Universal hierarchical behavior of citation networks

Enys Mones¹, Péter Pollner²,³ and Tamás Vicsek¹,²,³

¹ Department of Biological Physics, Eötvös Loránd University, Pázmány Péter stny. 1/A, H-1117 Budapest, Hungary
² MTA-ELTE Statistical and Biological Physics Research Group, Pázmány Péter stny. 1/A, H-1117 Budapest, Hungary
³ ELTE Faculty of Sciences, Regional Knowledge Centre Irányi Dániel u. 4, H-8000 Székesfehérvár, Hungary
E-mail: enys@hal.elte.hu, pollner@hal.elte.hu and vicsek@hal.elte.hu

Received 17 January 2014
Accepted for publication 8 April 2014
Published 22 May 2014

Abstract. Many of the essential features of the evolution of scientific research are imprinted in the structure of citation networks. Connections in these networks imply information about the transfer of knowledge among papers, or, in other words, edges describe the impact of papers on other publications. This inherent meaning of the edges implies that citation networks can exhibit hierarchical features that are typical of networks based on decision making. In this paper, we investigate the hierarchical structure of citation networks consisting of papers in the same field. We find that the majority of the networks follow a universal trend towards a highly hierarchical state, and the various fields display differences only concerning (i) their phase in life (distance from the ‘birth’ of a field) or (ii) the characteristic time according to which they are approaching the stationary state. We also show by a simple argument that the alterations in the behavior are related to and can be understood by the degree of specialization corresponding to the fields. Our results suggest that during the accumulation of knowledge in a given field, some papers are gradually becoming relatively more influential than most other papers.

Keywords: growth processes, network dynamics, communication, supply and information networks
1. Introduction

1.1. Citation networks

Citation networks are one of the simplest representations of the development of human knowledge, as they describe the connections and influences among academic papers in different fields [1, 2]. Due to their importance, these networks have been the subject of various studies, including the distribution of citations [3], community structures [4, 5], topology [6] or the history of single papers [7]. The number of available databases has shown spectacular growth recently, resulting in access to detailed bibliographic data dating back even for decades. These data provide the opportunity for accurate study of the structural evolution of academic publications over the accessible period of time [8, 9].

The connections in a citation network are inherently asymmetric: in most cases the citation is an act solely of the citing paper. There are, however, several reasons behind the inclusion of a publication in the reference list [10] and there are also signs of cases in which the cited paper is not even read by the authors before citing it. Nevertheless, it is reasonable to assume that a reference from paper B to paper A is related to some level of influence of paper A on paper B. In other words, the content of paper B is affected by that of paper A (at least in the choice of subject). Through the life of particular papers, the
network they form is permanently changing in time, resulting in a continuously developing structure. There are different approaches in the literature to characterize this evolution: by the distribution of times a paper is cited [11], the number of cited references [12], the change in the number of papers in time [13], etc. Most of these studies focus on local properties of the network, i.e., the small neighborhood of the nodes is considered, and they do not take into account the topology of the graph. However, the global structure of the network itself opens the door to the investigation of other properties, which span over the whole system and describe a global behavior.

In section 2 we introduce the central concept of this paper. We describe the hierarchical nature of networks and present a quantity, the global reaching centrality ($G_R$), from our earlier work for measuring the hierarchy. After determining the behavior of the $G_R$ in citation networks of several scientific fields, we consider a simple argument about the possible mechanism of the development of hierarchy in these networks. The model is based on the assumption that the evolution of hierarchy here is related to the level of generality of the field. Finally, we support our assumption by showing that the trend of the evolution of the hierarchy changes according to the proportion of external references, i.e., the average number of references to (or from) other fields divided by the number of inner references.

2. Concepts

2.1. Hierarchy in networks

Hierarchy is a global structural feature of many networks, and it is characteristic of a wide range of systems: animal groups [14], employees of a company [15] and several other social and technical networks [16]. Usually hierarchy is observed in heterogeneous systems, where nodes with different roles, dominance or histories are connected. According to the relations between the nodes of the network one can distinguish three types of hierarchy: the nested, the order and the flow hierarchy.

The nested hierarchy is applicable for networks, where different levels of generality coexist and containment relations are defined among the subgraphs of the network. In such systems the containing subgraphs are at higher levels in the hierarchy (e.g. the hierarchical clustering defines a nested hierarchy of subgraphs: if nodes are merged into clusters and clusters are merged into larger clusters, the nodes are at the bottom of the hierarchy and the merged clusters are at higher levels). Undirected networks can be hierarchic in this way.

The order hierarchy is usually defined by a variable, according to which nodes are ranked (e.g. authors can be ranked by cumulative impact factors, articles can be ranked by citations, etc). Order hierarchy can also be defined without any variable, if the network is a simple directed chain and the direction of the edges defines the rank order.

The third type, the flow hierarchy, is determined by asymmetric connections: if a node influences another, the latter is at a lower level in the hierarchy. In directed networks, flow hierarchy is defined as a layering of the nodes in which the direction of the edges obeys a ‘global flow’, e.g. edges point from upper levels towards lower levels [17] (see also figure 1 for an illustration). In systems of human–human interactions, the meaning
Universal hierarchical behavior of citation networks

Figure 1. Three classes of hierarchical structure in networks: (a) nested (inclusion) hierarchy, (b) order hierarchy and (c) flow hierarchy. Among these types, flow hierarchy is defined in directed networks and associates nodes into separated levels.

of flow hierarchy is most pronounced in networks of decision making, i.e. where an edge pointing from person $A$ to person $B$ represents the fact that the decision made by person $A$ is taken into consideration by $B$ when making their own decision. It has been shown that in a general framework, where greedy agents are supposed to solve consecutive tasks and optimize their success rate, a decision-following hierarchy emerges, if the solutions of agents are known to the other agents [18]. Citation networks can be considered as abstract decision-following networks among scientific publications, and the occurrence of a hierarchical feature would suggest that the phenomenon of hierarchy is not restricted to small groups of individuals, but also emerges at multiple scales.

2.2. Global reaching centrality

In order to compare networks according to their hierarchical structure one needs to define a quantity that appropriately measures the hierarchic nature. Several different metrics are available, but they either use a free parameter or measure the deviation from a tree [15, 19, 20]. The first class of measures is problematic for general networks, where the free parameter is usually not known. The latter measures are very specific by simply penalizing the existence of loops and multiple edges. We are using a hierarchy measure here, which is based on the ‘flow’ property of the hierarchy. We consider a network to be highly hierarchical if the influence of different nodes is very inhomogeneous. Since one node is able to influence another one if there is a directed path of consecutive edges from the first to the second, for a given node $i$ the influence ability is measured by the fraction of nodes in the graph that are reachable by directed pathes from node $i$. This node-specific quantity is called the local reaching centrality of node $i$ and we denote it here by $c_R(i)$.

The hierarchy of the network is measured by the heterogeneity of this local quantity. Therefore, we measure the level of flow hierarchy by the global reaching centrality ($G_R$) [21],

$$G_R = \frac{\sum_i c_R^M(i) - c_R(i)}{N - 1} = \frac{N}{N - 1} (c_R^M - \langle c_R \rangle),$$

(1)

where $c_R(i)$ is the fraction of nodes that node $i$ can reach via out-links and $N$ is the number of nodes in the graph. Here, $c_R^M$ denotes the maximal value of the fraction of reachable nodes and $\langle c_R \rangle$ is the average of the above quantity. In the case of a directed
tree, the levels and the hierarchical structure are rather simple and trivial. However, as the network includes a large number of edges, the simple structure is biased and the underlying hierarchical structure is less obvious to extract. This is addressed by $G_R$, which has been checked in classic networks and a corresponding synthetic model together with investigations in real networks, yielding results in agreement with the concept of ‘levels’ and flow hierarchy. For example, the most possible centralized network, the star graph, has $G_R = 1$. The most egalitarian network, where all nodes are threaded into a single loop, has a zero $G_R$. Since our hierarchy measure is a measure for heterogeneity, random networks without any structure show vanishing $G_R$ values, and the simple chain network due to its very symmetric structure takes $G_R = 0.5$.

3. Results

3.1. Universal trends

We address the issue of hierarchy in citation networks by considering 266 temporal networks, each defined by a category or a keyword given by Thomson Reuters’ Web of Science (WoS) database during the period 1975–2011 [22]. From the possible networks we have chosen all samples that are large enough and have an appropriate complexity for our analysis. The details of the generation of temporal networks are described in section 4.

We measured the time dependence of $G_R$ in 210 category and 56 keyword networks (which we will refer to as fields in the rest of the paper, as each category or keyword defines a related field). Figure 2 shows typical trends, together with the evolution of the corresponding average degree. It should be noted that the corresponding average degree is defined by the number of references where the cited papers are of the same field. The first and most striking observation is that the level of hierarchy is increasing in all cases, even though the average degrees corresponding to these fields are also growing (see the inset in figure 2). Studies have found that the extent of hierarchy is quickly vanishing for increasing average degrees in the configuration model [20, 23, 24]. This indicates that the evolved structure of the citations is nontrivial, and even networks with low $G_R$ values are considerably more hierarchical than most other real-world networks [21]. However, this continuous growth is the fingerprint of these networks, and it is strongly related to the fact that new edges do not emerge between nodes that are already present, but are only added between new nodes and old ones (aside from the negligible fraction of directed loops that are results of errors in the databases in most cases). The primary difference between the curves is the intensity of growth: the networks for the category cell biology and the keyword tumor-necrosis-factor show a rapid increase in the global reaching centrality, while the networks related to the other categories and keywords exhibit slower growth. The fields shown in figure 2 represent typical behaviors: the bottom and top curves represent the slowly and rapidly growing fields, and the middle trend illustrates those between the two extremes.

This similarity in the hierarchical curves suggests that the main mechanisms behind all categories have common features. This is also supported by figure 3, where we rescaled and plotted all curves to fit the universal trends defined by the category cell biology and keyword tumor-necrosis-factor. Details of the rescaling process are presented in section 4.
In figure 3, we considered only networks that have a final $G_R$ larger than 0.1, to ensure that the vertical scaling gives reasonable values (i.e. to obtain less noisy scaling). It is clearly visible that the hierarchical development of the different fields can be described by a single universal tendency (for both category and keyword networks), and the fields differ mostly in the timescale. Citation networks can reach different levels of final hierarchy that are specific to the fields, but in the long run all fields exhibit the same behavior. Although most of the curves can be fitted to the universal trend by linear transformations, there are 11 category and seven keyword networks that show a rather different behavior. However, these networks also show increasing hierarchy with a piecewise sigmoid-like shape (see the inset in figure 3, where the category engineering, electrical and electronic is plotted).
Figure 3. Collapsed trends in the two citation networks: (a) category and (b) keyword networks. The axes denote the rescaled time (horizontal) and $G_R$ (vertical) after the linear transformation (translation and scale) of the curves. The inset in (a) shows a typical curve that was rejected in the rescaling process due to the intermittent increase characteristics (it corresponds to the category engineering, electrical and electronic).

Also, some of the curves that reach their final state in a few years are continued after the last year shown in the collapsed data, but they remain at a constant value, or grow with a negligible rate; therefore these sections are not shown in the figures.

3.2. A simple model

In order to understand the differences seen in the trends, we present a simple dynamical model for the time evolution of the global reaching centrality. The model relies on two
assumptions about the structure of citation networks. First, $G_R$ is related to the difference between the largest and average reaching centralities (see equation (1)). As more detailed measurements on the hierarchy reveal, the observed trends are dominated by the $c_{RM}$ since $\langle c_R \rangle$ is negligible. This is rather clear: when new papers connect to the actual network, they only have in-edges (because edges point towards the citing papers). As the number of papers increases according to an exponential growth [13, 25], the increasing number of nodes without out-links reduces the average reaching centrality to near zero. The second assumption is less obvious and it can be described as follows. Each network is defined by a specific field, given by the corresponding category or keyword. Every field is characterized by several main research streams which are ignited by a handful of papers for each stream and the rest of the network is growing below these standard works. To illustrate this idea, in figure 4 we depict two scenarios corresponding to different levels of generality of the fields: specialized, which is more focused on a narrow topic, and general, which covers a broader range of topics. In the case of a more specialized field (figure 4(a)), the whole network is much more interconnected and the papers at the top of the hierarchy—those with the largest reaching centralities—have more or less the same number of nodes that they can reach. In other words, the nodes with the largest $c_R$ have similar values, thus more nodes have a local reaching centrality of nearly $c_{RM}$. On the other hand, in a general field including diverse research streams (figure 4(b)), top papers have varying $c_R$ values. Thus, the fraction of nodes having nearly $c_{RM}$ is much lower than in the case of a specialized field.

Based on the first observation, we will approximate the $G_R$ by $c_{RM}$ and construct an equation describing its rate of change. When a new paper appears, it is immediately attached to the network by its references. The probability that the node is attached to
the reachable set of the top node is proportional to the corresponding reaching centrality $c^M_R$. It can be shown that the increment of $c^M_R$ on the attachment of a single node is $(1 - c^M_R)/(N + 1)$. Also, the fraction of nodes with reaching centrality near $c^M_R$ is related to the generality of the field: if the category or keyword describes a field with few research streams (it is a rather specialized one), the probability of finding a node with a large local reaching centrality is high. In contrast, if the field is more general, this probability is reduced. Assuming that $\alpha N$ of the nodes have reaching centrality close to $c^M_R$, where $\alpha$ tunes the generality of the network, these points can be formulated in the following dynamical equation for $c^M_R$ in the large $N$ limit:

$$\frac{dc^M_R}{dt} = \alpha c^M_R (1 - c^M_R).$$

(2)

This logistic equation has the well-known solution

$$c^M_R(t) = \frac{1}{1 + (1/c^0_R - 1)e^{-\alpha t}}.$$ 

(3)

which describes a sigmoid with $c^0_R$ as its initial value. To illustrate the solution, we plotted equation (3) at $c^0_R = 0.01$ for different values of the parameter $\alpha$ in figure 5. As we can see, by varying the parameter $\alpha$, the solution gives the typical trends seen in figure 2, considering a finite interval of time. Figure 5 suggests that the differences among the categories and keywords might originate from the different levels of generality, i.e., in the number of research streams they cover.

### 3.3. Separation of trends

Motivated by the model, we now take a closer look at the curves seen in figures 2 and 3. In order to check the effect of generality on the time evolution trend of hierarchy, we considered the cross-category references among the papers in the category networks. More
Figure 6. Hierarchical trends in the category networks with different values of the external reference ratio \((E)\). The curves correspond to the averages taken over the networks with the indicated value of \(E\). The trends of the keyword networks are shown in the inset. In this case, they are separated into three groups of equal numbers of networks according to the values of \(E\). The error bars denote the standard deviation of the mean.

precisely, for each category \(C\), we calculated the number of references \((R_{C,C})\) between pairs of papers, both labeled by the same category \(C\). Similarly, we calculated the average number of references \((\langle R_{C,C'}\rangle_{C' \neq C})\) between papers of a given category \(C\) and papers of other categories. These two quantities describe the number of internal \((R_{C,C})\) and external \((\langle R_{C,C'}\rangle_{C' \neq C})\) references of category \(C\). Considering the generality of the category, we calculated the external reference ratio

\[
E(C) = \frac{\langle R_{C,C'}\rangle_{C' \neq C}}{R_{C,C}}. 
\]  

As an approximation of the problem of generality, we use this ratio to quantify how specialized a category is. If \(E(C)\) is very small, papers in the category \(C\) prefer to cite papers inside their own category, meaning that it is a rather specialized field. However, if the category covers many distinct subfields, it implies a higher fraction of external citations (and vice versa, papers in other fields cite papers in the category more frequently), resulting in a larger value of \(E\). In the 210 categories under consideration, the value of \(E(C)\) lies in the range \([0:0.1]\). We divided the networks into four groups by their external reference ratio, so that each group contained at least 27 categories, and calculated the average \(G_R\) trends inside the groups. The results are plotted in figure 6, showing well-separated hierarchical curves. The curves confirm the assumption of the model: the more specialized a category is, the sooner it develops a highly hierarchical structure. The inset of figure 6 shows the same curves for the keyword networks. Due to the small number of inner edges in these networks, the data are very noisy at the beginning, but the trends in later years are better separated.
Figure 7. The largest 1000 ranked $c_R$ values in the year 2011 for two category networks, sorted in descending order. The horizontal axis denotes the rank of $c_R$, starting with the maximum one at rank zero. On the vertical axis, the relative value of $c_R$ is plotted, normalized in such a way that the offset ($c_{1000}^M$, which is the centrality at rank 1000) is subtracted and the data are divided by the vertical extent ($\Delta c_R = c_{1000}^M - c_{1000}^L$) of the curves. The original curves are shown in the inset.

As an additional corroboration, we can calculate the values of $c_R$, to test whether more general categories are characterized by fewer nodes with near-maximal reaching centrality. For this, in figure 7 we plotted the first 1000 largest $c_R$ values against their ranks in the categories cell biology and biology, measured in the final state of the networks. To emphasize the difference between the two categories, we transformed the curves to a form that is easier to grasp, and we also provide the original data. The abscissa shows the rank of the $c_R$ value, starting with the largest one. On the ordinate, the normalized values are shown. From each value, we subtracted the last (i.e. the 1000th) data point ($c_{1000}^L$) and divided by the difference between the two extreme values ($\Delta c_R = c_{1000}^M - c_{1000}^L$). Clearly, the values in the category biology (having a larger value of $E$) vanish faster, meaning that the probability of having near-maximal reaching centrality is very low, compared to the category cell biology. Thus, according to the external reference ratio, the assumption of the model about the fraction of nodes with reachable set similar to the largest one is reasonable and shows good agreement with the data.

4. Methods

4.1. Temporal networks

The following two types of citation network are considered in this paper.

Category networks: these are defined by the subject categories given by the WoS database. The nodes of these networks are papers corresponding to one subject category, and an edge is linked from paper $A$ to paper $B$ if paper $B$ cited paper $A$.  

doi:10.1088/1742-5468/2014/05/P05023
Universal hierarchical behavior of citation networks

**Keyword networks**: these are based on specific keywords. All nodes of a keyword network are papers that are labeled, *inter alia*, by a selected keyword. As with the category networks, edges represent citations and are directed so that they point from the cited paper towards the citing paper.

The reason behind the above definition of edge directions is that we are interested in the network of influences among papers, and we assume that papers can only influence later papers. For each category, we generated temporal networks in the period $y = 1975, \ldots, 2011$, by aggregating the networks from year 1975 to year $y$. There are 251 subject categories in total; each paper in the database is labeled with at least one category. As an example, three categories are *acoustics*, *geography* and *robotics*. In order to detect whether there is a possible bias caused by the fact that papers in the WoS database are indexed only from 1975, in the case of keyword networks, we only consider keywords that were ‘born’ between the years 1990 and 1995. These keywords are defined as having a frequency of appearance ten times larger in 1995 than in 1990 and still being ‘alive’ in 2011 by appearing in at least 1000 papers.

4.2. Data filtering

Several filtering processes were applied to the data, due to the presence of categories and keywords with very small networks and very low connectivity. To decrease the level of noise, we first rejected networks with a final size of less than 1000. In other words, fields with at least 1000 papers in total are considered. Furthermore, we also rejected networks with a final average degree lower than 1, because these networks consist of disjoint small components and they cannot be characterized by a coherent structure. This is crucial, since we are aiming at the study of the evolution of a field as a whole, which is not true for the rejected networks, as any measured property is merely a mixture of the independent subgraphs (smaller subfields). Similarly, we also set a threshold on the size of the giant components: they must have a *giant weakly connected component* (GWCC) larger than 0.5, for the reason mentioned before. The GWCC is defined by the largest connected component of the graph. After the above filters, 210 category and 56 keyword networks remained.

4.3. Hierarchy

The measurement of the flow hierarchy in the citation networks is based on the global reaching centrality ($G_R$) given by equation (1). For this measure, one is supposed to determine the largest and the average local reaching centralities, which requires $O(N^2)$ operations using a simple depth-first search algorithm. In our case, the size of the networks increases quickly, and it is required to calculate the $G_R$ for graphs with $N \approx 10^6$ nodes and $N \approx 10^7$ edges. Therefore, we applied an approximation method for the calculation of the hierarchy measure. The method is divided into the following steps.

(i) Choose a node randomly.

(ii) Calculate the local reaching centrality $c_R(i)$ for the chosen node.
(iii) Update $c_M^H$ and if the node was chosen randomly update $\langle c_R \rangle$ as well.

(iv) With probability $p_{\text{jump}}$ go to step (i) (if $k_{\text{in},i} = 0$, set $p_{\text{jump}} = 1$ for this step).

(v) With probability $1 - p_{\text{jump}}$, choose an in-neighbor randomly and go to step (ii).

On the one hand, nodes are chosen randomly for a fraction $p_{\text{jump}}$ of the steps. Since these node samples are uncorrelated, the approximation of the average is unbiased. On the other hand, in a fraction $1 - p_{\text{jump}}$ of the steps, the method walks through in-edges, and therefore discovers nodes at higher levels of the hierarchy. The average relative errors of the estimations defined by

$$\langle \delta G_R \rangle = \left\langle \frac{G_{\text{exact}}^R - G_{\text{estimated}}^R}{G_{\text{exact}}^R} \right\rangle$$

are shown in figure 8, where we calculated $\langle \delta G_R \rangle$ for a hierarchical synthetic model that is introduced in [21]. The synthetic model is controlled by a parameter $p$, which tunes the extent of hierarchy. The networks have $N = 10000$ nodes and $\langle k \rangle = 3$ in each case. All dots are the average of 100 different samplings and also 100 different network realizations at $p_{\text{jump}} = 0.1$. Although the method has rather large errors at low hierarchy (>10%), as the network becomes more hierarchical, the error tends to decrease quickly to below an admissible level. However, it should be noted that networks with a negligible extent of hierarchy have a value of $G_R \ll 1$, and therefore any conclusions drawn from its value remain valid even at an error of 10%. In the calculations for real networks, we applied $p_{\text{jump}} = 0.1$ and a number of jumps equal to 10 000.
4.4. Data collapse

The collapsed data shown in figure 3 are obtained by the rescaling of the trends measured in the various fields. As figure 2(b) shows, when the networks consist of very few papers (less than 100), the global reaching centrality can take large values as well, which later vanish as the networks grow large enough. Therefore, we considered the $G_R(t)$ trends only from the year at which the networks have at least 100 nodes. Moreover, some of the networks gain a finite value of $G_R$ at the beginning and this value does not change over the next few years, since there is no significant change in the structure. Thus, we also removed the offset in the networks by subtracting $G_R(t = 0)$ from all other years.

We considered linear transformations of the trends in order to rescale them onto the base trends which are some of the most articulated ones (the networks related to cell biology and tumor-necrosis-factor). The transformation of the trend $T$ onto the base trend $T_{base}$ was carried out by minimizing the function $H = \langle R^2 \rangle / L$, where $\langle R^2 \rangle$ denotes the average squared difference of the data $T_{base}$ and $T$ in the overlapping region of the two data, and $L$ is the length of the overlap. We optimized $H$ by simulated annealing with respect to horizontal and vertical translations and scaling, and also the removal of some of the first data points (five at most).

5. Discussion

In this paper we have studied the evolution of the hierarchical structure of citation networks restricted to categories and keywords given by the Web of Science database. We measured the time dependence of flow hierarchy in the networks by a corresponding metric, and surprisingly found that a majority of the categories follow a unique growth tendency. It is known that hierarchical structures are less likely to appear and be present in networks with large average degree [20, 23, 24]. Although the growing citation networks are characterized by an increasing average degree, they also form a more hierarchical structure in time. The nature of hierarchy in the networks is further supported by the fact that most of the curves can be collapsed into one universal curve. To investigate the observed tendency, we proposed a simple model that is able to capture the main qualitative properties. By a few assumptions, we showed that the different pace of the hierarchy tendencies is related to the specialization level of the corresponding field (category or keyword). This is confirmed by a simple model with the aim of characterizing the dynamics of the hierarchy in the networks, and also through the separation of the data by a quantity that is based on additional information on the papers and captures the generality of the related categories.

An important comment should be made with regard to the model and especially equation (2). Obviously, the true dynamics in real networks will not be as simple as our description, and the purpose of the model is to indicate that the generality of the defining field of the network has an impact on the hierarchy. In this setup, when a new paper is born, it has multiple references, and thus the right-hand side of equation (2) should be multiplied by the average degree $\langle k \rangle$. Implicitly, in the derivation of the dynamical equation, we assumed that only the parameter $\alpha$ includes all further multiplicative factors, and therefore it counts for $\langle k \rangle$ as well. Furthermore, the solution given by equation (3)
has an asymptotic value of $c^M_R = 1$, independently of the initial value or the parameter $\alpha$, in contrast to real networks. This is related to the fact that real networks do not have the exact structural form described in the model: they are not simply tree-like subgraphs merged together, but much more complex objects, having nontrivial, and not even layered, inner structure. Thus, the number of nodes not living in the reachable set of the top node is not exactly $(1-c^M_R)N$, but somewhat lower, resulting in a smaller final level of hierarchy.

The growing level of hierarchy found in citation networks suggests that papers are organized in a highly ordered structure. As a hierarchy, it is characterized by a small fraction of papers that dominate the network through the citations they receive. Papers at the top of the hierarchy are also the root of smaller subgraphs (research streams), and as the network grows, these subgraphs merge together. Although the hierarchical structure can be understood by simple models, open questions remain. What are the special characteristics of papers at the top of the hierarchy? How do top papers evolve in time, and how is the success of a publication related to its position in the hierarchy? How does the structure of the neighborhood of a paper correlate with its success? As a further investigation, these questions could be addressed by a more detailed model and exploration of the local properties of citations.

Acknowledgments

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (Grant No. TÁMOP-4.2.2.C-11/1/KONV-2012-0013), by the Hungarian National Science Fund (OTKA K105447) and by the EU ERC FP7 COLLMOT Grant No. 227878.

References

[1] Garfield E, Citation indexes for science, 1955 Science 122 108
[2] de Solla Price D J, Networks of scientific papers, 1965 Science 149 510
[3] de Solla Price D J, A general theory of bibliometric and other cumulative advantage processes, 1976 J. Am. Soc. Inform. Sci. 27 292
[4] Lancichinetti A, Kivelä M, Saramäki J and Fortunato S, Characterizing the community structure of complex networks, 2010 PLoS One 5 e11976
[5] Chen P and Redner S, Community structure of the physical review citation network, 2010 J. Informetrics 4 278
[6] Karrer B and Newman M E J, Random graph models for directed acyclic networks, 2009 Phys. Rev. E 80 046110
[7] Wang D, Song C and Barabási A-L, Quantifying long-term scientific impact, 2013 Science 342 127
[8] Leicht E A, Clarkson G, Shedden K and Newman M E J, Large-scale structure of time evolving citation networks, 2007 Eur. Phys. J. B 59 75
[9] Eom Y-H and Fortunato S, Characterizing and modeling citation dynamics, 2011 PLoS One 6 e24926
[10] MacRoberts M H and MacRoberts B R, Problems of citation analysis: a critical review, 1996 Scientometrics 36 435
[11] Redner S, Citation statistics from more than a century of physical review, 2005 Phys. Today 58 49
[12] Vazquez A, Statistics of citation networks, 2001 arXiv:cond-mat/0105031
[13] de Solla Price D J, 1963 Little Science, Big Science... and Beyond (New York: Columbia University Press)
[14] Nagy M, Ákos Zs, Biro D and Vicsek T, Hierarchical group dynamics in pigeon flocks, 2010 Nature 464 890
[15] Rowe R, Creamer G, Hershkop S and Stolfo S J, Automated social hierarchy detection through email network analysis, 2007 Proc. 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (New York: ACM) pp 109–17

doi:10.1088/1742-5468/2014/05/P05023 15
Universal hierarchical behavior of citation networks

[16] Pumain D (ed) 2006 Hierarchy in Natural and Social Sciences (Dordrecht: Springer)
[17] Lane D, 2006 Hierarchy, Complexity, Society (Dordrecht: Springer)
[18] Nepusz T and Vicsek T, Hierarchical self-organization of non-cooperating individuals, 2013 PLoS One 8 e81449
[19] Krackhardt D, Graph theoretical dimensions of informal organizations, 1994 Computational Organizational Theory ed K Carley and M Prietula (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 89–111
[20] Luo J and Magee C L, Detecting evolving patterns of self-organizing networks by flow hierarchy measurement, 2011 Complexity 16 53
[21] Mones E, Vicsek L and Vicsek T, Hierarchy measure for complex networks, 2012 PLoS One 7 e33799
[22] www.thomsonscientific.com/cgi-bin/jrnlst/jlsubcatg.cgi
[23] Corominas-Murtra B, Goni J, Solé R V and Rodríguez-Caso C, On the origins of hierarchy in complex networks, 2013 Proc. Nat. Acad. Sci. USA 110 13316
[24] Suchecki K and Holyst J, Hierarchy depth in directed networks, 2013 arXiv:1303.2085
[25] Suh W H, Suslick K S, Stucky G D and Suh Y H, Nanotechnology, nanotoxicology, and neuroscience, 2009 Prog. Neurobiol. 87 133