A Letters Journal Exploring the Frontiers of Physics

OFFPRINT

Dynamics of regional multilinks in research innovation temporal networks

K. Angelou, M. Maragakis, K. Kosmidis and P. Argyrakis

EPL, **130** (2020) 28001

epl A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

epljournal.org

OVER
638,000
full-text downloads in 2017

Average submission to online publication
100 DAYS

21,500
citations in 2016

We greatly appreciate the efficient, professional and rapid processing of our paper by your team.

Cong Lin
Shanghai University

Four good reasons to publish with EPL

1. **International reach** – more than 3,300 institutions have access to EPL globally, enabling your work to be read by your peers in more than 90 countries.

2. **Exceptional peer review** – your paper will be handled by one of the 60+ co-editors, who are experts in their fields. They oversee the entire peer-review process, from selection of the referees to making all final acceptance decisions.

3. **Fast publication** – you will receive a quick and efficient service; the median time from submission to acceptance is 75 days, with an additional 20 days from acceptance to online publication.

4. **Green and gold open access** – your Letter in EPL will be published on a green open access basis. If you are required to publish using gold open access, we also offer this service for a one-off author payment. The Article Processing Charge (APC) is currently €1,400.

Details on preparing, submitting and tracking the progress of your manuscript from submission to acceptance are available on the EPL submission website, **epletters.net**.

If you would like further information about our author service or EPL in general, please visit **epljournal.org** or e-mail us at **info@epljournal.org**.

EPL is published in partnership with:

European Physical Society | Società Italiana di Fisica | EDP Sciences | IOP Publishing

epljournal.org

A Letters Journal Exploring the Frontiers of Physics

April 2020

EPL, **130** (2020) 28001 www.epljournal.org
doi: 10.1209/0295-5075/130/28001

Dynamics of regional multilinks in research innovation temporal networks

K. Angelou[1,2], M. Maragakis[1,2,3], K. Kosmidis[1,2] and P. Argyrakis[1,2]

[1] *Department of Physics, University of Thessaloniki - Thessaloniki, Greece*
[2] *Center of Complex Systems, University of Thessaloniki - Thessaloniki, Greece*
[3] *Department of Physics, International Hellenic University - Kavala, Greece*

received 20 February 2020; accepted in final form 27 April 2020
published online 14 May 2020

PACS 89.75.-k – Complex systems
PACS 64.60.aq – Networks
PACS 64.60.De – Statistical mechanics of model systems (Ising model, Potts model, field-theory models, Monte Carlo techniques, etc.)

Abstract – In this paper, we examine the evolution of a temporal multiplex innovation network, and develop a kinetic model that describes the dynamical process behind its growth. The multiplex consists of two collaboration network layers. The nodes of the first layer are the European regions of the participants of the EU Framework Programme (FP) projects, and the nodes of the other layer are the European regions of the patent inventors. A link between two regions exists, when scientists associated with those regions have collaborated in an FP project or inventors in a patent. The analysis has been conducted using the notion of multilinks, which essentially describes differences and similarities in the connectivity between identical nodes in both layers. A sliding windows method was employed in order to study the network in various time periods, and all multilinks were calculated in each window. All three types of multilinks were studied (Framework Programme $(1, 0)$, patents $(0, 1)$ and common ones $(1, 1)$, where links between the same regions exist in both layers). Results indicate that all multilink types exhibit a roughly similar growth pattern through the course of 16 years with few observable changes. The results also point out that patents are the driving force for the creation of common multilinks early on in time, while this is reversed later on. We suggest a simple kinetic model of 3 differential equations and 6 parameters that adequately describes the system dynamics. The parameter values exhibit a surprisingly small variation for large periods of time. We believe that this model could easily be extended to other systems with links which are added or removed (birth/death), or even to multiplex networks with more layers.

Introduction. – Networks are a tool for studying complex systems. Until recently, all studies on networks assumed that the individual components communicate with each other only through a single type of connection, which is a rather crude approximation. A more realistic approach would be to examine every possible relationship between the network nodes and not just the obvious ones. This would allow for the extraction of any hidden information and system characteristics, that a single kind of connection would simply not reveal. Such networks, whose nodes interact in various ways (or those whose each connection is represented in a different layer) are called multilayer networks [1,2]. The notion, however, of multiple interactions between nodes is not new [3], especially among sociologists who proposed to study social networks using various types of connections between the nodes [4,5].

Multilayer networks, like single networks, are widely used in various thematic areas. Pilosof *et al.* [6] study ecological systems using multilayer networks, Zitnic *et al.* [7] build a multilayer network with interactions of different human tissues, Muldoon *et al.* [8] describe the advantages of studying the brain dynamics using multilayer networks and Bargigli *et al.* [9] study the interbank system.

A sub-category of multilayer networks are the multiplex networks [1,10]. Their characteristic is that each layer contains the exact same nodes as the other layers and their difference is the type of interaction between them. Like multilayer networks, the notion of multiplexity is well established [11,12] and has also applications in various fields. Chodrow *et al.* [13] study an urban transportation system using multiplex network analysis, Stella *et al.* [14] model the mental lexicon of English-speaking toddlers as a multiplex network and De Domenico *et al.* [15] use multiplex networks to study the human brain.

The purpose of the current study is to understand the dynamic evolution of research and innovation multiplex networks, which consist of two layers of collaboration networks. Some examples of previous studies that have been conducted on temporal multiplex networks are by Liu *et al.* [16] who use time-varying multiple networks in order to find out how contagion is affected by social network features, Rakshit *et al.* [17] who study the synchronization in multiplex neuronal network, Starnini *et al.* [18] who study the temporal correlations between the layers of real social multiplex networks and Timme *et al.* [19] who use individual neurons as multiplex networks by considering their connections as time dependent. The main aim is to propose a model that describes the combined evolution of the two linked networks and extract any trends that may occur.

The section "Description of data" contains the data used for the analysis. The section "Model" describes the model developed for the analysis of the data. The section "Results and discussion" presents and analyses the results. Finally, we present the conclusions of this work.

Description of data. – In the current study we use two datasets to construct an innovation network made of patents and a research network made of the collaborative EU projects. The first dataset is that of patent data from the European Patent Office (EPO) for the period between 2000 and 2016, and it is provided by the Organization for Economic Co-Operation and Development (OECD). It includes information such as the name of each patent inventor, their stated geographic location at that time, and several other fields, some of which we will use in our study. More specifically, we are interested in the geographic aspect of the collaborations that occurred between inventors (Universities, Institutes, companies, individuals and others) for the creation of patents. Thus, we use the provided "Nomenclature of Territorial Units for Statistics" (NUTS)[1] codes, which is a classifier that indicates the region in the EU territory that each inventor belongs to. We will be using the NUTS2 level code that divides a country into basic regions (used primarily for the application of regional policies) of typical minimum population of 800000 and up to 3000000.

The second dataset contains the projects funded by the European Commission under the Framework Programmes 5–7 and Horizon 2020 (FP). These data were derived from the Community Research and Development Information Service (CORDIS).

CORDIS provides geographic details for each participant (such as the city, the address and the postal code), however, the NUTS codes were not included. In order to acquire the NUTS codes we used the geopy[2] package for python to convert each city to the respective coordinates. The coordinates were then used along with their respective shape files (provided by the PlanetData EU Network of Excellence[3]) to extract the NUTS code for each participant. Given the regional restrictions of the second dataset, we focus on the EU and Associated Countries.

In both datasets, the NUTS codes have been updated to their latest version, as there have been changes throughout the last 20 years. These two datasets are used to create two separate networks that have the same exact nodes. The nodes of these two networks represent the NUTS2 codes (regions) of the scientists/participants and an undirected link exists between two nodes when there is collaboration between them. For example, a collaboration between institutes of 6 different regions in the scientific collaboration network means that a full network of these 6 regions will be added (this means a total of 15 undirected links). The number of unique NUTS2 codes that exist in both networks is 330, and the number of links of the patent network is 4846 and that of FP collaboration network is 38262. The total number of possible connections is $N = n(n-1)/2 = 330 \times 329 = 54285$.

It should be noted that there is a significant difference between the two datasets. While we do have information for the birth and duration of a scientific/research collaboration, and we, thus, know that FP links last for a specific amount of time, we have no such information on patents. More specifically, the collaborations typically last for the period of the duration of the project (or slightly more) and they then generally terminate, which in our case means that the link dies out. On the other hand, patent links are considered to last from their registration date to the end of our study period, as we have no information on whether paid patent protection stops with the data we possess, and such information can only be purchased at high cost. However, the typical duration of protection for a patent, which can last up to 20 years, provided the fees are being paid, is significantly longer than that of an EU-project–based collaboration.

Model. – In the current study we use multilinks to analyse the multiplex network. A multilink is an indicator of whether two nodes are connected in all of the different layers of a multiplex network [20,21]. In a multiplex network with 2 layers (layer A and layer B) there are 4 possible different types of multilinks. Multilink $(1,1)$ exists when the nodes are connected in both layers, $(0,0)$ when the nodes are not connected in either of the two layers, and $(1,0)$ or $(0,1)$ indicate that two nodes are connected in only one of the two layers (layer A or layer B respectively). An example of a multilink is shown in fig. 1, where all possible combinations of multilinks are presented

[1] https://ec.europa.eu/eurostat/web/nuts/background.

[2] https://geopy.readthedocs.io/en/stable/.

[3] http://www.planet-data.eu/.

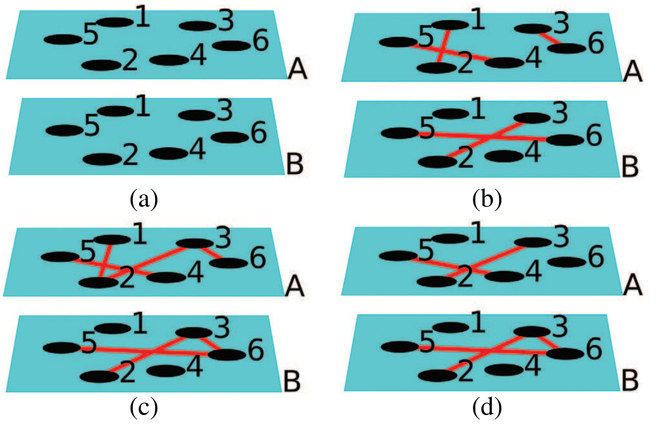

Fig. 1: The evolution of a multiplex network with two layers; A is the upper layer and B is the bottom layer. (a) The networks at their initial state. The nodes are exactly the same on the two layers and there are no links between them. Thus, there are 15 undirected multilinks $(0,0)$ and no multilinks $(1,1)$, $(0,1)$ and $(1,0)$. (b) Links are inserted into the networks. There are no common links, so multilinks $(1,1)=0$. Layer A has 3 links that do not exist in layer B (multilinks $(1,0)=3$) and layer B has 2 unique links (multilinks $(0,1)=2$). (c) More links are inserted into the two layers. Now, there are common links between the layers (*e.g.*, links (2-3) and (3-6)) resulting to multilinks $(1,1) = 2$. (d) Some links from layer A have been removed ("died") which resulted to the decrease of multilinks $(1,1)$ and $(1,0)$ and the increase of multilinks $(0,1)$. The graphs have been drawn using the pymnet library [1].

and, even more, a removal (death) of a link occurs, from fig. 1(c) to fig. 1(d).

Here, we suggest a simple kinetic model that describes the time evolution of the number of multilinks in our multiplex network. The model will be described by three ordinary differential equations; two for multilinks $(1,0)$ and $(0,1)$ that exist only in either of the two layers (eqs. (1), (2)), and one equation for multilinks $(1,1)$ whose nodes are connected in both layers (eq. (3)). There is no need for an equation that describes the case of multilinks $(0,0)$. The notion behind this model is quite simple. We assume that collaborations in either layer occur depending on the number of the remaining connections, where the remaining connections are the total possible connections minus the connections that have already been made. Multilinks in the patent layer (1, 0) can also increase in number, when a common multilink (1, 1) terminates, meaning that an FP project ends, and, thus, ends the collaboration between the participants. As a result, the previous common multilink (1, 1) now changes to patent multilink (1, 0). The decrement of the number of either multilink (1, 0) or (0, 1), resulting in their change to (1, 1) multilink, can also occur when a common link between the layers is created. The common links increase proportionally to the number of the existing collaborations of both networks and decrease when a connection dies. The set of equations that describe the current model are the following:

$$\frac{\mathrm{d}F}{\mathrm{d}t} = -w_F \cdot F \cdot P - d_F \cdot F + a_F \cdot (N - P - F - C), \quad (1)$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = -w_P \cdot F \cdot P + d_C \cdot C + a_P \cdot (N - P - F - C), \quad (2)$$

$$\frac{\mathrm{d}C}{\mathrm{d}t} = (w_P + w_F) \cdot F \cdot P - d_C \cdot C, \quad (3)$$

where F is the number of links on the EU Framework Programme layer, P is the number of links on the Patent layer, C is the number of common links between the two layers, N is the number of all possible links. This value depends only on the number of NUTS2 regions and thus it is a constant for this system, w_F is a rate constant to describe the number of links that stop being links only on the scientific collaboration layer and are now links to both layers, w_P is a rate constant to describe the number of links that stop being links only on the patent layer and are now links to both layers, d_F is a rate constant to describe the number of links that die from the scientific collaboration network, d_C is a rate constant to describe the number of links that die from the common layer and are added to the patent layer, a_F is a rate constant to describe the number of links that are created only on the scientific collaboration layer and a_P is a rate constant to describe the number of links that are created only on the patent layer.

For our study, the sliding windows method [22,23] will be used in order to divide the network into smaller subnetworks with various starting dates. More specifically, there are more than 2000 unique dates that indicate when patents are recorded in the Patent Offices or projects belonging to EU Framework Programmes start. We will use the first 1400 of these dates so that each window has a span of 8 years. Links will be inserted into the multiplex networks and at each time-step the number of multilinks $(1,1)$, $(1,0)$ and $(0,1)$ will be calculated.

Upon finding the evolution of the multilinks we apply the model on the results, so as to extract the parameters values that best converge to them. Figure 2, shows a small part of the multiplex network, where the nodes are placed on a map, according to their NUTS2 code.

Results and discussion. – We begin by performing a temporal analysis on the multilinks behaviour and evolution of the 1400 various multiplex networks. Figure 3 shows three examples of the multilinks evolution for three different starting dates, while all networks have a similar behaviour. All results of patents, FP and common multilinks, have been normalized over their maximum value on each network, so that they all have the same scale. Figure 3(a) shows the common multilinks $(1,1)$ that exist in both layers, fig. 3(b) shows the FP multilinks $(1,0)$ and fig. 3(c) the patent multilinks $(0,1)$. The results show that the number of multilinks that exist only on the FP layer $(1,0)$ is generally much larger than that of the patent

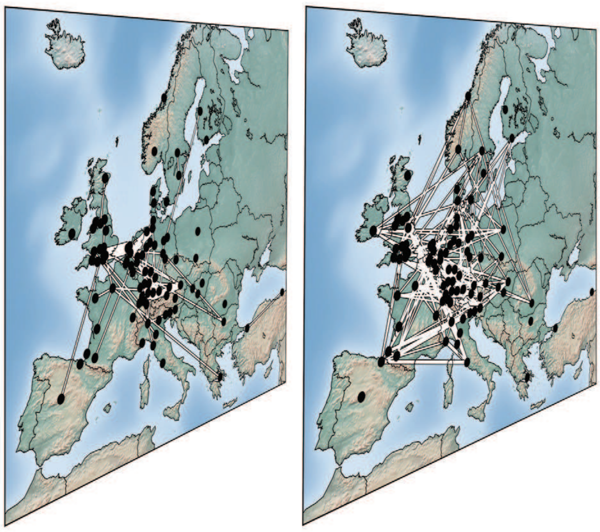

Fig. 2: Geographical representation of a small subset of the multiplex network. The left layer represents the patent layer while the right one represents the FP layer. One-to-one inter-layer connections exist between every NUTS2 region from one layer to the other. The thickness size of the links is proportional to their weight.

multilinks $(0,1)$ or the common multilinks $(1,1)$. Despite the fact that the FP layer links die at some point, this result seems reasonable given that the number of patent collaborations/links is significantly smaller than that of the FP collaborations, as indicated in the "Description of data" section. This is due to the fact that our data are analyzed spatially by their NUTS2 code regions. This contributes to getting far more links on the FP layer than on the patent layer. Whilst the EU has projects which promote large numbers of participants, very few patents have co-inventors from many different regions. Indicatively, the FP project with the highest number of different regions of its participants is 126, which translates to 7875 links, and the corresponding numbers for patents are just 13 regions, and 78 links. Thus, one scientific collaboration project that is large can add up to several thousands of links, while no patent adds more than a hundred links. It should be noted here that there is no way of determining which links have large or small importance in such collaborations, however, this convention has been used extensively in the literature.

Furthermore, although common multilinks $(1,1)$ are fewer in number when compared to the FP multilinks, they are still generally more in number than the patent ones. Thus, we observe that when there is a co-inventor collaboration between two regions, there often is a scientific collaboration as well. In simple terms, common multilinks $(1,1)$ form when either a patent multilink exists $(0,1)$ and a FP link forms between the same two nodes, or the opposite. The question as to which multilink $(0,1)$ or $(1,0)$ drives the formation of the common ones arises.

Results indicate that the 3 multilink types $(1,0)$, $(0,1)$, and $(1,1)$ of the 1400 networks, produced by the sliding window method, have similarities in their behaviour and evolution. Thus, we want to produce a simple model that will describe them. As mentioned in the "Model" section,

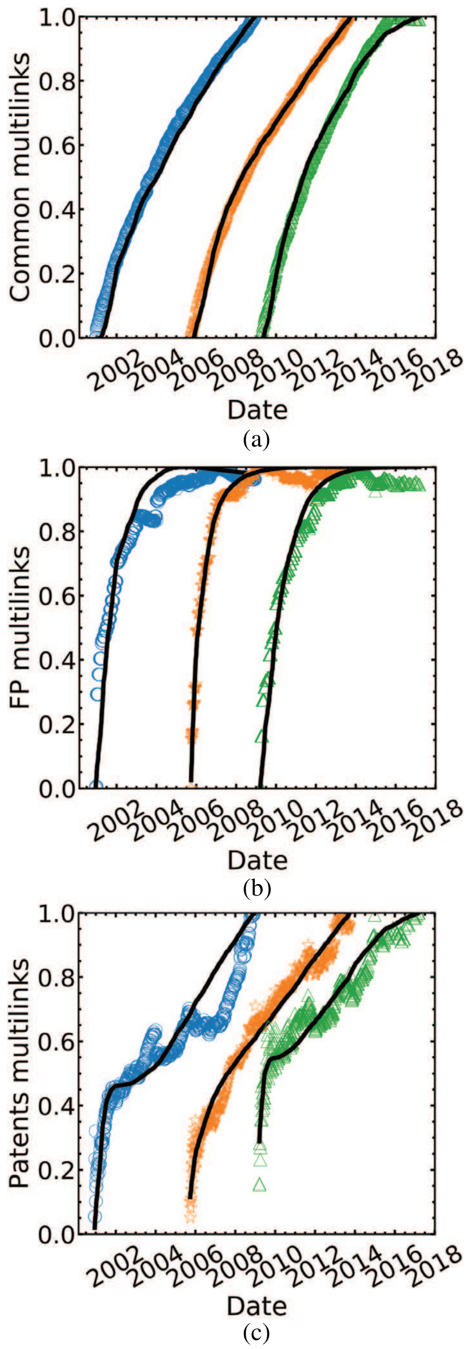

Fig. 3: The evolution over time of the multilinks number for three various cases. (a) Common links and (b) links in Framework Programmes layer show a similar behaviour through the years, while (c) links in the patent layer may show a different one. Black lines represent numerical solutions of the system ODEs (1)–(3), while blue circles, orange stars and green triangles are actual data from the 2001–2009, 2005–2013, and 2009–2017 time windows, respectively.

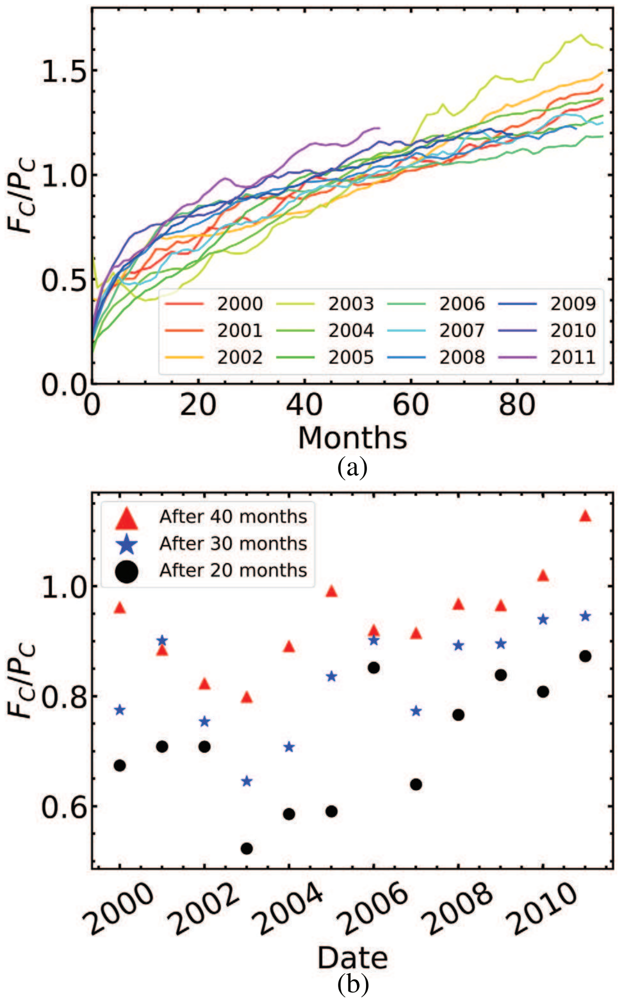

Fig. 4: (a) The ratio of normalized FP multilinks that become common, F_C, over normalized patent multilinks that become common, P_C, *vs.* time in months after which the system has started. Line colours represent specific starting years shown in the legend. (b) The ratio of normalized FP multilinks that become common, F_C, over normalized patent multilinks that become common, P_C, *vs.* year after which the system has started. Symbols represent months after which the ratio was calculated.

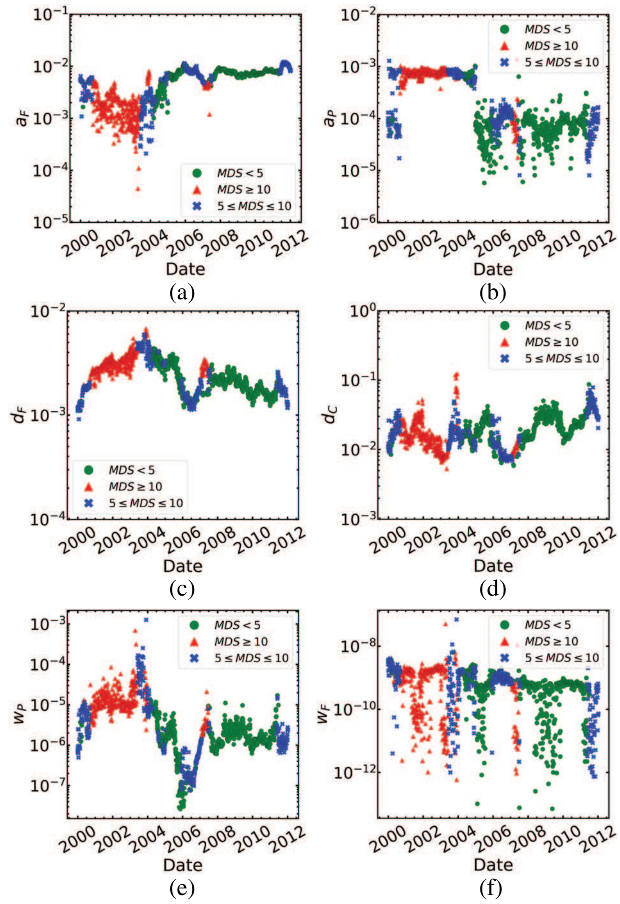

Fig. 5: Evolution of the parameters over time. The parameters have been calculated for all the 1400 various sub-networks and each dot represents the parameters in each one of them. The green color indicates that the fitting of the parameters to the actual values is pretty good, blue indicates lesser confidence in fitting and the red color indicates that the fitting is not quite good.

the model consists of 3 equations, which we apply on each of the 1400 results that occurred in order to extract the 6 parameters of eqs. (1), (2) and (3). The application of the model shows a satisfying agreement with the real data, as shown in fig. 3 with the black lines. The best fitting occurs on the common multilinks that have almost a linear evolution. FP multilinks also have a good fitting as the evolution is linear at the beginning, while later on it reaches a plateau. The patent multilinks behaviour is not fitted well, since their evolution presents small sharp transitions in several windows. However, its trend is followed satisfactorily by the model.

To measure which multilink (FP or patent one) is the driving force for the creation of the common multilinks, we calculated the ratio of normalized FP multilinks that become common, F_C, over the normalized patent multilinks that become common, P_C. The results are presented in fig. 4, where it is clear that for all years (fig. 4(a)) the ratio starts off with a value smaller than 1, meaning that patents play a more significant role in how the common multilinks will be formed. To be more specific, and given that patent multilinks are far fewer in absolute numbers, the patent multilinks formed in the first months of the multiplex network formation are much more likely to become common than their respective FP ones. As time progresses, and the network grows, this is reversed. This is due to the fact that all patent multilinks between "active" NUTS2 regions have already become common, while several FP multilinks between such regions have not yet collaborated to form a patent link.

Next, upon applying the model to the 1400 networks, we extract the parameter values which are shown in fig. 5. In order to acquire the values that will best converge we use the model sum of squares (MDS) [24], *i.e.*, the sum of the squared differences between the models predictions and the actual data, which measures the variation between the real data and the model. We observe the existence of periods where we were unable to determine a set of rate constants that leads to a model having a very good agreement with the data (*i.e.*, MDS not sufficiently low). These periods are met in the early years of study (2000–2004) and to a lesser extent around 2007 (red triangles). There are also cases where the convergence is not perfect, however it is within acceptable boundaries (blue × symbol). Finally, green dots indicate that the convergence is very good. We observe a significant amount of such examples (for almost half of the networks). The above results indicate that the model is quite resilient and that it can simulate the evolution and behaviour of the real data quite well. The resulting parameter values seem to have small variation around an average value for a significant time period (2005–2012), as seen in the Supplementary Material `Supplementarymaterial.pdf` (SM).

Further analysis of the distributions of values of each parameter (see the SM), shows that, by using the boxplot way of representing the distribution, half of the values of each parameter have small variations. By taking into account only patents and collaboration projects after 2005, these variations become even smaller. This is due to the fact that the period 2000 to 2004 includes projects from the end of FP5 and the start of FP6, which contain rather noisy data.

The set of equations describing this behavior can easily be extended to all kinds of data where we have a birth/death dynamics, or even multiplex networks of more than two layers.

Conclusions. – In summary, the present study quantifies the multilinks that exist in the multiplex network of patent co-inventions and scientific collaborations resulting from a spatial analysis of the real data. All research projects and co-invented patents are treated as fully connected components whose nodes are the geographic NUTS2 regions the researchers belong to. The constructed multiplex network of all NUTS2 EU regions exhibits, after some time, multilinks of all 3 types. The dynamics of these multilinks are analyzed.

We find that in early times the patent network acts as a driving force for the creation of the common multilinks. This means that it is more probable that the common multilink is created given that the corresponding patent link already exists. However, in later stages of the network growth, this situation is reversed.

To describe the changes observed in multilinks $(1,0)$, $(0,1)$, $(1,1)$, we constructed a model consisting of a set of 3 ordinary differential equations that depend on a total of 6 independent parameters. The model takes into account the creation of new links on both layers, the death of links in the scientific collaboration layer, and can be used to calculate dynamically the numbers of multilinks of all types. The model used can easily be extended to any type of data that have dynamics of the sort of birth and/or death, in two or more layers. The results from the best fitting (model sum of squares method) to the real data are obtained. The parameter values exhibit a surprisingly small variation for such a large period of time.

* * *

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

REFERENCES

[1] Kivela M., Arenas A., Barthelemy M., Gleeson J. P., Moreno Y. and Porter M. A., *SSRN Electron. J.*, **2** (2014) 203.

[2] Aleta A. and Moreno Y., *Annu. Rev. Condens. Matter Phys.*, **10** (2019) 45.

[3] Craven P. and Wellman B., *Sociol. Inq.*, **43** (1973) 57.

[4] Wasserman S. and Faust K., *Social Network Analysis* (Cambridge University Press, Cambridge) 1994, doi:10.1017/CBO9780511815478.

[5] Scott J. and Carrington P. J., *The SAGE Handbook of Social Network Analysis* (SAGE) 2011.

[6] Pilosof S., Porter M. A., Pascual M. and Kéfi S., *Nat. Ecol. Evol.*, **1** (2017) 0101.

[7] Zitnik M. and Leskovec J., *Bioinformatics*, **33** (2017) i190.

[8] Muldoon S. F. and Bassett D. S., *Philos. Sci.*, **83** (2016) 710.

[9] Bargigli L., di Iasio G., Infante L., Lillo F. and Pierobon F., *Quant. Finance*, **15** (2015) 673.

[10] Boccaletti S., Bianconi G., Criado R. and del Genio C. I., *Phys. Rep.*, **544** (2014) 1.

[11] Gluckman M., *The Judicial Process Among the Barotse of Northern Rhodesia* (Manchester University Press, Manchester) 1955.

[12] Verbrugge L. M., *Soc. Forces*, **57** (1979) 1286.

[13] Chodrow P. S., Al-Awwad Z., Jiang S. and González M. C., *PLoS ONE*, **11** (2016) e0161738.

[14] Stella M., Beckage N. M. and Brede M., *Sci. Rep.*, **7** (2017) 46730.

[15] De Domenico M., Sasai S. and Arenas A., *Front. Neurosci.*, **10** (2016) 326.

[16] Liu Q. H., Xiong X., Zhang Q. and Perra N., *Phys. Rev. E*, **98** (2018) 062303.

[17] Rakshit S., Bera B. K. and Ghosh D., *Phys. Rev. E*, **98** (2018) 032305.

[18] Starnini M., Baronchelli A. and Pastor-Satorras R., *Sci. Rep.*, **7** (2017) 1.

[19] Timme N., Ito S., Myroshnychenko M., Yeh F.-C., Hiolski E., Hottowy P. and Beggs J. M., *PLoS ONE*, **9** (2014) e115764.

[20] Bianconi G., *Phys. Rev. E*, **87** (2013) 062806.

[21] Menichetti G., Remondini D., Panzarasa P., Mondragón R. J. and Bianconi G., *PLoS ONE*, **9** (2014) e97857.

[22] Datar M., Gionis A., Indyk P. and Motwani R., *SIAM J. Comput.*, **31** (2002) 1794.

[23] Koskimäki H., Huikari V., Siirtola P., Laurinen P. and Röning J., *Activity recognition using a wrist-worn inertial measurement unit: A case study for industrial assembly lines*, in *2009 17th Mediterranean Conference on Control and Automation, MED 2009* (IEEE Computer Society) 2009, pp. 401–405, doi:10.1109/MED.2009.5164574.

[24] Maxwell S. E. and Delaney H. D., *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (Wadsworth/Thomson Learning, Belmont, Cal., USA) 1990.