Physica A 521 (2019) 476–483

Contents lists available at ScienceDirect

Physica A

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

A structural analysis of the patent citation network by the k-shell decomposition method

Konstantinos Angelou, Michael Maragakis, Panos Argyrakis *

Department of Physics and Center of Complex Systems, University of Thessaloniki, 54124 Thessaloniki, Greece

HIGHLIGHTS

- K-shell decomposition was applied to patent citation network to study its structure.
- Structure-wise highest degree patents can be less important than lower degree patents.
- Patents from N. America and Europe are the most important for the network's structure.
- Chemistry and Human Necessities are the dominant, among other equally sized sections.

ARTICLE INFO

Article history:
Received 8 August 2018
Received in revised form 1 November 2018
Available online 23 January 2019

Keywords:
Patent
Patent citation network
k-shell decomposition
Network structure

ABSTRACT

We study the patent citation network by means of structural analysis. Specifically, we use the k-shell decomposition method to reveal hidden structural properties. We aim to identify the characteristics of the most important nodes in the network, as well as extract any existing patterns in the connectedness of these nodes. We examine properties of the nodes such as the geographic origin and the thematic area they belong. Results yield unexpected information on which are the most important nodes in the network, and how these nodes are geographically and thematically distributed.

1. Introduction

A patent is the set of exclusive rights that some individual or company has on an invention. Patents provide information on the current technological and innovation trends and are used to protect inventions in both the industrial and academic field. Patent networks are created because of the co-authorship of patents. More specifically, these networks fall into the category of collaboration networks, where the inventors are the nodes. Two nodes are connected (in an undirected way), if the two inventors have worked together for the creation of a patent. Engelsman et al. map the technology by using patent data [1]. Research on patent networks aims to observe differences between research areas when it comes to establishing collaboration ties with local, national, or international partners, and to determine to what extent the collaboration can influence the patent transfer [2]. Hoekman et al. analyze the effects of geographical and institutional distance on research collaboration through the networks of scientific publications and patents [3] and Briggs et al. study the impact of the co-ownership of patents from multiple countries on patent quality [4].

Citation networks, are complex networks formed by the citation of one network node to another. For example, citation networks are found in research collaboration analysis where one scientific article refers to another [5–7]. The articles are the nodes and references to previously published articles are the (directed) connections. Other examples are blogs where one blogger refers to another [8], and movies where one refers to a previous one [9]. Radicchi et al. review the structural

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

https://doi.org/10.1016/j.physa.2019.01.063

properties of citation networks [10]. Subelj et al. compare the structural properties of various citation databases [11], and Eom et al. search for the type of function which best describes the observed citation distributions with data from papers published in journals of the American Physical Society [12].

Patent citation networks can be constructed from the reference of one patent to a previous one (prior art). These references are important to preserve the originality of an idea, and to mention its correlation with the one being referred to. In patent citation networks, patents are the nodes and citations to other patents are the links between the nodes. Patent citation networks, like all other citation networks, are always directed. They are important when one needs to identify patents that are highly cited and, thus, can be used to reflect on the importance of technological needs of times. Alcacer et al. and Criscuolo et al. examine the impact of citations added from the examiners [13,14], Maurseth et al. deal with knowledge flow through patent citations [15], and Karvonen et al. make use of patent citations to analyze technological convergence [16].

The main goal of the current study is to identify the underlying geographic and thematic relations between the individual patents in the patent citation network. Additionally, we seek to find the importance of geographic separation in a continental level with regard to the network formation mechanism.

The data used, and the main methodology followed, are shown in further detail in Section 2, Data and Methodology. The Results and Discussion Section 3 shows the results derived from the classical structural analysis and the k-shell decomposition method. Other than the general analysis, further geographical and thematic analysis conducted helps reveal hidden information on the importance of some geographic regions, as well as that of some thematic areas, for the network's structure. Conclusions of this work are summarized in Section 4.

2. Data and methodology

The data used for the analysis that follows are the citations found in patent documents and are provided by the Organization for Economic Co-operation and Development (OECD). The database contains citations of all patents filed at the European Patent Office (EPO) and at the Patent Cooperation Treaty (PCT), published from 1978 onwards. Both databases contain some common patents, but with different application numbers. To avoid duplicates, common patents are identified and then merged. They are stored in the database with their PCT application numbers. The total number of patents in the merged patent citation network examined in our study is 12,126,928, while the number of citations is 21,873,933. Patents are the nodes of the network and the citations are the directed links between them.

Data also contain information on the originating country of each patent that we categorize in continent level. However, some of the patents are filed in Offices that cannot be categorized in a country or continent level (Eurasian Patent Organization, International Bureau of WIPO, Soviet Union (USSR)) and, thus, are categorized as "Others".

The k-shell decomposition method [17–19] can be used to analyze and reveal structural properties where other measures (e.g. degree distribution) fail to or cannot help significantly on their own. More specifically, it is used to identify the hierarchical structures and properties of networks [20]. It also identifies the influential nodes and, thus, this method is commonly used as a measure of importance and well-connectedness for nodes in technological networks [21].

The method decomposes the network into partitions. This partitioning is a straightforward procedure. At first, all nodes with only one neighbor are pruned from the network, along with their connecting links. Because of this procedure, it is possible that some nodes, with originally higher degree (i.e. k=2), drop down to only one connection. These nodes require removal too at a second iteration. For example, a node in a chain has degree 1 if it is in the end of the chain, while its neighbor has originally degree 2. Once the end node is removed, its neighbor has now degree 1 and, thus, must be removed too. This procedure is repeated iteratively until there are only nodes with two or more connections in the network. The removed nodes are assigned at a k-shell with an index 1, representing their coreness in the network, and are considered to form k-shell=1.

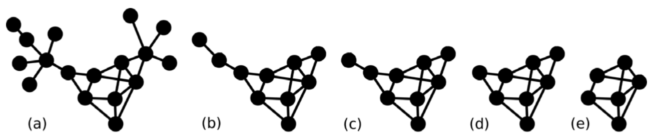

Fig. 1. (a) Initial network structure. (b) Network structure after nodes with one connection are removed. (c) New nodes with only one connection occur. Their removal takes place iteratively, until there are no nodes with one connection. (d) Status of network structure after all iterations have finished and shell 1 has been completely removed. (e) Nodes with two connections have been removed. Only nodes with 3 or more connections remain. They form the core of the network.

Subsequently, all nodes with 2 connections are removed. The process continues until there are only nodes with 3 or more connections. All the nodes removed at this step form k-shell=2. Then nodes with 3 connections are iteratively removed and so on, until there are no nodes left in the network. A schematic of the method is given in Fig. 1.

The result of this pruning process is the partitioning of the network into shells. For small k values the shells belong to the network periphery. As k increases, the size of the shells decreases, but the cores become more interlinked and cohesive [17]. The last k-shell, called the core or nucleus of the network, $k_{s,core}$, is the inner part of the network and is very robust.

The k partitioning from the periphery to the center of the network makes k an effective measure to detect the nodes' spreading capability, as the nodes through which information is disseminated the most are those in the higher k-shells of the network [20,22,23]. Such nodes usually exhibit the network's highest degree numbers.

3. Results and discussion

3.1. Degree distribution

Degree distributions reveal some of the structural properties of a network. The degree k of a node in a network is the number of its connections with other nodes (or the number of its direct neighbors). Therefore, the degree distribution P(k) gives the probability that a randomly chosen node has k connections.

In networks such as the patent citation network we have here, where the links between the nodes are directed (bidirectional links are practically non-existent), the nodes are characterized by two degree distributions; the in- and the out-degree distributions. In-degree of a node is the number of its incoming connections (patents citing this specific patent), while out-degree is the number of its outgoing connections (patents cited by this patent). Thus, the in-degree distribution $P(k_{in})$ is the probability that a randomly chosen node has in-degree k_{in} [similarly, for the out-degree distribution, $P(k_{out})$, it is k_{out}].

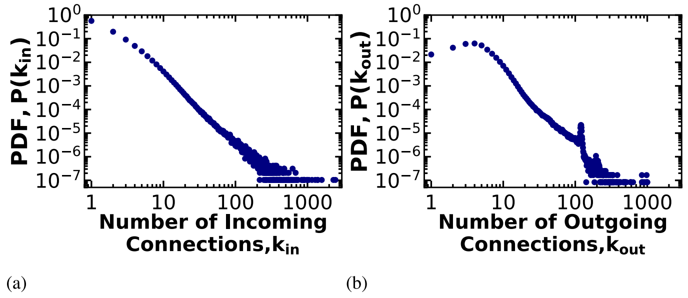

Fig. 2. The probability distribution function of connections of the patents. (a) The distribution of incoming connections follows an approximate power law with exponent $\gamma = 2.75$. (b) The distribution of outgoing connections has $\gamma = 2.94$. The horizontal axis displays the number of connections (incoming or outgoing) and the vertical axis represents the number of patents, normalized. Both axes are in a log–log scale. There are also 7,840,496 patents with no outgoing citations which cannot be shown in the logarithmic plot.

We derive the degree distributions using the patent citation network described earlier, and they are shown in Fig. 2. Both distributions follow a power-law, $P(k) \sim k^{-\gamma}$. The value of the γ exponent of the in-degree distribution is 2.75, while for the out-degree it is 2.94.

Other citation networks have shown similar results. For example, Redner et al. showed that the in-degree citation distribution of papers cataloged by the Institute for Scientific Information for the year 1981 and 20 years of publications in Physical Review D follows a power law with $\gamma \sim 3$ [5]. On the other hand, Vazquez et al. studied the out-degree citation distribution from journals in the period $1991 - 1999$ concluding that it has an exponential tail also with $\gamma \sim 3$ [6].

Power-law distribution suggests that there is a preferential attachment mechanism active in the patent citation network. It also means that the majority of patents have low degrees (small number of citations or references) and there is only a small fraction of patents that are highly cited or citing. Specifically, there are 5,468,112 and 257,636 patents with just one incoming and one outgoing connection respectively, while there are 7,840,496 patents with no outgoing connections, most of which were filed before 1978, and as such they are not available in the data. On the other hand, there is only one patent with 2444 in-degree and one patent with 1002 out-degree. The scale free properties that characterize the network imply that information is transferred quite fast within it and that it is not likely to collapse by the random removal of nodes. The degree distribution provides some useful information on the network. However, it does not fully reveal the complete structure of a network, so other structural measures should be applied, such as the k-shell decomposition method.

3.2. K-shell decomposition

In order to apply the k-shell decomposition method to the patent citation network we treat it as an undirected network [19,24]. The result that occurred is a power law behavior in the k-shell distribution, Fig. 3a. The number of patents decreases as the shell number increases. Noise, due to the finite system size, appears at k-shell=30 and above, with the maximum shell number reaching 90. Thus, according to this method, the innermost core or nucleus of the network is k-shell=90, which contains 226 patents. By observing the core, we see that it is unexpectedly comprised of patents with lower degree, Fig. 3b. In fact, degrees in the core range from 94 to 462, while the highest degree patents are located in much lower k-shell numbers. To observe this in detail we plot the degree distribution for each shell, example of k-shell=10 in Fig. 3c.

It is worth mentioning that the highest degree patents are not part of the k-core, as one would naively expect. Indeed, if we consider that high degree patents are those with more than 400 in and out connections (sum of both in and out), one sees that a large part of high degree patents is located in shells with numbers 57 to 60 and 70. In contrast, shells 86 and 90 have very few high degree patents. Furthermore, the highest degree patent (2444) is located in shell 70 and the second highest in shell 57. The above results have also been confirmed by plotting the average value of the in-degree of the neighbors of node i against the in-degree of node i. In fact, some nodes with very high degree are located in very low shell numbers (e.g. degree 945 in shell 11).

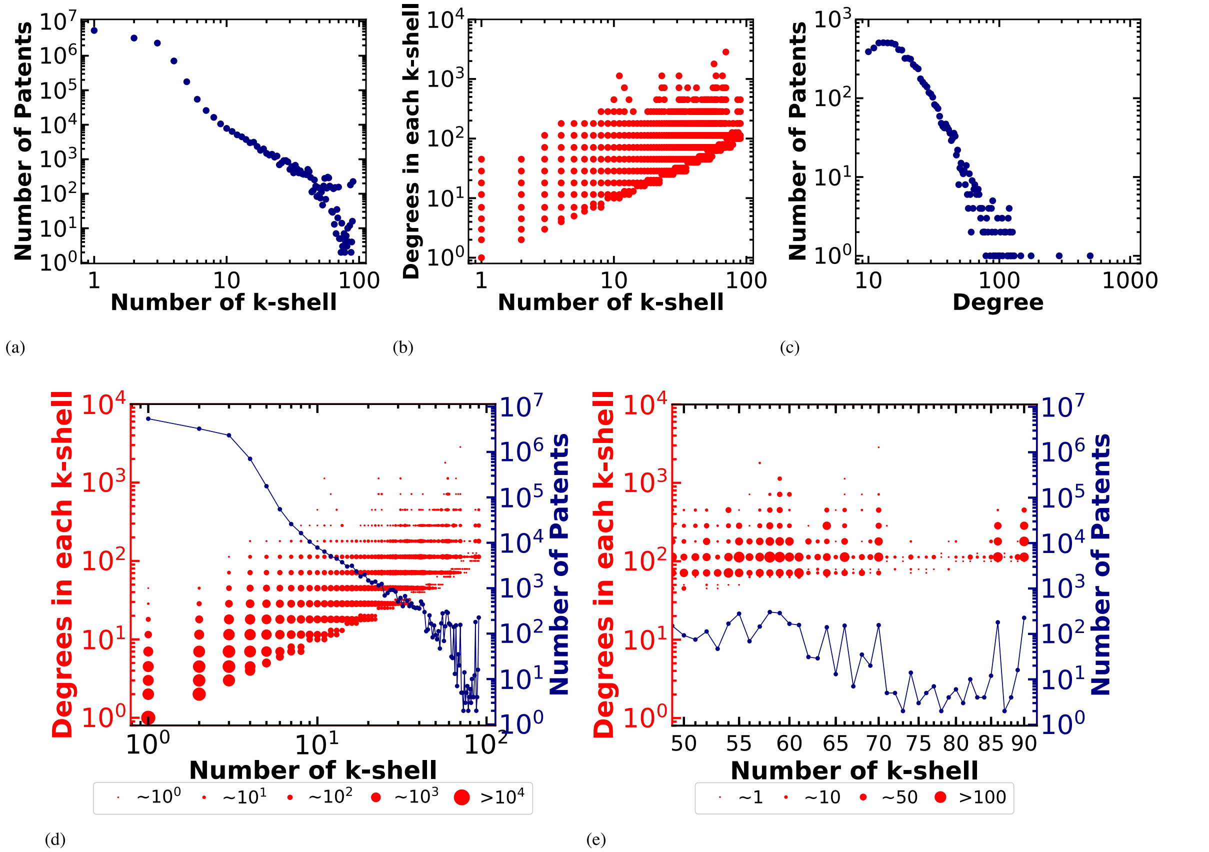

Fig. 3. (a) k-shell distribution (the number of patents in each k-shell) of the entire patent citation network. (b) The distribution of the patent degrees in each k-shell. (c) Example of the degree distribution of k-shell=10. (d) The combined graph of (a)(b)(c). (e) Zoom plot of part (d) in the region of 50 <k-shell< 90. Note that for viewing purposes the scale of the red dots is different in the two plots.

As a result, from a hierarchical and structural point of view, the highest degree patents are not that important for the network. Instead, patents that belong to the nucleus according to k-shell decomposition method have much lower degree.

The same analysis was also applied to shuffled versions of the patent citation network to properly quantify the significance of the finding. More specifically, the citations were shuffled so that the in- and out-degree distributions of each patent were preserved [25,26] . The k-shell decomposition method on these networks resulted to lower core shell number ($shell \approx 40$), but with more higher degree nodes in them. For instance, the cores of the shuffled networks contain many nodes of significantly higher degree (> 500), which is in sharp contrast with the analysis above.

3.3. Geographical analysis

A geographical analysis at a continental level was applied. The purpose of such an analysis was to find out which continents are dominant, whether shells exhibit a geographical preference, and if the highest shells come from the same or different continents.

We noticed that there are many higher shells that contain patents from mostly one continent (Fig. 4a). For example, the core of the network (shell 90) contains patents mostly from North America, shell 61 from Asia and shell 86 from Europe. There are also shells that have patents originating from more than one continent. For example, patents in shell 58 originate from Europe and North America at almost equal numbers. It is, however, uncommon in inner shells for patents to originate from all three continents (e.g. shell 59). This changes in outer shells, where the number of patents is much larger and shells do not belong to just one single thematic or geographical community.

The geographical analysis also shows that Asia, in the given dataset, has about 22% of the total patents. Yet, it is evident in Fig. 4b that apart from the lowest shell (shell 1), the patents that originate from Asia are fewer than those from North America and Europe with some exceptions (i.e. shell 61). Additionally, further temporal analysis of the patents originating from Asia show a significant increase in numbers after 2000 (results not included).

To clearly show that Europe and North America are dominant in higher shells after such an analysis, we placed on top of each continent (Fig. 4c) circles sized according to the number of patents in every shell above shell 50. It is evident that Asia does not have patents in many inner shells, while Oceania has only few in outer shells and Africa and South America have none.

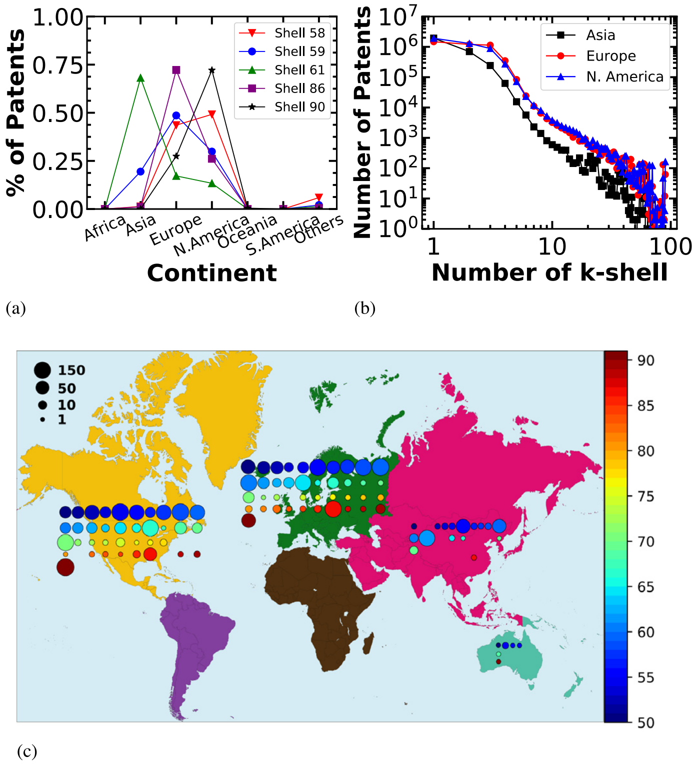

Fig. 4. Geographical analysis of k-shells. (a) Percentage of patents coming from each continent for some indicative higher shells where one, two or even three continents are found in the same shell. (b) Plot of the number of patents in each shell broken down by continent (only the three larger ones are mentioned). (c) Number of patents from each continent for the higher shells (50–90). Circle size is proportional to the number of patents on each shell that originate from the given continent. Indicative sizes, that correspond to specific numbers of patents in each shell, are shown in the upper left corner.

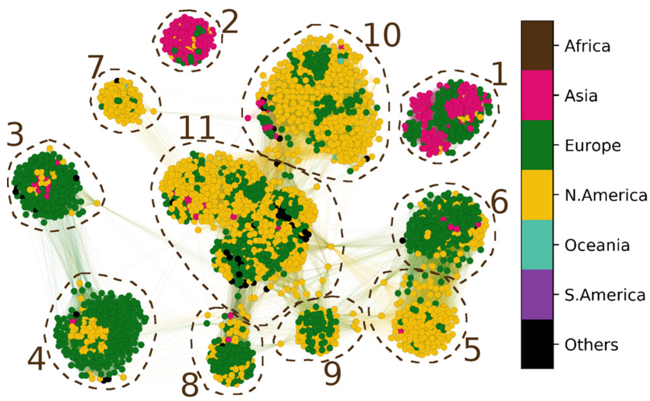

Fig. 5. (Color online) Graphic representation of the Louvain method community detection of the network, as it occurs with the k-shell decomposition method applied for 50 <k-shell< 90. Each continent is assigned a color (see legend). It is clear that large parts of a community often belong to an individual continent. A number is assigned randomly in each community, where dashed lines indicate the borders of each community.

To obtain a better picture of the network's structure, community detection was applied using the Louvain method [27]. This method partitions the network into communities of strongly connected nodes, the nodes of any community being less strongly connected with nodes of any other community. This is accomplished by dividing the network into partitions according to modularity, each one of them always having a local maximum.

Fig. 5 shows this illustratively, where by applying the Louvain method the inner shells (above shell 50) are separated in distinct communities. Colors are used to distinguish the continent of origin of the patent. It is obvious that there are communities where one continent is dominant (i.e. one community which is almost entirely Asia, one community mostly made of European patents, etc.), whereas in others several continents co-exist (i.e. in 9 and 11 Europe and North America). No community contains all three major (Europe, North America, Asia) continents in significant representation, which indicates

Table 1
The percentage of the continent of origin versus the community number.

	Asia	Europe	N.America	Oceania	Others
1	51.53	41.92	6.55	0.00	0.00
2	79.17	6.94	13.89	0.00	0.00
3	4.31	86.12	7.18	0.00	2.39
4	0.78	80.00	18.04	0.00	1.18
5	0.62	12.35	87.04	0.00	0.00
6	1.71	74.79	22.65	0.00	0.85
7	0.00	5.69	91.06	2.44	0.81
8	1.32	60.53	36.18	1.32	0.66
9	0.00	39.06	60.94	0.00	0.00
10	2.16	24.94	70.74	0.48	1.68
11	1.16	38.08	56.96	0.21	3.59

a strong geographic separation between them. There is also no community with patents originating from Asia and North America. Analysis of the continents where patents of each community lie is given in Table 1.

3.4. Thematic analysis

Patents have one or more IPC Codes which are either self assigned by the inventor, or assigned at the review process by a reviewer. An IPC Code corresponds to a scientific–technological area which covers the patent's theme. These IPC Codes classify every patent into one of the eight sections (or more than one for many patents) where the patent belongs to.

An analysis of the sections these patents belong to in every shell (Fig. 6a), shows that Chemistry and Human Necessities become dominant above shell 5, although they are almost equally represented in the entire network along with Performing Operations, Transporting, Physics and Electricity (all five categories differ less than 2.5%). Performing Operations and Physics show a noticeable decrease in importance after shell 5 (with a few exceptions at inner shells), while Electricity, Mechanical Engineering and Fixed Constructions are almost non-existent. It is counter-intuitive that these fields which seemingly have contributed much in modern society and technology have such low representation in the inner shells.

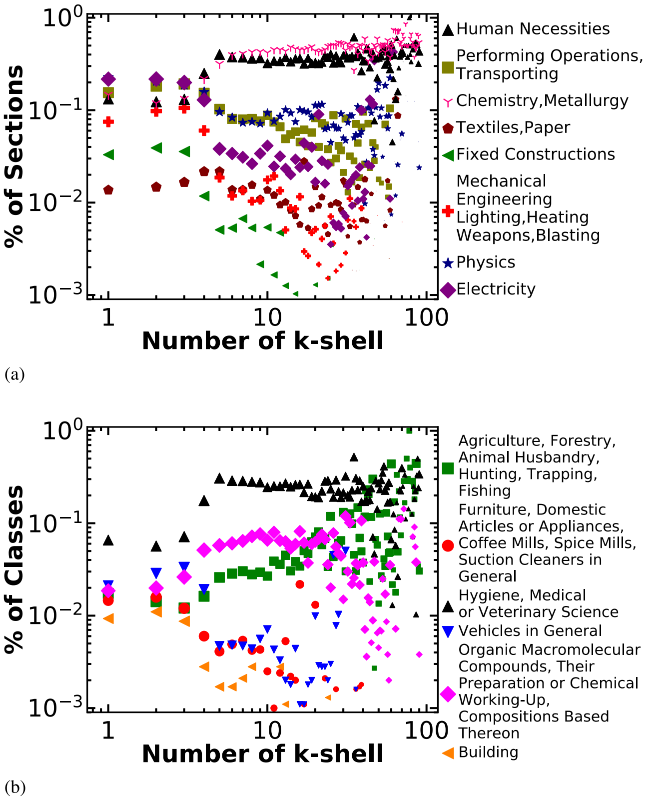

Fig. 6. Thematic analysis by section and class of k-shells. (a) Probability of a patent to belong to a specific section for every shell. The symbols sizes are indicative of the number of patents in that shell that belong in the given section. A logarithmic axis is used to show the dominance of Chemistry and Human Necessities in shells above shell 5. (b) Same as in (a) for classes. Only some indicative classes are shown. Most classes (about 100) have very low values of patents and are not shown.

Further analysis, at a class level (classes are more specialized thematic areas than sections) shows similar behavior, with some classes dominating over most. About 10 classes dominate in higher shells, again mostly from the Chemistry and Human Necessities sections, over more than 100 classes that exist mainly in very low shells and become non-existing at higher ones (Fig. 6b).

4. Conclusions

Structural analysis of the patent citation network with the use of the k-shell decomposition method has been used to identify the most important nodes. The analysis has yielded the surprising result that most well connected nodes of very high degree belong to outer shells, while others with relatively lower degree values belong to inner ones.

Furthermore, and at least for the given dataset and the given methodological approach, only patents originating from North America, Europe and, to a lesser extent, Asia, seem important enough to have patents in the higher shells. Indeed, although Asia has 22% of the total patents in the network, Europe and North America (which have about 43% and 33%, respectively) are dominating most inner shells.

On a community level analysis of the inner shells, several of the communities formed belong to just one (or mainly one) continent. Indeed out of the 11 communities shown in our results, only 4 seem to have two continents fighting for dominance, while no community seems to have all three major continents equally represented.

Similar analysis by thematic area, according to the patents section, was performed and revealed that only after the first 4 shells, two sections alone dominate all shells. Interestingly, these sections are not the only ones with many patents. Other sections with equal numbers of patents, Electricity, Physics and Transportation are well below those of Chemistry and Human Necessities after shell 4. Similarly, by looking further into the classes of patents we have seen that there are about 10 times more classes that show a significant decrease than increase in their importance in shells. As expected these classes belong to sections that are of the two dominating sections.

The results obtained raise important questions on why in the age of globalization, geographic separation (at least in the level of citations) still seems to exist, and on why sciences that have contributed much in modern technology (such as Physics, Electricity, Mechanical Engineering) seem to be under-represented.

Although this analysis focuses on the specific network of patent citations, the facts revealed show that similar analysis in other networks may unravel hidden information essential to the understanding of how such networks form and function.

Acknowledgments

Results presented in this work have been produced using the AUTH Compute Infrastructure and Resources. The authors would like to acknowledge the support provided by the Scientific Computing Office throughout the progress of this research work. This research was supported by European Commission FP7-FET project Multiplex No. 317532.

References

[1] E. Engelsman, A. van Raan, A patent-based cartography of technology, Res. Policy 23 (1) (1994) 1–26, http://dx.doi.org/10.1016/0048-7333(94)90024-8.
[2] J.L. Ortega, Collaboration patterns in patent networks and their relationship with the transfer of technology: the case study of the csic patents, Scientometrics 87 (3) (2011) 657–666, http://dx.doi.org/10.1007/s11192-011-0363-4.
[3] J. Hoekman, K. Frenken, F. van Oort, The geography of collaborative knowledge production in europe, Ann. Reg. Sci. 43 (3) (2009) 721–738, http://dx.doi.org/10.1007/s00168-008-0252-9.
[4] K. Briggs, Co-owner relationships conducive to high quality joint patents, Res. Policy 44 (8) (2015) 1566–1573, http://dx.doi.org/10.1016/j.respol.2015.05.011.
[5] S. Redner, How popular is your paper? an empirical study of the citation distribution, Eur. Phys. J. B 4 (2) (1998) 131–134, http://dx.doi.org/10.1007/s100510050359.
[6] A. Vazquez, Statistics of citation networks, arXiv cond-mat/0105031.
[7] T. Hammond, T. Hannay, B. Lund, J. Scott, Social bookmarking tools (i): A general review, D-Lib Magazine 11 (4) (2005) http://dx.doi.org/10.1045/april2005-hammond.
[8] Personal Information Management vs. Resource Shabelring: Towards a model of information behavior in social tagging systems, in: 3rd International AAAI Conference on Weblogs and Social Media (ICWSM), San Jose, California, USA, 2009.
[9] M. Wasserman, X.H.T. Zeng, L.A.N. Amaral, Cross-evaluation of metrics to estimate the significance of creative works, Proc. Natl. Acad. Sci. 112 (5) (2015) 1281–1286, http://dx.doi.org/10.1073/pnas.1412198112.
[10] F. Radicchi, S. Fortunato, A. Vespignani, Citation Networks, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 233–257, http://dx.doi.org/10.1007/978-3-642-23068-4_7.
[11] L. Šubelj, D. Fiala, M. Bajec, Network-based statistical comparison of citation topology of bibliographic databases, Sci. Rep. 4 (2014) 6496, http://dx.doi.org/10.1038/srep06496.
[12] Y.H. Eom, S. Fortunato, Characterizing and modeling citation dynamics, PLoS One 6 (9) (2011) 1–7, http://dx.doi.org/10.1371/journal.pone.0024926.
[13] J. Alcácer, M. Gittelman, Patent citations as a measure of knowledge flows: The influence of examiner citations, Rev. Econ. Stat. 88 (4) (2006) 774–779, http://dx.doi.org/10.1162/rest.88.4.774.
[14] P. Criscuolo, B. Verspagen, Does it matter where patent citations come from? inventor vs. examiner citations in european patents, Res. Policy 37 (10) (2008) 1892–1908, http://dx.doi.org/10.1016/j.respol.2008.07.011, special Section Knowledge Dynamics out of Balance: Knowledge Biased, Skewed and Unmatched.
[15] P.B. Maurseth, B. Verspagen, Knowledge spillovers in europe: A patent citations analysis, Scand. J. Econ. 104 (4) (2002) 531–545, http://dx.doi.org/10.1111/1467-9442.00300.

[16] M. Karvonen, T. Kässi, Patent citations as a tool for analysing the early stages of convergence, Technol. Forecast. Soc. Change 80 (6) (2013) 1094–1107, http://dx.doi.org/10.1016/j.techfore.2012.05.006.

[17] S.B. Seidman, Network structure and minimum degree, Social Networks 5 (3) (1983) 269–287, http://dx.doi.org/10.1016/0378-8733(83)90028-X.

[18] B. Pittel, J. Spencer, N. Wormald, Sudden emergence of a giant k-core in a random graph, J. Combin. Theory Ser. B 67 (1) (1996) 111–151, http://dx.doi.org/10.1006/jctb.1996.0036.

[19] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, A model of internet topology using k-shell decomposition, Proc. Natl. Acad. Sci. 104 (27) (2007) 11150–11154, http://dx.doi.org/10.1073/pnas.0701175104.

[20] Y. Liu, M. Tang, T. Zhou, D. Younghae, Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition, Sci. Rep. 5 (2015) 9602, http://dx.doi.org/10.1038/srep09602.

[21] J.I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, A. Vespignani, K-core decomposition of internet graphs: hierarchies, self-similarity and measurement biases, Netw. Heterogeneous Media 3 (2) (2008) 371–393, http://dx.doi.org/10.3934/nhm.2008.3.371.

[22] N. Azimi-Tafreshi, J. Gómez-Gardeñes, S.N. Dorogovtsev, K-core percolation on multiplex networks, Phys. Rev. E 90 (3) (2014) 032816, http://dx.doi.org/10.1103/PhysRevE.90.032816.

[23] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6 (11) (2010) 888–893, http://dx.doi.org/10.1038/nphys1746.

[24] X. He, H. Zhao, W. Cai, G.-G. Li, F.-D. Pei, Analyzing the structure of earthquake network by k-core decomposition, Physica A 421 (2015) 34–43, http://dx.doi.org/10.1016/j.physa.2014.11.022.

[25] S. Hakimi, On realizability of a set of integers as degrees of the vertices of a linear graph. i, J. Soc. Ind. Appl. Math. 10 (3) (1962) 496–506, http://dx.doi.org/10.1137/0110037.

[26] D. Kleitman, D. Wang, Algorithms for constructing graphs and digraphs with given valences and factors, Discrete Math. 6 (1) (1973) 79–88, http://dx.doi.org/10.1016/0012-365X(73)90037-X.

[27] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008, http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.