Journal of Informetrics 15 (2021) 101192

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

The evolution of triangular research and innovation collaborations in the European area

K. Angelou[a,b], M. Maragakis[a,b,c], K. Kosmidis[a,b], P. Argyrakis[a,b,*]

[a] Department of Physics, University of Thessaloniki, Thessaloniki, Greece
[b] Center of Complex Systems, University of Thessaloniki, Thessaloniki, Greece
[c] Department of Physics, International Hellenic University Kavala, Greece

ARTICLE INFO

Keywords:
Multiplex network
Triangles
Clustering coefficient
Patents
Subsidized research
Evolution

ABSTRACT

In the current study, we examine the multiplex network of patents and European Research Framework Programmes (FPs) aiming to uncover temporal variations in the formation patterns of triangles (a fully connected network between any three nodes). The multiplex network that we study consists of two layers whose nodes are the NUTS2 regions, and collaborations between scientists or inventors of different regions result to a link. We split the network temporally into 28 shorter sub-networks with a span of 6 years each, and calculate the number of triangles formed at the end of the 6-year period. Next, we shuffle the data creating 50 randomized networks for each of the 28 six-year sub-networks, in order to identify whether there is a hidden mechanism that favors a non-random behavior. Real and shuffled data are compared using a z-score, a measure of the differences of standard deviations between them. In addition, we repeat the same analysis using the clustering coefficient whose large value can signify a strong collaboration pattern for a given node. The results from the temporal analysis of real vs randomized multiplex networks, show that triangular FP collaborations tend to be favored over random ones, while in patents the case is strongly the opposite. Furthermore, results using triangles tend to be more comprehensive as opposed to those of the clustering coefficient. Finally, we identify which NUTS2 regions frequently exhibit a high clustering coefficient in either of the layers, and we present a map with these values for all regions. The results of this research can help policy making organizations understand the spatial dimension of subsidized research and patented innovation collaboration networks and perhaps to strengthen regional collaboration.

1. Introduction

Collaboration networks have been studied for many years, with works such as those by Scott (1991), Wasserman and Faust (1994), and Borgatti et al. (2009) standing out. The analysis of such networks is important, since it helps, for example, the study of relationships between individuals, extract patterns in the network's formation, or identify the most influential nodes.

In research, collaborations have been studied in the past focusing on the structure of the networks, like Newman (2001) and Abbasi et al. (2012). Broekel et al. (2015) examined the evolution of such networks, Scherngell and Barber (2009) examined their geographic origin and Ding (2011) examined whether there is a preference in the collaboration among scientists of the same interests and similar citation scores. Similarly, in innovation, efforts to study the structure of the patent network have attracted much interest, such as the study of Choe and Lee (2017), while Sun (2016) focuses on the spatial aspect. There are also studies, like those of Érdi et al. (2013) and You et al. (2017), that try to predict emerging technologies.

* Corresponding author.
E-mail address: panos@auth.gr (P. Argyrakis).

https://doi.org/10.1016/j.joi.2021.101192
Received 22 October 2020; Received in revised form 20 April 2021; Accepted 30 June 2021

The concept of multi-layer networks, i.e. networks formed by many layers, was introduced by Kivela et al. (2013). In multilayer networks each layer represents a network with different nodes and links, compared to any other layer, and each layer signifies a set of data having different nature. Studying the networks in such ways, allows us to investigate whether the evolution of one network, affects those of the other layers. A subcategory of multilayer networks are the multiplex ones. Their distinct characteristic is that nodes on different layers represent the exact same quantity. The type of link, or collaboration, and the nodes connected by this on each layer differ. For example, we may consider as a multiplex network the social interactions in Twitter and Facebook of a school's students (each social network corresponds to a layer of the multiplex network). The scientific interest of multiplex networks is significant, especially for social ones. Boccaletti et al. (2014) examined the structure and dynamics of multilayer networks, while Battiston et al. (2014) dealt with structural measures for multiplex ones. Mittal and Bhatia (2018) proposed a new way of calculating closeness centrality for multiplex social networks. In research and innovation, the literature is mostly limited in paper-patent studies, like that of Li et al. (2015), and does not generally use the notion of multiplex networks, but rather compares the separate layers characteristics, such as the studies of Magerman et al. (2015), Landini et al. (2015) and van der Pol and Rameshkoumar (2018). Few are the studies that treat knowledge and innovation networks as multiplex, like that of De Stefano and Zaccarin (2013) who analyzed the interactions among authors and inventors in the region of Trieste, Italy, and Pugliese et al. (2019) who built a tri-layered network of publications, patents and industrial production, and examined the interactions among them.

Triangles, essentially a triplet of nodes that are all connected to each other, is a notion introduced by Simmel (1908) and later by Kuper and Wolff (1951). They are essential in analyzing a network's structure. For example Ma and Ma (2019) suggested that the most important nodes may be identified as the nodes that participate in a significant number of triangles compared to others and, thus, they belong to the dense part of the network. Moreover, triangles are an indicator of the networks cohesion, as well as its homophily and transitivity, as shown by Seshadhri et al. (2014). Triangles have been used in various structural and dynamical studies, such as that of Dimitrova et al. (2020) who use them as part of their multiplex network's structure analysis and Antal et al. (2006) who study the dynamics of a social network by examining its balance based on triangles. In the current study, we will also examine the clustering coefficient, as introduced by Watts and Strogatz (1998), which contains the notion of triangles and has been used extensively in collaboration network studies. It has been applied on both static and dynamical networks. Kong et al. (2019) use it among other metrics to examine how collaborations affect the researchers influence. On the latter case, there are studies such as that of Perc (2010) who monitor the evolution of clustering coefficient in the Slovenian researchers collaboration network and that of Huang et al. (2012) who study patents with international collaboration of inventors.

The main aim of the current study is to uncover temporal variations in the formation patterns of triangles in the multiplex network of patents and European Framework Programmes (FPs). In our study, we take into consideration, both the multiplex network and the two layers individually. In addition, we examine the dynamic evolution of the respective clustering coefficients and we compare the results of the two metrics.

Scientific collaborations and discoveries that lead to scientific patents are not independent. They both have a common cause, i.e. the need for innovation and understanding. In order to deepen their knowledge and answer important questions scientists collaborate with each other. These same collaborations lead to discoveries which may have important practical applications and lead to patents. A way to study this interdependence is by representing collaboration networks and patent networks simultaneously as multiplex graphs. This sort of representation takes into account that the nodes of the collaboration and patent networks are of the same kind, as opposed to being two entirely separate networks. We will compare the real world networks structure with that occurring from comparable randomized processes imposed on the exact same data. The aim is to find if the mode of research subsidization in EU (multi-partner collaborations) is reflected in innovation and whether the dynamic approach taken reveals any changes in the period under study.

Concerning the subsidized research, the European Union (EU) has for the last decades continuously increased funding for research purposes, as indicated by Haegeman et al. (2015) and Reillon (2017). With these investments, the EU aims at increasing the research output of small and medium-sized enterprises/businesses (SMEs) and the private sector. Moreover, it has significantly favored commercial applications, as indicated by Kim and Yoo (2019). Despite this, one of the main EU objectives is to create cohesion throughout the entire EU, not just on a country level, but on a regional level as well[1]. To achieve this goal, it subsidizes research and research infrastructures by collaborating partners present in different regions both through the Funding Programmes, with the limitations in minimum countries involved, and other inter-regional Programmes, where partners must belong in different regions, see further on Varela-Vázquez et al. (2019) and Crescenzi and Giua (2020). By doing this, it aims to support quality employment, education, skills, and social inclusion, as indicated in Berkowitz et al. (2016) and Darvas et al. (2021).

The paper is structured as follows. Section 2 presents the data used for our study, Section 3 describes the steps followed for the analysis, Section 4 contains the results obtained, and finally, Section 5 draws the conclusions of this work and explains its potential applicability on other fields.

2. Data

The multiplex network that we study consists of two layers, patent and scientific collaborations. Our aim is to identify the evolution of collaborations in research and innovation at a regional level. The nodes of the two layers of the multiplex network are the 330 NUTS2 ("Nomenclature of Territorial Units for Statistics") regions of the European area, which are subdivisions of countries into

[1] https://ec.europa.eu/regional_policy/en/2021_2027/

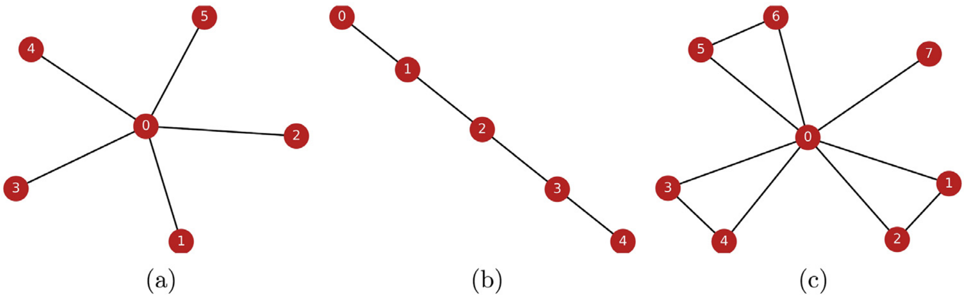

Fig. 1. Schematic of 3 networks with various types of collaboration between the nodes. a) Star network, where node 0 may be connected to many other nodes, however, these are all isolated from each other and, thus, its local clustering coefficient is 0 (same for nodes $1-5$). b) Linear network, where all the nodes have local clustering coefficient 0. c) Triangles are formed in this type of network, so all nodes have a local clustering coefficient greater than 0, apart for node 7 which does not participate in any triangle. More specifically, nodes $1-6$ clustering coefficient is 1, and node's 0 is 0.14.

smaller regions of typically about 800,000 to 3,000,000 population. In our network approach, links between two individual NUTS2 codes are introduced, when there is collaboration between inventors or scientists, located in two different such regions, for the creation of a patent or in a European Framework Programme (FP5 − 7, and Horizon2020) project. Thus, each patent or project can add several links, assuming that there are several partners from several different regions in it. In fact, since we assume that each projects partner is connected to all other partners (fully connected sub-network), in a typical project many links are indeed created between different regions.

The first layer of the multiplex network is constructed using data from subsidized research, namely from the FP5 − 7 and Horizon2020 Framework Programme projects. Although FP5 started in 1998, we begin our study from the year 2000 because data prior to that year are quite sparse, and could prove misleading. More specifically, there are only 307 links formed before 2000 at a 2 year range, a very small number which is surpassed within a single month in data after 2000. This also defines the maximally available time period for the study of the multiplex network. FP data have been extracted by the Community Research and Development Information Service (CORDIS) and the aggregate network contains in total 34,061 projects, 330 different regions and 40,032 unique links. There are also 307 unique self loops (collaborations within the same region) which have been omitted for the purposes of this study.

The contents of the patent layer have been derived by the European Patent Office (EPO) and contain patents for the years 1978 to 2020. However, and in order to be in compliance with the FP database, we utilize only patents registered since 2000 and onwards. The total number of patents that have been registered in any NUTS2 region by at least two inventors since then is 48,466. The maximum number of regions participating in the aggregate network is equal to 330, while the unique links are 5,692 in total and the self loops, which are ignored as above, are 297.

It should be noted here that while the number of patents is larger than the number of projects, the number of unique links in the patent layer is less than the number of unique links in the projects layer. This is an indication that the two layers have differences in their connectivity. Indeed, given that the number of unique links is much smaller in the patents than in the FPs, one expects that this layer is less dense that the FP one.

As analyzed in the methodology the multiplex network is studied in smaller time periods, and obviously, for each time period the number of nodes and links is smaller than the total one. On average the number of nodes is 291 for patents and 324 for FP, while the average number of links is 2,669 and 27,948, respectively.

Both databases contain geographic information about the scientists/inventors location. However, at some occasions the NUTS2 codes had to be extracted manually and were not listed as a separate field. The databases also contain a field that lists the duration of the patents and the FP projects. It should be noted that about 20% of the links' duration data are missing here for the patent layer (although such data in some cases may be non-existent because the patent protection still exists), while for FP projects practically almost all collaborations have a "death" (link removal) time listed. This information is used for the temporal analysis in order to include the death of a link, which reflects to the end of an existing collaboration, e.g due to lack of further funding. This information is used for the temporal analysis to remove links. Link removal reflects the end of an existing collaboration, e.g. due to lack of further funding.

Further information on the datasets and the code used can be found in the supplementary material.

3. Methodology

As mentioned in Section 1, triangles have been used in various studies for the structural analysis of networks. Our focus is on identifying collaboration triangles that exist in the multiplex network over long periods of time and, thus, regions that may have persistent collaboration patterns. We also seek to find those time points where structural changes lead to the significant increase, or decrease of such forms of inter-regional collaborations.

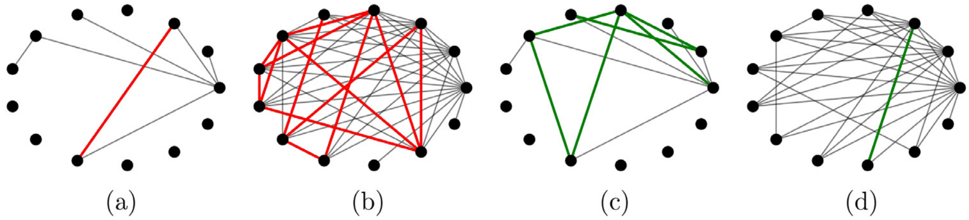

Fig. 2. Schematic of the evolution of a multiplex network with temporal data and links that "die" at some point. Figures a) and b) are the two layers of the multiplex network that have grown until a specific date, and figures c) and d) are the same layers at a later point of time. Red lines indicate the links that are about to be removed, while green lines are the new links that have been inserted into the network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Consider, for example, the two limiting cases of a star-like graph and of a linear chain, presented in Fig. 1. Both cases correspond to states without triangles. In a star-like network the removal of the central node will leave all partner nodes isolated. In a chain any removal fragments the chain. On the contrary, if nodes with high clustering are removed, bridges between the remaining nodes still exist. We consider this to be a sign of a more robust collaboration in a network overall (see Li and O'Riordan (2013) and Bordons et al. (2015)), as opposed to the case of "dominance" observed in the star-like network or the case of "localization" shown in the linear chain. The existence, or absence, of triangles in the multiplex networks is the result of the collaboration dynamics between scientific, technological or other groups. The clustering coefficient has been used as a method to characterize the structure of similar networks by Liu et al. (2019). Thus, we are motivated in studying triangles and clustering coefficient time series generated from collections of multiplex networks formed by research and patent collaborations.

Indeed, for a star-like network, according to the definition of the local clustering coefficient of node i, given by Watts and Strogatz (1998):

$$C_i = \frac{\lambda_G(v)}{\tau_G(v)},$$

where $\lambda_G(v)$ is the number of triangles, one of which is v, and $\tau_G(v)$ is the number of triples (2 links and 3 nodes), where node v is incident to both edges, the resulting value for the central node is 0 and, all other nodes also exhibit a C_i value of 0. The same applies to a linear network, Fig. 1b, where no nodes participate in any triangle and C_i is again 0 for all nodes. In contrast to these very specific types of collaboration, there can be a star-like network, whose peripheral nodes are connected to each other, such as the one in Fig. 1c. In this type of structures the nodes form clusters that increase connectivity. This is shown by their local clustering coefficient, which is 1 for all nodes, except for 0 ($C_0 = 0.14$) and 7 ($C_7 = 0$). An important limiting case is that of a fully connected network where all nodes connect to all other nodes (not shown schematically). In this case the local clustering coefficients would be $C_i = 1$ for all i. Generally speaking, a node with high clustering coefficient is a node that collaborates "well" with others. We believe that this aspect can be used as a valuable indicator in the informetrics community.

In our study, we want to quantitatively find out which type of collaboration and how often a triangle is met in a multiplex research and innovation collaboration network, as opposed to the case of a similar randomized multiplex network. This helps us to classify qualitatively the network topology. Such an approach has also been used in the study of research and innovation networks by Tahmooresnejad and Beaudry (2018).

We use the sliding windows method as indicated by Datar et al. (2002), and similarly by Angelou et al. (2020) that allows for the division of an evolving network into smaller, in temporal terms, networks. To be more specific, each sub-network initiates at the beginning of every year (January 1st) or in the middle of it (July 1st) for all years since 2000, and up to 2014, resulting to 28 windows. We allow for a growth period of exactly 6 years for the sub-networks to reach a relative plateau in their evolution, fig. S1. At each date that patents or FP projects are registered, links are inserted into the multiplex networks, or removed if a collaboration has reached the "death" date. Figure 2 shows an illustration of the patent and the FP layer, at two different points of time, that links have been removed or inserted. At the end of each 6-year period the number of existing triangles is calculated. In addition, we calculate the local clustering coefficient C_i of each node, and subsequently the average value of the local clustering coefficient, $\overline{C_i}$, for all nodes, i. We, thus, make a qualitative comparison between the results of the triangles approach and the clustering coefficient one.

Given the methodology of sliding windows that we use, the creation and death of a link is always captured over many consecutive time windows. Indeed, given the time length of the window, it is possible that most such events will be captured for more than 5 windows (due to the 6 months time displacement between consecutive windows), before having to be born and die in the same window. Our results imprint and show the status of the network after some time, and not the events leading up to that state. The sliding windows method simply visualizes how this state has changed dynamically, and ends up being at the windows end date.

The quantity of interest in our study is the clustering coefficient, which depends on 2 different factors. The first is the number of links in the network, and consequently the degree distribution, and the second is the specific dynamics that drive the collaborations between institutions, countries etc. The second factor is of particular interest to us. Thus, we randomly shuffle the network links, by maintaining exactly the same degree distribution and we repeat the entire analysis for the shuffled versions of the network. We aim to see if the observed clustering is similar to what one would observe if collaborations between institutions were random, given the restriction of the degree distribution. To check that we compare the clustering coefficient of the data to that of the randomly shuffled case.

To be more specific on the shuffling procedure that we followed, we randomly shuffle the links, within each time window, while the degree distribution remains the same in both layers and reconstruct the multiplex network. We take care that the number of projects/patents per sliding window remains the same. However, we shuffle the dates that projects and patents are inserted into the network so as to ensure greater randomization, even during the networks' evolution.

In order to compare the results between the real and the shuffled data we use the standard score or z-score, described in Spiegel (2017), given by:

$$z = \frac{x - \overline{x}}{s}$$

where x is the real value, $\overline{x}$ is the average value of the shuffled data, and s their standard deviation. What this metric does is to calculate how many standard deviations the real data differ from the mean shuffled data. Randomly occurring networks with no preferential attachment in the creation process would have produced a z-score value around 0. The comparison will take place both in triangles and the local clustering coefficient analysis.

To avoid any problems with in the comparison process we aggregate the data over each time window and use the same exact sets for real data, as well as for the randomized ones. Thus, we keep the data for real and randomized results in each time window comparable. Additionally, we have verified that the results remain robust when probing smaller time windows, even ones where data do not reach a plateau.

4. Results and discussion

As mentioned in Section 3, the sliding window method creates 28 multiplex sub-networks, i.e. 28 patent sub-networks and 28 FP sub-networks. For each set of sub-networks we identify and calculate the patent triangles and the FP triangles. We then remove from each layer their common links, those existing between the same two nodes in both layers, and place them into a new network, the common multiplex one.

We follow the same procedure for the 50 shuffled networks that we have created for each of the 28 real sub-networks. The results show that the real world patent layer, Fig. 3a, and the common network, Fig. 3c, are much more likely to form more isolated collaborations (simple links) when compared to their shuffled versions. In fact, shuffled data tend to form many more triangles (up to about 10 times more in patents and 4 in the common network) for the entire time period studied.

On the other hand, real FP data tend to form more triangles than their corresponding shuffled data, Fig. 3b. This effect is possibly due to the specifics of EU funding rules, which promote in most funding calls a collaboration of three or more partners. Thus, triangular collaborations are favored and successful ones are most often those with even more than 3 partners. The z-score values of all three cases prove that these results are not random as they range far from 3 typical standard deviations. In fact the patents and the common network, Fig. 3d, show negative z-scores and the FP positive ones, agreeing with the observations of the previous plots. This behavior is, therefore, non random and hides an inherent preference in the network creation process.

It is worth noting that the patent triangles are practically zero in the real data and few even in the shuffled ones. This is due to the very high number of triangles in the FP layer, which means that practically any inter-regional collaboration link in the patent layer created will very likely have a corresponding one in the FP layer. Thus, and owing to the approach used, almost all patent links are transformed to common ones and only few remain with no corresponding ones. The exact same analysis was also conducted for shuffled versions of the entire data and not within the windows, with the results being quite similar, as shown in supplementary material fig. S2.

Next, we repeat the same process by calculating each node's local clustering coefficient of each layer, including their common network. This multiplex network is again formed by removing the links that exist in both the patent and the FP layers and inserting them into the common one. We then calculate the average value of the local clustering coefficients of each layer and of the common network, Fig. 4. Real data in patents, Fig. 4a, and the common network, Fig. 4c, show that they clearly have smaller averaged local clustering coefficients and, thus, when compared to shuffled data, do not tend to form triangles. As for the FP layer, Fig. 4b, we notice that real data exhibit a higher averaged local clustering coefficient, as compared to the shuffled data. The z-score of the local clustering coefficient, Fig. 4d, clears up the question whether such results could be randomly obtained. The values shown point to a non randomized process for all three cases, and the existence of an underlying preferential type of mechanism for the growth of these systems.

Figure 4, when compared with Fig. 3, shows some similarities in a qualitative sense, as both show the same type of preference in real over shuffled triangle formation for FPs and the opposite for patents and the common network. However, the quantitative difference is significant as the triangles approach shows much more strongly the existence of a non random process. It shows a $20 - 50\%$ preference in real over shuffled FP triangles, while real $\overline{C_i}$ values are only $15 - 20\%$ higher than their respective shuffled ones. Similarly, for the patents and the common network the number of triangles is about 10 and $3 - 4$ times higher in the shuffled data than in the real ones, while $\overline{C_i}$ values are only $3 - 4$ and less than 2 times more, respectively. z-score values are similar in both figures with the triangle ones, Fig. 3d, being again larger than the averaged local clustering coefficient ones, Fig. 4d.

Our next goal is to identify at each layer the nodes with the higher average rank (rank_{aver}), according to their local clustering coefficient rank for the 28 windows (rank_{sb}). This will help us pinpoint the key nodes in triangular collaborations over time. More specifically, for each window we rank the nodes according to their local clustering coefficient, rank_{sb}. We allow for two, or more,

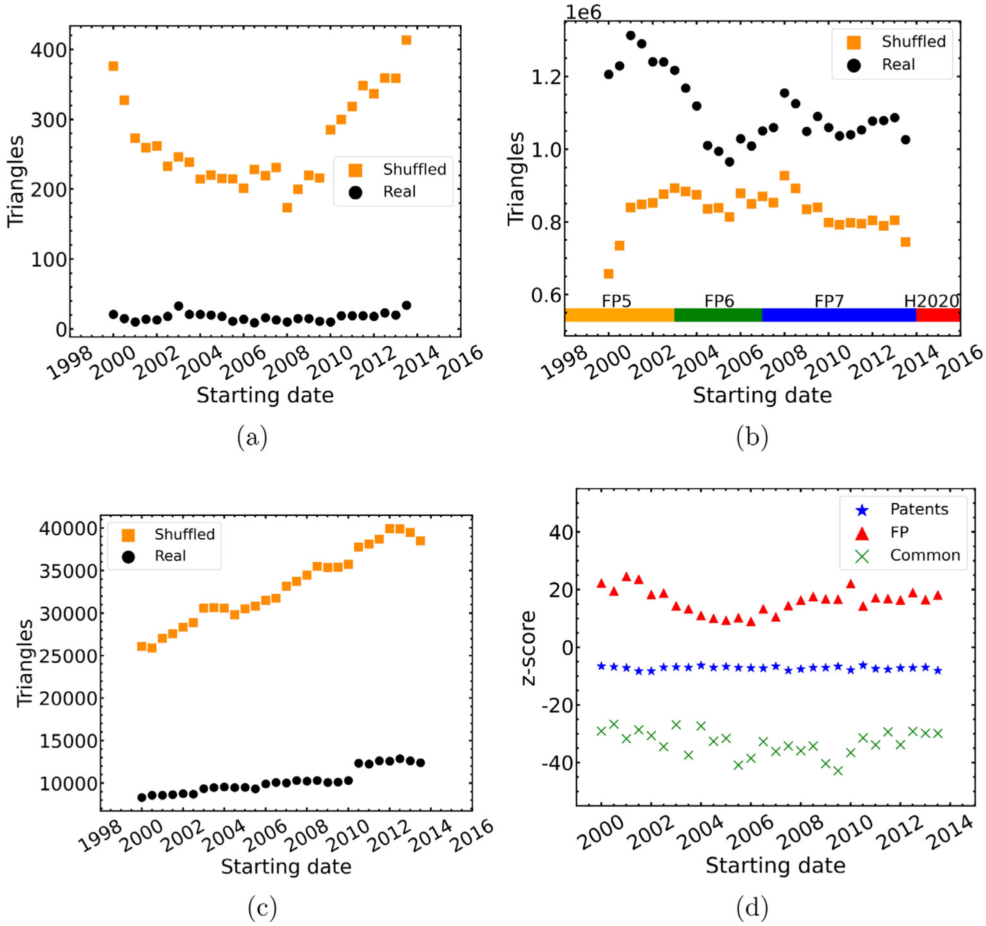

Fig. 3. (Color online) The number of triangles at the end of the 28 windows, versus their starting dates for a) patents, b) FP, and c) their common network. Black dots represent the real data, and orange squares represent the average of 50 shuffled networks. The rectangular colored areas (colors are random) at the bottom of figure b show the FPs duration of each. d) z-score of the number of triangles for the patent (stars), FP (triangles) and Common (cross) networks versus time. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

nodes to have the same rank if needed. We then average over all the 28 ranking positions of each node in each sub-network, and then re-rank the averaged data, $rank_{aver}$. Figure 5 shows the boxplots of the top 10 NUTS2 regions for all 3 cases (patents, FPs, common network), which are sorted from left (higher $rank_{aver}$) to right (lower $rank_{aver}$) according to the average value of their rank. We notice that there are very few NUTS2 regions that exist in all 3 layers.

Finally, we present 3 European maps, one for each case, colored according to the averaged (out of the 28 windows) local clustering coefficient of each NUTS2 region, Fig. 6. We notice that regions with intense scientific history, as for example Paris, Rome, Madrid, Barcelona, and others, are not among the highest regions in the FPs and the common network. Although this may not seem reasonable, it is due to the fact that we examine the existence of triangles (through the local clustering coefficient), and not that of links which may actually be too many (in this case in Paris there are 324 links, which ranks Paris as the first region in number of links). We also notice that there are regions, or even entire countries, that may be part of triangular type of scientific collaborations in the FPs at a much higher than expected rate. However, their part in the actual number of patents is relatively small as compared to that of other regions (Iceland, and most of Turkey).

In the supplementary material we also show the same figure, fig. S5, produced with a different goal in mind. Specifically, we include the common triangles in the patent and FP layers rather than remove them, in order to have a clearer view of these two layers only, each on their own, and not the common one. The results show that the patent layer is almost identical to the results of the common network, Fig. 6c, while the FP layer shows even higher local clustering coefficient values for most regions.

The supplementary material also presents the results for the top-ten NUTS2 regions ranked in order of participation in collaboration triangles for FPs, patents and their common network in fig. S6. The average number of triangles, shown in fig. S7, is normalized over the maximum value of the average number of triangles, while the averaging is done over the same 28 windows as before. The results paint a slightly different picture than that of the local clustering coefficient. They emphasize on the role of major urbanized city

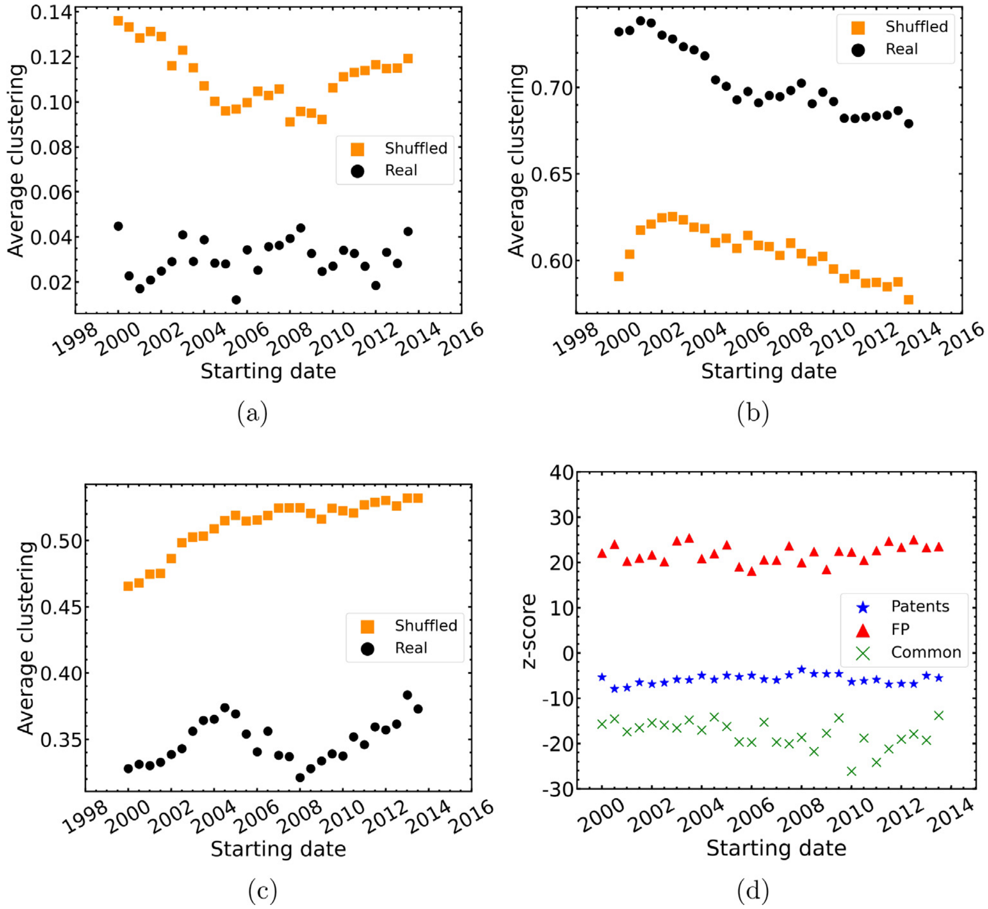

Fig. 4. The average local clustering coefficient calculated at the end of the 28 windows versus their starting dates for a) patents, b) FP, and c) their common network. Black dots represent the real data and orange squares represent the average of 50 shuffled networks. d) z-score of the average local clustering coefficient for the patent (stars), FP (triangles), and Common (cross) network versus time.

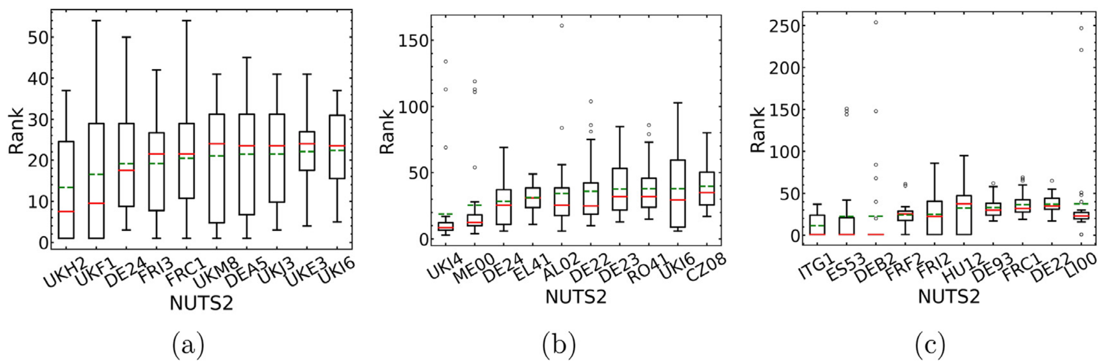

Fig. 5. Boxplot of the top 10 NUTS2 regions for a) patents, b) FPs, and c) their common network in terms of average ranking, according to the local clustering coefficient ranking of the 28 windows (green dashed lines). Red lines show the median value of their ranking for the 28 windows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

centers found on all EU countries which simply gather much larger numbers in links belonging to triangles in the FP layer and the common network. One can notice the major differences between regions such as Sicily, Paris, Rome, Madrid, and many more in Fig. 6 as opposed to fig. S7. Such a representation of collaborations is closer to those of studies doing a simple statistical analysis of regions in either patents or subsidized research.

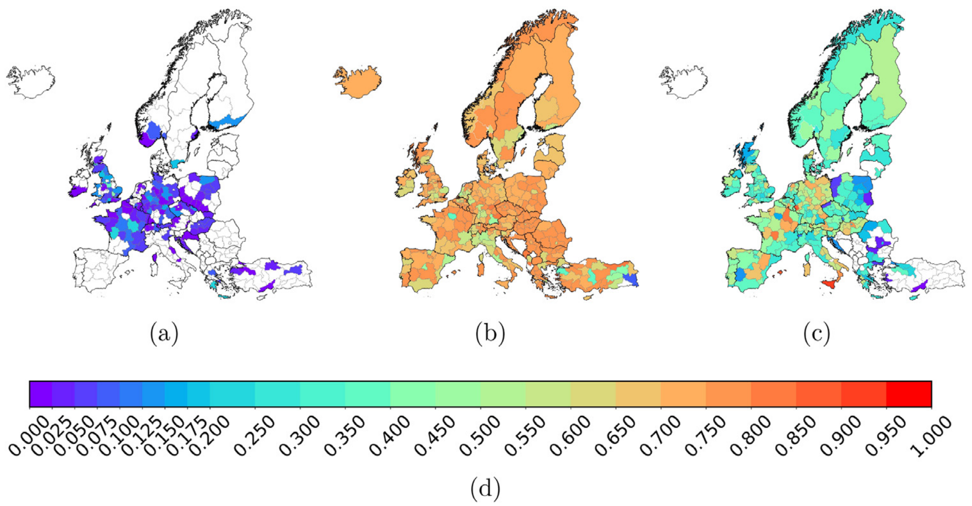

Fig. 6. (Color online) European color-based map of the averaged local clustering coefficient of each NUTS2 region. Real data for a) patents, b) FPs, and c) their common layer. As the average value of the local clustering coefficient increases, the colors change according to the d) colorbar (from left to right). White color indicates regions with a zero value in the clustering coefficient, which corresponds to zero triangles in that layer after the removal of the common triangles, or zero common triangles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

In summary, we study the evolution of triangles in a multiplex network consisting of patents and European Framework Programmes to uncover any preference in forming extended collaborations of triangular form, rather than dual ones of simple link type. In addition, we compare the results to those of the local clustering coefficient in order to identify any differences. Our approach is that of analyzing the triangle and clustering coefficient time series which are generated from collections of the multiplex network formed by the FP - patent graphs.

The results show that in the patent layer, the number of triangles formed only on that specific layer is extremely small in comparison to those of the FP layer and the common network. In addition, when comparing the real data to its shuffled versions, one concludes that in such a real system there is a strong preference to form isolated links rather than entire triangles. In fact, the number of triangles in the shuffled data is many times larger (up to 10 times). Similar behavior is noticed in the common network, where the shuffled data form significantly more triangles than the real ones (up to 4 times). On the contrary, FP collaborations show a stronger preference to form triangles in the real, rather than the shuffled, layers. We also study the local clustering coefficient, which yields qualitatively quite similar results, yet in all cases with much less distinguishable differences. Indeed, by studying the number of triangles one can emphasize on the differences between real and shuffled networks when compared to their local clustering coefficient counterparts. The main point observed in the system of research and innovation studied here is that triangles elucidate more easily than the clustering coefficient the preference (or avoidance) in triangular forms of collaboration over simple dual ones.

The analysis presented in our paper aspires to show that European level Framework Programmes tend to favor the creation of multi-partner collaborations and, thus, one would expect them to be able to spark innovative research leading to patents. Such patents should reasonably be multi-region as well. It is, however, surprising that the opposite holds true. In fact, in both the patent layer and the common network a random re-organization of the links forming them would lead to more triangles than those actually formed. The exact cause for this has not, to the best of our knowledge, been explained yet in the literature. However, a plausible cause for the preference on such a structural dynamic in the FP layer are the funding rules of the EU which favor multi country collaborations. Similarly, in the patent layer and the common network, the potential commercial value of patents may inhibit the creation of multi-partner collaborations. It would be interesting if future work were to focus on understanding the mechanisms behind this behavior.

Next, we want to identify the NUTS2 regions that have constantly, for all time windows, proven to have a high local clustering coefficient, and can, thus, be considered significant for the creation of their networks. What we see is that for each layer the top 10 regions are mostly different, and there are very few which are common in all three layers. Furthermore, according to the averaged value of the local clustering coefficient from all the 28 windows, we notice some regions with counter-intuitive behavior. There are regions of high scientific activity with low local clustering coefficient, and other regions that are not so scientifically active which have a higher local clustering coefficient value. The same result can be seen clearly by comparing Fig. 6 with fig. S7. This could be due to the fact that the latter regions may have a very small number of collaborations, but possibly the exact same regions over time, thus forming relatively high numbers of triangles. On the contrary, scientifically active regions may have many more collaborations, but these do not always form enough triangles, resulting on lower local clustering coefficient values.

Our research results and the combination of existing methodologies used can help funding authorities and policy makers decide on whether specific regional actions need to be taken to support specific geographical areas. The EU, in order to increase the cohesion of its Member States, has long implemented funding programmes that act with cross border or regional characteristics. These Interregional programmes often fund actions such as research infrastructure actions for the public domain, actions that help innovative private sector companies, SME incubators etc. There are also national or regional (cross border/cross national) level actions that require partners to belong in specific NUTS2 level regions. Besides this, the mere fact that the Framework Programmes of EU typically requires a minimum of three partners, which must come from at least 2 Member States, favors the collaboration of researchers from more than 2 regions. Although these collaborations are not on a NUTS2 level, in many NUTS2 regions there are relatively few research organizations that dominate the entire region in terms of research production. Regarding the patent layer, there is no well established and centrally controlled funding organization at the EU level to promote patenting between researchers of different regions, nor is there simply one Research and Innovation epicentre per NUTS2 region in most cases. However, and given that FPs and other regional programmes assist in patent increase, we expect that well targeted changes in research funding policies have the ability to affect patenting as well.

The results of this study could also prove valuable even to other systems, since a new criterion has been added that can in some cases identify differences in real versus randomized networks much easier. For example, in cases of real social multiplex networks, it can perhaps help identify in a new way whether there is a preference or not for the friend of a friend to be a friend. In any case, since the results hold true for both a sparse (patent) and a dense (FP) layer, as well as their common multiplex network, it is possible that many other such systems can use our results and the methodology that we followed.

CRediT authorship contribution statement

Konstantinos Angelou: Software, Formal analysis, Data curation, Visualization, Writing - Original Draft. **Michael Maragakis**: Conceptualization, Writing - Original Draft, Project administration, Methodology, Validation, Writing - Review & Editing. **Kosmas Kosmidis**: Conceptualization, Methodology, Writing - Review & Editing. **Panos Argyrakis**: Writing - Review & Editing, Project administration.

Acknowledgments

Results presented in this work have been produced using the Aristotle University of Thessaloniki (AUTh) High Performance Computing Infrastructure and Resources.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2021.101192.

References

Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics, 6*(3), 403–412. 10.1016/j.joi.2012.01.002.

Angelou, K., Maragakis, M., Kosmidis, K., & Argyrakis, P. (2020). A hybrid model for the patent citation network structure. *Physica A: Statistical Mechanics and its Applications, 541*. 10.1016/j.physa.2019.123363.

Antal, T., Krapivsky, P. L., & Redner, S. (2006). Social balance on networks: The dynamics of friendship and enmity. *Physica D: Nonlinear Phenomena, 224*(1–2), 130–136. 10.1016/j.physd.2006.09.028.

Battiston, F., Nicosia, V., & Latora, V. (2014). Structural measures for multiplex networks. *Physical Review E, 89*(3), 032804. 10.1103/PhysRevE.89.032804.

Berkowitz, P., Hardy, S., Muravska, T., & Bachtler, J. (2016). *EU cohesion policy: Reassessing performance and direction*. Routledge. 10.4324/9781315401867.

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., & Zanin, M. (2014). The structure and dynamics of multilayer networks. 10.1016/j.physrep.2014.07.001

Bordons, M., Aparicio, J., Gonzlez-Albo, B., & Daz-Faes, A. A. (2015). The relationship between the research performance of scientists and their position in co-authorship networks in three fields. *Journal of Informetrics, 9*(1), 135–144. 10.1016/j.joi.2014.12.001.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. 10.1126/science.1165821

Broekel, T., Fornahl, D., & Morrison, A. (2015). Another cluster premium: Innovation subsidies and R&D collaboration networks. *Research Policy, 44*(8), 1431–1444. 10.1016/j.respol.2015.05.002.

Choe, H., & Lee, D. H. (2017). The structure and change of the research collaboration network in Korea (2000–2011): Network analysis of joint patents. *Scientometrics, 111*(2), 917–939. 10.1007/s11192-017-2321-2.

Crescenzi, R., & Giua, M. (2020). One or many cohesion policies of the european union? On the differential economic impacts of cohesion policy across member states. *Regional Studies, 54*(1), 10–20. 10.1080/00343404.2019.1665174.

Darvas, Z., Mazza, J., & Mides, C. (2021). European union cohesion project characteristics and regional economic growth. *Working Papers 40970*. Bruegel.

Datar, M., Gionis, A., Indyk, P., & Motwani, R. (2002). Maintaining stream statistics over sliding windows. *SIAM Journal on Computing, 31*(6), 1794–1813. 10.1137/S0097539701398363.

De Stefano, D., & Zaccarin, S. (2013). Modelling multiple interactions in science and technology networks. *Industry & Innovation, 20*(3), 221–240. 10.1080/13662716.2013.791130.

Dimitrova, T., Petrovski, K., & Kocarev, L. (2020). Graphlets in multiplex networks. *Scientific Reports, 10*(1), 1–13. 10.1038/s41598-020-57609-3.

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics, 5*(1), 187–203. 10.1016/j.joi.2010.10.008.

Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., & Zalányi, L. (2013). Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics, 95*(1), 225–242. 10.1007/s11192-012-0796-4.

Haegeman, K., Doussineau, M., & Boden, M. (2015). *Measuring progress in transnational coordination of research programming in Europe*. Publications Office of the European Union https://doi.org/10.2791/953432.

Huang, M. H., Dong, H. R., & Chen, D. Z. (2012). Globalization of collaborative creativity through cross-border patent activities. *Journal of Informetrics, 6*(2), 226–236. 10.1016/j.joi.2011.10.003.

Kim, J., & Yoo, J. (2019). Science and technology policy research in the EU: From framework programme to HORIZON 2020. *Social Sciences, 8*(5), 153. 10.3390/socsci8050153.

Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2013). Multilayer networks. *SSRN Electronic Journal*. 10.2139/ssrn.2341334.

Kong, X., Mao, M., Jiang, H., Yu, S., & Wan, L. (2019). How does collaboration affect researchers' positions in co-authorship networks? *Journal of Informetrics, 13*(3), 887–900. 10.1016/j.joi.2019.07.005.

Kuper, L., & Wolff, K. (1951). The sociology of Georg Simmel. *The British Journal of Sociology, 2*(3), 260. 10.2307/586725.

Landini, F., Malerba, F., & Mavilia, R. (2015). The structure and dynamics of networks of scientific collaborations in Northern Africa. *Scientometrics, 105*(3), 1787–1807. 10.1007/s11192-015-1635-1.

Li, D., Wei, Y. D., & Wang, T. (2015). Spatial and temporal evolution of urban innovation network in China. *Habitat International, 49*, 484–496. 10.1016/j.habitatint.2015.05.031.

Li, M., & O'Riordan, C. (2013). The effect of clustering coefficient and node degree on the robustness of cooperation. In *2013 IEEE congress on evolutionary computation* (pp. 2833–2839). 10.1109/CEC.2013.6557913.

Liu, W., Tao, Y., Yang, Z., & Bi, K. (2019). Exploring and visualizing the patent collaboration network: A case study of smart grid field in China. *Sustainability, 11*(2). 10.3390/su11020465.

Ma, X., & Ma, Y. (2019). The local triangle structure centrality method to rank nodes in networks. *Complexity, 2019*. 10.1155/2019/9057194.

Magerman, T., Van Looy, B., & Debackere, K. (2015). Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology. *Research Policy, 44*(9), 1702–1713. 10.1016/j.respol.2015.06.005.

Mittal, R., & Bhatia, M. P. (2018). Cross-layer closeness centrality in multiplex social networks. *2018 9th International conference on computing, communication and networking technologies, ICCCNT 2018*. Institute of Electrical and Electronics Engineers Inc.. 10.1109/ICCCNT.2018.8494042.

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404–409. 10.1073/pnas.98.2.404.

Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics, 4*(4), 475–482. 10.1016/j.joi.2010.04.003.

van der Pol, J., & Rameshkoumar, J. P. (2018). The co-evolution of knowledge and collaboration networks: The role of the technology life-cycle. *Scientometrics, 114*(1), 307–323. 10.1007/s11192-017-2579-4.

Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., & Gabrielli, A. (2019). Unfolding the innovation system for the development of countries: Coevolution of science, technology and production. *Scientific Reports, 9*(1), 16440. 10.1038/s41598-019-52767-5.

Reillon, V. (2017). EU framework program for research and innovation. Evolution and key data from FP1 to horizon. In *2020 in view of FP9.* (p. 38). 10.2861/60724.

Scherngell, T., & Barber, M. J. (2009). Spatial interaction modelling of cross-region R&D collaborations: Empirical evidence from the 5th EU framework programme*. *Papers in Regional Science, 88*(3), 531–546. 10.1111/j.1435-5957.2008.00215.x.

Scott, J. (1991). *Social network analysis: A handbook.*. Thousand Oaks, CA, US: Sage Publications, Inc.

Seshadhri, C., Pinar, A., & Kolda, T. G. (2014). Wedge sampling for computing clustering coefficients and triangle counts on large graphs. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 7*(4), 294–307. 10.1002/sam.11224.

Simmel, G. (1908). *Soziologie. Untersuchungen über die formen der vergesellschaftung*. Leipzig: Verlag von Duncker & Humblot. 10.3790/978-3-428-53725-9.

Spiegel, M. (2017). *Schaum's outline of statistics* (6th ed.). McGraw-Hill Education.

Sun, Y. (2016). The structure and dynamics of intra- and inter-regional research collaborative networks: The case of China (1985–2008). *Technological Forecasting and Social Change, 108*, 70–82. 10.1016/j.techfore.2016.04.017.

Tahmooresnejad, L., & Beaudry, C. (2018). The importance of collaborative networks in Canadian scientific research. *Industry and Innovation, 25*(10), 990–1029. 10.1080/13662716.2017.1421913.

Varela-Vázquez, P., González-López, M., & del Carmen Sánchez-Carreira, M. (2019). The uneven regional distribution of projects funded by the EU framework programmes. *Journal of Entrepreneurship, Management and Innovation, 15*(3), 45–72. 10.7341/20191532.

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge University Press. 10.1017/CBO9780511815478.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world9 networks. *Nature, 393*(6684), 440–442. 10.1038/30918.

You, H., Li, M., Jiang, J., Ge, B., & Zhang, X. (2017). Evolution monitoring for innovation sources using patent cluster analysis. *Scientometrics, 111*(2), 693–715. 10.1007/s11192-017-2318-x.