Scientometrics (2017) 111:1385–1413
DOI 10.1007/s11192-017-2350-x

The impact of individual collaborative activities on knowledge creation and transmission

Nuha Zamzami[1,2] · Andrea Schiffauerova[1,3]

Received: 18 May 2016 / Published online: 18 March 2017

Abstract Collaboration is a major factor in the knowledge and innovation creation in emerging science-driven industries where the technology is rapidly changing and constantly evolving, such as nanotechnology. The objective of this work is to investigate the role of individual scientists and their collaborations in enhancing the knowledge flows, and consequently the scientific production. The methodology involves two main phases. First, the data on all the nanotechnology journal publications in Canada was extracted from the SCOPUS database to create the co-authorship network, and then employ statistical data mining techniques to analyze the scientists' research performance and partnership history. Also, a questionnaire was sent directly to the researchers selected from our database seeking the predominant properties that make a scientist sufficiently attractive to be selected as a research partner. In the second phase, an agent-based model using Netlogo has been developed to study the network in its dynamic context where several factors could be controlled. It was found that scientists in centralized positions in such networks have a considerable positive impact on the knowledge flows, while loyalty and strong connections within a dense local research group negatively affect the knowledge transmission. Star scientists appear to play a substitutive role in the network and are selected when the usual collaborators, i.e., most famous, and trustable partners are scarce or missing.

Keywords Scientific collaboration · Partnership · Productivity · Knowledge flows · Social network · Network structure

✉ Nuha Zamzami
n_zamz@encs.concordia.ca

[1] Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

[2] Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

[3] Department of Engineering Systems and Management, Masdar Institute of Science and Technology, Abu Dhabi, UAE

Introduction and overview

In today's rapidly growing technological fields the sources of knowledge are widely distributed. Solving new rising issues and answering many complex and multidisciplinary research questions requires higher level of skills and comprehensive knowledge. This leads to the need for collaborative knowledge sharing whose ability to address more complex and critical research problems has already been demonstrated in the literature (Sonnenwald 2007). Moreover, a significant increase in research productivity as result of scientific collaboration has been suggested by several scholars (e.g. Price and Beaver 1966; Zuckerman 1967; Glänzel and Winterhager 1992; Landry et al. 1996)

The collaborative activities can be mapped as a complex network, where its nodes represent the collaborators and their partnerships form the links. In such networks, the knowledge is created and transmitted by socially connected individuals whose collaborations shape the links of the networks. In this work, it is the individual researchers who are the network nodes and their co-authorship of journal articles are the partnership linkages between these nodes, i.e., partners co-creating knowledge through their scientific collaboration.

The knowledge creation network is a dynamic social network where the behavior of collaborators is influenced by their interactions with others over an interval of time. Scholars have analyzed the performance of such networks in the last decades in order to derive policy implications, and to enhance the institutional and governmental decision-making in the area of innovation policy. The existing research studies based on the dynamic social networks approach mainly focused on the firm level analysis, for example analyzing different categories of firms, such as firm leaders or startups (Nagpaul 2002), or studying the roles of various institutions such as academic, industry and governmental ones (Scholz et al. 2010; Triulzi et al. 2011), but much less of the research has been carried out at individual level (Wang 2013; Tajaddod Alizadeh et al. 2015).

As the overall productivity of the network depends on the performance of its actors, the quantity and speed of knowledge diffusion is greatly affected by the individual collaborative activities (Pyka et al. 2002). That is, individuals with certain characteristics would facilitate the network's productivity while the behavior of others might have more negative effects on the knowledge creation and transmission within the network. For example, star scientists, researchers with high impact on innovation and knowledge development reflected by their considerably higher productivity comparing to their colleagues and competitors, are important in the process of technology transfer. The importance of star scientists is not only because of their outstanding scientific knowledge and performance that contribute significantly to the success of firms (Zucker and Darby 2005) but also because they actually act as knowledge circulation improvers in the scientific networks (Schiffauerova and Beaudry 2012). Moreover, gatekeepers, the nodes with highest betweenness centrality, are the influential individuals who are responsible for the knowledge transfer in the network where they interconnect different individuals or bridge separate research groups, and help this way in merging various existing ideas (Gould and Fernandez 1989).

Furthermore, popular researchers, those who are connected to a greater number of collaborators, are critically important for sharing the knowledge considering their ability to access a significant amount of fresh and new knowledge and to bring that knowledge to their colleagues for further collaborative activities which will positively affect the productivity (Henderson and Cockburn 1996). Besides, a better performance for scientist can be detected in the case of maintaining strong ties with previous partners, i.e. loyal scientists, rather than having several co-authorship relationships with multiple ones (Abbasi and Altmann 2011). On the other hand, scientists who are much willing to collaborate with the neighbors of their neighbors, i.e. embedded scientists, are more likely to be deeply involved in a local network of collaboration (his/her own research group) which will discourage the knowledge transmission within the network (Breschi and Lissoni 2006).

The presence of star scientists, gatekeepers, loyal scientists or popular scientists seems to have an impact on the knowledge flows within the network and, consequently, it can be assumed that they also influence the scientific productivity of the other network actors and of the network as a whole. This work therefore aims to first analyze and understand the collaborative knowledge sharing behavior in the real world, and then to simulate the Canadian nanotechnology knowledge-based network and evaluate its dynamics. Our concern is to study the impact of specific groups of researchers on the collaborative behavior of others within the network as well as on the structure, productivity and efficiency of the whole network. We have created the network at individual level based on the co-authorship relationships between Canadian scientists publishing in the field of nanotechnology. The network is then studied in a dynamic context using an agent-based modeling approach. The contribution of this work is twofold: First; It provides a detailed examination of the distinguished individual scientists' roles which invloves the study of their impact on the productivity, on the network structure and also on the perfromance of other sceintists in the network. Second; There is also a methodological contribution, because this work takes a dynamic perspective while basing the whole study on the real data on the scientists collected by the authors using various methods (such as statistical analysis, data mining and survey).

The paper is organized as follows. "Review of the literature" section reviews the literature on the networks of collaborators and their research performance. "Research methods" section describes the data upon which our analyses are based and presents the methodology used. "Results" section discusses the analysis performed and reports the results. "Discussion and concluding remarks" section concludes the findings and proposes some new research opportunities.

Review of the literature

Scientific performance of individuals

Several studies on collaboration suggested significant increase in research productivity related to the collaborative activity (e.g. Price and Beaver 1966; Zuckerman 1967; Glänzel and Winterhager 1992; Landry et al. 1996). Analyzing scientific papers over an interval of time while considering the percentage of articles written by co-authors showed a positive relationship between collaboration and higher productivity, which shows that collaborative knowledge creation played an essential positive role in the scientific performance (Beaver and Rosen 1979; Allen 1983; Drejer and Vinding 2006; Manley et al. 2009).

Most of the scientific output is typically produced by the top 1% or 2% of scientists working in a specific area. These most productive scientists are generally called star scientists. According to Zucker and Darby (1996) the most productive bio-scientists ("stars") have extraordinary value due to the union of still scarce knowledge of the new research techniques and genius and vision to apply them in novel, valuable ways. Their productivity is almost 22 times higher than that of the average GenBank scientist, as they are the authors of 17.3% of the published articles. Hess and Rothaermel (2011) empirically tracked the innovative performance of 108 global pharmaceutical firms over three decades (1974–2003) and found that the top 1% of productive and cited authors; star scientists; account for almost 40% of all publications.

Various definitions, frameworks and contexts of gatekeepers can be found in the literature. For example, Keller (1991) points out the role of gatekeepers by introducing the following characteristics based on an empirical study in U.S. and Mexican organizations: concentration and proximity of gatekeepers in strong organizations, higher performance than usual employees (i.e., higher number of patents and publications), and similarity of action in various industries. According to Sosa and Gero (2005), gatekeeper can be defined as an opinion leader who manages the process of innovation by controlling the selection, feedback and assessment of the new ideas. They believe that in societies with strong ties, only a small group of experts is always playing the gatekeeper role, while in weakly-tied societies the gatekeeper role is rather distributed among the agents and does not represent a consistent behavior. Hence, the effective power of gatekeepers, and consequently the sensitivity of the network to their presence can be determined by its ties and links. Graf (2011), on the other hand, stresses the fact that gatekeepers are the actors that generate novelty by drawing on local and external knowledge, and it is the existence of the gatekeepers which characterizes successful sub-networks.

Some researchers were interested in assessing the strength of the ties, i.e., keeping the same collaborating partners. For example, Van Segbroeck et al. (2009) studied a dynamic graph where they could adjust the behavior and the social ties, and observed that the scientists prefer to keep collaborating with the same partner even in the case when an alternative is available. The authors however suggested that being committed to limited social ties could negatively affect the scientific evolution, which corresponds to the results of Beaudry and Schiffauerova (2011) who found that repeated collaborations with the same partners negatively affect the quality of the inventions they create. In contrast, the study of Abbasi and Altmann (2011) showed that maintaining a strong tie with a previous partner leads to a better performance than having several co-authorship relationships with multiple ones.

Knowledge-based networks

A number of studies assessed the impact of the individuals' network positions and their collaboration patterns on the performance of the researchers within static co-authorship or co-invention networks. From the more recent studies, for example, Yan and Ding (2009) examined authors in the co-authorship network in the library and information science field while focusing on four centrality measures (betweenness centrality, degree centrality, closeness centrality and eigenvector). They concluded that there was a positive relation between these centrality measures and the number of citations of the articles. Similarly, Abbasi et al. (2011) found a positive relation between both eigenvector and degree centralities and the performance of the scholars in the field of information systems using citation based performance indicators. Kumar and Jan (2014) later evaluated the impact of the network variables on performance of researchers in the field of energy fuels in Turkey and Malaysia. They observed that popularity, position and prestige of the researchers assessed through the network centrality indicators have a positive impact on their research performance, where eigenvector was found to be the most influential centrality measure. More recently, the international knowledge diffusion structure and its evolution have been investigated using patent citation networks with the help of measures from social network analysis (Chen and Guan 2016; Guan and Yan 2016).

The impact of the network measures has been intensively studied in Canada focusing on nanotechnology (Beaudry and Schiffauerova 2011; Beaudry and Allaoui 2012; Beaudry and Kananian 2013; Tahmooresnejad et al. 2015), biotechnology (Eslami et al. 2013; Beaudry and Kananian 2013) or health sciences (Contandriopoulos et al. 2016). These studies most often found positive relations between the centrality measures and performance. For example, (Beaudry and Schiffauerova 2011) have shown that the presence of star scientists and other central nodes, i.e. the ones with high betweenness centrality, in the research team has a positive influence on patents quality, while the results for the betweenness centrality of researchers show its significant positive impact on scientific performance in terms of the number of publications, article quality or researchers' h-index (Abbasi et al. 2012; Beaudry and Allaoui 2012; Tahmooresnejad et al. 2015; Contandriopoulos et al. 2016; Guan et al. 2016).

In contrast, Eslami et al. (2013) have found no significant effect of betweenness on the scientific performance and even a negative influence in their patents model where its effect on the technological productivity was estimated. The cliquishness, or clustering coefficient, has also been reported to have an influence on scientific output, although the results vary considerably. In some studies, cliquish networks are found to augment scientific productivity and efficiency of scientists (Tahmooresnejad et al. 2015; Guan et al. 2016) or inventors (Beaudry and Kananian 2013). However, the results of Eslami et al. (2013) suggest that high level of cliquishness, on one hand, hinders the knowledge productivity in scientific communities, while on the other hand, it facilitates the efforts of scientists leading to patenting of their inventions. Moreover, Beaudry and Allaoui (2012) observe that the relationship between cliquishness and scientific productivity exhibits diminishing returns, and some level of fragmentation in the immediate surrounding network of a scientist is hence desirable.

The literature has also discussed the impact of the network connections of the individuals, especially the importance of being connected to distinguished, central or highly performing nodes in the network. For example, Moody (2004) suggests that new researchers tend to get connected to highly reputable authors and with many collaborators. The results of Abbasi et al. (2012) who investigated how new researchers search for collaborators confirm this finding. They highlight the importance of betweenness and degree centrality of an existing researcher in attracting new researchers entering the network. Some studies indicate that researchers who get connected to central or high performing nodes may in fact directly improve their own research productivity (Abbasi et al. 2011; Eslami et al. 2013; Guan et al. 2016). At firms level, the direct ties showed to have an inverted U-shaped effect on the organization's exploitative innovation (Guan and Liu 2016). Moreover, it was suggested that getting connected to productive researchers with a good control over the collaboration network and the flow of information will not only enhance the scientific performance of the connected researcher (Ebadi and Schiffauerova 2015b), but it can also improve the chances of obtaining higher funding for that researcher in future (Ebadi and Schiffauerova 2015a). The importance of the researchers' involvement in large research teams in order to increase the size of the scientific output or to produce high quality publications is also highlighted in some works (e.g. Ebadi and Schiffauerova 2015a, 2016). All of this underscores the role of structural network positions and connections in scientific networks.

Dynamic approach to collaboration networks

Considering that the network of collaborators is a dynamic system where a collection of interacting entities produces some form of behavior that can be observed over an interval of time, modeling and simulation are essential tools to gain a deeper understanding of the system or to provide a root for managerial decision making in order to improve its performance (Bonabeau 2002; Glahn and Ruth 2003). Computer simulation has been primarily recognized as a crucial tool for analyzing complex systems as the most accurate manner to describe what is actually happening in the real world (Banks 1998; Bonabeau 2002; Hao et al. 2008). There are few studies that tried to understand the structure of the system and the behavior of the nodes in large-scale networks under a variety of conditions, as well as, to contribute in predicting the influence of various assumptions and initial conditions to the current behavior (Axelrod 1997; Pyka et al. 2002; Fujimoto et al. 2003; Hao et al. 2008).

The use of simulation approach in studying the complex innovation networks has been started when Gilbert and Troitzsch (1999) developed a simulation platform in The Self-organizing Innovation Networks (SEIN) project to investigate the structure and dynamics of technological collaborations using computational experiments. Pyka et al. (2004) later developed a multi-agent model for Simulating Knowledge Dynamics in Innovation Networks (SKIN) containing heterogeneous innovative firms in a complex environment. The SKIN is modeling the market and the firms' behavior in exchanging knowledge, cooperating and networking with others in order to improve their innovation performance and sales. SKIN allows the investigation of different industries where different strategies have an impact on the firms' productivity with altering several parameters and describing an industry's cooperative behavior. Several experiments have been conducted on SKIN trying to illustrate the impact of different learning activities and emphasize the significance of innovation and learning (Pyka et al. 2007).

These simulation experiments were the start that shows the possibility of investigating the complex relationships between firm and sector success and organizational learning through carrying out experiments on a model that would be impossible to perform in the real world. Several simulation attempts to evaluate the performance of the collaborators' networks have been carried out later for analyzing the network at the firm level, where different categories of firms or the organization type such as academic, industry and government relationships are examined (Scholz et al. 2010; Triulzi et al. 2011; Schrempf et al. 2013 and others).

Nevertheless, there are few recent studies that explored the dynamics of innovation networks at individual level. Wang (2013) for example, constructed an agent dynamics agent-based model that simulates and explains the knowledge transfer activities of individuals. Moreover, Tajaddod Alizadeh et al. (2015) studied the role of the star scientists and gatekeepers in the innovation network, as well as the impact of loyalty based on the link age. Both were only initial explorative studies with many real world issues simplified and with no exactly defined researchers' roles and with a limited study of the data behind the researchers' collaboration.

Given the novelty of this research avenue, several research gaps could be recognized. Although various simulation attempts have been carried out recently to analyze the performance of the innovation networks at the firm level, the individual level has not been much explored in the literature yet. There is also a lack of research on the dynamics of combination of various factors in such networks at individual level, especially with the use of real data. In fact, the major novelty of this work stems from the fact that the developed simulation model is fully based on the real data and on the observed behavior pattern of the scientists in knowledge-based network, which should make it more realistic than any existing simulation model.

Research methods

Field of study and data source

This paper presents the results of a dynamic study of Canadian nanotechnology co-authorship network. The main approach is the exploitation of the large amount of information related to articles, authors and collaboration activities history in the field of nanotechnology. Based on the comparison of different digital libraries and online databases, SCOPUS has been found to be the most reliable and comprehensive source of data in terms of the diversity of fields, the completeness of authors' and address information, and number of articles that can be retrieved. Nanotechnology is very multidisciplinary field, which covers a wide range of nanotechnology disciplines, materials and systems. Meanwhile, there is no formal categorization for nanotechnology in the databases of scientific articles. For these reasons, some sets of specialized keywords have been used by the scholars to distinguish the nano-related articles (Fitzgibbons and McNiven 2006; Zucker and Darby 2005; Porter et al. 2008).

The used combined collection of keywords (see "Appendix") has been created based on seven different sources and was then consulted with nanotechnology experts (Moazami et al. 2015). In order to get benefits from the additional information that Scopus provides, and to still be able to search the full text of the articles, they develop a new data extraction methodology which involves a combined use of Google Scholar and Scopus. The main idea is to use the full text available search in Google Scholar with the help of software called "Publish or Perish[1]" and then to search these results in the Scopus database. A total of 81,727 records contain information about articles published between 1996 and 2011 have been extracted from the database using the automated extraction program.

The present work is based on the nanotechnology database created by Moazami et al. (2015). Each article in our data set contains one or more of the specialized keywords related only to nanotechnology and has at least one of the coauthors affiliated to a Canadian institution. The total number of coauthors is 21,498 including those from outside Canada who are collaborating with Canadian scientists. Research activities information about each co-author, such as his/her publications count, co-authorships count and h-index as in 2012, are then used as inputs for data mining procedure to detect the patterns of different group of actors' behavior.

Network representation and structural analysis

In co-authorship network, the knowledge is created and transmitted by socially connected scientists whose collaborations shape the links of the network. Although there are other forms of collaborations between scientists taking place for various purposes, our main focus is on these co-authorship links, because they are the means of the knowledge transmission in the network. They create a complex net of knowledge-based relationships

[1] Publish or Perish is a software program that retrieves and analyzes academic citations.

and thereby greatly contribute to the production of scientific publications. In such networks, each node of the graph represents a researcher and the links between any node i and node j indicate the collaboration relationships between nodes and represent the co-authorships of researchers on publications.

To visualize the Canadian nanotechnology innovation network and represent the co-authorship relationships among scientists we have implemented social network analysis using Pajek software. Social network analysis is a diagnostic method based on graph theory for studying the mechanisms of communication and collaboration between members in different groups (Racherla and Hu 2010). The value of analyzing social networks consists in its ability to assist with understanding of how to share professional and scientific knowledge efficiently and with evaluating the performance of individuals, groups, or the entire social network (Abbasi et al. 2010).

The method used is to categorize the scientists into groups characterized by their position (centrality) in the networks in order to understand the value, importance, and influence of each group of actors. Betweenness centrality, for example, is an indicator of an actor's potential control of communication within the network. The betweenness of a node x within an undirected graph is computed as follows:

$$B_x = \sum_{i \neq j \neq x} \frac{\sigma_{ij}(x)}{\sigma_{ij}}$$

where σ_{ij} represents all the shortest paths from i to j, and $\sigma_{ij}(x)$ is the number of those paths that pass through the node x. The highest betweenness centrality suggests the most central vertices who are expected to be responsible for the knowledge transfer (gatekeepers). Gatekeepers symbolize those individuals who are bridging the information flows between two or more geographically separate clusters by making connections between them. Since there is no exact definition of gatekeepers percentage in the literature, we have assumed that the top 5% of all scientists in our network with the highest betweenness centrality to represent the gatekeepers.

Similarly, we used the degree centrality of each vertex (author) as an indicator for their number of connections, and have considered the top 5% of scientists with the highest degree centrality as popular scientists who are sought-after collaborators and probably also very well known in the field of nanotechnology. Degree centrality indicates the number of connections within the network and thus reflects an actor's communication activity (Chung and Hossain 2009; Abbasi and Altmann 2011). The degree centrality d_i of node i is given as:

$$d_i = \sum_j a_{ij}$$

where a_{ij} indicates the existence or none-existence of a link between node i and node j. If there is any link between node i and node j, $a_{ij} = 1$ and otherwise, $a_{ij} = 0$. The more links a scientist has to outside sources of knowledge, the more extensive amount of fresh and new knowledge he/she can access and bring to his/her colleagues for further collaborative activities. Those researchers who are connected to a greater number of collaborators (popular scientists) are critically important for sharing the knowledge, which leads to better scientific performance (Henderson and Cockburn 1996).

Moreover, according to the theory of the 'Strength of Ties' introduced by Granovetter (1973) we defined the strength of a tie (link) between node i and node j (i.e., the number of co-authorships between two scholars), as the weight of the link w_{ij} between those nodes. Therefore, we indicate the loyalty of a node (researcher) by its weighted degree centrality calculated as the average of weights of an actor's co-authorships (links). The degree centrality for the weighted graph (weighted degree centrality) d'_i is expressed as follows:

$$d'_i = \frac{\sum_j w_{ij}}{n}$$

where w_{ij} represents the weights of the links between node i and node j, and n represents the number of nodes. In other words, the loyalty of a researcher is represented by dividing the sum of links' weights (total number of co-authorships) by the number of co-authors. Scholars with a strong relationship (frequent co-authorship with the same partner represented by high weighted degree centrality) are considered as loyal ones (Abbasi and Altmann 2011). As for the previously defined groups, we have considered the top 5% with the highest weighted degree as the most loyal scientists among our database.

Furthermore, we suggest that scientists with the highest clustering coefficient are the most willing collaborators to work within their cluster and thus create more cliques and act as embedded scientists in our study. The clustering coefficient (CC) of a vertex (node) in a network graph quantifies how close its neighbors are to being a clique[2] (complete graph). Here, clustering coefficient is defined as the fraction of connections that are realized between the neighbors of a node i, as follows:

$$CC_i = \frac{2n_i}{k_i(k_i - 1)}$$

where n_i denotes the number of links connecting the k_i neighbors of node i. This measure shows how related each scientist is to his/her neighbors, and the probability that they become a closed research group. In other words, the clustering coefficient of an actor indicates how much they are willing to collaborate with the neighbors of their neighbors. Embedded scientists are characterized by their high clustering coefficient which means that they are more likely to be deeply involved in a local network of collaboration (their research group) (Breschi and Lissoni 2006).

Historical data analysis

An extensive analysis of the real world has been conducted in order to understand the behavior of scientists in real world and to detect a pattern for each group of scientists in our database in terms of research performance and collaboration activities. This analysis assisted us to set some assumptions that agents will share regarding their collaboration behavior for building the conceptual model. Statistical data mining was performed through exploratory data analysis, extreme value, hypothesis testing and statistical distribution. The maximum number of potential partners, for example, has been determined referring to the degree probability analysis of the database. Based on the probability density function, we have found that the highest likelihood is to have no more than 10 partners. Accordingly, we have assigned 10 as the maximum allowable candidates that an author will search for, while each will have an actual partnership with the preferable number the model learned from the collaboration history. Moreover, giving the change rate in the publications and patents volume over the study period, the model represents the evolving trend by increasing the number of starters by a random percentage between 1.34 and 2.54 every

[2] Based on the graph theory; a clique in an undirected graph is a subset of its vertices such that every two vertices in the subset are connected by an edge.

year, thus the outcome will be increased by a ratio corresponding to reality. Furthermore, average H-index for the researchers with the highest academic standings and best scientific performance is greater than or equal 17. In our model, the active agents are usually gatekeepers, star scientists and those who have H-index value comparable to this result. On the other hand, as a complementary data collection approach we ran a survey sent to active researchers identified in our database as having scientific collaborations. The main objective was to elucidate the personal preferences to be considered while seeking potential collaborators for conducting a research project. The questions included 18 factors regarding the potential partner's affiliation, research attitude, collaboration history and personal/cultural background.

Within the nanotechnology peer-reviewed papers of authors and coauthors that are affiliated to Canadian institutions, 1500 researchers were randomly selected to be surveyed. This sample contains researchers from different provinces, from both firms and universities while having various research performance. Participants were recruited on a voluntary basis through email and the response rate for this survey was 20% which was unexpectedly high. The findings show that the most critical factors to be considered while selecting the partners are: their academic reputation, their experience in a complementary field, the resources and funding accessibility, the previous collaboration relation with them and its strength.

Simulation model building

We have designed and developed a simulation model using NetLogo (v. 5.0.4), a multi-agent programmable modeling environment (Wilensky 1999). Our model simulates the knowledge creation and exchange interactions among a set of agents that act in a complex and changing environment, given some rules and initial conditions. Its agents are the scientists identified in our database as the ones who have published in nanotechnology at least once with a Canadian affiliation during our study period. The scientists try to interact with others who are also seeking partners to conduct collaborative research projects and publish new articles. Both the behaviors of agents and their software implementations have been verified and validated (internally and externally) to ensure that the program code faithfully reflects the behavior of the conceptual model.

There are two phases in each model, namely SETUP and GO. In the SETUP phase, the initial values for a set of agents' parameters will be loaded into the model through reading text files. In the GO phase, the model will assign a random number of nodes that will be acting as starters who will initiate the partnership process by searching for candidates to collaborate with. An agent can be involved in more than one collaboration activity at the same time with a maximum number of partners for each involvement. In the next sections, the elements and processes of our model are described in further details.

Model's agents and links

Around 14,000 Canadian scientists in our database, who are also the nodes of the network, act as the individual agents of our model and are characterized by a set of parameters reflecting their research performance, scientific collaboration activities and network properties as in 2012. The initial values for these parameters and information about the collaborative activities history will be loaded into the model through reading text files created based on proper SQL queries from our database. The co-authorship relationships between each two scientists will represent links in our model. Each link has a weight reflecting the strength of their collaboration relationships based on how many times they have coauthored an article together. All scientists who have a prior collaboration with a researcher will be stored as his/her previous partners agent set. Considering that this is a two-way relationship, the pair of scientists at both ends of each link will be added to be referred to while seeking partners for new collaboration. Table 1 presents all model agents' parameters and their description:

The environment

Within the model, there are two groups of global variables for setting the environment. The values for the first group of variables will be given using the sliders on the interface and they determine the percentage of scientists in each group to the whole population. The default value for each group is 5%, while we will decrease and increase this percentage in different scenarios for analyzing the effect of this change on the structure and efficiency of the network. The initially given value for each of the status parameters (i.e. Star?, Gatekeeper?, etc.) is false, and will be changed to true for a ratio of the scientists with the highest values for the associated parameters.

The interface of our model has switches used for setting the second group of variables to represent the existence of each group in the world. All switches are set to ON by default, which means scientists belonging to all groups exist unless other settings are specified. When a switch gives a false value, the nodes representing the scientists in the associated group will die (the node will be removed completely from the world along with the collaborative links the author entertained). The purpose of using this setting is to examine the role of each group of scientists by investigating the impact of their absence on both

Table 1 List of model parameters owned by its agents (authors)

Category	Parameter	Description
Identification and status	Node-ID	The author identification number as in SCOPUS
	Firm	The category of author's affiliation as in 2012
	Star?	True when this author is a star scientists
	Gatekeeper?	True when this author is a gatekeeper
	Popular?	True when this author is a popular scientist
	Loyal?	True when this author is a loyal scientist
	Embedded?	True when this author is well connected to others in the cluster
Research performance	Nano-articles	Number of the author's publications which contain the specialized keywords in nanotechnology
	All-articles	Number of all articles that the author has in SCOPUS
	Citation-count	Total number of citations this authors' articles received
	H-index	The H-index considering SCOPUS articles published after 1996
Collaboration	Max-partners	The maximum number of potential partners the author may search for
	Previous-partners	Agent-set of authors with whom the author has previously partnered
Network properties	Betweenness	Betweenness centrality of this node in the network
	Degree	Degree centrality of this node in the network
	wDegree	Weighted degree centrality of this node in the network
	CC	Clustering coefficient of this node in the network

network structure and productivity. While the model is running, these two settings (the ratio of each group and whether they exist or not) will be implemented at each model's iteration considering the updated agents' variables from the previous iteration.

The partnerships

An agent in the model may consider partnerships and start seeking potential partners to collaborate with in order to complement their knowledge and consequently publish a new article. For each iteration (time unit), a random number of nodes will be acting as starters who will initiate the partnership process by searching for candidates to collaborate with. That is, the total number of articles resulting in that iteration depends on the performance and initial number of starters. In experimenting with the model, starters will follow different strategies for seeking their partners while another starter can select them as well. In other words, an agent can be involved in more than one collaboration activity at the same time with a maximum number of partners for each involvement.

Several factors were considered for forming the collaboration ties within the model. These were determined based on the results of the extensive analysis for the historical data collected from SCOPUS and the questionnaire as discussed earlier ("Historical data analysis" section). There are three different strategies for seeking potential partners. First, an agent who has a prior satisfactory collaboration experience with an author will most likely attract him/her for a new one. This is reflected in the model by the *repeated collaboration* function: to find a partner, a starter will seek among previous partners' agent set and assign some as candidates. Second, finding a partner who is well known and has a high academic reputation shows a high importance level for all groups of scientists and especially for those with minimal level of experience. Star scientists, gatekeepers, and those with good academic profile (represented by H-index) will be more frequently selected than others to act as potential partners for new collaborations. That is, author's superior productivity gives him/her a higher probability to act as a starter or to be among the most attractive scientists for a potential new partnership. Lastly, a random search strategy is used for finding more candidates for new collaboration till the maximum number of partners is reached.

After finding the candidates, the partnership relationship will be established, where for some of them it will be based on the preferable number of partners according to past collaboration. If this is the first time for a pair of scientists to collaborate a new link will be created between them and a value of 1 will be given to its strength. Alternatively, if the collaboration tie between them already exists, its strength will be incremented by 1. We are assuming that each collaboration activity is resulting in a new publication coauthored by the involved scientists. Thus, the variable (Nano articles) for each of these agents will be also increased by 1. Besides, the actual partners will be added to previous partners agent set, if they are not already there, for a future collaboration that might occur in the next iterations.

The networks

Only agents that have participated in any collaboration activity during this step (iteration) will be given an age value equal to the step number x. Those agents will form the new network which its structure and productivity will be examined. For all nodes with (age = x) we will recalculate the values of variables related to network measurements. In other words, the idle agents, those who did not publish at this time, will be available for future collaboration but will not be part of the reconstructed network.

The NetLogo NW[3] extension for network analysis have been integrated with our model to reanalyze the network constructed in each iteration based on the new collaboration activities. The degree centrality, betweenness centrality and clustering coefficient for each node in the new network will be updated as values for the associated variables. After updating the values, the structure measurements for the whole network will be calculated by averaging the values of individual participants. Before moving to the next iteration randomly selected agents who were a part of this network will be completely removed from the network. This represents the behavior in the real world where some scientists publish only once and quit the network after.

According to the changes in the performance and centrality of scientists involved in lately formed network they might have different status and act as new or different member of the identified groups. That will be verified by implementing the set up world functions at the beginning of each iteration. That will find the agents with the highest values for the associated variables and change their status parameters to true and remove the scientists in specific group if any of the switches is set to OFF.

The flowchart (Fig. 1) below describes the sequence of the process in the developed model.

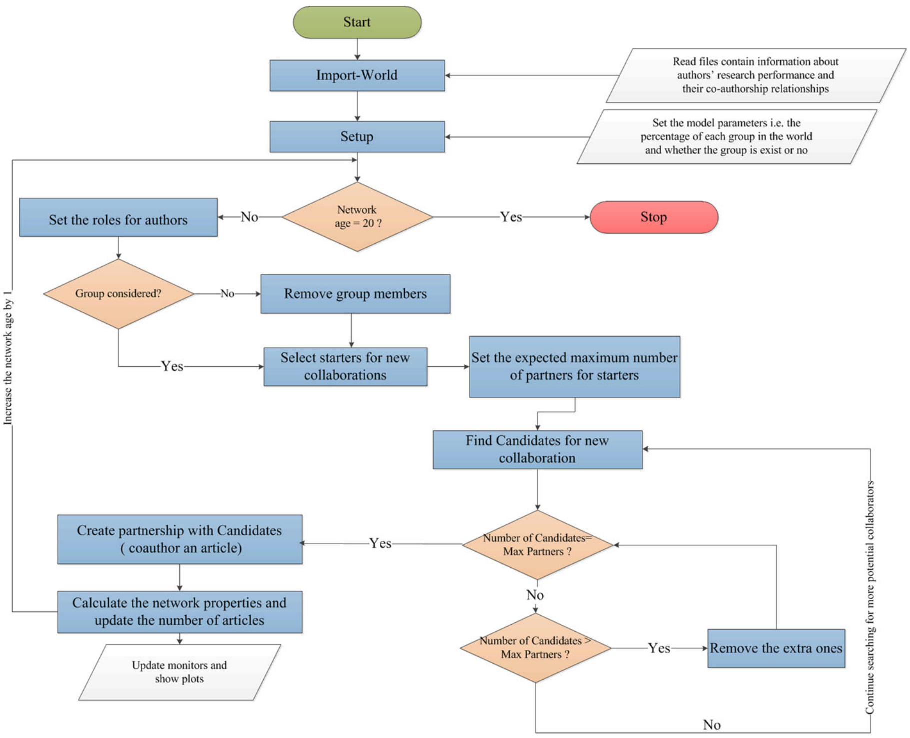

Fig. 1 Flowchart of the developed simulation model

3 NW is an extended library that can be integrated with models developed in NetLogo to perform the social network analysis. More information and the downloadable files are available at: https://github.com/NetLogo/NW-Extension.

Experimental scenarios

The nanotechnology scientists are part of the knowledge-based networks, that is, they appear and grow in the networks naturally. Accordingly, the hypotheses regarding their number as well as their absence from the network can be validated only through simulated scenarios and not by real evidences. The parameter variability analysis is implemented by carrying out several experiments to examine the effect of changing the values of the input and internal parameters of the model upon the model‘s behavior or output. Various scenarios are simulated to study the role of each group of scientists first by removing them completely from the network and then by increasing or decreasing their ratio to the population.

Using BehaviorSpace[4] we have run the model 10 times for each scenario and the results reported are the average from these runs. Beside the basic scenario where each group of interest (i.e., star scientists, gatekeepers, popular scientists, loyal, and embedded scientists) is present as 5% of the population, several experimental scenarios are carried out using two values for the switches (true and false) reflecting the existence and absence of each group respectively. The objective of these scenarios is to examine the role of scientists representing each group and how their absence will affect both the productivity and structure of the network, as well as, the collaborative behavior of other scientists used to interconnect with them. In this set of scenarios, the tested group and their co-authorship relationships will be completely removed from the network. In other words, in each scenario we have removed the nodes that act as specific group along with their links (i.e., their collaboration ties will be removed also, but their partners will remain in the network open for new partnerships).

The other set of experiments used four different values for each slider reflecting the increase and decrease of the group’s ratio to the population (2 scenarios each). Since we work with 5% as default setting, we used 1%, 3%, 7% and 9% as testing values. In each scenario, we have examined the change of one value only while the rest of the settings remain the same. For comparing and evaluating the scenarios we were mainly concerned about the performance and the structure of the whole network.

As for the performance, the total number of publications for the whole network and the average number of the articles published by each group are used as indicators of the productivity. On the other hand, we have examined the structure of the network as it plays the key role in the diffusion of knowledge and production of innovation. The network structure properties have been calculated by averaging the values of the corresponding variables for all nodes that the network consists of. Degree centrality, betweenness centrality, clustering coefficient and network density have been calculated and compared in different scenarios to evaluate the impact of the changed setting.

Results

The role of individual researchers in co-authorship network

In this section, the performance and structure of knowledge-creation networks are numerically analyzed for five different scenarios where each involves the absence of one of

[4] BehaviorSpace is a software tool integrated with NetLogo that allows you to perform experiments with models.

the previously introduced groups. The results of each scenario have been always compared to the basic one that exemplify the real world where all groups are included with a 0.05 ratio to the population. Our findings are based on different sets of analysis which can be categorized into three main groups; the overall scientific production performance, the knowledge transmission efficiency and the productivity of each individual group of scientists. The results of each set of analysis are discussed in the following sections:

The overall network productivity

The performance of Canadian nanotechnology network is analyzed in the presence and absence of each group while considering two indicators: (1) the performance of scientists (measured by the average number of publications/author), (2) the network productivity (measured by the average number of publications/year in the network). Figure 2 below illustrates corresponding results of the average productivity of the scientists in scenarios where star scientists were excluded, and when gatekeepers were excluded comparing to the real world scenario (all groups are included).

The figure shows that the average productivity of the network is almost 25% less than its amount in the absence of star scientists in the network and reduces over time to almost one-third of the one with gatekeepers' existence. In addition, the average productivity of scientists in the network is 1.66 article/author in the first scenario (when all groups are included), which is considerably higher than the average of 1.26 articles/author in the case of star scientists' or gatekeepers' absence. In fact, the roles of stars and gatekeepers are complementary, and their existence is critical for the knowledge production in the whole network. Similar results have been proposed by Abbasi et al. (2012), Tajaddod Alizadeh et al. (2015) and Guan et al. (2016). While star scientists are extremely productive by themselves and attract other knowledge and resources, gatekeepers may not necessarily be

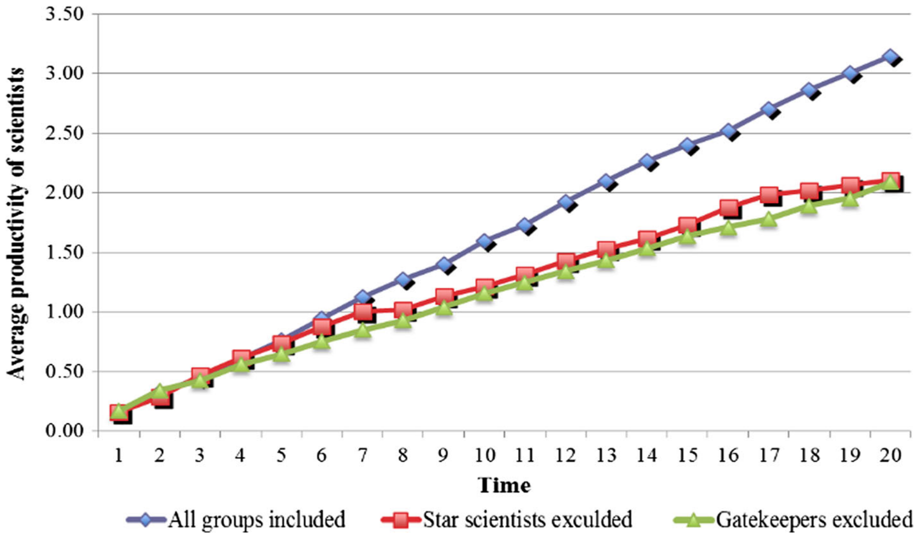

Fig. 2 Average number of publications per scientist when star scientists or gatekeepers are excluded

very productive in terms of publications but they provide other fellow scientists with significant connections to fresh new knowledge.

On the other hand, popular scientists, as defined in this work, are expected to play a role in the knowledge sharing and transmission within the network enabled by the fact that they know large number of collaborators. However, by excluding this group of scientists from the network we observe inconsiderable difference in the average individual productivity in the network. The average performance for the scientists in case of popular scientists-excluded network is 1.58 article/author comparing to 1.66 article/author in their presence (Fig. 3). That is, although popular scientists are greatly responsible for the knowledge sharing and involvement of the nodes within the network due their high number of connection, the overall productivity is not much affected by their absence.

Figure 3 shows that the loyalty of scientists, measured by their high weighted degree centrality, is positively associated with the individual performance of scientists and consequently the overall network performance. That is, the total number of publications per year (3436 publications/year) is affected by the frequency of repeated collaborations and drops to 2676 publications/year in the scenario where loyal scientists were excluded. Therefore, we can state that researchers who have strong ties to co-authors (i.e., repeated co-authorships, given by high weighted degree centrality) have a better research performance than others, which enhances the overall performance of the network. Although some studies suggested a negative impact of loyalty on the researchers' performance (Van Segbroeck et al. 2009) or the organization's ability to innovate (Guan and Liu 2016), our analysis supports the findings suggesting that maintaining a strong co-authorship relationship with same co-author(s) leads to a better individual performance as well as network efficiency (Abbasi et al. 2012; Tajaddod Alizadeh et al. 2015). In fact, the increasing number of scientists who already have satisfactory collaboration experiences would motivate them to renew the partnership and to be involved in new research activities together. This consequently affects the individuals productivity as well as the overall network efficiency.

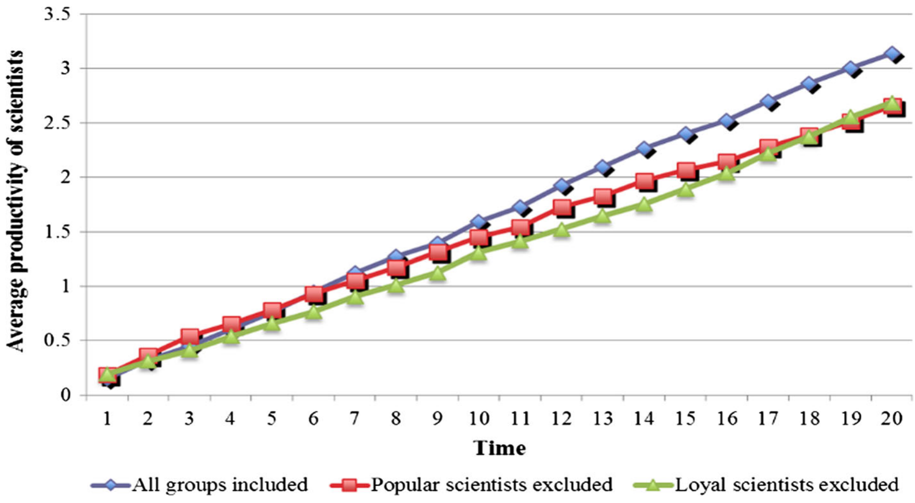

Fig. 3 Average number of publications per scientist when popular or loyal scientists are excluded

Moreover, both the overall performance of the network and average number of publications per scientist show that the network scientific productivity slightly improves when embedded scientists (i.e., those with high clustering coefficient) are excluded. Several researchers have studied the impact of clustering coefficient in networks and their conclusions are not consistent, as both positive and negative effects have been reported. Our model experiments support the finding that the high clustering coefficient of the co-authorship network limits the knowledge creation and thus affects the research productivity negatively (for example Fleming et al. 2007; Gilsing et al. 2008; He and Fallah 2009; Eslami et al. 2013). We can therefore conclude, based on the simulation results, that the presence of embedded scientists has negative impact on the average productivity per scientist in the network. Embedded scientists are usually engaged in research activities within limited closed groups, therefore excluding these scientists from the network results in a better individual performance by opening new opportunities for the scientists to collaborate with new partners outside their research group.

The network structure and knowledge transmission efficiency

The structure of the Canadian nanotechnology network has been also analyzed in the scenarios where each group was excluded from the network and the results were then compared to the original one (all groups included). In all the scenarios, we have calculated the average of degree centrality, betweenness centrality, and clustering coefficient (CC) as well as the network density. Even though the density is related to the size of the network, and the removal of some nodes (one or more of the defined groups) will lead to a smaller network size, the change in this measure for both scenarios was found to be relatively inconsiderable. That is, the proportion of ties in a network is comparable to the total potential ties in the real world. The network structure properties under the 6 tested scenarios are summarized in Table 2.

According to the literature, star scientists in most cases are more likely to repetitively collaborate with the same scientists (Zucker and Darby 1996), which could consequently result in a less socialized network context that reduces the transmission of knowledge among other scientists. The transmission of knowledge within the network seems to be affected by the absence of the stars as they have been considered the main sources of knowledge. Gatekeepers, on the other hand, have a significant impact not only on the productivity of the networks, but also on the improvement of the performance of individuals connected to them in the network. Their role as controllers of the connections and resources in the network can even affect the direction of the research (Heikkinen et al.

Table 2 Summary results for the network structure properties in different scenarios

Scenario	Avg. betweenness centrality	Avg. degree centrality	Avg. clustering coefficient	Network density
All groups included	0.0032	6.58	0.47	1.28
Star scientists excluded	0.0027	6.61	0.53	1.20
Gatekeepers excluded	0.0020	6.61	0.67	1.17
Popular scientists excluded	0.0028	5.53	0.51	1.24
Loyal scientists excluded	0.0029	6.66	0.40	1.26
Embedded scientists excluded	0.0040	6.56	0.35	1.31

2007). Since gatekeepers facilitate the communication in such networks, the scientists with no direct connection to gatekeepers will have lower chance to be involved in collaboration activates with others, which may affect the overall network productivity and knowledge transmission as well.

Due to the centralized positions of the star scientists and gatekeepers, it is expected that their absence would decrease the degree centralization of the whole network. However, our results show that in all scenarios, scientists will be engaged in collaborative activities with an equivalent opportunity to find partners (not necessarily star scientists or gatekeepers). However, other network properties are slightly affected and changed when we excluded these groups. The average betweenness centrality, for example, reduces from 0.0032 to 0.0027 when star scientists are not there, and considerably drops from 0.0032 to 0.0020 when gatekeepers were excluded. Hence, given their centralized positions, the overall centralization of the network will be negatively affected by removing the star scientists.

The change in network centralization will obviously affect the knowledge transmission among its nodes. The presence of star scientists with high average betweenness centrality creates opportunities for potentially more flows of knowledge between different network clusters. Similarly, gatekeepers act as connection points in the network by having shortest paths running through them. That is, if such centralized position researchers are not present in the network, others will probably start collaborating more within their own groups (clusters) (Schiffauerova and Beaudry 2011). Consequently, the cliquishness of the network, represented by the average clustering coefficient, increased when nodes with central positions were removed. Our results show that the 47% likelihood for two individuals with a common collaborator to also have partnership together when all groups were included has increased to almost 53% in case of star scientists' absence and to over 67% when gatekeepers are not there. That is, the central position of star scientists and gatekeepers resulting in higher number of connections (compared to other scientists) will cause that many scholars in the network would become isolated if their connection to the centralized ones was lost for any reason.

As expected, the average degree centrality of the network, as an indicator of the average number of collaborators per node, has decreased from 6.58 to 5.53 in the scenario where popular scientists were excluded. The fact that the portion of authors with extremely high degree centrality is very small comparing to our large population makes this relatively significant change even more considerable. Apparently, the transmission of knowledge in the network is very much affected by sharing the knowledge between the nodes through collaboration activities. In addition, a small change has been noticed in the average degree centrality of 6.66 reflects an insignificantly higher number of connections for each node in case of excluding the loyal scientists comparing to 6.58 in case of their presence. In a network without loyal scientists there is a slightly higher possibility of improving the performance as a result of accessing fresh knowledge through having partnership with new collaborators.

The average betweenness centrality, however, has been negatively affected by the absence of both popular and loyal scientists. That is, the average betweenness of the network representing the real world is 0.0032, but this decreases to 0.0028 after popular scientists removal and to 0.0029 when loyal scientists where absent. Mathematically, a network with both higher average degree centrality and higher average betweenness centrality is more centralized and theoretically supports better flow of knowledge. Thus, we can conclude that by including the popular scientists in the network (researchers with high number of collaborators) and loyal scientists (researchers with repeated co-authorship relationships), the overall centralization of the network slightly improves which make the network more cohesive, and may subsequently enhance the knowledge transmission within the network.

Moreover, repeating the scientific collaboration with same or limited number of co-authors shows an increase in the average clustering coefficient indicating higher network cliquishness. This result supports the argument stated by Mat et al. (2009) that loyalty, i.e., maintaining strong collaboration ties, causes the structure of collaborative networks to become embedded. High network clustering (embeddedness) affects the knowledge transmission within the network negatively; as more clustered groups have many redundant links bearing the same knowledge and little fresh knowledge flowing to the cluster. We can thus conclude that having strong collaboration ties with limited number of partners within the network has negative impact on the network structure and on the transmission of knowledge.

Furthermore, the absence of embedded scientists would result in lower cliquishness and improved knowledge transmission performance of the whole network. Since the embedded scientists are identified in the network by a high clustering coefficient, their removal from the network must obviously cause the network to become less clustered. The lower average clustering coefficient of 0.35 suggests a lower probability of two individuals with a common collaborator to also have partnership together comparing the real-world scenario whereas this probability is 0.47. Consequently, the researchers will have more chances to gain external knowledge instead of being limited within a closed research group. A positive impact of their exclusion on the network structure can be also observed through the increase in the average betweenness centrality to 0.0040 versus 0.0032 for the network that includes the embedded scientists. By excluding the embedded scientists from the network, it becomes more centralized in terms of its betweenness and less clustered, which will reduce the number of closed research groups within the network and support the knowledge transmission among clusters.

The collaborative behavior and productivity of the groups

The last set of analyses we have conducted was related to the performance of individual scholars in case of excluding the scientists belonging to a specific group in order to investigate how other scientists would behave in case of their absence. In fact, it is expected that there is some overlapping between scholars in the different introduced groups. Star scientists, for example, usually occupy more central positions in the network with high number of connections. It is expected that some of these individuals will also be gatekeepers and popular scientists who have highest degree and betweenness centrality. Consequently, removing the star scientists from the network will also affect the performance of these groups negatively as the average number of articles coauthored by scientists belonging to gatekeepers and popular scientists groups reduces to around 50% comparing to their average performance when the star scientists are included. Moreover, our survey results (discussed earlier) suggest that stars, as the scientists with highest reputation in the field, represent usually more attractive potential collaborators and can be selected by more than one scientist at the same time which will result in a considerable increase of the average number of articles coauthored by each scientist in the network. That is, the absence of star scientists will give similar opportunities for all the scientists to be selected as partners, and therefore the share of productivity is more evenly distributed among the scientists.

Besides, our results show a slightly negative impact of excluding the gatekeepers on the performance of all other groups except for the star scientists. The average research productivity for star scientists raised from 25 articles/author to around 28 articles/author and to almost 29 articles/author when gatekeepers and popular scientists are removed respectively. That is, when regular collaborators are not present in the network, other scholars still need to get the access to the knowledge and turn to the most trustable and well-known ones within their cluster. According to Kollock (1994), there is a relation between trust and reputation in the formation of cooperative and exchange structures. The academic reputation of star scientists hence makes them more trustable than other scientists in the network, which may result in the overemphasis on the trust and in the neglect of other factors during the selection of potential collaboration partners. Star scientists are usually not only well connected but they are also attractive partners by themselves. As a result, they also play the role of a substitute for the missing partners.

Since we assume that the loyal scientists are among the productive researchers, their removal from the network will leave their former collaboration partners in need for some active productive researchers to collaborate with. Our results show that star scientists will again play the role of substitutes and create thus many new fruitful ties, which is reflected by the great hike in the average performance to around 33 articles/author when loyal scientists are excluded.

In addition, when reducing the network cliquishness by the removal of the highly clustered nodes, we observe that our results are in accordance with the findings of some previous studies mentioned earlier (Fleming et al. 2007; Gilsing et al. 2008; He and Fallah 2009; Eslami et al. 2013) who observed an improved performance of less clustered networks. We found a positive impact of the lower degree of cliquishness on the average research performance of almost all other groups. Excluding scientists with highest clustering coefficient from the network will destroy the network cliques and thus open new opportunities for the scientists to collaborate with new partners outside their research group, which leads to a better individual performance (Fig. 4).

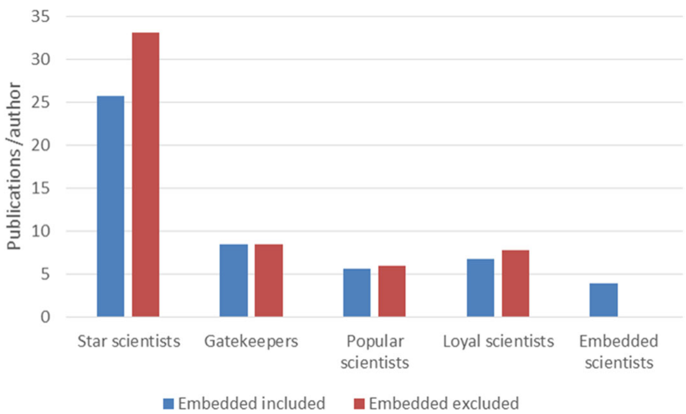

Fig. 4 The impact of embedded scientists on the performance of other groups

Effect of changing the ratio of the researchers' groups to the population

Seeing that the removal of each group of scientists from the network affects the average performance so significantly, our next set of scenarios aimed to investigate how the network will be affected by removing only some of them and keeping their various portions. For each group, we have conducted four different experiments using 1%, 3%, 7% and 9% as values for group's ratio to the population and then compared the results to the basic scenario when each group represented 5% of the population.

Surprisingly, the lower number of star scientists within the scientists' population in fact shows a better performance of the network, and with an increasing number of the star scientists in the network the average productivity decreases. Also, both lower and higher than 5% ratio of gatekeepers to the population decreases the overall network productivity (see Table 3). It seems that if the lower number of star scientists existed in the network it would give the researchers higher chance to be selected over and over again and consequently enhance their individual performance, which would positively contribute toward improving the overall productivity. According to Zucker and Darby (1996) star scientists are the most productive researchers in the network and thus this result is rather surprising as it was expected that having more star scientists in the network will increase the productivity of the system.

Besides, we can say that when more scientists establish partnerships through the same gatekeepers it will result in enhancing the knowledge transmission within the network by improving its centrality and thus leading to better research performance, which is compatible with the findings of Abbasi et al. (2012). Our result can be explained by the network properties as if there are only very few individuals performing extremely well, the network becomes much more centralized which improves its overall knowledge transmission properties, which consequently has a positive impact on its performance. We therefore suggest that if there is any "optimal" percentage of star scientists and gatekeepers in the network to achieve the highest possible productivity it would be somewhere around 5%. However, this issue has never been addressed in the existing literature and further research is needed.

On the other hand, our experiments showed an increasing efficiency of the network with more popular or loyal scientists included. The results of the scenarios where we changed the ratio of these groups to the population present a considerable improvement in the research performance in correlation with the increase of popular scientists (see Table 3).

Table 3 The average and standard deviation of individuals research performance

Groups	The percentage of researchers group to the population									
	1%		3%		5%		7%		9%	
	AVG	SD	AVG	SD	AVG	SD	AVG	SD	AVG	SD
Star scientists	1.69	0.27	1.67	0.32	1.66	0.28	1.61	0.23	1.58	0.21
Gatekeepers	1.60	0.32	1.60	0.18	1.66	0.28	1.50	0.26	1.49	0.28
Popular scientists	1.74	0.30	1.61	0.37	1.66	0.28	1.62	0.39	1.77	0.29
Loyal scientists	1.68	0.23	1.61	0.30	1.66	0.28	1.59	0.24	1.71	0.29
Embedded scientists	1.57	0.22	1.74	0.25	1.66	0.28	1.60	0.22	1.63	0.25

That is, the better connected scientists we have in the network the more cohesive the network becomes and the more the knowledge transmission is enhanced. This result supports the finding of a study conducted by Ahuja (2000) regarding the negative impact of increasing structural holes on innovation in collaboration network. Likewise, the increasing number of scientists who have already satisfactory collaboration experiences would motivate them to renew the partnership and involve in new research activities together (Mat et al. 2009). This consequently affects the individual productivity as well as the overall network efficiency.

Furthermore, as the efficiency of the network correlates inversely with its cliquishness, the increasing number of embedded scientists leads to the presence of more clusters (closed research groups) which is expected to negatively affect the knowledge transmission among the scientists (Eslami et al. 2013). Table 3 shows the result of different scenarios including lower and higher values of the percentage of embedded scientists comparing to the default setting of 5%. The observed finding supports the hypothesis that the lower number of embedded scientists within the scientists' population results in a better performance of the network, and thus increases the average individual productivity.

Discussion and concluding remarks

The aim of the study is to evaluate the knowledge transmission within the Canadian nanotechnology co-authorship network based on the individual scholars' behavior. A comprehensive dataset of nano-related articles has been first analyzed to detect the collaborative behavior of scientists in real world and then modeled in a system to be tested with control of various parameters.

We could characterize and categorize the individual researchers into five groups based on their research performance and their positions in the network. The introduced groups are star scientists, gatekeepers, popular scientists, loyal scientists and embedded scientists. The highest values for number of publications, betweenness centrality, degree centrality, weighted degree centrality and clustering coefficient have been used as criteria to identify the scientists belonging to each group respectively.

Our results related to network's productivity confirm the general evidence of the positive association between research performance and actors' position in the network (Abbasi and Altmann 2011; Abbasi et al. 2012; Eslami et al. 2013; Contandriopoulos et al. 2016). Our results demonstrated the critical role of the star scientists and gatekeepers in enriching both the scientific production and knowledge transmission of Canadian nanotechnology network due to their high individual performance as well as their centralized positions. Star scientists are also very active partners, and therefore they are attractive for other scientists to be selected as collaborators, which however reduces the chance for other scientists to establish partnerships. Gatekeepers are the influential individuals who are responsible for the dynamics of knowledge transmission in the network, and they are also valuable for merging different existing ideas that are held by various disconnected or otherwise isolated research groups.

Nevertheless, there is some literature suggesting that more homogenous power distribution in the network in fact provides more productive environment where the central structure of the network reduces the overall knowledge spillovers among the scientists, resulting in less productivity in the upcoming year (Chung and Hossain 2009; Eslami et al. 2013). Hence, this indirectly suggests that too many stars in the system can in fact hinder the performance of the whole probably by its impact on other researchers. Our results support this point as we found that although the complete removal of star scientists from the network resulted in a poor performance, including a high percentage of them did not prove to be good for the network performance either as the network becomes more centralized which affects the productivity negatively. We can say that there is an inverted-U relationship between network centralization and productivity, where too many or too few centrally positioned scholars will not lead to the best performance.

We suggest that for each research field some optimum percentage, the value of which is not very high, seems to exist. We find specific percentage values for each group in our system. In case of the gatekeepers it is at around 5% when they appear to have most beneficial effect on the transmission of knowledge in the network. For star scientists, however, it seems to be at 1% or even less, as even though they have a profound effect on the publication count themselves, the negative network effects are probably much greater than their positive individual effect. These numbers will most likely differ in other systems since the studied effects will have different size in different environments and under various conditions. Therefore, further research is needed.

We expected that the flow of knowledge within the network would also be affected by popular and loyal scientists. The high numbers of connections that popular scientists have provide them with a unique role in increasing the speed of the knowledge sharing and transmission, enhancing connectivity within the network and decreasing its embeddedness (Henderson and Cockburn 1996). Consequently, it was surprising to find that the overall productivity of the network was not affected much by increasing or decreasing the number of partners involved in each collaboration activity. Loyalty, i.e., maintaining strong collaboration ties, has shown to have a considerable impact on the transmission of knowledge. Although our findings support the previous evidences of the better individual research performance achieved by the strong collaboration ties (Abbasi et al. 2012; Tajaddod Alizadeh et al. 2015), the results also suggest that maintaining the collaboration relationship with the same partners negatively affects the network structure over the time. That is, repeated collaboration activities would make the network more embedded and consequently worsen the knowledge transmission.

Embedded scientists, i.e., scholars with highest clustering coefficients, provide higher chance for their collaborators to be involved deeply in closed research groups (Eslami et al. 2013) and also to collaborate less with new partners outside their team. The results show the negative impacts of embedded scientists by making the network less centralized and more embedded. With lower average betweenness centrality, and higher average clustering coefficient the knowledge flow among clusters is slow and the number of closed research groups within the network increases.

It is interesting that Eslami et al. (2013) argue that even though high degree of embededness in scientific communities hinders knowledge productivity, they also found that it in fact facilitates scientific efforts leading to possible applications in industrial context due to the importance of increased confidence and trust in industrial settings. This means that the impact of cliquishness in different environments and under various conditions may vary. Furthermore, the effects of cliquishness were in this work studied in separation from other effects. Small worlds networks, which are the networks which combine a high degree of clustering (cliquishness) and short network distance (shortest path length), have been proposed to lead to greater innovative productivity (Schilling and Phelps 2007) and to higher quality of articles in terms of citations or journal impact factor (Ebadi and Schiffauerova 2015a).

It is thus possible that high cliquishness itself does not have positive effects on the network productivity but if the highly cliquish network has at the same time short path length the productivity may be positively affected. In this case high level of loyalty and trust among the partners will still exist but the established research groups collaborating repetitively will be close to other researchers. The short path length in the network will enable them to bring fresh knowldge from other communities and research groups. This still needs further investigation. We were unable to measure path length in the network because our network is disconnected and we would need to limit our investigation to the greatest component only.

We have observed a great overlap between scholars in the different defined groups. For example, the more centrally positioned scholars, e.g., star scientists, have also high number of connections, thus, the star scientists are also part of gatekeepers and popular scientists groups which have highest degree and betweenness centrality. We find a negative impact of the performance of all groups when the star scientists, gatekeepers or popular scientists are not present in the network. However, star scientists appear to play a substitutive role in the network, i.e., they are the ones most likely to be selected as potential partners if the usual collaborators are missing. This role leads to an increasing productivity of the star scientists group in case that any other group is excluded.

The results of this work could be used by governmental agencies and other institutions for improving the research and technology polices. We suggest to provide support for star scientists and gatekeepers, as they are important and they are also complementary. However, since it was shown that the existence of a few big star scientists and well connected agents in the network will lead to higher network productivity than if there are more of the smaller ones, specific granting opportunities should be designed for these scientists, such as strategic, targeted and high priority funding programs for high caliber researchers.

Besides, the funding opportunities should involve programs for existing research groups such as renewal grants in order to support the established research groups within which the trust and confidence have already been developed. However, we need to be careful here, because, as discussed previously, these existing groups with constant repetitive collaboration partnerships can get closed upon themselves with limited access to new knowledge from outside of this community. Therefore, the inflow of new knowledge should be supported. This could be done through the support of various collaborations between these funded research groups or through the requirement of new partners in the groups stipulated by the funding agencies.

Limitations and directions for future research

The contributions of this research were the essential first steps towards studying the performance of knowledge-based networks at the individual level. Many real-world problems were simplified or ignored due the need for more data or because their solutions were outside the scope of this research. In this section, the limitations of this study will be summarized and the opportunities for future research will be outlined accordingly.

First of all, although this work is mainly concerning the nanotechnology sector in Canada, our developed model is sufficiently flexible to be used for extending the results of this research into the global level and/or comparing the findings to the ones from other high-tech industries in Canada. On the other hand, further research could use more comprehensive database(s) where more information about the field of expertise, research interests and funding amount each scientist receives could be collected to improve the partner‘s selection mechanism in the model and reduce the level of randomness.

Moreover, the analysis of network performance in our simulation model considers only quantity of the knowledge diffusion and transmission in nanotechnology field, i.e., average number of publications, while their quality is ignored. Research performance indicators for individual scientists, such as the H-index, and for the research society, such as the RC-index and the CC-index, should be included for quality evaluation.

Furthermore, our results suggested that the absence of both star scientists and gatekeepers negatively affects the network performance, but, at the same time, we observed that their presence produces negative effects in the network as well. This could be an interesting issue for the scholarly investigation to determine whether any optimal portion of this group exists which would allow to achieve the highest efficiency of the network.

Lastly, other types of centralities such as closeness or eigenvector could be taken into consideration for more extensive network structure analysis. It would be interesting and more realistic also to consider some details about the scientists' research career, e.g., change in their positions and/or mobility between different firms or organization. These changes might affect their productivity and open new opportunities for scientific partnerships.

Appendix: List of nanotechnology keywords (based on Moazami et al. 2015)

Search term	Search queries
Nano* terms	"nano assembly", "nano computer", "nano cubic technology", "nano molecular machine", "nano optic", "nano optical tweezers", "nano warfare", "nanoarray", "nanoassembler", "nanobarcode", "nanobarcodes particle", "nanobioprocess", "nanobot", "nanobotics", "nanobots", "nanobubble", "nanobusiness alliance", "nanobusiness company", "nanocatalysis", "nanoceramic", "nanochemistry", "nanochip", "nanocircle", "nanocluster", "nanocomputer", "nanocone", "nanocontact", "nanocrystal", "nanocrystal antenna", "nanodefense", "nanodentistry", "nanodetect", "nanodevice", "nanodiamond", "nanodisaster", "nanodot", "nanoelectrospray", "nanoengineering", "nanofacture", "nanofacty", "nanofiber", "nanofibre", "nanofiltration", "nanofluidic", "nanofoam", "nanogate", "nanogear", "nanogenomic", "nanoimaging", "nanoimprint lithography", "nanoimprint machine", "nanoimprinting", "nanolabel", "nanolithography", "nanomachine", "nanomagnet", "nanomanipulat", "nanomanipulation", "nanomanufacturing", "nanomaterial", "nanomechanical", "nanomot", "nanoparticles",nanowire", "nanope", "nanope", "nanopharmaceutical", "nanophotonic", "nanophysic", "nanoplumbing", "nanoprism", "nano-ring", "nanoscale self assembly", "nanoscale synthesis", "nanoscience", "nanoscopic scale", "nanoscopic scale", "nanosens", "nanosheet", "nanoshell", "nanosource", "nanostructure", "nanostructured", "nanosurgery", "nanosystem", "nanotechism", "nanotechnology", "nanotube", "nanotube bundle", "nanowalker", "nanowetting"
Quantum terms	"quantum cascade laser", "quantum coherence", "quantum computation", "quantum compute", "quantum computer", "quantum 116 computing", "quantum conduct", "quantum conductance", "quantum conductivity", "quantum confine", "quantum device", "quantum dot", "quantum gate", "quantum information", "quantum information process", "quantum mirage", "quantum nanophysics", "quantum nanomechanics", "quantum system", "quantum well"

Search term	Search queries
Molecular* terms	"molecular assembler", "molecular machine", "molecular nanogenerat", "molecular nanotechnology", "molecular robotic", "molecular scale manufacturing", "molecular systems engineering", "molecular technology"
Self assembly terms	"fluidic self assembly", "nanoscale self assembly", "self assembled"
Atomic terms	"atomic manipulation", "atomic nanostructure"
Other terms	"biofabrication", "biomedical nanotechnology", "biomimetic synthesis", "biomolecular assembly", "biomolecular nanoscale computing", "biomolecular nanotechnology", "bionems", "brownian assembly", "buckminsterfullerene", "buckyball", "buckytube", "c60 molecule", "carbon nanotubes", "conductance quantization", "dna chip", "electron beam lithography", "epitaxial film", "epitaxy", "fat fingers problem", "ganic led", "glyconanotechnology", "grey.goo", "immune machine", "khaki goo", "laser tweezer", "limited assembler", "military nanotech.", "moletronic", "naneplicat", "nanite", "optical trapping", "protein design", "protein engineering", "proximal probe", "rotaxane", "single cell manipulation", "spin coating", "stewart platfm", "sticky fingers problem", "textronic", "universal assembler", "utility fog", "zettatechnology"

References

Abbasi, A., & Altmann, J. (2011). On the correlation between research performance and social network analysis measures applied to research collaboration networks. In *44th Hawaii international conference on systems science (HICSS-44)*. Hawaii, USA.

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics, 5*(4), 594–607.

Abbasi, A., Altmann, J., & Hwang, J. (2010). Evaluating scholars based on their academic collaboration activities: Two indices, the Rc-index and the Cc-index, for quantifying collaboration activities of researchers and scientific communities. *Scientometrics, 83*(1), 1–13.

Abbasi, A., Chung, K. S. K., & Hossain, L. (2012). Egocentric analysis of co-authorship network structure, position and performance. *Journal of Information Processing & Management, 48*(4), 671–679.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly, 45*(3), 425–455.

Allen, R. (1983). Collective invention. *Journal of Economic Behavior & Organization, 4*(1), 1–24.

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.

Banks, J. (1998). *Handbook of simulation: Principles, methodology, advances, applications, and practice*. New York, NY: Wiley.

Beaudry, C., & Allaoui, S. (2012). Impact of public and private research funding scientific production: The case of nanotechnology. *Research Policy, 41*(9), 1589–1606.

Beaudry, C., & Kananian, T. S. R. (2013). Follow the (industry) money—the impact of science networks and industry-to-university contracts on academic patenting in nanotechnology and biotechnology. *Industry and Innovation, 20*(3), 241–260.

Beaudry, C., & Schiffauerova, A. (2011). Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. *European Management Journal, 29*(5), 362–376.

Beaver, D., & Rosen, R. (1979). Studies in scientific collaboration part III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics, 1*(3), 231–245.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America, 99*(3), 7280–7287.

Breschi, S., & Lissoni, F. (2006). Mobility of inventors and the geography of knowledge spillovers. New evidence on US data. In *CESPRI conference*. Milan, Italy.

Chen, Z., & Guan, J. (2016). The core-peripheral structure of international knowledge flows: Evidence from patent citation data. *R&D Management, 46*(1), 62–79.

Chung, K. S. K., & Hossain, L. (2009). Measuring performance of knowledge intensive workgroups through social networks. *Project Management Journal, 40*(2), 34–58.

Contandriopoulos, D., Duhoux, A., Larouche, C., & Perroux, M. (2016). The impact of a researcher's structural position on scientific performance: An empirical analysis. *PLoS ONE, 11*(8), e0161281.

Drejer, I., & Vinding, A. L. (2006). Organisation, "anchoring" of knowledge, and innovative activity in construction. *Construction Management and Economics, 24*(9), 921–931.

Ebadi, A., & Schiffauerova, A. (2015a). How to become an important player in scientific collaboration networks? *Journal of Informetrics, 9*(4), 809–825.

Ebadi, A., & Schiffauerova, A. (2015b). On the relation between the small world structure and scientific activities. *PLoS ONE, 10*(3), e0121129.

Ebadi, A., & Schiffauerova, A. (2016). How to boost scientific production? A statistical analysis of research funding and other influencing factors. *Scientometrics, 106*(3), 1093–1116.

Eslami, H., Ebadi, A., & Schiffauerova, A. (2013). Effect of collaboration network structure on knowledge creation and technological performance: The case of biotechnology in Canada. *Scientometrics, 97*(1), 99–119.

Fitzgibbons, K., & McNiven, C. (2006). Towards a nanotechnology statistical framework. In *Blue sky indicators conference II* (pp. 25–27). Ottawa, Canada.

Fleming, L., King, C., III, & Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science, 18*(6), 938–954.

Fujimoto, R. M., Perumalla, K., Park, A., Wu, H., Ammar, M. H., Riley, G. F. (2003). Large-scale network simulation: how big? How fast? In *Modeling, analysis and simulation of computer telecommunications systems, 2003. MASCOTS 2003*: *11th IEEE/ACM international symposium.* Atlanta, USA. doi: 10.1109/MASCOT.2003.1240649.

Gilbert, N., & Troitzsch, K. (1999). *Simulation for the social scientist*. Buckingham: Open University Press.

Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., & Aan Den Oord, A. (2008). Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy, 37*(10), 1717–1731.

Glahn, H. R., & Ruth, D. P. (2003). The new digital forecast database of the National Weather Service. *Bulletin of the American Meteorological Society, 84*(2), 195–201.

Glänzel, W., & Winterhager, M. (1992). International collaboration of three east European countries with Germany in the sciences, 1980–1989. *Scientometrics, 25*(2), 219–227.

Gould, R. V., & Fernandez, R. M. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology, 19*(1989), 89–126.

Graf, H. (2011). Gatekeepers in regional networks of innovators. *Cambridge Journal of Economics, 35*(1), 173–198.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.

Guan, J., & Liu, N. (2016). Exploitative and exploratory innovations in knowledge network and collaboration network: A patent analysis in the technological field of nano-energy. *Research Policy, 45*(1), 97–112.

Guan, J. C., & Yan, Y. (2016). Technological proximity and recombinative innovation in the alternative energy field. *Research Policy, 45*(7), 1460–1473.

Guan, J., Zuo, K., Chen, K., & Yam, R. C. (2016). Does country-level R&D efficiency benefit from the collaboration network structure? *Research Policy, 45*(4), 770–784.

Hao, Z., Yun, X., & Zhang, H. (2008). An efficient routing mechanism in network simulation. *Journal of Simulation, 84*(10–11), 511–520.

Harzing, A. W. (2007). Publish or Perish. http://www.harzing.com/pop.htm.

He, J., & Fallah, M. H. (2009). Is inventor network structure a predictor of cluster evolution? *Technological Forecasting and Social Change, 76*(1), 91–106.

Heikkinen, M. T., Mainela, T., Still, J., & Tähtinen, J. (2007). Roles for managing in mobile service development nets. *Industrial Marketing Management, 36*(7), 909–925.

Henderson, R., & Cockburn, I. (1996). Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *The Rand Journal of Economics, 27*(1), 32–59.

Hess, A. M., & Rothaermel, F. T. (2011). When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry. *Strategic Management Journal, 32*(8), 895–909.

Keller, R. T. (1991). Gatekeeper communication networks and technological innovation: A study of U.S. and Mexican R&D organizations. *The Journal of High Technology Management Research, 2*(1), 1–13.

Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology, 100*(2), 313–345.

Kumar, S., & Jan, J. M. (2014). Research collaboration networks of two OIC nations: Comparative study between Turkey and Malaysia in the field of 'Energy fuels', 2009–2011. *Scientometrics, 98*(1), 387–414.

Landry, R., Traore, N., & Godin, B. (1996). An econometric analysis of the effect of collaboration on academic research productivity. *Higher Education, 32*(3), 283–301.

Manley, K., Mcfallan, S., & Kajewski, S. (2009). Relationship between construction firm strategies and innovation outcomes. *Journal of Construction Engineering and Management, 135*(8), 764–771.

Mat, N. C., Cheung, Y., Scheepers, H. (2009). Partner selection: Criteria for successful collaborative network. In *20th Australian conference on information systems*. Melbourne, Australia.

Moazami, A., Ebadi, A., & Schiffauerova, A. (2015). A network perspective of academiaindustry nanotechnology collaboration: A comparison of Canada and the United States. *Collnet Journal of Scientometrics and Information Management, 9*(2), 263–293. doi:10.1080/09737766.2015.1069966.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociology Review, 69*(2), 138–213.

Nagpaul, P. S. (2002). Visualizing cooperation networks of Elite institutions in India. *Scientometrics, 54*(2), 213–228.

Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research, 10*(5), 715–728.

Price, D., & Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist, 21*(11), 1011.

Pyka, A., Ebersberger, B., & Hanusch, H. (2004). A conceptual framework to model long-run qualitative change in the energy system. In J. S. Metcalfe & J. Foster (Eds.), *Evolution and economic complexity* (pp. 191–213). Cheltenham: Edward Elgar.

Pyka, A., Gilbert, N., & Ahrweiler, P. (2002). Simulating innovation networks. In A. Pyka & G. Kuppers (Eds.), *Innovation networks: Theory and practice* (pp. 169–196). Cheltenham: Edward Elgar.

Pyka, A., Gilbert, N., & Ahrweiler, P. (2007). Simulating knowledge-generation and distribution processes in innovation collaborations and networks. *Cybernetics and Systems: An International Journal, 38*(7), 667–693.

Racherla, P., & Hu, C. (2010). A social network perspective of tourism research collaborations. *Annals of Tourism Research, 37*(4), 1012–1034.

Schiffauerova, A., & Beaudry, C. (2011). Star scientists and their positions in the Canadian biotechnology network. *Economics of Innovation and New Technology, 20*(4), 343–366.

Schiffauerova, A., & Beaudry, C. (2012). Collaboration spaces in Canadian biotechnology: A search for gatekeepers. *Journal of Engineering and Technology Management, 29*(2), 281–306.

Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science, 53*(7), 1113–1126.

Scholz, R., Nokkala, T., Ahrweiler, P., Pyka, A., & Gilbert, N. (2010). *The agent-based Nemo model (SKEIN)—simulating european framework programmes. Innovation in complex social systems* (pp. 300–314). London: Routledge.

Schrempf, B., Kaplan, D., & Schroeder, D. (2013). National, regional, and sectoral systems of innovation—an overview. *Report for FP7 Project" Progress". European Comission*. https://www.google.com.co/url.

Sonnenwald, D. (2007). Scientific collaboration: A synthesis of challenges and strategies. *Annual Review of Information Science and Technology, 41,* 643–681.

Sosa, R., & Gero, J. S. (2005). A computational study of creativity in design: The role of society. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing Journal, 19*(04), 229–244.

Tahmooresnejad, L., Beaudry, C., & Schiffauerova, A. (2015). The role of public funding in nanotechnology scientific production: Where Canada stands in comparison to the United States. *Scientometrics, 102,* 753–787.

Tajaddod Alizadeh, D., Ghiasi, G., Schiffauerova, A. (2015) The role of individuals in innovation networks: A simulation approach in canadian biotechnology network. In *11^e Congres International de Genie Industriel-CIGI2015*. Québec, Canada.

Triulzi, G., Pyka, A., & Scholz, R. (2011). R&D and knowledge dynamics in university-industry relationships in biotech and pharmaceuticals: An agent-based model. *International Journal of Biotechnology, 13*(1–3), 137–179.

Van Segbroeck, S., Santos, F. C., Nowé, A., Pacheco, J. M., Lenaerts, T. (2009). The coevolution of loyalty and cooperation. In *2009 IEEE congress on evolutionary computation, 2009.* Trondheim, Norway: IEEE.

Wang, X. (2013). Forming mechanisms and structures of a knowledge transfer network: Theoretical and simulation research. *Journal of Knowledge Management, 17*(2), 278–289.

Wilensky, U. (1999). *Center for connected learning and computer-based modeling*. Evanston, IL: Northwestern University.

Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology, 60*(10), 2107–2118.

Zucker, L. G., & Darby, M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences, 93*(23), 12709–12716.

Zucker, L. G., & Darby, M. R. (2005). *Socio-economic impact of nanoscale science: Initial results and Nanobank*. Washington: National Bureau of Economic Research Inc.

Zuckerman, H. (1967). Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *American Sociological Review, 32*(3), 391–403.