



Dynamics of collaboration among high-growth firms: results from an agent-based policy simulation

Anna Varga-Csajkás^{1,3} · Tamás Sebestyén^{1,2} · Attila Varga^{1,2}

Received: 6 January 2020 / Accepted: 18 May 2022 / Published online: 23 June 2022
© The Author(s) 2022

Abstract

This study aims to show how supporting new business formation could promote the emergence of new ties in the knowledge network. Stimulating start-ups is conducive to the innovation system and economic development, however, we show how it has additional beneficial effects by the increasing the density of the innovation network. An innovation network is a complex socioeconomic phenomenon, which emerges from the decisions of many heterogeneous agents, that justifies the use of an agent-based model (ABM). We introduce an application of an ABM that is appropriate for simulating network formation among organizations. Agents are placed in the two-dimensional abstract social space where they are moving toward each other to find cooperation partners, following the gravity principle. The attraction is determined by the mass and the distance of agents. These parameters are specified by regression analysis, the result of which shows that the geographical, social, and technological distance has a negative impact on innovation-related cooperation. For the empirical underpinning of the model, we used survey data on the Hungarian high-growth firms' (known as *gazelles*) egocentric network that contains information about innovation-purpose cooperation.

JEL Classification C21 · C63 · D85 · O20 · O30

✉ Anna Varga-Csajkás
csajkas.anna@ktk.pte.hu

¹ MTA-PTE Innovation and Economic Growth Research Group, ELRN, Pécs, Hungary
² Faculty of Business and Economics, Department of Economics and Econometrics, University of Pécs, Pécs, Hungary
³ Faculty of Business and Economics, Business and Economics Excellence Center, University of Pécs, Pécs, Hungary

1 Introduction

Multiple studies have shown an empirical relationship between geographical proximity and cooperation, hence the knowledge transfer among economic actors. Marshall (1890) was the first to highlight the positive external effects of the spatial concentration of economic activity, but the role of geographical proximity is still a widely researched topic (Inoue et al. 2019; Abramo et al. 2020). Following Marshall, Arrow (1962), and Romer (1986) (together, they form the MAR framework) deals with this positive effect among enterprises in the same industry. Porter (1990) also supported the idea that geographical specialization facilitates growth, but unlike MAR, he argued for the importance of local competition. In contrast to Porter and the MAR argument, Jacobs (1969) shed light on the relevance of knowledge transfer between different industries, which is also supported by Glaeser et al. (1992), who found a positive relationship between knowledge spillover across industries and economic growth. As we can see, the role of geographical proximity is a rich research topic, however, other types of proximity could also enhance cooperation and knowledge flow between organizations. According to Boschma (2005), spatial proximity is neither a necessary nor a sufficient condition for inter-organizational learning, and other forms of proximity may function as substitutes or as complements to the geographical one. During the last decades, it has been demonstrated in many empirical studies that other types of proximities are conducive to successful partnerships and knowledge flows (Autant-Bernard et al. 2007; Cantner and Meder 2007; Broekel and Boschma 2012; D'Este et al. 2013; Marrocu et al. 2013; Cassi and Plunket 2015; Caragliu and Nijkamp 2016; Hansen 2015; Usai et al. 2017; Capone and Lazzeretti 2018; Gui et al. 2018; Bagley 2019; Ghinoi et al. 2021). Most of the studies in the proximity literature use econometric analysis for investigating the effects of different proximity dimensions on specific economic or innovation indicators, e.g., innovation collaborations. Balland et al. (2014) emphasize that the relationship between proximity and knowledge network formation should be analyzed dynamically. Accordingly, not only does proximity affects inter-organizational relationships, but also these networks re-affect the proximity dimensions. This feedback mechanism cannot be captured solely by an econometric estimation, but an agent-based simulation is appropriate to model this phenomenon.

This study investigates the network of Hungarian high-growth enterprises that arises from their innovation-related collaborations among each other and with further organizations from the country. We apply a network approach by focusing not on the entities but on their relationships and the characteristic of the network that arises. In the frame of social network analysis, different units can be nodes, such as firms or individuals, and different types of relations can be considered, e.g., similarities, social relations, or interactions (Borgatti and Ofem 2010). In this case, nodes represent organizations (firms or universities), and edges are professional relations. Since we do not know the whole population, an ego network approach is adapted; accordingly, a selected subset of firms and their partners are investigated. During a snowball sampling method, a predefined set of Hungarian

high-tech gazelles were asked to provide the names and characteristics of the organizations with whom they collaborated during their innovation activities. This sampling was repeated by asking the partners of the gazelles to identify their collaborators. Partnership in this context means that the respondent stated that her organization collaborated with the given organization during its innovation activities. These relationships may manifest in the form of information exchange, R&D cooperation, tender cooperation, education for innovation, or other kinds of joint innovation-related activities. We assume that, if two organizations collaborate at a given point in time, there is a relationship between them that results in an undirected and unweighted ego network. We will simulate how this network possibly evolves in time within two different scenarios.

While explaining the emergence of ties, we do not concentrate only on geographical, but also on technological, social, and institutional proximity. Technological or cognitive proximity may facilitate the establishment of relationships as a certain set of common knowledge is required for collaboration (Cohen and Levinthal 1990). However, this set must be different enough to gain some novelty from cooperation (Boschma and Frenken 2010). The significance of social proximity lies within the fact that personal relationships assist the emergence of trust and reduce the risk of opportunistic behavior (Boschma 2005). Institutional proximity means similar rules and norms, that promote cooperation, hence the flow of knowledge (Ponds et al. 2007; Usai et al. 2017).

Although there is a rich body of literature on proximity dimensions and their effect on knowledge networks, most of the papers deal with only a few specific types of formal cooperation. Since EU Framework Programs and co-patents are well-documented forms of knowledge relationships, they can be used effectively in quantitative analysis (Scherngell and Barber 2009; Varga and Sebestyén 2017). In contrast, observing informal cooperation is more difficult and it requires primary data collection (Capone and Lazzarotti 2018). Despite the argument for the necessity of dynamic analysis of proximity and knowledge networks (Balland et al. 2014), the static approach is still dominant. In the last decade, however, there are examples of dynamic analysis as well (Balland 2012; Gui et al. 2018) and some of them use simulation methods for modeling network formation (Sebestyén and Varga 2019). Our paper is rooted in the work of Sebestyén and Varga (2019) who originally developed their model for the European NUTS 2 regions' Framework Program collaboration network. We borrow the mechanism from this model, but it is applied in a different setup: We use this approach in an inter-organizational context, and we initialize and calibrate it with survey data that allows us to investigate a wide range of formal and informal cooperation.

Our results contribute to the quantitative research on high-growth firms by showing how their innovation network evolves along different proximity dimensions. The regression analysis suggests that geographical, social, and technological proximity has a positive impact on innovation-related cooperation, while organizational proximity was not significant in our study. The simulation exercise pointed out that a successful entrepreneurship policy could significantly increase the number of relationships between the organizations. However, this effect is limited along social and technological proximity.

The remaining of the study is structured as follows: In Sect. 2, we introduce the term *gazelles* and give a short literature review of their relation to networks. In Sect. 3, we describe the empirical data, in Sect. 4, we introduce the agent-based model (ABM). In Sect. 5, we estimate a gravity-like regression equation, which provides parameters for the ABM. In Sect. 6, a calibration procedure is used to pin down other parameters of the model. Given the estimated and calibrated model parameters, in Sect. 7, we show a simulation exercise with the purpose of illustrating the capabilities of the model in analyzing the potential effects of policy interventions on network formation. In the end, we draw conclusions and address the limitations of the study.

2 High-growth firms

Birch and Medoff (1994) pointed out that a few rapidly growing firms created the majority of jobs; therefore, they named them *gazelles*. However, most of the firm population is composed of *small mice* and *elephants*. The former ones are small firms that are unable or not willing to grow, while the latter ones are big companies with slower growth. Due to their low number and noteworthy economic significance, gazelles shall be taken into consideration and put under scrutiny. Although there is no generally accepted definition for gazelles, such separations are carried out based on some growth indicators that are related to the number of employees and sales revenues. The factors behind the growth of gazelles have been investigated in multiple studies, and some of them included networking as an explanatory variable.

It is generally assumed that being part of a network is beneficial for small- and medium-sized firms because they are provided with access to knowledge and other resources. This positive relationship can be empirically demonstrated by different measurement methods. Schoonjans et al. (2013) for instance found that participation in formal business networks has a positive effect on the value-added and assets of the firm, however, it triggers no significant increase in employment. Havnes and Senneseth (2001) demonstrated only one positive correlation out of three indicators regarding network membership and growth. The volume of sales and number of employees was not higher in the case of companies with a higher networking index, but their market expansion was higher, contributing to growth in the long term. Zeng et al. (2010) carried out an analysis of the impact of networking on innovation in the case of Chinese small- and medium-sized enterprises. They found that the greatest positive impact on innovation performance was the cooperation between firms, however, collaboration with intermediating and research institutions also has a positive effect.

Lechner and Dowling (2003) carried out a qualitative analysis of the egocentric network of fast-growing enterprises. They evaluated the importance of different kinds of networks on the different stages of the firms' development. They found that each firm establishes its own relational mix that facilitates expansion the most. According to their results, knowledge, technology, and innovation networks are important at all levels of development as opposed to some other types

of networks. All in all, network membership does not contribute to every growth indicator, but there is no doubt that it plays an important role in the development of gazelles.

In the past decade, multiple pieces of research were carried out to unearth the characteristics of the Hungarian gazelles (Papanek 2010; Csapó 2011; Némethné 2010; Békés and Muraközy 2012; Szerb et al. 2017). They support the findings of international studies (Coad et al. 2014) about how difficult is to standardize rapidly growing firms and forecast which enterprises may become gazelles based on the firms' reports as they occur in all sectors and regions (Békés and Muraközy 2012). Moreover, the Hungarian gazelles often lack positive features such as innovation, export-oriented mentality, or better competitiveness that are related to rapid growth, in accordance with the international literature (Szerb et al. 2017).

From the literature, we can see those collaborations, especially participation in knowledge, technology, and innovation networks, that are conducive to the growth of gazelles. However, in this field of research, qualitative studies are dominant, and even among quantitative analysis, a dynamic approach is relatively rare. The current paper adds new insights into the gazelles' innovation-related collaborations by showing how their innovation network may evolve in time along different proximity dimensions. With the help of an agent-based simulation method, the emergent nature of this network can be demonstrated.

3 Empirical data

The data were collected in the frame of the "Examination of the Hungarian gazelle's willingness of cooperation" research program. This research was conducted by the MTA-PTE Innovation and Economic Development Research Group at the University of Pécs complemented by sociologists from the same university. The aim was to shed light on the characteristics of Hungarian high-growth enterprises and to unfold the connection between cooperation in innovation activities and innovation performance. The data collection was performed in three rounds between 2014 and 2016. The Szocio-Gráf Opinion Research Institute carried out the survey in the first round, and the researchers themselves collected the data in the second and third rounds. During the first round, a representative sample of Hungarian gazelles was identified and interviewed, then in the second and third rounds, the aim was to explore the ego network of a subset of these gazelles. The second round identified co-operative partners of the Hungarian high-tech gazelles, and then, in the third run, these partners were asked to specify their partner organizations. Partnership in this context means that the respondent stated that her organization collaborated with the given organization during its innovation activities. Information exchange, R&D cooperation, tender cooperation, education for innovation, or other kinds of activity could be the content of cooperation. The results of the survey were supplemented with other organizational data, which were collected from firm reports (<http://e-beszamolo.im.gov.hu>), and the organizations' websites. A firm is considered a gazelle if it meets the following two conditions:

- The average annualized growth rate of net sales revenues exceeds 20 percent per annum, over a three-year period
- At least five employees in each given year.

This interpretation is based on the Eurostat-OECD Manual on Business Demography Statistics (2007) definition of high-growth enterprises with one difference: The original definition includes 10 or more employees, but according to the researchers who designed the survey, the five employees threshold suited better to Hungarian conditions (Szerb et al. 2017).

As the goal of the survey was to measure the domestic high-growth enterprises, two additional properties were needed to be sampled: Hungarian-based firm with a minimum of 75% Hungarian ownership. In the database provided by Opten Informatics Ltd., 4037 firms met this definition. From this population, 404 firms were sampled during the layered sampling performed according to agglomeration areas. A cluster analysis from the results of this first round can be found in Szerb et al. (2017). In the following steps of the research, the aim was to better know the high-tech gazelles' cooperation behavior, so the sample was reduced according to two aspects. On the one hand, firms were filtered out that did not report any connection to external organizations during their innovation activity, and on the other hand, those firms that did not belong to a high-tech sector.¹ As a result, a sample of 80 high-tech gazelles was generated. In the second round, 55 of the 80 firms finally gave valuable responses. The respondents identified 94 organizations that we call the primary partners. In the third round, these partners were questioned, and 53 of them gave a valid answer. The respondents reported a total of 183 partners, who form the group of secondary partners of the gazelles. Bodor et al. (2019) used these data to analyze the role of social capital in the innovation activity of the Hungarian high-tech gazelles.

As a result of the survey, we got a graph, where nodes are the agents (firms or universities) and the links between them are their innovation-related collaborations. Theoretically, we may get a directed graph with the explicit direction of knowledge flows, but we treat the network of gazelles as an undirected graph. If one party states that there is a relationship between them, then the direction and the strength of the relationship are not interpreted, we only record that the tie is established. As a result, the gazelles' network is demonstrated by a binary symmetric matrix, where the elements represent the existence of the relationships of the organizations.

The set of partners was restricted to Hungarian firms and higher education institutions since the necessary additional information was not available for all types of organizations. In the case of universities, partners were given at different organizational levels (institute, department, faculty, university), which we aggregated to the level of the university, so the data could become comparable. Thus, in the model, an agent is an organization, that can be either a firm or a university. A total of 207 agents remained in the examination, of which 102 form a sparsely connected component, while the other

¹ The "high-tech" term refers to firms whose primary activity is one of the high- and medium-high technology manufacturing or knowledge-intensive service sectors according to Eurostat's classification.

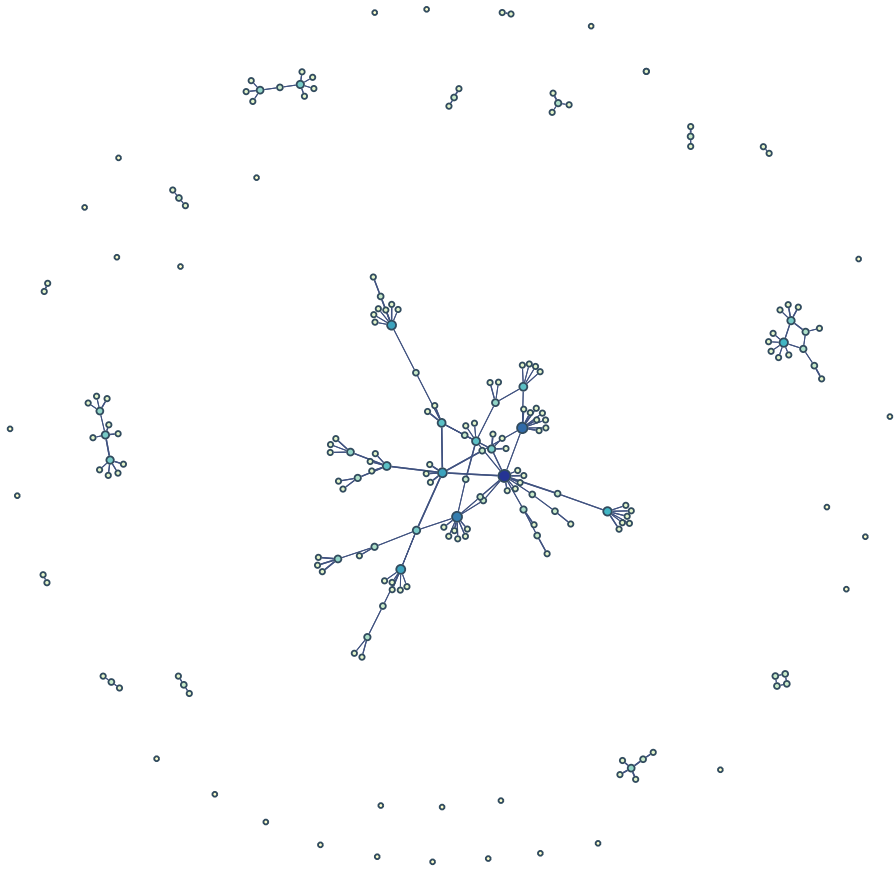


Fig 1 The innovation network of Hungarian high-tech gazelles. Note: The network was visualized in Gephi software with Yifan Hu proportional layout algorithm

organizations are located in smaller separate groups or are isolated, as shown in Fig. 1. For technical reasons, we restricted the sample to the connected part of the network, so we applied the model to 102 agents. It was necessary because the social distance is a basic concept of the model, and we are not able to interpret this distance between unconnected agents.

The results of the survey reveal that in most cases, the content of the relationship was information exchange, there was a smaller number of cases of R&D cooperation or tender cooperation, while only a couple of respondents indicated that innovation-purpose education was the content of their cooperation.

4 The agent-based model

Agent-based modeling is one of the potential techniques for modeling network formation besides random graph models (Erdős and Rényi 1959; Watts and Strogatz 1998; Barabási and Albert 1999) and strategic models of network formation (Jackson 2005). The first one takes a probabilistic view on network formation and is able to explain many phenomena observed in network topology. Strategic models are based on individual incentives for link-formation which interact in shaping the emerging network structures. While taking into account individual choice, these models remain stylized. The main advantage of agent-based models over the former two types is their ability to be empirically calibrated and validated which makes them appropriate for ex-ante policy simulations. The SKIN model (Gilbert et al. 2001; Ahrweiler et al. 2004; Pyka et al. 2007) is a well-known agent-based model that contains network formation. It was the base of many empirically calibrated studies that include the whole innovation system, but the network of actors is of secondary interest (Korber and Paier 2013) like in other agent-based innovation models (Pyka and Saviotti 2002; Heshmati and Lenz-Cesar 2013; Paier et al. 2017). It is argued that ABMs are particularly suitable for modeling complex systems where agents are heterogeneous, and their local (peer-to-peer) interactions build up emergent phenomena at the system level. Ponsiglione et al. (2018) demonstrate how innovation systems can be treated as complex adaptive systems (CAS) and effectively modeled by an ABM. These systems are characterized by several heterogeneous interacting agents that follow individual rules and goals, constituting a self-organizing system where the involved actors adapt their behaviors to the changes in the environment. While this complexity can be captured via sets of differential equations, these remain dominantly unsolvable, providing a comparative advantage for ABMs relying on simulation techniques. Also, their relative richness in detail which is allowed by this simulation approach renders ABMs more suitable for empirical work where models need to closely resemble real-world systems.

The current study builds on the work of Sebestyén and Varga (2019) whose original model was developed for the European NUTS 2 regions' knowledge network. This work, just like the SKIN model, uses an agent-based method to model the dynamics of innovation network. While the SKIN model captures the production of innovation and the arising network is only one aspect of it, Sebestyén and Varga (2019) specifically focuses on link formation.

The major elements of the model are the social space where agents are located and the gravitational force that drives their motion. Moreover, agents have heterogeneous attributes that also affect their behavior thereby the emerging network. Agents are placed in the social space according to their position in the initial innovation network. The distance between them is measured by the length of the shortest path which can be regarded as social distance. These multidimensional network distances can be represented in two dimensions with the help of an appropriate algorithm; therefore, we use multidimensional scaling in order to get these 2D positions. From the initial positions, agents start to move toward each other according to their pairwise attraction to find cooperation partners.

Together with the social distance, the mass (size) and the proximity of actors are assumed to affect their mutual attractiveness which expresses their willingness to cooperate. The attraction force between two agents i and j in period t ($A_{i,j,t}$) is determined by the gravity equation which contains the mass of the agents ($M_{i,t}$ and $M_{j,t}$) and the pairwise distance between them in geographical ($GD_{i,j,t}$), technological ($TP_{i,j,t}$), social ($SD_{i,j,t}$) and institutional ($IP_{i,j,t}$) respects.

$$A_{i,j,t} = f(M_{i,t}, M_{j,t}, GD_{i,j,t}, TP_{i,j,t}, SD_{i,j,t}, IP_{i,j,t}) \quad (1)$$

The attraction force is an abstract variable. During the parameter estimation (see Sect. 5), it is captured with a dichotomous variable, the existence of a relationship between two agents. We interpret mass as the number of edges, connected to the node (node degree). The geographical distance is simply the Euclidian distance between the two headquarters. Technological proximity is captured by four dummy variables based on the economic activities of the agent pair. Social distance shows how far agents are in the 2D social space. Institutional proximity is measured by a dummy variable, which is equal to one if both organizations are universities or both are firms, and it is zero if they belong to different categories. A detailed description of the variables can be found in Sect. 5.

According to the gravity principle, the mass has a positive and the distance has a negative impact on the attraction force between two agents. The only endogenous variable on the right-hand side is the social distance. During the simulation, the attraction force changes only if the positions of agents in the social space change. If the attraction force between two organizations reaches the threshold, they will link up and they will remain connected until the attraction is higher than this threshold.

When agents choose their target position, they consider only a subset of the potential partners because of their cognitive limitations. It is represented by the length of the partner list that expresses the number of potential partners that agents follow when they choose their target positions. The length of the partner list is not firm-specific, but potentially every agent can follow different partners which brings heterogeneity into the model. Besides the attraction force, there is a counterforce that works exactly in the opposite direction. Preferential attachment (Barabási and Albert 1999) suggests that new players tend to connect to other players that already have a greater number of relationships. However, the maintenance of a connection has a cost, that makes an additional relationship less desirable. The counterforce represents this cost of forming and maintaining a relationship that weakens the original attraction force. Its technical role is to ensure the stationarity of the model.

The target position is not achieved automatically but agents start to move toward that point with an agent-specific constant speed. In the model, speed means the distance traveled by the agent per one timestep in the 2D social space. The agent-specific speed could be different according to the size of the agent reflecting the idea that finding cooperating partners may be more or less desirable for smaller and bigger agents. To describe the speed of agents' movement, two parameters are needed: A basic parameter (\bar{S}) which expresses the average speed

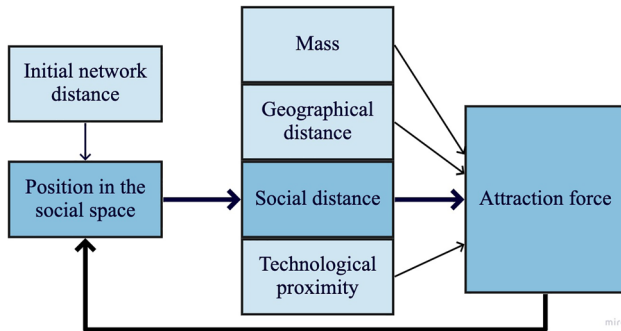


Fig. 2 How the simulation model works

level, and an elasticity parameter (SR), which shows how the speed depends on the agent's size (W_i), which is a transformed version of the mass where the value of the average mass is equal to one. Based on these, the speed can be given by the following formula:

$$S_i = \left(\bar{S} - SR \right) + SR \cdot W_i \quad (2)$$

Finally, we assume that agents are exposed to different link formation costs, also depending on the size of the agents as described in the following formula:

$$BFP_i = \left(\overline{BF} - BR \right) + BR \cdot W_i \quad (3)$$

where \overline{BF} expresses the degree of average cost (counterforce) and BR is the elasticity parameter, which shows how the degree of link formation cost (counterforce) depends on the size of the agent. The economic meaning of this elasticity parameter is that the cost of link formation may be different for smaller and bigger agents.

The parameters of the gravity equation are estimated with a regression introduced in Sect. 5, and the heterogeneity parameters, such as the length of the partner list, and the parameters of speed and counterforce are determined through a calibration process that is described in Sect 6.

The mechanisms of the simulation model are shown in Fig. 2. We can sum up them as follows:

1. Agents' initial positions in the social space are defined according to their observed network distances. These positions determine the social distance in the first timestep, but the network distance does not have any effect later.
2. The innovative mass, the geographical distance, the technological proximity, and the social distance determine the attractiveness values through the gravity equation. Equal counterforces are calibrated so that agents are static in their initial positions.
3. If any of these variables change, agents start to move, which feeds back into the social distance through changing positions.

4. The modified social distances again affect attractiveness, and the system keeps moving for a while.
5. As agents approach their target positions, the attraction force loses its strength and the counterforce starts to dominate, hence the model settles down in a new stationary state.
6. Social distances in the new stationary state are translated into network connections.

5 The gravity equation

Regression analyses were conducted to determine the parameters of the gravity equation. Newton's gravitational law originally describes the attraction force between physical bodies, but the principle of gravitation can also be found in the social and economic processes. It became an analytical tool of economics by the explanation of international trade, as it is a clear idea that the volume of international trade is positively influenced by the economic size of the two countries, while the distance between them has a negative effect (Brun et al. 2005). This analogy can be useful also in other contexts, such as migration (Karemera et al. 2000), tourism (Morley et al. 2014), or research and development collaboration (Frenken et al. 2009; Hoekman et al. 2009; Montobbio and Sterzi 2013). This principle indirectly appears in most of the studies examining the proximity dimensions, but in some articles, a gravitational equation is explicitly specified. For example, Maggioni and Uberti (2009) and Scherngell and Barber (2009) both examined the R&D cooperation among European regions with the help of an econometric model based on a gravity equation. Gravity models in economics are usually estimated in an international context but we reduced our analysis to one country. There are examples of national-level distance-based approaches analyzing collaboration in knowledge creation (Inoue et al. 2019) and gravity models using national data for collaborative knowledge production (Scherngell and Hu 2011). Admittedly, these are done in countries such as Japan or China that are much larger than Hungary. However, we argue that the gravity principle can be valid also within a smaller country, especially in the case of non-geographical proximities.

The dependent variable is the *innovation-related cooperation* (Conn) which has two possible values. It is equal to one if there is a reported relationship between the two organizations and zero if they are not connected directly.

Mass (M): The gravitational force is higher if the mass of the two bodies is larger. In this case, we interpret the number of existing links of a node as a proxy of mass. We suppose that the higher the number of partners, the more attractive will be the agent. It is in line with the *preferential attachment* phenomenon (Barabási and Albert 1999), producing scale-free networks when new nodes tend to link to the more connected ones.

Geographical distance (GD): Geographical distance is captured by the Euclidean distance between the organizations' headquarters.² In the case of high-tech gazelles

² An alternative measure of geographical distance would be travel time (Ponds et al. 2007). In that case, we express distance in terms of the time needed to get from A to B using the public roads. For this method, much more information and computation would be needed, and we thought that it would not improve our results significantly. Besides, in most of the related studies, geographical distance or proximity is measured in kilometers (e.g., Autant-Bernard et al. 2007; Broekel and Boschma 2012).

and their primary partners, the address of the headquarters is known from the responses. The secondary partners have not been questioned, but either the postal address or the website had to be given by the nominator, so their addresses could be specified as well. Using a geocoding program, we identified the latitude and longitude coordinates, from which we calculated Euclidean distance, and this was included in the regression model.

Technological proximity (TP): Technological or cognitive proximity is usually measured by the overlap in patent portfolios (Cantner and Meder 2007; Cassi and Plunket 2015), or with entropy measures (Frenken et al. 2007) but in some studies, this dimension is captured by the similarity of the economic activity of the two agents (Usai et al. 2017). Since there were many organizations in the sample that did not have a patent, we chose the latter solution. We use the codes of economic activities, similarly to the entropy measure of related and unrelated variety in a region (Frenken et al. 2007; Boschma and Iammarino 2009). Technological proximity is expressed by four dummy variables in our study. The value is 1 if the two organizations are in the same category according to the 1,2,3 or 4-digits NACE (Nomenclature statistique des activités économiques dans la Communauté européenne) codes. Accordingly, if all 4 digits in the NACE code of the primary activities of the two organizations are the same (TP1 is equal to 1), then the technological proximity is very strong between them, and it is the weakest if only the first digit is the same (TP4 is equal to 1). The reference group is when they are different even on the one-digit level. This measurement method provides an opportunity to demonstrate the proximity paradox, according to which technological proximity has a positive impact on cooperation but only to a certain extent. After that point, it may hinder innovation (Broekel and Boschma 2012) so the motivation to cooperate.

Social distance (SD): Social distance is measured in accordance with a position in a social network emerging on the basis of earlier innovation cooperation (Autant-Bernard et al. 2007; Balland 2012; Usai et al. 2017). If a tie is established in 2014 or before, the two organizations are considered to have a common history. From this network, we calculated the geodesic distance per pair, which is the length of the shortest path between the two edges. For example, in Fig. 3, between nodes A and E, the shortest path which contains the minimum number of edges goes through node C. It contains two edges, so the geodesic distance between A and E is equal to two. We use multidimensional scaling to convert these geodesic distances to the social space.

Institutional proximity (IP): Institutional proximity has a positive impact on innovation cooperation and knowledge flows, since establishing and managing collaboration are easier given the same institutional environment. An interpretation of the institutional proximity is that organizations with the same status are closer to one another (Ponds et al. 2007; Balland 2012; Cassi and Plunket 2015; Usai et al. 2017). In the current study, belonging to the same organization type is considered institutional proximity and it is measured by a dummy variable. We consider two organization types: firms and universities. The value of the variable is one if both agents are firms or universities and it is zero if they are different.

There are two main approaches in the literature to reveal the relationship between different proximity dimensions and innovation-purpose cooperation. If the number

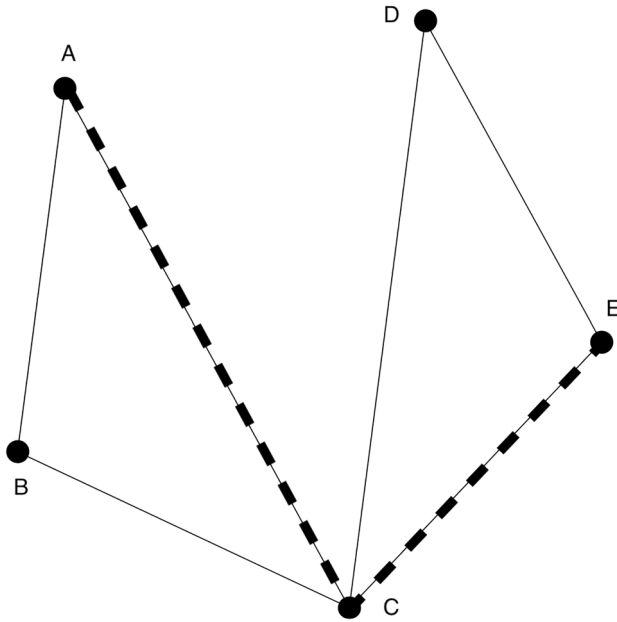


Fig. 3 An example of geodesic distance

of cooperative projects is known, a Poisson or binomial count data model can be used (Hoekman et al. 2009). When, however, the dependent variable cannot be counted, but only the existence of the connection or its intensity is known, then discrete choice models are applied (Autant-Bernard et al. 2007; Paier and Scherngell 2011). In our case, the unit of observation is the organization pair, and the dependent variable is the relationship between the two organizations, so we have a binary choice model. The likelihood of the emergence of a relationship is explained by a binary logit model:

$$\begin{aligned}
 Conn_{i,j} = P_i = \frac{1}{1 + e^{-z_{i,j}}} z_{i,j} = & \beta_0 + \beta_1 \cdot (M_i + M_j) \\
 & + \beta_2 \cdot GD_{i,j} + \beta_3 \cdot TP1_{i,j} + \beta_4 \cdot TP2_{i,j} + \beta_5 \\
 & \cdot TP3_{i,j} + \beta_6 \cdot TP4_{i,j} + \beta_7 \cdot SD_{i,j} + \beta_8 \cdot IP_{i,j}
 \end{aligned} \tag{4}$$

As a starting point, we included only the mass variable then we introduced the different proximity and distance dimensions after one another. Table 1 presents a summary of the regression results.

The values of the coefficients were determined by a maximum likelihood estimate and below them, in brackets, the standard errors are shown. The absolute value of the coefficients is not informative in the logit model, but its sign and significance can be interpreted similarly to the estimated results of the ordinary least squares method. We have chosen model (4) for further investigation. It has the highest R-square (0.32) which is considered moderate explanatory power. The results show

Table 1 Regression results from estimating the gravity model. Dependent variable: innovation-related cooperation between agent-pairs. *** significant at the 0.001 significance level, ** significant at the 0.01 level, * significant at the 0.05 level

	(1)	(2)	(3)	(4)	(5)
Number of observations	10,404	10,404	10,404	10,404	10,404
constant	-5.4136*** (0.135484)	-4.94132*** (0.153972)	-5.23328*** (0.167774)	-2.47979*** (0.210755)	-2.50442*** (0.284908)
M_i	0.262912*** (0.0180924)	0.25913*** (0.0182462)	0.263822*** (0.0186455)	0.231947*** (0.0204122)	0.232636*** (0.0211122)
M_j	0.262912*** (0.0180924)	0.25913*** (0.0182462)	0.263822*** (0.0186455)	0.231947*** (0.0204122)	0.232636*** (0.0211122)
GD_{ij}		-0.40340*** (0.0778016)	-0.3685*** (0.0782538)	-0.42813*** (0.078039)	-0.42820*** (0.0780462)
$TP1_{ij}$			1.54396*** (0.23086)	-0.0546157 (0.261105)	-0.0608071 (0.265484)
$TP2_{ij}$			1.95864*** (0.450445)	1.88331*** (0.497809)	1.8787*** (0.499292)
$TP3_{ij}$			-0.315881 (0.73058)	-1.12480 (0.753739)	-1.11038 (0.762052)
$TP4_{ij}$			0.594376*** (0.205043)	0.588778*** (0.216449)	0.585361*** (0.218058)
SD_{ij}				-1.16957*** (0.088498)	-1.16983*** (0.0885152)
IP_{ij}					0.026437 (0.205724)
Adjusted McFadden R^2	0.151275	0.164662	0.184875	0.321731	0.320787
Log-likelihood	-888.7306	-873.6646	-703.6292	-703.6375	-703.6292

that geographical, social, and technological distance/proximity has an impact on innovation-related cooperation. As expected, we found that the closer they are in the sense of different dimensions, the higher the chance for cooperation between them. Even it is possible that too strong technological proximity hinders innovation, as a result, it reduces the likelihood of cooperation, we found no support for this kind of relationship. The negative value of the coefficient of TP1, which stands for the strongest proximity, may indicate this relation, but it was not significant. Only two of the technological proximity measures were significant, and both have a positive sign meaning that if two organizations are more similar in their economic activity (they are closer in a technological sense), they will cooperate with higher probability. So technological proximity fosters innovation cooperation. Organizational proximity was the only investigated dimension that was not significant in our analysis. The mass variable, measured by the degree of the node, was positive and significant which supports the preferential attachment principle (Barabási and Albert 1999). It

Table 2 The values of the calibrated parameters

Description	Parameter	Range	Optimal value
Common speed parameter	\bar{S}	[0;0.1]	0.076802
Common counterforce parameter	\overline{BF}	[0;1]	0.64178
Length of partner list	AP	[0;101]	5
Speed elasticity parameter	SR	[-1;1]	0.010456
Counterforce elasticity parameter	BR	[-1;1]	-0.084521

leads to the situation when a small number of nodes have a high number of links, and a high number of nodes have a small number of edges.

The constant term is negative and significant. It depicts the mean response value when all predictors are equal to zero, but as in many cases in regression analysis, it has no economic meaning. Theoretically, it is the response value if two organizations do not have any collaboration partners, located exactly in the same place, their economic activity is totally different, and they are located in the same place even in the two-dimensional social space.

It should be noted that the regression was conducted on a selective sample; thus, our econometric results may be biased, therefore, results could not be generalized. Nonetheless, it helps us to determine the parameters of the agent-based model. Multicollinearity could be another limitation of the regression results since it influences the statistical significance of the independent variables. In the case of multicollinearity, the coefficient values could be too sensitive for changes in the model, but except for the technological proximity, there are no big differences in the coefficients across the model specifications.

6 Model calibration

The model contains a series of parameters that need to be numerically calibrated in order to execute empirically valid simulations. These parameters are found partly in the key gravity equation of the model (Eqs. 1 and 4) describing the static attraction between agents, and in Eqs. (2) and (3) further specifying their motion. As shown in the previous section, the gravity equation and its parameters have a solid empirical ground as they are set as a result of the regression estimations summarized in Table 1. However, the parameters describing the motion of agents are not that straightforward to estimate, so we employed a standard calibration procedure in order to ensure that the model is empirically valid.

These “heterogeneity parameters” are listed in Table 2 and are responsible for agent-specific dynamics in the model and are set in a way that a no-intervention simulation with the model replicates observed (past) dynamics as close as possible. In this study, we apply the simulated minimal distance approach (Grazzini and Richiardi 2015), which is one of the most popular calibration methods (Fagiolo et al. 2019). This method sets up an objective function, measuring the distance between

the simulated and observed values of a set of target variables. Then, an optimization process is run in order to find that parameter combination that minimizes the distance, so that the chosen set of (endogenous) model variables are most closely resembled by the simulations (Platt 2020).

This process was started by constructing an initial position for agents: This was done by mapping those relationships that already existed in 2010 (before the survey was conducted), into the two-dimensional social space. Then, the model was simulated a multitude of times with different parameter combinations in order to find that combination which brings the resulting steady state network as close to the observed one as possible. The latter was the network mapped through the survey, and the parameters listed in Table 2 are simultaneously modified to reach the optimal parametrization, while the parameters of the gravity equations are kept fixed at their estimated values.

The objective function of the calibration process measures the distance of the simulated network from the observed one. This function consists of two parts in the present calibration procedure. On the one hand, we require the links in the observed and simulated network to match as closely as possible. This is measured by the following expression, reflecting the sum of differences between the observed and simulated adjacency matrices:

$$F_{link} = \frac{\sum_{i=1}^n \sum_{j=1}^n |c_{ij} - a_{ij}|}{n^2} \quad (5)$$

where c_{ij} is an element of the observed relationship matrix and a_{ij} is an element of the simulated matrix. Given that in a relatively sparse network with many zeros in the adjacency matrix, a relatively good fit can be achieved by simulating an empty network, we also require that the number of simulated links gets as close as possible to the number of observed links. Formally:

$$F_{sum} = \frac{\left| \sum_{i=1}^n \sum_{j=1}^n c_{ij} - \sum_{i=1}^n \sum_{j=1}^n a_{ij} \right|}{\sum_{i=1}^n \sum_{j=1}^n c_{ij}} \quad (6)$$

These two indicators F_{link} and F_{sum} are considered with equal weight during the optimization process, so the final objective function was simply

$$F = F_{link} + F_{sum} \quad (7)$$

Minimizing F , we obtain the parameter combination at which the network structure simulated in the steady state of the model best fits the observed one. Table 2 shows the range at which the optimization took place for each parameter and the optimal values.

The calibration results show that agents take into consideration only the five most attractive potential partners when they choose their target position in the social space. The speed elasticity parameter is positive, which means that bigger agents move faster. The size of agents is based on the node degree, so this result implies that organizations that have already more connections can more easily reach

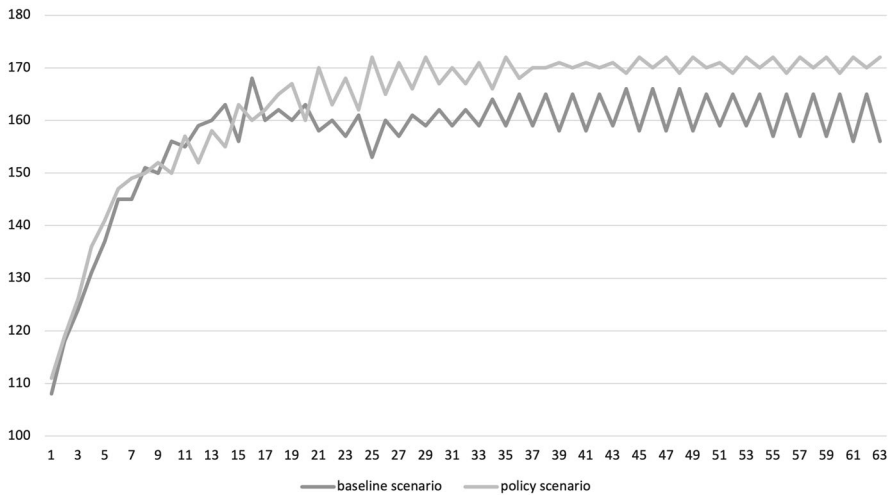


Fig. 4 The number of ties in the network of the Hungarian high-tech gazelles

their target, therefore, connect to the most attractive partners. In contrast, agents' size influences negatively the counterforce, which indicates that link formation is more costly for smaller agents, who have fewer links. Both elasticity parameters strengthen the effect of preferential attachment since those who already have more partners can get new ones relatively easily.

7 The simulation exercise

The aim of this simulation exercise is to illustrate through a simple example how the proposed agent-based model can reflect the potential network effects of an innovation policy that supports new business formation. Stimulating start-ups is conducive for the innovation system and economic development (Gilbert et al. 2004; Fritsch and Mueller 2007). However, we would like to investigate whether it has additional beneficial effects through increasing the density of the innovation network. A denser innovation network allows more opportunity for knowledge spillover, and firms that participate in collaborations can gain advantages in many ways.

The simulation starts from the observed position in 2016, and it predicts the number of ties in the future in a new steady state. As the point of intervention, we chose the capital region of Hungary, Budapest, and Pest County, because here we can find both university and firm agents with diverse economic activities and distinct degrees of connectedness. We ran two scenarios: one baseline scenario, where only the 102 observed organizations are included, and one policy scenario, where we simulate the effects of a successful entrepreneurship policy which results in three new firms entering the architectural and engineering activities sector. These new firms are assumed to be located in Budapest, and each of them has one partner from the same sector and the same region. Thus, they can be interpreted as spin-offs of the latter

Table 3 The characteristics of the innovation network in the baseline and the policy scenario. Note: N is the number of nodes, E is the number of edges between them, E_{\max} is the number of possible edges, D is the density of the network. In brackets, the percentage changes are shown between the beginning and the end of the simulation. The end of the simulation data are the mean values after the 35th period

	At the beginning of the simulation	At the end of the simulation
Baseline scenario	$N = 102$	$N = 102$
	$E = 108$	$E = 161 (+49\%)$
	$E_{\max} = 5151$	$E_{\max} = 5151$
	$D = 0.02097$	$D = 0.03126 (+49\%)$
Policy scenario	$N = 105$	$N = 105$
	$E = 111$	$E = 171 (+55\%)$
	$E_{\max} = 5460$	$E_{\max} = 5460$
	$D = 0.02033$	$D = 0.03132 (+54\%)$

firms. We have chosen architectural and engineering activities as a priority area because this is a well-embedded sector in the region. The NACE code of the primary activity of the spin-offs is 7112 Engineering activities, technical consultancy. There are several firms in our sample with this activity, and other technologically proximate firms are present. From Fig. 4, we can see that the entry of three new firms increases the number of connections in the long run. There is no exact equilibrium point, but a fluctuation is observable near the new stationary position. This phenomenon is a standard characteristic of CASs that often operate out of equilibrium position.

Since new entrants also affect the number of potential connections, an increase in the number of connections cannot be automatically considered as an increase in network density. The network density (D) can be defined as the quotient of the number of actual edges (E) and the number of total possible edges (E_{\max}):

$$D = \frac{E}{E_{\max}} \quad (8)$$

where the number of theoretically possible edges is given by the number of nodes (N) as follows:

$$E_{\max} = \frac{(N \cdot (N - 1))}{2} \quad (9)$$

Table 3 summarizes the numeric results of the two simulation scenarios. There is network dynamics observed already in the baseline scenario. The effects of the policy scenario can thus be interpreted relative to this baseline scenario. The results show that the density already increases quite considerably (by 49%, from 0.02097 to 0.03126 already in the baseline scenario.) This is due to the “built-in” dynamics of the model, coming from the fact that its calibration was done in a way that it captures observed dynamics over a given time period. In contrast, the policy scenario shows a 54% increase in network density which is clearly larger than that in

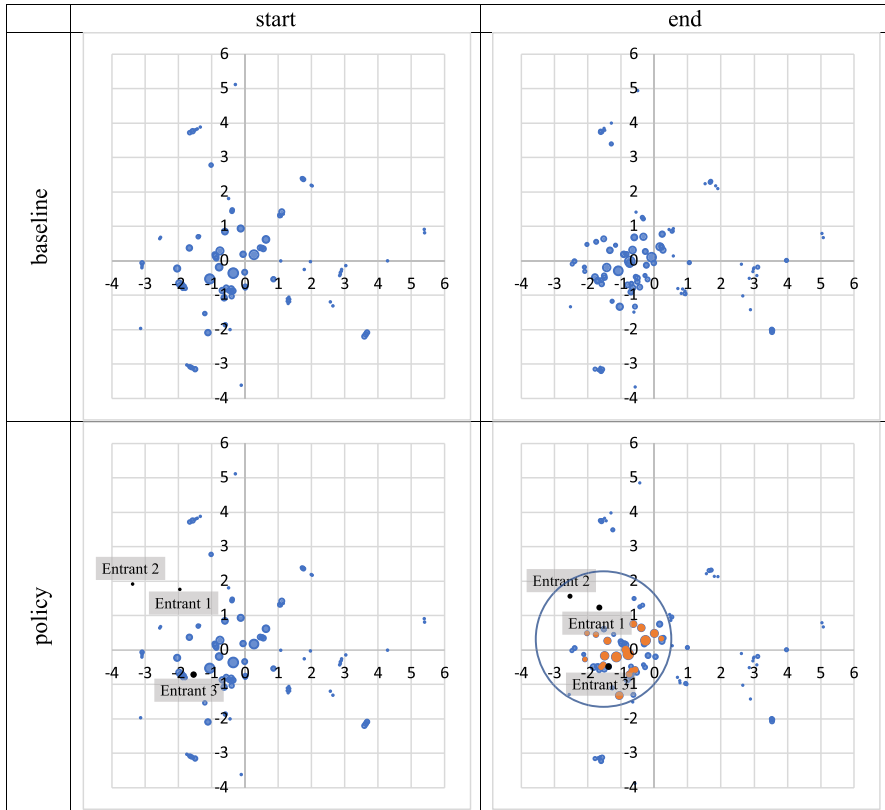


Fig. 5 The position of agents in the social space. The top row of the figure represents the initial and final positions of the baseline, the bottom row represents the initial and final positions of the policy scenario. The size of the dots is proportional to the square root of the number of relationships (degree) of agents. Black dots represent the new entrants in the policy scenario, while the orange dots represent the agents that have more relationships in the final position of the policy scenario compared to the baseline

the baseline scenario. Although there is a base effect resulting from the naturally smaller density in the beginning (three nodes are added to the network but only three more edges are added parallel to that, which lowers density by definition), still the density at the end of the simulation is slightly higher in the policy scenario than in the baseline scenario.

The entry of new agents, representing spin-off firms, directly affects the agents to whom these entrants are connected but it also has an indirect effect through the possibly changing positions of any other agents. The result suggests that during the policy scenario, agents started to move in the social space in a way that brought them closer to each other on average, resulting in shorter social distances. Since the other variables that determine attraction force are constant during the simulation, it results in stronger attraction forces thereby more connections compared to the baseline scenario. The actual number of ties increased more than proportionately than

Table 4 The absolute difference in the number of ties between the baseline and the policy scenarios

Absolute change in the number of ties	Number of agents
-1	2
0	82
1	14
2	2
3	4
4	1

the possible number of ties. Thus, network density increased more in the policy scenario than in the baseline. This illustrative simulation exercise thus shows that supporting business formation in a well-established sector in the region increases the density of the innovation network, which leads to more opportunities for knowledge spillovers in the sector and in the wider innovation ecosystem.

Figure 5 shows the positions of agents in the social space at the initial and final periods of the baseline and policy scenarios. It stems from the two-dimensional mapping algorithm that agents with more connections at the beginning have more central locations. One can notice a trend that agents get closer to each other and to the center of the social space in both the baseline and policy scenarios. This reflects stronger attraction, thus more connections between them over time. The black dots in the policy scenario (bottom panels) represent the new entrants. Although all three new firms have only one partner (connection) according to the initialization, their distance from the center of the social space is different, as they are connected to different other agents. For instance, Entrant 3 has a very central partner with eight connections at the beginning, while the partners of Entrants 1 and 2 have much less connections initially. The orange circles in the final positions represent those agents that have more relationships in the new steady state (final positions) of the policy scenario than in the baseline, so they are the organizations that benefited from the policy measure. Most of these organizations have a central position in the network and already have many ties in the baseline scenario. They are located close to the new entrants which shows that the effect of the policy has a limitation in the social space.

Comparing the degree (number of connections) of individual agents in the two new steady states (final positions), one can see that out of the 105 agents, 21 had more ties in the policy scenario than in the baseline, and only two of them lost any. One agent gained a maximum of four additional relationships from the intervention. Although the majority of the agents are not affected by the policy measure, far more organizations won with it than lost (Table 4).

Comparing the new steady states in the baseline and the policy scenarios, it is visible that the number of ties is 10% higher in the case of the policy measure than without the intervention. Looking at the breakdown of this change, we see that the difference is more pronounced for firms (11%) than for universities (4%). In the sector (according to the 2-digit NACE codes), which is directly affected by the policy

Table 5 The percentage difference of the number of ties between the baseline and the policy scenario

	General	Universities	Firms	Firms in NACE 71	Firms in NACE 7112
Baseline-policy difference	10%	4%	11%	23%	29%
Difference without the entrants	9%	4%	10%	15%	17%

measure, the effect is 23% which is much larger than the overall increase in ties. If we only consider the subset of firms whose primary activity is the same-according to 4-digit NACE codes-as that of the new entrants, the response to the intervention is even stronger (29%). It shows that the technological proximity promotes partnerships in the model. Part of the difference naturally stems from the connections of the new entrants, but the gap is still visible if the ties of these spin-off companies are not counted (second row of Table 5). These results reflect the mechanisms that the model intended to grab: The intervention that supports the creation of spin-offs in a certain sector has the largest effect on the given industry, but it has a spillover effect on other sectors as well. It creates more opportunities for knowledge exchange not only within the sector of intervention but between sectors as well. Although this latter effect is less pronounced, it is present and can serve as the basis of cross-fertilization.

8 Conclusions and the limitations of the study

In the current study, we introduced an application of an agent-based model that is appropriate for modeling the dynamics of network formation based on different proximity dimensions. With the help of unique survey data about the Hungarian gazelles, we have conducted our analysis on a broad range of formal and informal cooperation, which allows a more in-depth understanding of collaboration in innovation. Our results contribute to the studies of gazelles, by showing how the innovation network of gazelles evolves in time. Part of the agent-based simulation parameters has been determined by regression analysis, the result of which shows that the geographical, social, and technological distance has an impact on innovation-related cooperation. As expected, we found that the closer they are in the sense of different dimensions, the higher the chance for cooperation between them. Organizational proximity was the only investigated proximity dimension that was not significant in our analysis. The main added value of the paper compared to earlier empirical studies on proximity dimensions is that we treat the relationship between social proximity and network formation in a dynamic way. As the network evolves in time, agents get in touch with new partners and earlier relationships may dissolve. It feeds back

on social proximity that re-affects the formation of ties in the network. With the help of this model, we demonstrated how policy simulation can be used in the context of network formation. It pointed out that a successful entrepreneurship policy, that induces spin-offs in a given sector, could significantly increase the number of relationships between the organizations. Our findings show that spin-off formation is conducive not only for the sector concerned but it has a spillover effect on other industries. It results in more opportunities for knowledge exchange not only within the sector but between sectors, serving as a basis of cross-fertilization between different technologies. The simulation results also show that the effect of this kind of policy is limited along social and technological proximities of the agents. On the one hand, only those organizations gained additional partners in the policy scenario that were close enough to the new entrants in the social space. On the other hand, the positive effect was stronger for agents that were technologically more proximate to the new entrants.

The limitation of the study is that only cross-sectional data were available therefore we had to construct a hypothetical starting position for the calibration. Besides, we conducted the regression analysis on a selective sample thus, our econometric results are probably biased. With representative sample and panel data, we could gain generalizable results and more relevant policy simulations.

Funding Open access funding provided by University of Pécs. Project no. TKP2021-NKTA-19 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NKTA funding scheme.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramo G, D'Angelo CA, Di Costa F (2020) The role of geographical proximity in knowledge diffusion, measured by citations to scientific literature. *J Informet* 14(1):101010
- Ahrweiler P, Pyka A, Gilbert N (2004) Simulating knowledge dynamics in innovation networks (SKIN). The agent-based computational economics approach, In *Industry and labor dynamics*, pp 284–296
- Arrow KJ (1962) The Economic Implications of Learning by Doing. *Rev Econ Stud* 29(3):155–173. <https://doi.org/10.2307/2295952>
- Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. *Pap Reg Sci* 86(3):495–519. <https://doi.org/10.1111/j.1435-5957.2007.00132.x>
- Bagley MJ (2019) Networks, geography, and the survival of the firm. *J Evol Econ* 29(4):1173–1209

- Balland PA (2012) Proximity and the evolution of collaboration networks: evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Reg Stud* 46(6):741–756
- Balland PA, Boschma R, Frenken K (2014) Proximity and innovation: From statics to dynamics. *Reg Stud* 49(6):907–920
- Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512
- Békés G, Muraközy B (2012) Magyar gazellák A gyors növekedésű vállalatok jellemzői és kialakulásuk elemzése [Hungarian gazelles: what makes a high-growth firm in Hungary?]. *Hung gazelles, Econ Rev-mon Hung Acad Sci* 59(3):233–262
- Birch DL, Medoff J (1994) Gazelles. In: Solomon L, Levenson A, editors. *Labor Markets, Employment Policy and Job Creation*. Economics and Education, 159–167.
- Bodor Á, Füzér K, Szerb L, Varga A (2019) A társadalmi tőke szerepe a magyarországi gyors növekedésű vállalatok innovációs tevékenységében. *Replika* 111(2):9–22
- Borgatti, S. P., Ofem, B. (2010). Social network theory analysis. *Social network theory and educational change*, 17–29
- Boschma R (2005) Proximity and innovation: a critical assessment. *Reg Stud* 39(1):61–74
- Boschma R, Frenken K (2010) The spatial evolution of innovation networks: a proximity perspective. Edward Elgar Publishing. In *The handbook of evolutionary economic geography*
- Boschma R, Iammarino S (2009) Related variety, trade linkages, and regional growth in Italy. *Econ Geogr* 85(3):289–311
- Broekel T, Boschma R (2012) Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography* 12(2):409–433
- Brun JF, Carrère C, Guillaumont P, De Melo J (2005) Has distance died evidence from a panel gravity model. *The World Bank Econ Rev* 19(1):99–120
- Cantner U, Meder A (2007) Technological proximity and the choice of cooperation partner. *J Econ Interac Coord* 2(1):45–65
- Capone F, Lazzeretti L (2018) The different roles of proximity in multiple informal network relationships: evidence from the cluster of high technology applied to cultural goods in Tuscany. *Ind Innov* 25(9):897–917
- Caragliu A, Nijkamp P (2016) Space and knowledge spillovers in European regions: the impact of different forms of proximity on spatial knowledge diffusion. *J Econ Geogr* 16(3):749–774
- Cassi L, Plunket A (2015) Research collaboration in co-inventor networks: combining closure, bridging and proximities. *Reg Stud* 49(6):936–954
- Coad A, Daunfeldt SO, Hölzl W, Johansson D, Nightingale P (2014) High-growth firms: introduction to the special section. *Ind Corp Chang* 23(1):91–112
- Cohen Wesley M, Levinthal Daniel A (1990) Absorptive capacity: A new perspective on learning and innovation. *Administrative Sci Q* 35(1):128. <https://doi.org/10.2307/2393553>
- Csapó K (2011) The Characteristics and Development Possibilities of Fast-Growing Small and Medium-Sized Companies in Hungary. *Erenet Profile* 6(3):16–27
- D’Este P, Guy F, Iammarino S (2013) Shaping the formation of university–industry research collaborations: what type of proximity does really matter? *J Econ Geogr* 13(4):537–558
- Erdős P, Rényi A (1959) On random graphs. I. *Publicationes Math* 6:290–297
- Eurostat-OECD (2007) Eurostat-OECD Manual on Business Demography Statistics. Office for Official Publications of the European Communities, Luxembourg
- Fagiolo G, Guerini M, Lamperti F, Moneta A, Roventini A (2019) Validation of agent-based models in economics and finance. *Computer simulation validation*. Springer, Cham, pp 763–787
- Frenken K, Van Oort F, Verburg T (2007) Related variety, unrelated variety and regional economic growth. *Reg Stud* 41(5):685–697
- Frenken K, Hoekman J, Kok S, Ponds R, Oort FV, Vliet JV (2009) Death of distance in science A gravity approach to research collaboration. *Innovation networks*. Springer, Berlin, Heidelberg, pp 43–57
- Fritsch M, Mueller P (2007) The persistence of regional new business formation-activity over time—assessing the potential of policy promotion programs. *J Evol Econ* 17(3):299–315
- Ghinoi Stefano, Steiner Bodo, Makkonen Teemu (2021) The role of proximity in stakeholder networks for smart specialisation: a sparsely populated area case study. *Innovation: Eur J Social Sci Res*. <https://doi.org/10.1080/13511610.2021.1879631>
- Gilbert N, Pyka A, Ahrweiler P (2001) Innovation networks – A simulation approach. *J Artif Soc Res Simul* 4:1–13

- Gilbert BA, Audretsch DB, McDougall PP (2004) The emergence of entrepreneurship policy. *Small Bus Econ* 22(3):313–323
- Glaeser EL, Kallal HD, Scheinkman JA, Shleifer A (1992) Growth in cities. *J Polit Econ* 100(6):1126–1152
- Grazzini J, Richiardi M (2015) Estimation of ergodic agent-based models by simulated minimum distance. *J Econ Dyn Control* 51:148–165
- Gui Q, Liu C, Du D (2018) International knowledge flows and the role of proximity. *Growth Chang* 49(3):532–547
- Hansen T (2015) Substitution or overlap? The relations between geographical and non-spatial proximity dimensions in collaborative innovation projects. *Reg Stud* 49(10):1672–1684
- Havnæs PA, Senneseth K (2001) A panel study of firm growth among SMEs in networks. *Small Bus Econ* 16(4):293–302
- Heshmati A, Lenz-Cesar F (2013) Agent-based simulation of cooperative innovation in R&D. *Res Eval* 22(1):15–29
- Hoekman J, Frenken K, Van Oort F (2009) The geography of collaborative knowledge production in Europe. *Ann Reg Sci* 43(3):721–738
- Inoue H, Nakajima K, Saito YU (2019) Localization of collaborations in knowledge creation. *Ann Reg Sci* 62(1):119–140
- Jackson MO (2005) A survey of models of network formation: Stability and efficiency. In: Demange G, Wooders M (eds) *Group formation in economics: Networks, clubs, and coalitions*. Cambridge University Press, Cambridge, pp 11–49
- Jacobs, J. (1969) *The economy of cities* New York: Vintage Books
- Karemera D, Oguledo VI, Davis B (2000) A gravity model analysis of international migration to North America. *Appl Econ* 32(13):1745–1755
- Korber M, Paier M (2013) Effects of competence centres on regional knowledge production: An agent-based simulation of the Vienna life sciences innovation system. In: Scherngell T (ed) *The geography of networks and R&D collaborations*. Springer, Berlin, pp 353–371
- Lechner C, Dowling M (2003) Firm networks: external relationships as sources for the growth and competitiveness of entrepreneurial firms. *Entrep Reg Dev* 15(1):1–26
- Maggiioni MA, Uberti TE (2009) Knowledge networks across Europe: which distance matters? *Ann Reg Sci* 43(3):691–720
- Marrocu E, Paci R, Usai S (2013) Proximity, networking and knowledge production in Europe: What lessons for innovation policy? *Technol Forecast Soc Chang* 80(8):1484–1498
- Marshall A (1890) *Principles of Economics*. Macmillan, London, New York
- Montobbio F, Sterzi V (2013) The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations. *World Dev* 44:281–299
- Morley C, Rosselló J, Santana-Gallego M (2014) Gravity models for tourism demand: theory and use. *Ann Tour Res* 48:1–10
- Némethné P (2010) Hol szökellnek a magyar gazellák? –A dinamikusan növekvő kis-és középvállalatok néhány jellemzője. *Vezetéstudomány/budapest Management Review* 41(4):32–44
- Paier M, Dünser M, Scherngell T, Martin, S (2017) Knowledge creation and research policy in science-based industries: an empirical agent-based model. In *Innovation Networks for Regional Development* (pp. 153–183). Springer, Cham.
- Paier M, Scherngell T (2011) Determinants of collaboration in European R&D networks: empirical evidence from a discrete choice model. *Ind Innov* 18(1):89–104
- Papanek G. (2010) A gyorsan növekvő magyar kis-és középvállalatok a gazdaság motorjai. *Közgazdasági Szemle (Economic Review-monthly of the Hungarian Academy of Sciences)*, 57(4), 354–370.
- Platt D (2020) A comparison of economic agent-based model calibration methods. *J Econ Dyn Control* 113:103859
- Ponds R, Van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration. *Pap Reg Sci* 86(3):423–443
- Ponsiglione C, Quinto I, Zollo G (2018) Regional innovation systems as complex adaptive systems: The case of lagging European regions. *Sustainability* 10(8):2862
- Porter, M. (1990) Porter, Michael E. The competitive advantage of nations. *Harvard Business Review* 68, (2), 73.
- Pyka A, Gilbert N, Ahrweiler P (2007) Simulating knowledge generation and distribution processes in innovation collaborations and networks. *Cybern Syst* 38:667–693. <https://doi.org/10.1080/01969720701534059>

- Pyka A, Saviotti P (2002) Innovation networks in the biotechnology-based sectors. *Innovation networks: Theory and Practice* (pp. 75–107). Cheltenham: Edward Elgar.
- Romer PM (1986) Increasing returns and long-run growth. *J Polit Econ* 94(5):1002–1037
- Scherngell T, Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. *Pap Reg Sci* 88(3):531–546
- Scherngell T, Hu Y (2011) Collaborative knowledge production in China: Regional evidence from a gravity model approach. *Reg Stud* 45(6):755–772
- Schoonjans B, Van Cauwenberge P, Vander Bauwhede H (2013) Formal business networking and SME growth. *Small Bus Econ* 41(1):169–181
- Sebestyén T, Varga A (2019) Knowledge networks in regional development: an agent-based model and its application. *Reg Stud* 53(9):1333–1343
- Szerb L, Komlósi É, Varga A (2017) Gyors növekedésű vállalatok Magyarországon: Az innovatív, a rejtélyes és a virtuális gazellák. *Közgazdasági Szemle* 64(5):476–506
- Usai S, Marrocu E, Paci R (2017) Networks, Proximities, and Interfirm Knowledge Exchanges. *Int Reg Sci Rev* 40(4):377–404
- Varga A, Sebestyén T (2017) Does EU Framework Program participation affect regional innovation? The differentiating role of economic development. *Int Reg Sci Rev* 40(4):405–439
- Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393:440–442. <https://doi.org/10.1038/30918>
- Zeng SX, Xie XM, Tam CM (2010) Relationship between cooperation networks and innovation performance of SMEs. *Technovation* 30(3):181–194

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.