensated for (see Eq. 4) by the spline function (see Fig. 2); nevertheless, the best fit ($\epsilon = 0.0080$) is produced by the proposed GEN model.

When applying Eq. (6), the indicators with VIF > 2.5 were removed from the model. The adjusted R^2 was decreased only slightly (compare Tables 3 and 6 in "Appendix"); thus, the assumptions of normality, homogeneity and independence were satisfied for the remaining model.

Table 3 shows the results of the proposed GEN model in the years studied. In addition, it summarizes the coefficients and their significance, as well as the absolute error of the fit of the prediction and centralities between the original and predicted network by the gravity model. For example, $\beta_{\mathrm{PI}_{i,2018}} = -0.0233$ means that if GDP decreases 1% from the source (i) region, the number of owners is expected to increase by 0.0233%. Positive (negative) significant coefficients on the source side indicate that an increase in the components may increase (decrease) ownership relations. Similarly, positive (negative) significant coefficients on the host site of the NUTS 3 regions show that an increase in such components may increase (decrease) investments and the development of new corporate sites.

Table 3(a) shows that the applied model (see Eq. 6) is significant, and the adjusted R^2 is slightly greater than 0.4. Among the independent variables examined, the strongly significant coefficients of fixed assets (FA) are positive on the source side (i)

Table 3 Summary of gravity models

	2010	2011	2012	2013	2014	2015	2016	2017	2018
Coefficients	β	β	β	β	β	β	β	β	β
(a) Regression table									
(Intercept)	1.5829***	1.6187***	1.5723***	1.5766***	1.6713***	1.8964***	1.6534***	2.0972***	2.1456***
D_{ij}	− 0.4726***	− 0.4725***	− 0.4727***	− 0.4729***	− 0.4735***	− 0.4717***	− 0.4718***	− 0.4711***	− 0.4725***
SR_i	− 0.1155***	− 0.1193***	− 0.0897***	− 0.1106***	− 0.1283***	− 0.1316***	− 0.1113***	− 0.1105***	− 0.0969***
RB_i	− 0.0717***	− 0.0551***	− 0.0695***	− 0.0603***	− 0.0616***	− 0.0670***	− 0.0910***	− 0.0747***	− 0.0866***
RCB_i	0.0086	0.0079	− 0.0035	− 0.0062	− 0.0214***	0.0047	0.0001	− 0.0020	0.0071
FA_i	0.0206***	0.0183***	0.0159***	0.0163***	0.0128***	0.0210***	0.0177***	0.0149***	0.0111***
CR_i	0.1236***	0.1099***	0.1219***	0.1689***	0.1986***	0.1280***	0.1765***	0.1138***	0.0795***
CO_i	0.2193***	0.2172***	0.2215***	0.2202***	0.2209***	0.2255***	0.2287***	0.2281***	0.2299***
GDP_i	− 0.0023*	− 0.0019*	− 0.0016	− 0.0013	− 0.0023*	− 0.0390***	− 0.0258***	− 0.0285***	− 0.0223**
PI_i	− 0.0042**	− 0.0030*	− 0.0040**	− 0.0044**	− 0.0029*	− 0.0021	− 0.0031*	− 0.0009	0.0026
SR_j	− 0.0054	− 0.0099	0.0043	0.0017	0.0158	0.0594***	0.0758***	0.0260	0.0058
RB_j	− 0.0057	− 0.0112	− 0.0291***	− 0.0351***	− 0.0459***	− 0.0430***	− 0.0548***	− 0.0473***	− 0.0599***
RCB_j	− 0.0152**	− 0.0130**	− 0.0197***	− 0.0213***	− 0.0326***	− 0.0178***	− 0.0160***	− 0.0229***	− 0.0134***
FA_j	− 0.0122***	− 0.0120***	− 0.0114***	− 0.0111***	− 0.0123***	− 0.0086***	− 0.0107***	− 0.0159***	− 0.0199***
CR_j	0.0464 **	0.0537***	0.0514***	0.0726***	0.0811***	0.0100	− 0.0210	− 0.0661***	− 0.0583***
CO_j	0.2218***	0.2205***	0.2216***	0.2223***	0.2244***	0.2304***	0.2326***	0.2327***	0.2344***
GDP_j	− 0.0076***	− 0.0080***	− 0.0085***	− 0.0082***	− 0.0086***	− 0.0139**	− 0.0062	− 0.0163**	− 0.0152*
PI_j	− 0.0105***	− 0.0096***	− 0.0091***	− 0.0101***	− 0.0095***	− 0.0139***	− 0.0139***	− 0.0091***	− 0.0089***
Adj. R^2	0.4034***	0.4029***	0.4029***	0.4032***	0.4041***	0.4026***	0.4032***	0.4019***	0.4030***
ϵ^{grav}	0.0078	0.0073	0.0078	0.0077	0.0076	0.0080	0.0081	0.0080	0.0082

Table 3 (continued)

Errors	2010	2011	2012	2013	2014	2015	2016	2017	2018
(b) Absolute errors of estimated centralities									
$\epsilon_{C_D^+}$	3.0414	5.7021	3.3355	3.5763	4.2312	3.8885	4.6273	6.5212	4.7929
$\epsilon_{C_D^-}$	4.5173	3.6933	5.1694	4.9229	4.9568	5.5349	4.1927	6.3334	5.5961
ϵ_{C_B}	142.7669	183.9388	139.3517	160.2476	161.9900	147.5215	167.2923	185.0560	182.4788
$\epsilon_{C_C^+}$	2.72E−06	3.25E−06	2.46E−06	2.22E−06	1.75E−06	2.77E−06	2.99E−06	1.91E−06	2.43E−06
$\epsilon_{C_C^-}$	4.62E−06	1.97E−06	4.71E−06	4.06E−06	3.60E−06	4.59E−06	2.77E−06	2.26E−06	3.77E−06
ϵ_{C_H}	1.98E−05	1.67E−05	2.06E−05	1.94E−05	1.85E−05	2.13E−05	1.81E−05	1.64E−05	2.06E−05
ϵ_{C_A}	1.42E−05	1.93E−05	1.28E−05	1.38E−05	1.39E−05	1.47E−05	1.45E−05	1.22E−05	1.53E−05
ϵ_{C_P}	2.83E−05	3.45E−05	2.62E−05	3.15E−05	3.18E−05	3.25E−05	2.23E−05	2.05E−05	3.00E−05

Values are significant at: *$p = 0.05$; **$p = 0.01$; ***$p = 0.001$levels

Table 4 Prediction error of centralities

Prediction error of centralities	Random	Spatial	Gravity
In-degree, $\epsilon_{C_D^+}$	33.36681896	32.90683376	4.79292979
Out-degree, $\epsilon_{C_D^-}$	33.40637236	32.94787645	5.59613725
Betweenness, ϵ_{C_B}	170.04457052	169.95647946	182.47883742
In-closeness, $\epsilon_{C_C^+}$	0.00000940	0.00000919	0.00000243
Out-closeness, $\epsilon_{C_C^-}$	0.00000938	0.00000917	0.00000377
Hubs, ϵ_{C_H}	0.00002001	0.00002003	0.00002061
Authority, ϵ_{C_A}	0.00001703	0.00001704	0.00001532
PageRank, ϵ_{C_P}	0.00001782	0.00001781	0.00003000

and negative on the host side (*j*). This result indicates that parent companies typically own higher FA than their subsidiaries.

The current ratio (CR) is a liquidity measure that represents the quotient of current assets and liabilities. The coefficients of this variable are high and significant on the source side, but they are smaller and positive only until 2015 on the host side. The coefficients of the solvency ratio (SR) show the opposite effect to that of CR. This finding suggests that parent companies are typically much more liquid and less solvent than their subsidiaries.

The financial metrics return on capital employed (ROCE), denoted RCB in Table 3(a), can be applied to gauge companies' operational efficiency. The significant coefficients of these variables have a negative sign regardless of side, but this negative effect is greater and much more significant for subsidiaries.

To examine the effect of the economic and technological development of the NUTS 3 regions on the formation of ownership relations, GDP per capita (GDP) and the annual number of patents (PI) are applied. The coefficients of these indicators are negative on both sides; however, they are smaller and less significant on the source side. This observation shows that, in contrast to parent companies, subsidiaries are typically related to NUTS 3 regions that have smaller GDP and fewer patent applications.

Furthermore, we applied the number of companies (CO) within the given NUTS 3 region to control the size of the regions. In the case of this indicator, we obtained coefficients corresponding to our preliminary assumptions since the coefficients of the number of companies are highly positive, regardless of the side. The coefficients of the distance between parent and subsidiary companies are negative and relatively constant over the examined time period.

Finally, we predicted the original network based on the applied model (see Eq. 6); then, we calculated the mean absolute deviation between centrality measures calculated from the original and the predicted network. As shown in Table 3(b), these deviations are relatively stable over the examined time period.

Predicting network properties

Since null models predict links between nodes, they can predict networks as well. Therefore, centralities can also be calculated for the predicted networks. A good fit of the centralities assumes good link predictions. However, the differences between

Table 5 In-degree centralities to top 5 NUTS 3 region (2018)

C_D^-	Original network		GEN model			Expert et al. (2011)'s model		
Rank	NUTS 3	Name	NUTS 3	Name	Rank	NUTS 3	Name	Rank
1	ITC4C	Milan	ITC4C	Milan	(1)	'DK014'	Bornholm	(1291)
2	ITC11	Turin	PL911	Warsaw	(3)	'DK050'	Nordjylland	(589)
3	PL911	Warsaw	ES300	Madrid	(5)	'EE004'	Lääne-Eesti	(634)
4	ES511	Barcelona	ES511	Barcelona	(4)	'EE007'	Kirde-Eesti	(1081)
5	ES300	Madrid	ITI43	Roma	(7)	'EE008'	Lõuna-Eesti	(455)

the real and predicted parameters provide new insight into the structure of corporate networks.

Table 4 shows the mean absolute error of centralities between the original and predicted networks.

Three kinds of networks can be predicted. Newman and Girvan (2004)'s method provides a random network, where the links are predicted via Eq. (3). Thus, no organizing force is assumed. Edges between two regions are estimated as a proportion of incoming and outgoing edges. The spatial network is specified by the Expert et al. (2011) method based on the model of Eq. (4), which compensates for the distance dependency between nodes. Although the prediction errors are lower ($\epsilon^{NG} = 0.0191, \epsilon^{spa} = 0.0112$), the absolute differences between centralities are very similar. The relevant changes are provided by only the gravity model. Since the mean absolute error of link prediction is $\epsilon^{grab} = 0.0080$ lower, the degree centrality error is much lower. Furthermore, the prediction error of betweenness centrality and PageRank centrality is greater.

Table 5 shows an example of the power of modeling centralities. Better fits occur in the case of in-/out-degree centralities. Table 5 shows the top 5 regions with high indegree centralities. In other words, these regions are the most attractive regions to establish a new subsidiary company.

Table 5 shows that the gravity (GEN) model better predicts the ranks of the top 5 regions than the distance-dependent model. These counties are often capitals (e.g., Rome, Warsaw, Madrid) or larger cities (e.g., Milan, Barcelona, Turin). The distance-dependent models, as expected based on the errors (ϵ), estimate the regions poorly. This indicates that in addition to geographical distances, economic, technological and financial indicators should be included in the null model to predict top role players.

Figure 3 shows the in-degree centralities of NUTS 3 regions. Figure 3a shows the original network, and Fig. 3b–d show the predicted networks. All the predicted networks use the same color bars for a clear comparison of the results.

All network predictions indicate low in-degree centrality for Germany, Benelux countries (Belgium, Netherlands, and Luxembourg), and the UK. Furthermore, the original network contains fewer high in-degree centrality nodes than do the predicted networks. The in-degree centralities are overestimated, especially in the case of random networks and spatial network models, for both South and Central European countries. Thus, the amount of investment (characterized by the making of a corporate site) for Southern Europe and Central Europe is much less than that predicted by any of the models. The investment is much less than allowed by the economic

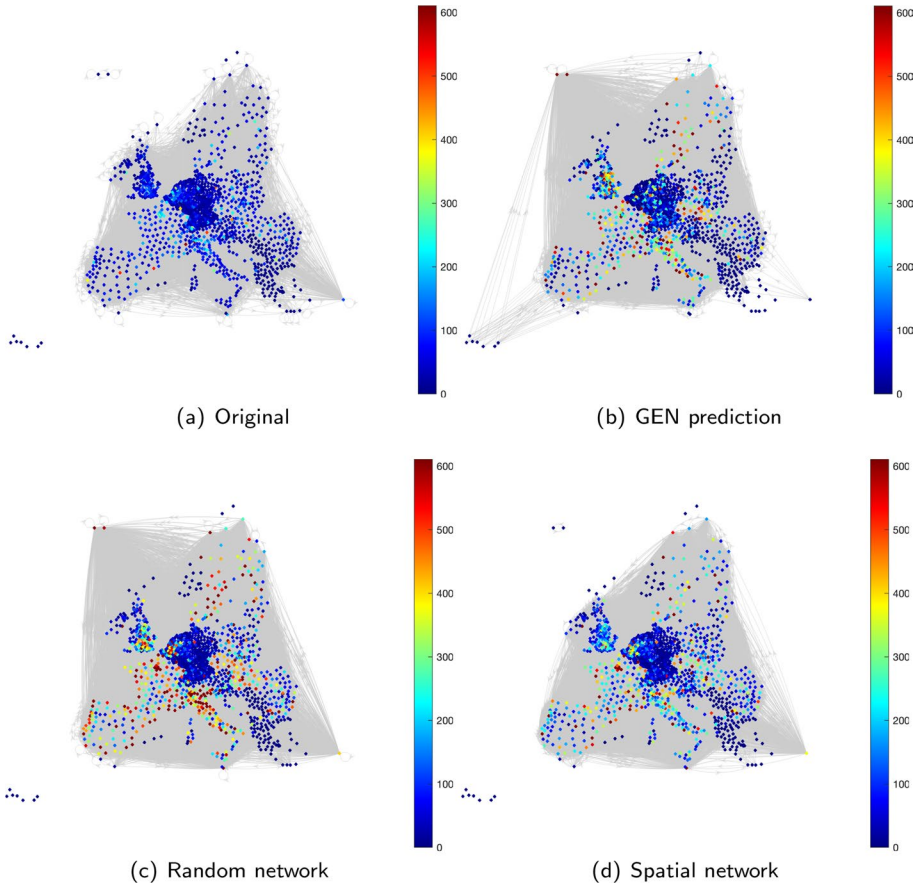


Fig. 3 Predicted network structures and their in-degree centralities (2018)

opportunities, including the spatial, technological and economic distances between regions.

Figure 4 shows the closeness centralities of NUTS 3 regions. Figure 4a shows the original network, and Fig. 3b–d show the predicted networks. In this case, the predicted networks use the same color bars to ensure an easy comparison of the results.

Figure 4 shows that the best predictions are provided by the gravity model. Both the random network and the spatial network models overestimate the in-closeness centralities. Both the original and gravity models indicate important roles for several eastern German, southern England, and northern French regions. In addition, all models and the original network indicate low in-closeness centralities for the Serbian NUTS 3 regions.

Figure 5 shows the in-closeness centralities by year of a multilayer network predicted by the proposed GEN model.

Figure 5 shows the changes in the roles of the NUTS 3 regions by investments. The changes in the yearly in-closeness centralities predict an increased role of Germany and the UK for investments.

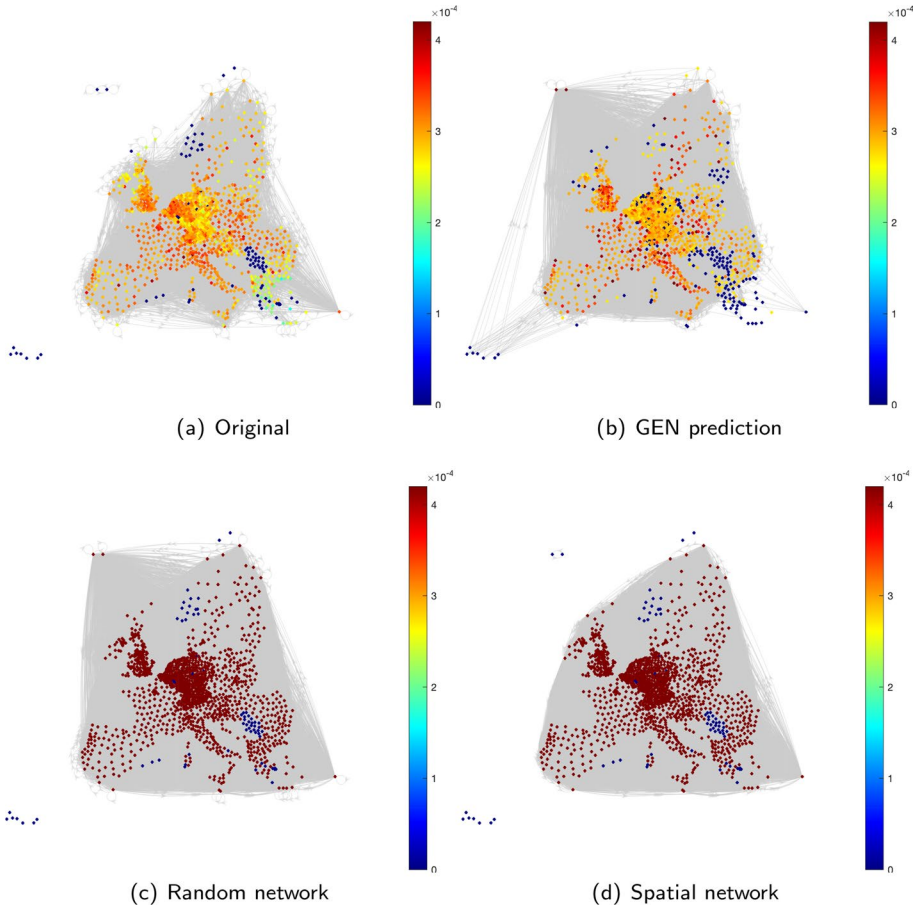


Fig. 4 Predicted network structures and their in-closeness centralities (2018)

Economic communities

In the case of module seeking, the group of nodes that are more connected than the predicted model estimates are identified. In the case of Newman and Girvan (2004)'s model (see Fig. 6a), the modules represent the set of NUTS 3 regions that are more connected to each other within a module than between two different modules. The spatial (see Fig. 6b) and proposed GEN model (see Fig. 6c) already consider spatial and economic properties during the prediction. Therefore, a distance-dependent module indicates communities in which there are stronger connections between regions within a module than are predicted by the distant-dependent model. In the case of seeking modules based on GEN prediction models, the economic communities are specified, where the connections are stronger than would be justified by geographical distance or economic and technological factors.

Figure 6 shows modules by different null models. Modules with lower values (reddish regions) contain more regions, while (blueish) modules with fewer regions have a higher number. Modules marked in black contain only 1-1 regions. Since the employed database had no data for Turkey, NUTS 3 regions within Turkey are shown in white.

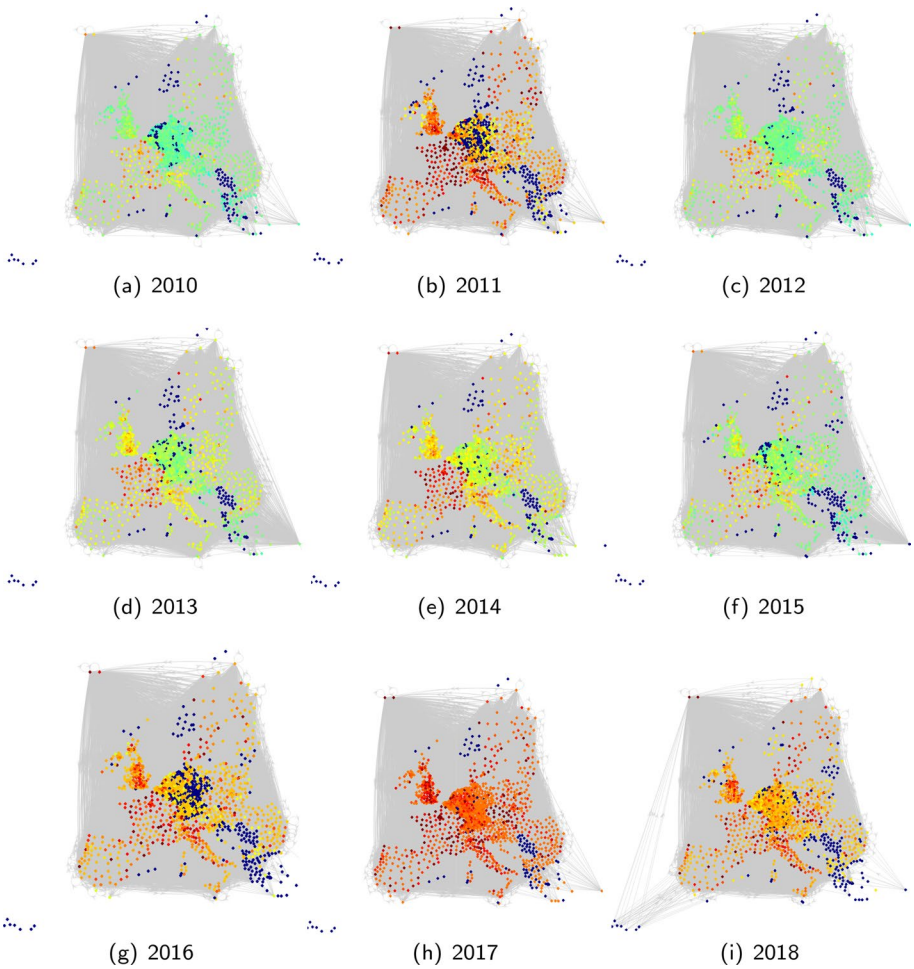
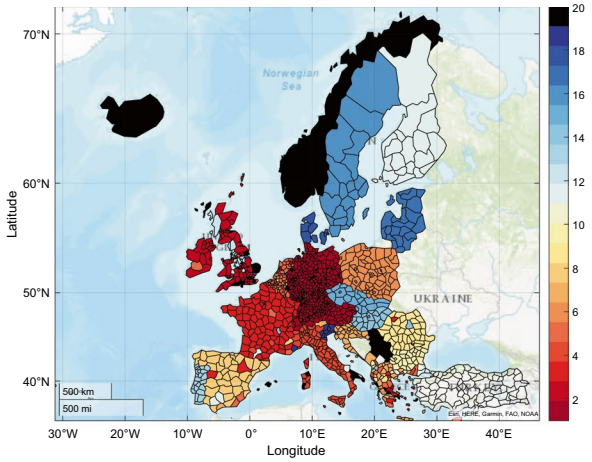


Fig. 5 Centralities of GEN predictions 2010–2018

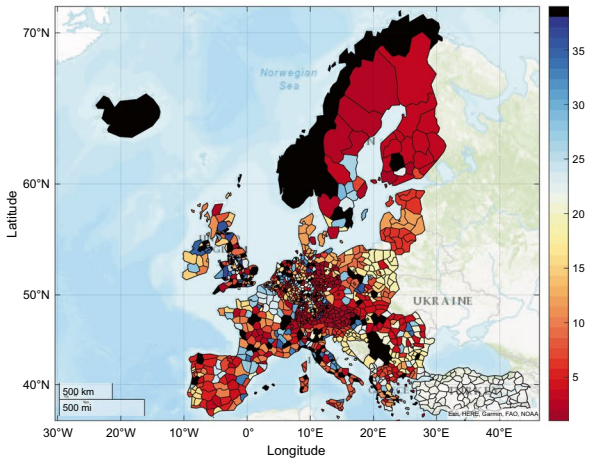
As Gadár et al. (2018) showed, if Newman and Girvan (2004)'s method is employed to search for modules for a spatial network, the modules return the borders of the higher region area (in this case NUTS 1 regions, i.e., countries).

The compensation of distance dependency (see Fig. 6b) changes the shapes of the modules: the locations of modules appear more random. Modules are more separated, and there are more smaller modules. This structure can also be treated as organizing modules without spatial distance.

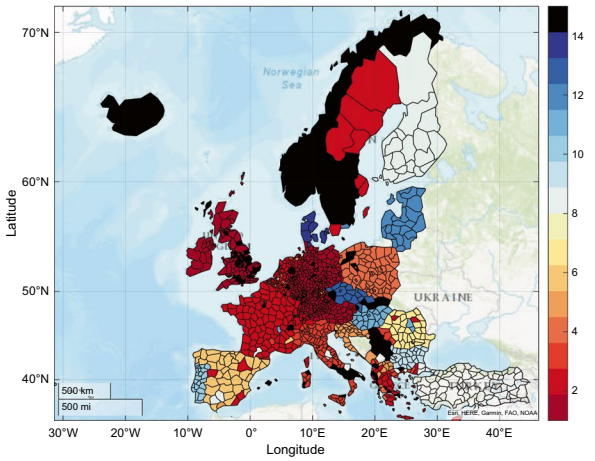
Gadár et al. (2018) showed that gravity modules can specify investment catchment areas that can extend beyond administrative boundaries. Moreover, the fact that the gravity model primarily returned administrative boundaries, the results indicate that the established parent-subsidiary companies still have administrative borders. Economic communities are formed mainly within countries. Furthermore, most regions in Great Britain and Germany and in France and northern Italy form 1-1 economic blocks.



(a) Newman and Girvan (2004)'s modules



(b) Distance-dependent modules



(c) Economic communities (2018)

Fig. 6 Modules of NUTS 3 regions

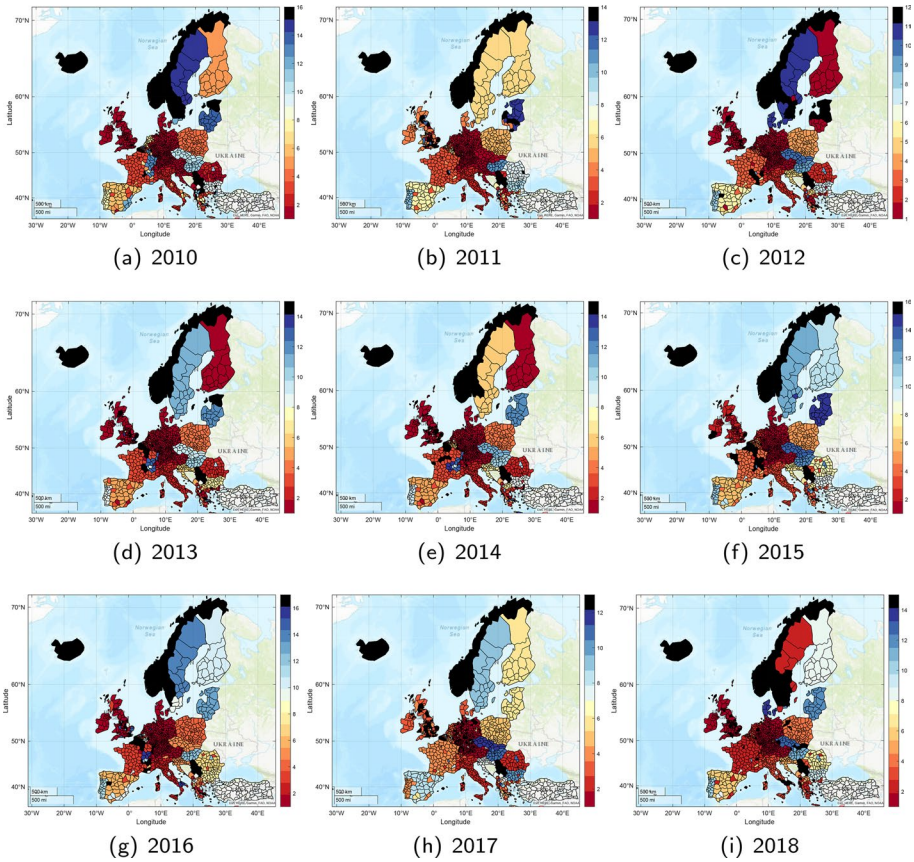


Fig. 7 Layers of economic modules 2010–2018

Figure 7 shows the economic modules in the multilayer network, where the layers specify years.

Figure 7 shows that administrative boundaries, which make it difficult for economic communities to form, can be observed every year. The largest block with the most regions remains the core of the European economy, Germany, Britain, France and northern Italy.

Discussion

The establishment of subsidiaries itself can be considered a kind of investment since the parent company establishes a subsidiary company in another region or country. The company takes the required technology to the new site and creates new jobs. Thus, the study of such networks is important. However, very few databases fully cover both economic and ownership relationships with firms. The paper formulates four research aims (see A_1-A_4), five research questions (see RQ_1-RQ_5) and two research hypotheses (see RH_1-RH_2).

The proposed yearly gravity model has shown that company establishments are driven by technological and economic inequalities (see Table 3 and RH_1): capital flows from economically and technologically more developed regions to less developed ones. Moreover, the integration with network models has shown that this investment is primarily

domestic (see Figs. 6 and 7, and RH_2). Although the European Union continues to push for higher integration, administrative boundaries still have a significant impact on ownership network formation (compare subfigures in Fig. 6, and see RQ_3-RQ_4) and have remained broadly unchanged over the period under review (Fig. 7, and RQ_5).

The proposed (yearly) GEN models best explain the formation (see Fig. 1, and A_1, RQ_1) of the corporate ownership network (see centrality estimates in Table 4, and A_2, RQ_2). GEN also predicts the most attractive regions for investments well; see Table 5. By combining the proposed GEN model with the module search procedures, so-called economic communities can be defined and explained. In this way, the combined models show that several core countries, such as France, Germany, Great Britain and the Benelux countries, play a key role in ownership formation (see Figs. 3, 4, 5 and 6). Of particular interest is the strength of Great Britain's European integration (see, e.g., Fig. 5) of the core countries to create an economic community (see Fig. 6, and A_3–A_4, RQ_3–RQ_5).

Summary and conclusion

This study attempts to combine networks using descriptive and economic factors in explanatory models, providing an opportunity to exploit the strengths of approaches (see A_1-A_2). This paper proposes a generalized yearly gravity-based economic null (GEN) model to predict the spatial network of corporate ownership. Gravity models provide good estimates of the entire network properties, such as centrality, compared to the Newman and Girvan (2004) and Expert et al. (2011) model (see A_1-A_2). The proposed gravity-based modularity provides economic communities (see A_3-A_4). In addition, the proposed yearly model offers an analysis of the changes in economic communities (see RQ_5). The GEN model provides an opportunity to find a better explanation for the formation of networks.

Limitations and future works

In this research, the authors focused on European organizations; however, the corporate, patent, and GDP data are available globally. Nevertheless, NUTS 3 regions can only be interpreted within Europe. This study has shown that Great Britain, especially England, is connected to the European Union by a thousand strands. However, it may be interesting to examine the network structure after Brexit. In the proposed GEN model, we are currently undertaking a yearly estimation of a multilayer network, namely, the ownership network. Furthermore, the gravity model can be extended to estimate the connections of several networks organized into a so-called multiplex network. Finally, an industry-level analysis of ownership structure may provide additional information on the formation of parent-subsidiary relationships.[9]

Appendix

The indicators described in Table 1 are defined as follows:[10]

[9] Abbreviated from the French version: Nomenclature des Unit's Territoriales Statistiques.

[10] Sources: https://investopedia.com, https://help.bvdinfo.com/LearningZone/Products/orbis4.1/en-us/Content/I_Data/UnderstandData.htm, retrieved: 7 July 2022.

m_1 *Total assets (thousand €)*: Total assets = (fixed assets + current assets).

m_2 *Solvency ratio (asset based) (%)*: (shareholder funds/total assets) * 100.

m_3 *Shareholders' funds (thousand €):* total equity (capital + other shareholders funds).

m_4 *ROE using P/L before tax (%):* return on equity (ROE) is a measure of financial performance calculated by dividing net income by shareholders' equity. Because shareholders' equity is equal to a company's assets minus its debt, ROE is considered the return on net assets. (Profit before tax/shareholder funds) * 100.

m_5 *ROCE using P/L before tax (%):* return on capital employed (ROCE) is a financial ratio that can be used to assess a company's profitability and capital efficiency. In other words, this ratio can help to understand how well a company is generating profits from its capital as it is put to use. (Profit before tax + interest paid)/(shareholders' funds + noncurrent liabilities) * 100.

m_6 *Profit margin (%):* (profit before tax/operating revenue) * 100.

m_7 *P/L for period (thousand €):* net income.

m_8 *P/L before tax (thousand €):* operating profit + financial profit.

m_9 *Operating revenue (turnover) (thousand €):* total operating revenues (net sales + other operating revenues + stock variations). The figures do not include VAT. Local differences may occur regarding excises taxes and similar obligatory payments for specific markets of tobacco and alcoholic beverage industries.

m_{10} *Fixed Assets (thousand €):* total amount (after depreciation) of noncurrent assets (intangible assets + tangible assets + other fixed assets).

m_{11} *Number of employees (employee)*: the number of employees.

m_{12} *Current ratio (-):* The current ratio is a liquidity ratio that measures a company's ability to pay short-term obligations or those due within 1 year. Current assets/current liabilities.

m_{13} *Cash Flow (thousand €):* The term cash flow refers to the net amount of cash and cash equivalents being transferred in and out of a company. Cash received represents inflows, while money spent represents outflows. (Profit for period + depreciation).

m_{14} *Number of companies:* The accumulated number of companies in the NUTS3 region.

m_{14} *Number of companies:* The accumulated number of companies in the NUTS3 region.

m_{15} *GDP per capita in purchasing power priority (thousand €):* One popular macroeconomic analysis metric to compare economic productivity and standards of living between countries is PPP. PPP is an economic theory that compares different countries' currencies through a "basket of goods" approach, not to be confused with the Paycheck Protection Program created by the CARES Act.

m_{16} *Patents:* Patents protect technical inventions in all fields of technology. They are valid in individual countries for a specified period. Patents give holders the right to prevent third parties from commercially exploiting their invention. In return, applicants must fully disclose their invention. Patent applications and granted patents are published, which makes them a prime source of technical information.

See Table 6.

Table 6 Summary table of complete gravity models

	2010	2011	2012	2013	2014	2015	2016	2017	2018
Coefficients	β	β	β	β	β	β	β	β	β
(Intercept)	1.5886***	1.5838***	1.5350***	1.5797***	1.5889***	1.5754***	1.2931***	2.0835***	2.1312***
$D_{i,j}$	− 0.4730***	− 0.4731***	− 0.4734***	− 0.4741***	− 0.4759***	− 0.4739***	− 0.4740***	− 0.4729***	− 0.4737***
TA_i	− 0.0510***	− 0.0515***	− 0.0900***	− 0.1057***	− 0.0856***	− 0.1212***	− 0.1363***	− 0.1592***	− 0.1773***
SR_i	− 0.1107***	− 0.1053***	− 0.0792***	− 0.1055***	− 0.1252***	− 0.1277***	− 0.1572***	− 0.1548***	− 0.1538***
SH_i	− 0.0227*	− 0.0228	− 0.0265*	0.0079	− 0.0207	0.0073	0.0301*	0.0265*	0.0678***
RB_i	− 0.1127***	− 0.0825***	− 0.1015***	− 0.0978***	− 0.1223***	− 0.1166***	− 0.1453***	− 0.1253***	− 0.1268***
RCB_i	− 0.0114	− 0.0066	− 0.0237***	− 0.0244***	− 0.0262***	− 0.0176***	− 0.0160**	− 0.0174**	− 0.0151**
PM_i	0.0937***	0.0514***	0.0840***	0.0980***	0.0932***	0.1134***	0.1312***	0.1120***	0.1141***
PLF_i	− 0.0331***	− 0.0412***	− 0.0418***	− 0.0925***	− 0.0591***	− 0.0965***	− 0.0472***	− 0.0460***	− 0.0355**
PLB_i	0.0115	0.0209	0.0081	0.0451***	0.0305*	0.0448**	− 0.0184	0.0426**	0.0090
OR_i	0.0269***	0.0223**	0.0381***	0.0283***	0.0190*	0.0374***	0.0794***	0.0476***	0.0256**
FA_i	0.0838***	0.0879***	0.1118***	0.1173***	0.1129***	0.1173***	0.1068***	0.1019***	0.1046***
EN_i	0.0005	0.0070	0.0137***	0.0240***	0.0449***	0.0420***	0.0117*	0.0020	0.0075
CR_i	0.1097***	0.1041***	0.0853***	0.1207***	0.1483***	0.1034***	0.1724***	0.0986***	0.0644***
CF_i	0.0064	− 0.0014	0.0117*	0.0042	− 0.0141**	0.0039	0.0054	0.0087	0.0050
CO_i	0.2042***	0.2058***	0.2085***	0.2082***	0.2133***	0.2184***	0.2159***	0.2145***	0.2074***
GDP_i	− 0.0014	− 0.0011	− 0.0001	− 0.0004	− 0.0004	− 0.0222***	− 0.0131**	− 0.0157**	0.0214*
PI_i	0.0028	0.0031*	0.0019	0.0007	0.0020	0.0012	0.0026	0.0041**	0.0127***
TA_j	− 0.0264*	− 0.0400**	− 0.0267*	0.0218	0.0543***	0.0577***	0.0302*	0.0242	− 0.0058
SR_j	− 0.0429***	− 0.0754***	− 0.0523***	− 0.0604***	− 0.0200	0.0386*	0.0139	− 0.0296	− 0.0567***
SH_j	0.0777***	0.0961***	0.0744***	0.0740***	0.0673***	0.0372**	0.0629***	0.0579***	0.0693***
RB_j	− 0.0252**	− 0.0240**	− 0.0408***	− 0.0468***	− 0.0475***	− 0.0550***	− 0.0694***	− 0.0826***	− 0.0777***
RCB_j	− 0.0152**	− 0.0130*	− 0.0222***	− 0.0172***	− 0.0138**	0.0044	0.0032	− 0.0069	− 0.0028
PM_j	0.0566***	0.0638***	0.0578***	0.0517***	0.0237*	0.0249*	0.0482***	0.0608***	0.0529***
PLF_j	− 0.0192*	− 0.0374***	− 0.0497***	− 0.0620***	0.0321**	0.0034	0.0287*	0.0208	0.0584***
PLB_j	0.0026	− 0.0072	0.0055	0.0147	− 0.0954***	− 0.0564***	− 0.1063***	− 0.0778***	− 0.1027***
OR_j	0.0240***	0.0407***	0.0403***	0.0098	0.0206*	0.0314***	0.0729***	0.0500***	0.0374***
FA_j	− 0.0369***	− 0.0338***	− 0.0284***	− 0.0414***	− 0.0535***	− 0.0385***	− 0.0376***	− 0.0348***	− 0.0286***
EN_j	− 0.0359***	− 0.0246***	− 0.0262***	− 0.0135**	− 0.0125**	− 0.0194***	− 0.0466***	− 0.0536***	− 0.0483***
CR_j	0.0542***	0.0805***	0.0655***	0.0807***	0.0946***	0.0231	0.0333*	− 0.0411**	− 0.0239
CF_j	− 0.0067	− 0.0104*	− 0.0064	− 0.0199***	− 0.0308***	− 0.0342***	− 0.0296***	− 0.0193***	− 0.0144***
CO_j	0.2152***	0.2139***	0.2168***	0.2182***	0.2218***	0.2280***	0.2222***	0.2222***	0.2202***
GDP_j	− 0.0085***	− 0.0091***	− 0.0089***	− 0.0099***	− 0.0093***	− 0.0163**	− 0.0059	− 0.0162**	− 0.0157*
PI_j	− 0.0067***	− 0.0054***	− 0.0051***	− 0.0073***	− 0.0072***	− 0.0106***	− 0.0083***	− 0.0050**	− 0.0007
Adj. R^2	0.4061***	0.4057***	0.4062***	0.4073***	0.4087***	0.4072***	0.4082***	0.4058***	0.4067***
ϵ^{grav}	0.0078	0.0073	0.0078	0.0077	0.0076	0.0080	0.0081	0.0080	0.0082
$\epsilon_{C_D^+}$	3.0414	5.7021	3.3355	3.5763	4.2312	3.8885	4.6273	6.5212	4.7929
$\epsilon_{C_D^-}$	4.5173	3.6933	5.1694	4.9229	4.9568	5.5349	4.1927	6.3334	5.5961
ϵ_{C_B}	142.7669	183.9388	139.3517	160.2476	161.9900	147.5215	167.2923	185.0560	182.4788
$\epsilon_{C_C^+}$	2.72E−06	3.25E−06	2.46E−06	2.22E−06	1.75E−06	2.77E−06	2.99E−06	1.91E−06	2.43E−06
$\epsilon_{C_C^-}$	4.62E−06	1.97E−06	4.71E−06	4.06E−06	3.60E−06	4.59E−06	2.77E−06	2.26E−06	3.77E−06
ϵ_{C_H}	1.98E−05	1.67E−05	2.06E−05	1.94E−05	1.85E−05	2.13E−05	1.81E−05	1.64E−05	2.06E−05
ϵ_{C_A}	1.42E−05	1.93E−05	1.28E−05	1.38E−05	1.39E−05	1.47E−05	1.45E−05	1.22E−05	1.53E−05
ϵ_{C_P}	2.83E−05	3.45E−05	2.62E−05	3.15E−05	3.18E−05	3.25E−05	2.23E−05	2.05E−05	3.00E−05

Values are significant at: *$p = 0.05$; **$p = 0.01$; ***$p = 0.001$levels

Table 7 shows the correlation table between variables.

Table 7 Correlation table of all variables

	Yij	TAi	SRi	SHi	RBi	RCBi	PMi	PLFi	PLBi	ORi	Fai	Eni	CRi	CFi	Coi	GDPi	Pli	TAj	SRj	SHj	RBj	RCBj	PMj	PLFj	PLBj	ORj	FAj	ENj	CRj	CFj	Coj	GDPj	Plj	Dij
Yij	1,00	-0,02	0,99	-0,13	-0,21	-0,10	0,96	0,97	0,96	0,95	0,89	-0,16	0,89	-0,11	0,72	0,50	0,30	0,00	0,30	-0,12	-0,11	-0,12	0,29	0,30	0,31	0,27	0,31	-0,08	0,27	-0,12	0,24	0,14	-0,02	-0,08
TAi	-0,02	1,00	0,02	0,32	0,15	0,50	0,03	0,00	-0,04	-0,01	0,03	0,49	0,06	-0,01	-0,23	-0,20	-0,01	0,18	0,00	0,14	0,02	0,18	0,01	0,00	-0,03	0,02	-0,01	0,09	-0,02	0,05	-0,07	-0,09	-0,03	0,01
SRi	0,99	0,02	1,00	-0,13	-0,23	-0,08	0,97	0,97	0,94	0,95	0,88	-0,14	0,88	-0,11	0,70	0,48	0,29	0,02	0,30	-0,11	-0,12	-0,12	0,29	0,30	0,31	0,26	0,31	-0,08	0,26	-0,12	0,23	0,13	-0,02	-0,08
SHi	-0,13	0,32	-0,13	1,00	0,45	0,77	-0,05	-0,09	-0,22	-0,03	-0,23	0,24	-0,14	0,48	-0,09	-0,15	-0,07	0,10	-0,07	0,26	0,11	0,27	-0,04	-0,07	-0,10	-0,04	-0,10	0,04	-0,09	0,12	-0,09	-0,12	0,12	0,10
RBi	-0,21	0,15	-0,23	0,45	1,00	0,37	-0,19	-0,20	-0,24	-0,13	-0,23	0,32	0,03	0,29	0,01	0,04	-0,07	0,00	-0,08	0,10	0,19	0,11	-0,07	-0,07	-0,09	-0,06	-0,10	0,03	-0,03	0,07	-0,02	-0,06	0,08	0,06
RCBi	-0,10	0,50	-0,08	0,77	0,37	1,00	0,00	-0,06	-0,24	0,00	-0,25	0,35	-0,10	0,54	-0,11	-0,23	-0,08	0,13	-0,08	0,28	0,12	0,34	-0,04	-0,07	-0,12	-0,04	-0,12	0,06	-0,09	0,15	-0,09	-0,16	0,16	0,13
PMi	0,96	0,03	0,97	-0,05	-0,19	0,00	1,00	0,99	0,92	0,92	0,86	-0,13	0,88	-0,07	0,69	0,46	0,29	0,02	0,29	-0,08	-0,10	-0,08	0,29	0,30	0,30	0,26	0,30	-0,07	0,25	-0,10	0,22	0,12	0,00	-0,07
PLFi	0,97	0,00	0,97	-0,09	-0,20	-0,06	0,99	1,00	0,94	0,92	0,88	-0,14	0,89	-0,11	0,71	0,49	0,29	0,01	0,30	-0,11	-0,11	-0,11	0,29	0,30	0,31	0,26	0,31	-0,08	0,26	-0,12	0,23	0,14	-0,01	-0,08
PLBi	0,96	-0,04	0,94	-0,22	-0,24	-0,24	0,92	0,94	1,00	0,87	0,94	-0,23	0,87	-0,25	0,69	0,50	0,30	-0,01	0,30	-0,15	-0,12	-0,17	0,29	0,31	0,32	0,26	0,32	-0,09	0,27	-0,14	0,25	0,16	-0,06	-0,11
ORi	0,95	-0,01	0,95	-0,03	-0,13	0,00	0,92	0,92	0,87	1,00	0,80	-0,09	0,83	0,02	0,69	0,47	0,27	0,02	0,27	-0,09	-0,09	-0,08	0,27	0,27	0,28	0,25	0,28	-0,07	0,24	-0,11	0,22	0,11	0,02	-0,04
Fai	0,89	0,03	0,88	-0,23	-0,23	-0,25	0,86	0,88	0,94	0,80	1,00	-0,17	0,84	-0,32	0,58	0,47	0,28	0,01	0,29	-0,14	-0,12	-0,16	0,28	0,29	0,30	0,25	0,32	-0,06	0,26	-0,14	0,21	0,16	-0,09	-0,12
Eni	-0,16	0,49	-0,14	0,24	0,32	0,35	-0,13	-0,14	-0,23	-0,09	-0,17	1,00	-0,02	0,24	-0,16	0,01	-0,08	0,08	-0,08	0,06	0,05	0,09	-0,07	-0,08	-0,10	-0,07	-0,07	0,14	-0,05	0,02	-0,11	-0,05	0,05	0,07
CRi	0,89	0,06	0,88	-0,14	0,03	-0,10	0,88	0,89	0,87	0,83	0,84	-0,02	1,00	-0,14	0,69	0,52	0,27	0,00	0,27	-0,13	-0,08	-0,13	0,26	0,28	0,28	0,24	0,29	-0,06	0,27	-0,12	0,23	0,14	-0,03	-0,08
CFi	-0,11	-0,01	-0,11	0,48	0,29	0,54	-0,07	-0,11	-0,25	0,02	-0,32	0,24	-0,14	1,00	0,15	0,18	-0,09	0,03	-0,10	0,13	0,07	0,16	-0,08	-0,09	-0,11	-0,08	-0,12	0,01	-0,10	0,08	-0,07	-0,11	0,29	0,20
Coi	0,72	-0,23	0,70	-0,09	0,01	-0,11	0,69	0,71	0,69	0,69	0,58	-0,16	0,69	0,15	1,00	0,58	0,25	-0,06	0,25	-0,14	-0,07	-0,14	0,24	0,25	0,27	0,22	0,24	-0,11	0,24	-0,10	0,29	0,15	0,08	-0,03
GDPi	0,50	-0,20	0,48	-0,15	0,04	-0,23	0,46	0,49	0,50	0,47	0,47	0,01	0,52	0,18	0,58	1,00	0,13	-0,07	0,13	-0,16	-0,09	-0,19	0,12	0,14	0,16	0,09	0,17	-0,05	0,14	-0,13	0,12	0,14	0,06	0,00
Pli	0,30	-0,01	0,29	-0,07	-0,07	-0,08	0,29	0,29	0,30	0,27	0,28	-0,08	0,27	-0,09	0,25	0,13	1,00	0,08	0,99	-0,10	-0,16	-0,05	0,96	0,97	0,96	0,94	0,88	-0,17	0,87	-0,07	0,73	0,51	-0,09	-0,04
TAj	0,00	0,18	0,02	0,10	0,00	0,13	0,02	0,01	-0,01	0,02	0,01	0,08	0,00	0,03	-0,06	-0,07	0,08	1,00	0,11	0,38	0,12	0,52	0,11	0,08	0,08	0,12	0,11	0,52	0,12	0,04	-0,14	-0,16	0,02	-0,01
SRj	0,30	0,00	0,30	-0,07	-0,08	-0,08	0,29	0,30	0,30	0,27	0,29	-0,08	0,27	-0,10	0,25	0,13	0,99	0,11	1,00	-0,10	-0,19	-0,04	0,96	0,97	0,95	0,93	0,88	-0,16	0,86	-0,08	0,70	0,49	-0,09	-0,04
SHj	-0,12	0,14	-0,11	0,26	0,10	0,28	-0,08	-0,11	-0,15	-0,09	-0,14	0,06	-0,13	0,13	-0,14	-0,16	-0,10	0,38	-0,10	1,00	0,43	0,78	-0,02	-0,06	-0,16	0,03	-0,16	0,29	-0,12	0,47	-0,12	-0,15	0,15	0,08
RBj	-0,11	0,02	-0,12	0,11	0,19	0,12	-0,10	-0,11	-0,12	-0,09	-0,12	0,05	-0,08	0,07	-0,07	-0,09	-0,16	0,12	-0,19	0,43	1,00	0,33	-0,14	-0,15	-0,17	-0,07	-0,19	0,29	0,07	0,28	0,03	0,04	0,05	0,07
RCBj	-0,12	0,18	-0,12	0,27	0,11	0,34	-0,08	-0,11	-0,17	-0,08	-0,16	0,09	-0,13	0,16	-0,14	-0,19	-0,05	0,52	-0,04	0,78	0,33	1,00	0,04	-0,02	-0,16	0,07	-0,16	0,39	-0,05	0,51	-0,11	-0,21	0,19	0,10
PMj	0,29	0,01	0,29	-0,04	-0,07	-0,04	0,29	0,29	0,29	0,27	0,28	-0,07	0,26	-0,08	0,24	0,12	0,96	0,11	0,96	-0,02	-0,14	0,04	1,00	0,99	0,92	0,90	0,86	-0,15	0,86	-0,04	0,69	0,47	-0,07	-0,04
PLFj	0,30	0,00	0,30	-0,07	-0,07	-0,07	0,30	0,30	0,31	0,27	0,29	-0,08	0,28	-0,09	0,25	0,14	0,97	0,08	0,97	-0,06	-0,15	-0,02	0,99	1,00	0,94	0,90	0,88	-0,16	0,87	-0,07	0,71	0,50	-0,08	-0,05
PLBj	0,31	-0,03	0,31	-0,10	-0,09	-0,12	0,30	0,31	0,32	0,28	0,30	-0,10	0,28	-0,11	0,27	0,16	0,96	0,08	0,95	-0,16	-0,17	-0,16	0,92	0,94	1,00	0,86	0,92	-0,23	0,85	-0,18	0,71	0,50	-0,12	-0,06
ORj	0,27	0,02	0,26	-0,04	-0,06	-0,04	0,26	0,26	0,26	0,25	0,25	-0,07	0,24	-0,08	0,22	0,09	0,94	0,12	0,93	0,03	-0,07	0,07	0,90	0,90	0,86	1,00	0,78	-0,07	0,79	0,07	0,70	0,47	-0,06	-0,01
FAj	0,31	-0,01	0,31	-0,10	-0,10	-0,12	0,30	0,31	0,32	0,28	0,32	-0,07	0,29	-0,12	0,24	0,17	0,88	0,11	0,88	-0,16	-0,19	-0,16	0,86	0,88	0,92	0,78	1,00	-0,16	0,80	-0,26	0,57	0,47	-0,14	-0,09
ENj	-0,08	0,09	-0,08	0,04	0,03	0,06	-0,07	-0,08	-0,09	-0,07	-0,06	0,14	-0,06	0,01	-0,11	-0,05	-0,17	0,52	-0,16	0,29	0,29	0,39	-0,15	-0,16	-0,23	-0,07	-0,16	1,00	-0,03	0,26	-0,19	-0,02	0,08	0,03
CRj	0,27	-0,02	0,26	-0,09	-0,03	-0,09	0,25	0,26	0,27	0,24	0,26	-0,05	0,27	-0,10	0,24	0,14	0,87	0,12	0,86	-0,12	0,07	-0,05	0,86	0,87	0,85	0,79	0,80	-0,03	1,00	-0,08	0,69	0,52	-0,10	-0,04
CFj	-0,12	0,05	-0,12	0,12	0,07	0,15	-0,10	-0,12	-0,14	-0,11	-0,14	0,02	-0,12	0,08	-0,10	-0,13	-0,07	0,04	-0,08	0,47	0,28	0,51	-0,04	-0,07	-0,18	0,07	-0,26	0,26	-0,08	1,00	0,15	0,19	0,29	0,19
Coj	0,24	-0,07	0,23	-0,09	-0,02	-0,09	0,22	0,23	0,25	0,22	0,21	-0,11	0,23	-0,07	0,29	0,12	0,73	-0,14	0,70	-0,12	0,03	-0,11	0,69	0,71	0,71	0,70	0,57	-0,19	0,69	0,15	1,00	0,58	-0,05	0,03
GDPj	0,14	-0,09	0,13	-0,12	-0,06	-0,16	0,12	0,14	0,16	0,11	0,16	-0,05	0,14	-0,11	0,15	0,14	0,51	-0,16	0,49	-0,15	0,04	-0,21	0,47	0,50	0,50	0,47	0,47	-0,02	0,52	0,19	0,58	1,00	-0,03	0,02
Plj	-0,02	-0,03	-0,02	0,12	0,08	0,16	0,00	-0,01	-0,06	0,02	-0,09	0,05	-0,03	0,29	0,08	0,06	-0,09	0,02	-0,09	0,15	0,05	0,19	-0,07	-0,08	-0,12	-0,06	-0,14	0,08	-0,10	0,29	-0,05	-0,03	1,00	-0,42
Dij	-0,08	0,01	-0,08	0,10	0,06	0,13	-0,07	-0,08	-0,11	-0,04	-0,12	0,07	-0,08	0,20	-0,03	0,00	-0,04	-0,01	-0,04	0,08	0,07	0,10	-0,04	-0,05	-0,06	-0,01	-0,09	0,03	-0,04	0,19	0,03	0,02	-0,42	1,00

Abbreviations

ACOM	Aggregated companies' ownership matrix
COM	Companies' ownership matrix
CON	Companies' ownership network
GDP	Gross Domestic Product
GDP (PPP)	Gross domestic product based on purchasing power parity
GEN	Gravity-based economic null model
EIC	Economic-Investment Community
NUTS	Nomenclature of Territorial Units for Statistics
ROCE	Return on Capital Employed
SNA	Social Network Analysis

Acknowledgements
This work has been implemented by the TKP2020-NKA-10 project with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2020 Thematic Excellence Programme funding scheme and by the Research Centre at Faculty of Business and Economics (No. PE-GTK-GSKK A095000000-1) of the University of Pannonia (Veszprém, Hungary). Prepared with professional support of the Doctoral Student Scholarship Program of the Co-Operative Doctoral Program of the Ministry of Innovation and Technology financed by the National Research, Development and Innovation Fund. The authors would like to thank Tünde Király for her help in collecting data.

Author contributions
ZTK designed the study. FK collected and cleaned the data, and MTK conducted the literature review. All authors were involved in writing the paper. All authors read and approved the final manuscript.

Funding
Open access funding provided by University of Pannonia.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 26 May 2022 Accepted: 15 August 2022
Published online: 29 August 2022

References

Abonyi J, Czvetkó T, Honti GM (2020) Are regions prepared for industry 4.0?: the industry 4.0+ indicator system for assessment. Springer Nature, Cham

Abrham J, Vosta M (2010) Regional differentiation, agglomeration and clusters within the EU

Arenas A, Fernandez A, Gomez S (2008) Analysis of the structure of complex networks at different resolution levels. New J Phys 10(5):053039

Asero V, Gozzo S, Tomaselli V (2016) Building tourism networks through tourist mobility. J Travel Res 55(6):751–763

Barthélemy M (2011) Spatial networks. Phys Rep 499(1–3):1–101

Bavelas A (1950) Communication patterns in task-oriented groups. J Acoust Soc Am 22(6):725–730

Bhattacharya K, Mukherjee G, Saramäki J, Kaski K, Manna SS (2008) The international trade network: weighted network analysis and modelling. J Stat Mech: Theory Exp 2008(2):P02002

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theory Exp 2008(10):P10008

Burger M, van Oort F, Linders G-J (2009) On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. Spat Econ Anal 4(2):167–190

Czvetkó T, Honti G, Abonyi J (2021) Regional development potentials of industry 4.0: open data indicators of the industry 4.0+ model. PLoS ONE 16(4):e0250247

D'Agata R, Gozzo S, Tomaselli V (2013) Network analysis approach to map tourism mobility. Qual Quant 47(6):3167–3184

Dahesh MB, Tabarsa G, Zandieh M, Hamidizadeh M (2020) Reviewing the intellectual structure and evolution of the innovation systems approach: a social network analysis. Technol Soc 63:101399

Dijk B (2018) Source: orbis, bureau van dijk

Expert P, Evans TS, Blondel VD, Lambiotte R (2011) Uncovering space-independent communities in spatial networks. Proc Natl Acad Sci 108(19):7663–7668

Gadár L, Kosztyán ZT, Abonyi J (2018) The settlement structure is reflected in personal investments: distance-dependent network modularity-based measurement of regional attractiveness. Complexity 2018:1–17

Heidbreder EG (2022) Federalism in the European Union. In: Keil S, Kropp S (eds) Emerging federal structures in the post-cold war era. Springer, Cham, pp 277–299

Hui EC, Li X, Chen T, Lang W (2020) Deciphering the spatial structure of china's megacity region: a new bay area-the Guangdong–Hong Kong–Macao greater bay area in the making. Cities 105:102168

Johnston R, Jones K, Manley D (2018) Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Qual Quant 52(4):1957–1976

Khalife S, Read J, Vazirgiannis M (2021) Structure and influence in a global capital–ownership network. Appl Netw Sci 6(1):1–21

Kurt Y, Kurt M (2020) Social network analysis in international business research: an assessment of the current state of play and future research directions. Int Bus Rev 29(2):101633

Liu F, Zhang J, Zhang J, Chen D, Liu Z, Lu S (2012a) Roles and functions of tourism destinations in tourism region of South Anhui: a tourist flow network perspective. Chin Geogr Sci 22(6):755–764

Liu Z, Mu R, Hu S, Li M, Wang L (2018) The method and application of graphic recognition of the social network structure of urban agglomeration. Wirel Pers Commun 103(1):447–480

Liu X, Murata T, Wakita K (2012b) Extending modularity by incorporating distance functions in the null model, pp 1–12. CoRR, arxiv: abs/1210.4007

Mao M, Cheng X (2019) Evolution analysis of foreign trade network structructure based on complex network SNA. In: Proceedings of the 2019 2nd international conference on e-business, information management and computer science, pp 1–5

Mizuno T, Doi S, Kurizaki S (2020) The power of corporate control in the global ownership network. PLoS ONE 15(8):e0237862

Morrison A (2008) Gatekeepers of knowledge within industrial districts: who they are, how they interact. Reg Stud 42(6):817–835

Mou N, Zheng Y, Makkonen T, Yang T, Tang JJ, Song Y (2020) Tourists' digital footprint: the spatial patterns of tourist flows in Qingdao, China. Tour Manag 81:104151

Nakamoto T, Chakraborty A, Ikeda Y (2019) Identification of key companies for international profit shifting in the global ownership network. Appl Netw Sci 4(1):1–26

Newman M (2010) Networks: an introduction. OUP Oxford. Google-Books-ID: q7HVtpYVfC0C

Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2):026113

Paas T, Tafenau E, Scannell NJ (2008) Gravity equation analysis in the context of international trade: model specification implications in the case of the European Union. East Eur Econ 46(5):92–113

Saurugger S (2018) The European Union and federalism: possibilities and limits. In: Forms of Europe. Paris: Economica, pp 173–200

Searle G, Sigler T, Martinus K (2018) Firm evolution and cluster specialization: a social network analysis of resource industry change in two Australian cities. Reg Stud Reg Sci 5(1):369–387

Sebestyén T, Varga A (2013) Research productivity and the quality of interregional knowledge networks. Ann Reg Sci 51(1):155–189

Seok H, Barnett GA, Nam Y (2021) A social network analysis of international tourism flow. Qual Quant 55(2):419–439

Takes FW, Kosters WA, Witte B, Heemskerk EM (2018) Multiplex network motifs as building blocks of corporate networks. Appl Netw Sci 3(1):1–22

Traag VA, Waltman L, van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9(1):5233

Van Meeteren M, Neal Z, Derudder B (2016) Disentangling agglomeration and network externalities: a conceptual typology. Pap Reg Sci 95(1):61–80

Vitali S, Glattfelder JB, Battiston S (2011) The network of global corporate control. PLoS ONE 6(10):e25995

Weidenfeld A, Makkonen T, Clifton N (2021) From interregional knowledge networks to systems. Technol Forecast Soc Change 171:120904
Yang J, Leskovec J (2015) Defining and evaluating network communities based on ground-truth. Knowl Inf Syst 42(1):181–213
Ye M, Mao W, et al (2022) The spatial structure of regional logistics and influencing factors: an empirical analysis based on Sichuan Province, China

Publisher's Note