Density-based Influence Metrics for Research Papers

Christos Giatsidis[1,*], **Giannis Nikolentzos**[1], **Chenhui Zhang**[2], **Jie Tang**[2], **and Michalis Vazirgiannis**[1]

[1]Computer Science Laboratory, École Polytechnique, Palaiseau, 91120, France
[2]Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China
[*]giatsidis@lix.polytechnique.fr
[+]these authors contributed equally to this work

ABSTRACT

Traditionally, citation count has served as the main evaluation measure for a paper's importance and influence. In turn, many evaluations of authors, institutions and journals are based on aggregations upon papers (e.g. h-index) In this work, we explore measures defined on the citation graph that offer a more intuitive insight into the impact of a paper than the superficial count of citations. Our main argument is focused on the identification of influence as an expression of the citation density in the subgraph of citations built for each paper. We propose two measures that capitalize on the notion of density providing researchers alternative evaluations of their work. While the general idea of impact for a paper can be viewed as how many have shown interest to a piece of work, the proposed measures are based on the hypothesis that a piece of work may have influenced some papers even if they do not contain references to that piece of work. The proposed measures are also extended to researchers and journals.

Introduction

The use of citation counts for evaluating the importance of scientific work has a long history, going back to 1927 and the work of Gross and Gross[1]. Since then, the number of citations received by a paper has been established as the major tool for evaluating its impact and for measuring its influence in the scientific community. A citation is considered as an acknowledgement from the citing paper to the cited one. Hence, it is generally assumed that the importance of a research paper is reflected by the number of citations it has received. This evaluation scheme can be seen as a form of crowd-sourcing since the quality of a paper is evaluated by the whole scientific community. Hence, a high-quality paper is very likely to receive more citations compared to a low-quality paper. It is not thus surprising why all bibliographic databases have adopted citation counts as their main metric for evaluating research work.

The same unit of measure (i.e. citation count) is commonly used as the basis for quantitative evaluation of work produced by research institutes, research teams and individual researchers. A plethora of metrics have been defined based on citation counts. Typical metrics include the total number of citations, the number of citations per published paper, the h-index[2] and the g-index[3]; citation counts are also at the core of measures of journal impact[4]. Although it is generally agreed that these metrics are problematic[5,6], they offer an easy and straightforward way of measuring an individual's, an institution's or a journal's influence in the research community. Furthermore, since the task of assessing the importance of each paper has been assigned to the scientific community itself, this can be considered a very transparent procedure. Besides the criticism against them, these metrics have been established as an indispensable tool for evaluating scientific impact and administrators take them into serious consideration when evaluating grant proposals or when hiring new faculty members.

Citations are thus the basic building block for metrics that compute the impact of a researcher, an institute or a journal. However, citations themselves have received a lot of criticism[7–10]; and it has been recognized that not all citations should be counted equally[11]. For example, most people would agree that self-citations should count less than foreign citations. Furthermore, one may expect cited papers, that substantially influenced the citing sources, to get a higher utility compared to the others. To this end, there have been attempts to represent citation counts relatively to some other factor (either by normalization or by taking a subset of the citations- e. g. by a window in time)[12]. A general normalization factor is that of the expected number of citations (overall or specifically by a subset e. g. by papers of a journal[13] or a field[14]). In this normalization, each citation counts inversely to the expected number of citations. Normalization factors have also been determined based on the paper's field as this is captured by its co-citation neighbourhood in the citation network[15].

Due to their limitations, it is not clear in the scientific community if citations are the appropriate unit of measure to evaluate a piece of research work. This is evident by various explorations of alternative measures proposed in the past. One example of such a measure computes the centrality of a paper in the citation network[16], while a second one counts the number of lead authors that have been influenced by the paper[17]. One may also consider meta-data (e.g. number of downloads) from journal and publishing platforms as an alternative evaluation of the impact of a paper. Nevertheless, citation count in its raw format is the dominant metric in utilization.

This article provides a new approach for evaluating the impact of a research paper with novel metrics which assess the academic influence it has on subsequently published papers. The main idea behind the measure is that influence is not reflected solely by the set of citing papers, but also by the papers that cite them (the citing papers) and so on. The emerging set of papers corresponds to a subgraph of the citation network. To that end, we extract the subgraph that is directly related to a publication, and we then define a measure of scientific quality based on this network. The proposed approach offers a set of attractive characteristics. For example, we show in our results that (i) it is capable of identifying commonly accepted influential papers and (ii) it ranks as its top results papers that do not necessarily correspond to those having received the largest number of citations. Existing higher-level performance metrics based on citation count may benefit from the proposed measure as it can be used as the basis for measuring the impact of research institutes, research teams and individual researchers. For instance, existing indexes such as the h-index and the g-index can be adapted to take into account the new measure and it can also be employed to produce journal impact factors. By no means the proposed measure is designed to address all the problems of citation counts. Currently, there is no "silver bullet" to the problem of measuring scientific quality and impact. In fact, it is generally agreed that no single measure of research performance will ever reveal more than a small part of the multi-dimensional picture[18]. Hence, the proposed measure is not intended to replace existing measures, but to complement them. People interested in evaluating research performance would then be able to draw more credible conclusions. Overall, the proposed measure can serve as a useful addition to the list of current measures of scientific influence.

Results

Evaluating Paper Impact

A set of papers can be naturally represented as a network where each vertex corresponds to a paper and there is a directed edge from vertex A to vertex B if paper A cites paper B. Such networks of papers are known as *citation networks*. Due to the nature of citations (one paper can/should cite only those that precede it chronologically), the formed graph belongs to a fundamental class of graphs: *directed acyclic graphs* (DAGs).

Citation networks have been studied for several decades in the information sciences[19,20] and more recently in physics[21,22] and in computational linguistics[23] for various properties that are similar to other real world networks (e. g. clustering into communities). When it comes to influence (or similar concepts like impact), it is widely accepted in practice that the number of citations reflects the paper's importance/strength and is used as an evaluation measurement. The number of citations of a paper corresponds to its in-degree in the afformentioned network. Although there has been substantial work on the identification of influential vertices in general graphs[24], most of them cannot be directly applied to our setting due to the special structure of citation networks.However, if we were to consider the entire network instead of the shallow count of immediate neighbours, we would see that there are other more interesting measures to be extracted from the above network.

To this end, for each paper, we define its *rooted citation graph* (RCG) as follows: Given a paper v of a citation network $G=(V,E)$, the RCG of v is the subgraph of G induced by a set of vertices S which contains v and all the vertices which can reach v by a directed path. That is, $u \in V(v)$ if there is a directed path from vertex u to vertex v. Then, for any paper $u \notin V(v)$, it holds that u was published before v or the two papers are not related to each other. The resulting DAG contains all the papers that were influenced (directly or implicitly) by the paper corresponding to vertex v. We should note here that the number of references included in a paper is limited. Therefore, it is impossible for an author to cite all the scientific articles relevant to his topic and thus one would pick the most important ones.

Given the RCG of a paper, we can define various approaches for measuring the importance of the paper inside the network. For example, one may consider the average degree of the entire RCG as an evaluation of the paper which is arguably as simple to compute as the degree of the root. To motivate this approach, we provide an example in Figure 1 as an intuition of how the average degree can represent the influence of a paper. The paper shown in Example B of Figure 1 has received many citations, but the papers that have cited it are independent from one another and have not received any citations at all. The former would indicate that their areas of research are different (co-citation indicates similarity in the fields the papers belong to[15]). Conversely, the paper shown in Example C of Figure 1 has received less citations, and the papers that have cited it form a densely connected graph (which would indicate that their topics are in the same field or very close). Papers whose shortest path distance from the root is equal to 2 have also been influenced by the paper corresponding to the root vertex. We can see in this example how the paper on of Example C would be considered more influential than the one of Example B as the papers "under" as the first forms a small sub-field of papers. This shows potentially a higher level of interest between the authors of the publications (and the cited papers). On the other hand, the paper on the left is referenced by many but those do not seem to

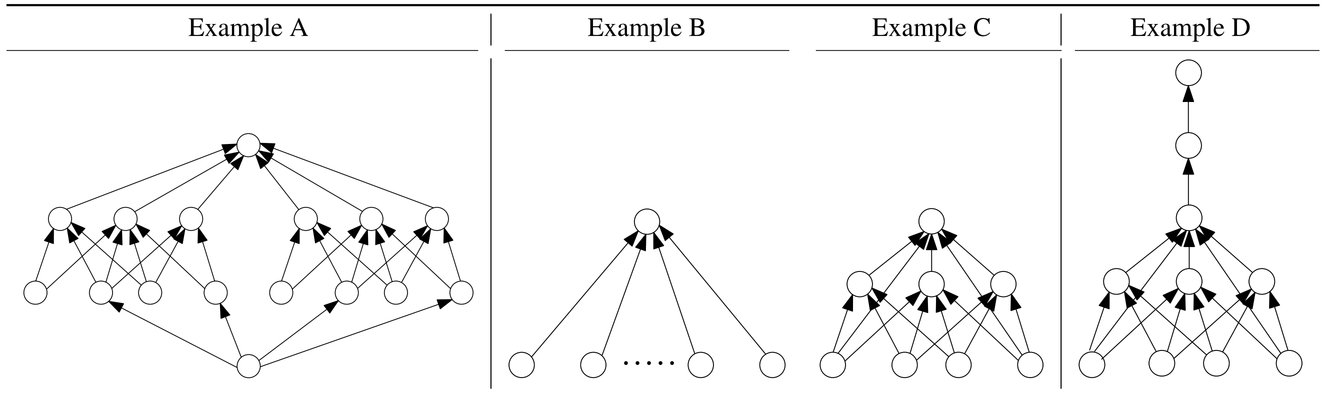

Figure 1. Toy examples illustrating the RCGs of different research papers. Example A shows the RCG of a paper that has given rise to two independent areas of research which are then brought together by the paper shown at the bottom of the example. Example B illustrates the RCG of a paper that has received many citations, but the papers that have cited it have not themselves received any citations at all. Example C shows the RCG of a paper that has received less citations than that of Example B, but the papers in the RCG form a densely connected graph. Finally, Example D corresponds to the RCG of a paper that has received a single citation by a very influential paper.

transfer that influence further on. We can see that citation networks exhibit several interesting properties which we will try to uncover here in order to evaluate research work.

Unfortunately, average degree is not always a good indicator of a paper's importance. We next display a generic case where it would fail to measure the true impact of a paper. Example D of Figure 1 illustrates the RCG of a paper v. It is clear that the in-degree of the root v is equal to 1. Furthermore, we assume that the paper u that is adjacent to the root is a very influential paper. All the papers that belong to the RCG of u also belong to the RCG of v since there is a directed path from each of these papers to u and then a directed path from u to v by definition. Nevertheless, the papers contained in the DAG rooted at u are very likely not to be related to v. As a real example of the aforementioned limitation of average degree, we consider the paper *"An introductory problem in symbol manipulation for the student"* from our computer science dataset (described later on this section). The average degree of the paper is one of the highest average degrees in that dataset. However, its in-degree (i. e. number of citations) is just equal to 1. It turns out that the high average degree of the paper stems from the single paper that cites it (much like in the afformentioned toy examples). The only difference between the RCG's of the two papers is the directed link from the latter to the former. Hence, the average degree of the former is almost equal to that of the latter. Thusly, a main challenge in our setting is how to distinguish in the RCGs papers that attribute influence to the root paper instead of some subsequent publication.

To this end and in order to avoid the pitfall of Example D, we capitalize on the k-core decomposition of networks, a powerful tool for network analysis which is commonly used as a measure of well connectedness for vertices in a broad spectrum of applications. The notion of k-core was first introduced by Seidman[25] for the study of cohesion in social networks. For an undirected graph $G = (V,E)$, the k-core of G is a maximal subgraph (of G) in which all vertices have degree at least k. The *core number* (a.k.a. coreness) $c(v)$ of a vertex v is equal to the highest-order core that v belongs to. Hence, vertex v belongs to a $c(v)$-core but not to a $(c(v)+1)$-core. We next show how k-core can be applied to measure a paper's impact:

Definition 1 (core influence) *Let $G = (V,E)$ be a citation network and $G(v)$ be the rooted citation graph of paper v. The core influence of paper v is the core number $c(v)$ of vertex v in $G(v)$ after transforming it to an undirected graph.*

We consider the core number of the root vertex $c(v)$ as the impact of the corresponding paper. If the root v is a member of a tightly interlinked group of vertices, then its core number will be high. Practically, this means that the root is cited by numerous papers that are also highly cited. Furthermore, we can see that the core number of the root is independent of the size and/or the density of its subgraph. Hence, even if there exists a large and dense subgraph that is the result of another paper in the rooted DAG, it will not affect the core number of the root. It is clear that core influence avoids many of the shortcomings of average degree. For example, in Figure 1, the root's core influence in the RCG of Example D is only 1.

Next, we present evaluation results of the core influence and the average degree in comparison to the in-degree of a paper. For our experiments, we employed the following two datasets in our experiments: (i) a dataset that is focused on the field of computer science and (ii) a dataset that is focused on physics. Regarding the field of computer science, we used a preselected

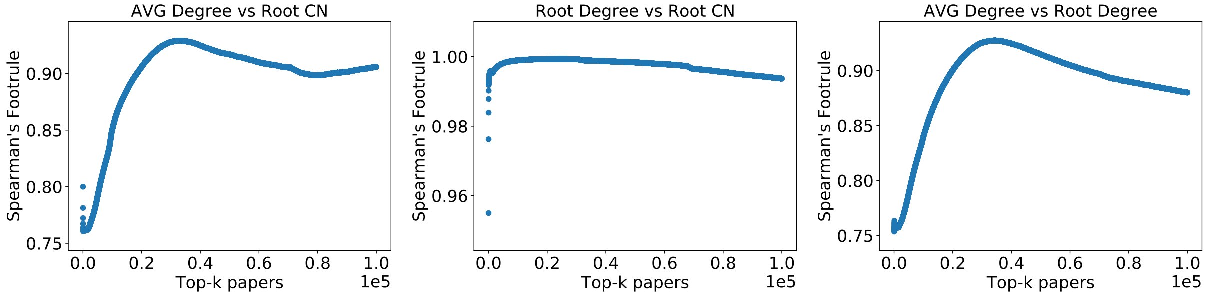

Figure 2. Spearman's footrule distance between average degree and core influence (Left), citation counts and core influence (Middle), average degree and citation counts (Right) on the physics dataset.

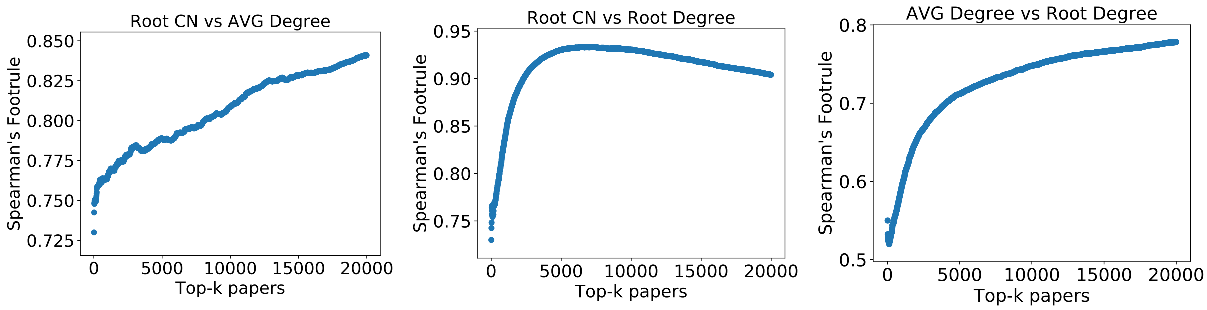

Figure 3. Spearman's footrule distance between average degree and core influence (Left), citation counts and core influence (Middle), average degree and citation counts (Right) on the computer science dataset.

set of papers from AMiner (publicly available on AMiner[1]) which contains $\sim$ 2.3M papers and $\sim$ 10.4M citation relationships. For the physics data, we utilized the Microsoft Academic Graph data[2] which span over a multitude of scientific fields but we focused only on papers coming from Physics Journals[3]. As the amount of emerging data is considerably large, we limited our set of papers only on those that were published after 1980. This resulted to a citation network of $\sim$ 3M papers and $\sim$ 15M citations.

For our results we focus on three measurements for each paper: a) number of citations/in-degree b) average degree of the RCG (AVGD) and c) core influence (i. e. core number of the root of the undirected RCG) (CN). The average degree is used as a way to evaluate the density of the RCG directly. Even though we illustrated some of the disadvantages of AVGD above, those disadvantages are not that frequent in practice for the top ranked papers. Moreover, as shown in the following, it provides unique results and -when compared with an external list of influential papers- meaningful as well.

We begin our analysis by displaying that the 3 measures uncover distinct properties of the RCG. Specifically, we compare the rankings of the papers based on each measure at the top-k elements with the *footrule distance with location parameter* l[26] where $l = k+1$. For the physics data, Figure 2 displays the disagreement for $k = 0,\ldots,10^5$ and Figure 3 is the equivalent one for the computer science data for $k = 0,\ldots,2\cdot 10^4$. It should be noted that for partial lists (like the top-k), the location parameter can be considered as a default value for the distance when an item does not exist in one of the lists. We see in both cases that the distance starts at relatively high values and keeps increasing. This shows that that the top-k items are very different. In fact, for small values of k (< 1000) the lists have little overlap among them which indicates that different properties are captured by them.

To verify that the proposed measure can identify papers that have had an important impact on their corresponding research communities, we focus on papers that have received the *Test of Time* award. Several major computer science conferences such as SIGIR, SIGMOD and KDD provide this award to outstanding papers that have had long-lasting influence. We thus expect these papers to be ranked high with respect to the three measures in comparison to all the other papers that were published in the same conferences and in the same years. Next, we report some interesting findings; as an example to demonstrate the value of density based influence metrics, the paper "*Development of the Domain Name System*" which was published in the SIGCOMM conference in 1988 has received only 81 citations in our computer science dataset, while the most cited paper

[1] https://aminer.org/citation
[2] https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/
[3] https://en.wikipedia.org/wiki/List_of_physics_journals

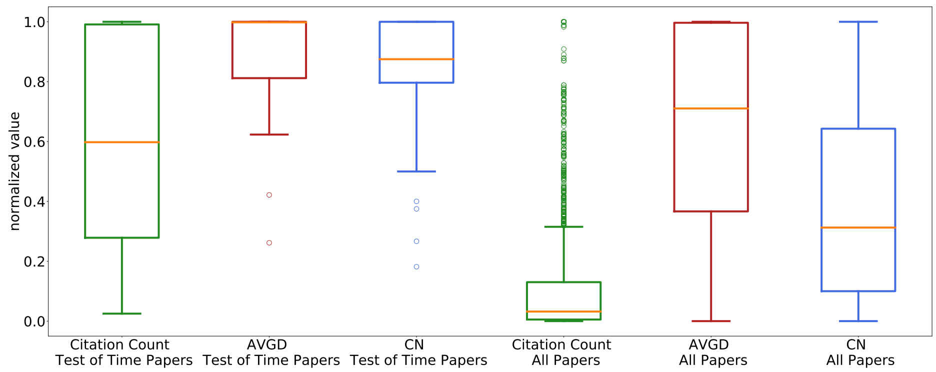

Figure 4. Citation counts, average degrees and core influences of 48 Test of Time papers and of all the papers that were published in the same conferences and in the same years. Given a paper and a measure, the value of the paper according to that measure has been normalized by dividing with the maximum value among the papers published in the same conference and in the same year.

published in the same conference and in the same year has recieved 1080 citations. Conversely, the core influence of that paper is the highest among the papers that were published in SIGCOMM that year (equal to 15). In a similar example, the paper "*An Analysis of BGP Convergence Properties*" was published in the same conference in 1999, and has received only 100 citations, 645 less that the most cited paper of that conference. The core influence of this paper is again the highest among papers of that conference (equal to 13). We computed the three measures for 48 Test of Time papers and for the rest of the papers published in the same conferences, and in Figure 4 we illustrate the distribution of the three measures when considering the Test of Time papers and when considering all the papers published in these conferences. Given a paper and a measure, we have normalized the value of the paper by dividing with the maximum value of the measure attained by papers published in the same conference. Hence, a value equal to 1 indicates the top paper according to the considered measure, while a value equal to 0.5 indicates that a paper exhibits half the value of the top paper. It is clear that the Test of Time papers have very high average degrees compared to the other papers that were published in the same conference. This is not surprising since, as mentioned above, we expect average degree to take high values for very influential papers. However, it can also take high values for low-impact papers which is its major weakness. The core influence of the Test of Time papers is also much higher than that of the other papers that were published in the same conferences. Specifically, the average core influence of the Test of Time papers is approximately $9/10$ of the maximum possible average core influence. On the other hand, the number of citations that the Test of Time papers have received is generally much lower than those received by the top-cited papers of the same conference. However, we should note that Test of Time papers have received as well a greater number of citation than those published in the corresponding conference and year.

In order to further evaluate the three measures of influence, we used a list[4] -build by a general consensus- of the most influential papers in the domain of computer science so as to compare how the metrics evaluate them in comparison to those that attain the highest values according to the three measures in the dataset. The list, as found on Wikipedia, could match 64 papers from the dataset of computer science from which an equivalent number of most cited, highest average degree and highest core influence papers was also selected. Figure 5 shows the comparisons of all metrics for the different selections of papers. From Figure 5 (Left), it is clear that the majority of the influential papers from Wikipedia and almost all the papers with highest core influence also have high average degrees. Conversely, the average degree of the RCGs of several of the top-cited papers is low. In Figure 5 (Middle), we can see that most of the influential papers from Wikipedia and the majority of the most-cited papers attain relatively high values of core influence, which is not the case for the top papers according to the average degree. However, the core influences of these papers are still somewhat lower than those of the top papers according to core influence. Regarding the citation count measure, it is important to note that there are several papers whose number of citations is much larger than the number of citations of most influential papers. For example, there is a paper which has received more than $8,000$ citations, while the majority of the influential papers has received less than $1,000$ citations. Figure 5 (Right) shows how many of the 64 influential papers from Wikipedia are included in the top-k lists of the three measures for different values of k. In general, for k up to 10^3, the top-k list according to citation counts contains many more influential papers than the corresponding lists of core influence and average degree. However, for larger values of k the core influence gets very close to citation counts in terms of the number of influential papers the top-k lists of the two measures contain. Conversely, average degree fails to compete with the other two measures except for very high values of k ($> 10^5$).

[4] https://en.wikipedia.org/wiki/List_of_important_publications_in_computer_science

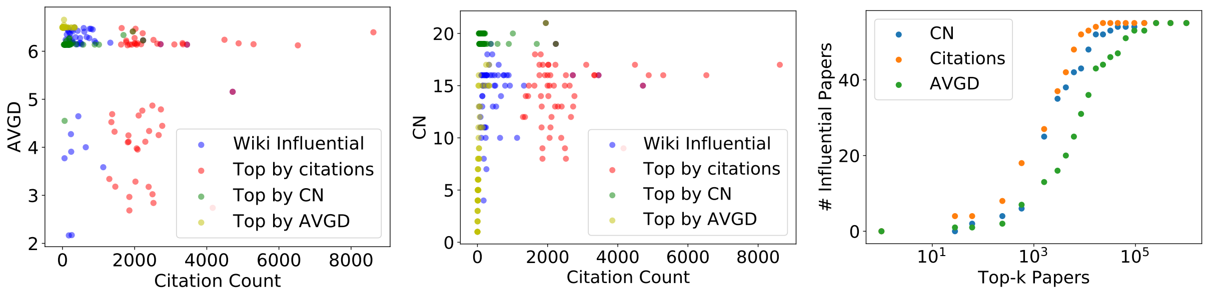

Figure 5. Comparison of the average degrees and citation counts (Left), and core influences and citation counts (Middle) of the 64 influential papers from Wikipedia and the top-64 papers as ranked by the three measures. The Right plot illustrates how many of the 64 influential papers from Wikipedia are included in the top-k lists of the three measures for different values of k.

From the previous analysis, it is clear that citation counts and core influence exhibit similar behavior with respect to the influential papers from Wikipedia. However, Figures 2 and 3 show that the top-k items according to one measure are very different from the top-k items according to another measure. This is an interesting observation which shows that although the two measures evaluate the influential papers almost equivalently, their results do not in general agree with each other. This suggests that core influence captures some properties of research papers that is not captured by citation counts. Therefore, it would be meaningful core influence to be used in parallel to citation counts, since the two measures seem to complement each other, and could thus quantify scientific impact more accurately.

Next, we identify the highest impact papers of our physics dataset according to the three measures. Specifically, Table 1 illustrates for each of the three measures, the top 10 papers of our physics dataset. Interestingly, there are no common papers between the three sets nor between any two of them. The majority of the top papers according to citation counts come from the field of chemical physics, while the top papers according to core influence and average degree come from the field of theoretical physics. We also identify distinguished scholars in the list of authors of the top papers. In the case of citation counts, from a total of 20 authors, Andre Geim and Konstantin Novoselov have received the Nobel Prize in Physics for their work on the graphene material. However, none of the rest 18 authors has won any prestigious prize or medal. In the case of core influence, Edward Witten, the author of the top paper, has won among others a Fields medal and a Dirac medal. From the 15 scientists that have co-authored the remaining 9 papers, 4 of them have also won a Dirac medal. Finally, from the 13 researchers appearing in the authorship list of the top papers according to average degree, only Cumrun Vafa has won a Dirac medal.

Evaluating Author Impact

In order to provide an evaluation at the author level, we employ the well-known h-index metric[2]. The h-index measures the productivity and the quality of the work of a researcher. Hence, it considers both the number of publications and their impact (i. e. number of citation received). Specifically, a researcher has an h-index equal to k if she has published k papers each of which has received at least k citations. More formally, if S is the sequence of papers of an author A ordered in decreasing order of their in-degrees in their corresponding RCGs, then, her h-index is defined as $h\text{-index}(A) = \max_{i \in \{1,\dots,|S|\}} \min(d_{in}(S_i), i)$ where $d_{in}(v)$ is the in-degree of a paper v, and S_i the i^{th} paper of sequence S. The h-index can be directly extended to our setting by replacing the in-degrees of the roots of the RCGs in sequence S with their core numbers. Of course, sequence S has then to be re-ordered based on the core numbers it contains. The h-index of an author is now equal to $h\text{-index}(A) = \max_{i \in \{1,\dots,|S|\}} \min(c(S_i), i)$. The h-index can also be computed using average degree as its basic component.

Table 2 illustrates the 10 authors from the computer science dataset with highest h-index with respect to the three measures. The best researchers according to the in-degree, average degree and core number are shown at the left, middle and right of the Table, respectively. All the top authors identified by the three measures are indeed considered to be among the most prestigious scientists, and have conducted pioneering work in their fields of research. Quite surprisingly, the top authors according to the average degree are those that have won the most Turing awards, the most prestigious award in computer science. Specifically, Robert W. Floyd, A.J. Perlis, Niklaus Wirth, Donald E. Knuth and C.A.R. Hoare have all received that award. Conversely, none of the top authors according to citation counts has received a Turing award, while Michael Stonebraker is the only researcher from the top authors according to core influence that has won such an award. The top authors according to the citation count and the core influence appear to be younger than those returned by the average degree. To support this claim, we computed the average year of first publication for the best 100 researchers with respect to citation counts, average degree and core influence, and this turned out to be the years 1988, 1968 and 1983, respectively. Hence, it is clear that there is a large difference in the age of the top authors according to average degree and those according to citation counts and core influence. Another interesting

Citation counts	Density-functional thermochemistry. III. The role of exact exchange
	Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set
	Generalized Gradient Approximation Made Simple
	Comparison of simple potential functions for simulating liquid water
	Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions
	Molecular dynamics with coupling to an external bath
	Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density
	A fifth-order perturbation comparison of electron correlation theories
	The rise of graphene
	Fast parallel algorithms for short-range molecular dynamics
Average degree	Orbifolds and Solitons
	Branes and N = 1 duality in string theory
	String Vacua and Orbifoldized L-G Models
	Fusion rings and geometry
	On classification of $N = 2$ supersymmetric theories
	Topological orbifold models and quantum cohomology rings
	Exact Results for Supersymmetric Sigma Models
	Singularity Theory and $N = 2$ Supersymmetry
	Nonlinear partial difference equations for the two-dimensional Ising model
	Ising field theory: Quadratic difference equations for the n-point Green's functions on the lattice
Core influence	Anti-de Sitter space and holography
	The Large N Limit of Superconformal Field Theories and Supergravity
	M theory as a matrix model: A Conjecture
	Gauge theory correlators from noncritical string theory
	An Alternative to Compactification
	Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity
	Microscopic origin of the Bekenstein-Hawking entropy
	A Large mass hierarchy from a small extra dimension
	Wilson Loops in Large N Field Theories
	RR Flux on Calabi-Yau and Partial Supersymmetry Breaking

Table 1. Top 10 research papers in physics according to citation counts, average degree and core influence.

Citation counts		Average degree		Core influence	
Hector Garcia-Molina	Jennifer Widom	Robert W. Floyd	Zohar Manna	Donald P. Greenberg	Michael Stonebraker
Jiawei Han	Philip S. Yu	Seymour Ginsburg	C. A. R. Hoare	David H. Salesin	Rakesh Agrawal
Scott Shenker	Deborah Estrin	A. J. Perlis	George E. Forsythe	Hugues Hoppe	Jeffrey D. Ullman
Christos Faloutsos	Anil K. Jain	Niklaus Wirth	Peter J. Denning	Pat Hanrahan	Michael F. Cohen
Hari Balakrishnan	W. Bruce Croft	Donald E. Knuth	Jeffrey D. Ullman	David J. DeWitt	Christos Faloutsos

Table 2. Top 10 researchers in computer science according to the h-indexes based on citation counts, average degree and core influence.

observation is that although average degree suffers from the afformentioned weakness, when used as part of the h-index to evaluate the research work of an author, it can identify very influential scientists. This is because an author may have published a paper whose RCG is accidentally dense and its average degree is high, however, unless the author is very influential, she cannot have published multiple such papers. Finally, it is interesting to note that the field of study of some of the top authors according to core influence is computer graphics, and although these authors have not been awarded with a Turing award, they have won some Academy Awards (a.k.a. Oscars) instead. For instance, Pat Hanrahan has received three Academy Awards for his work in rendering and computer graphics research.

Evaluating Journal Impact

The journal impact factor is by no doubt the most popular tool used to measure the importance of journals[4]. Although more than 60 years have passed since the journal impact factor was first proposed, it still serves as the major evaluation metric of a journal's impact on the scientific community. Journals with higher impact factors are often deemed to be more important and influential than those with lower impact factors. Hence, it is not surprising why researchers take impact factors seriously into account when deciding to which journal to submit their work. For a given year, the impact factor of a journal is equal to the average number of citations in that year for articles published in the past two years in that journal. In our setting, this translates to the average in-degree of the roots of the RCGs corresponding to the articles published in the journal in the past two years, but only for edges from papers published in the current year. In other words, the impact factor of a journal is equal to the average

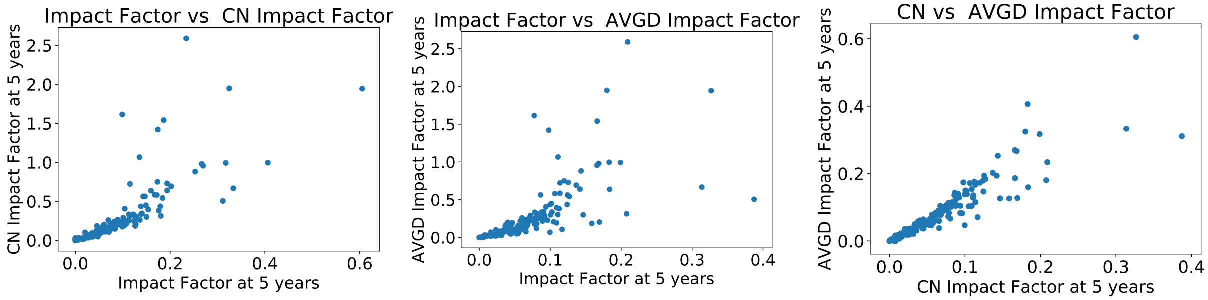

Figure 6. Comparison of the five-year impact factors of all physics journals based on core influence and citation counts (Left), average degree and citation counts (Middle), average degree and core influence (Right).

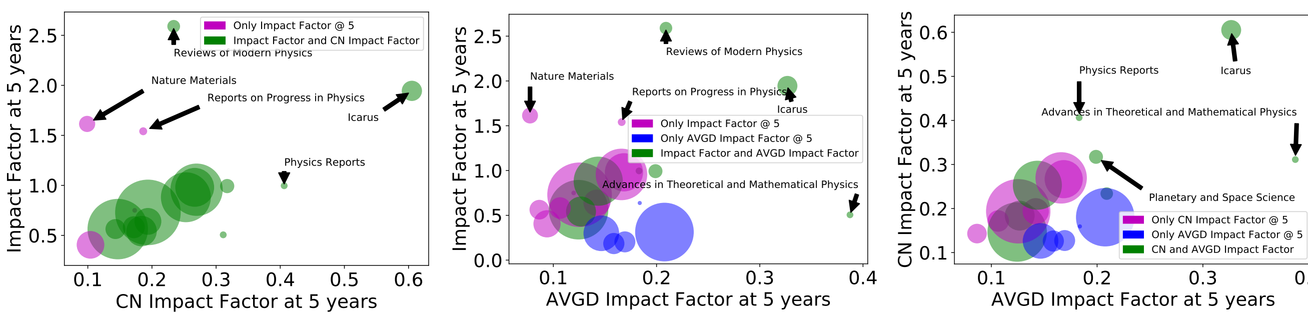

Figure 7. Comparison of the five-year impact factors of a subset of physics journals based on citation counts and core influence (Left), citation counts and average degree (Middle), core influence and average degree (Right). Only journals that are consistently in the top 30 journals according to the three impact factor from 2005 to 2011 are included in the plots.

in-degree of the roots of these RCGs if we remove from each RCG all the vertices (and their adjacent edges) corresponding to papers that were not published in the current year. Note that the impact factor is not necessarily limited to a period of two years, but it is possible to calculate it for any desired period. Many bibliographic databases include a five-year impact factor which is computed on a five-year window of citation data. This longer period of time smooths out the volatility of the values.

As in the case of the h-index, the proposed core influence and average degree measures can be directly used as the building block of the journal impact factor. However, since a unit in the case of core influence and average degree is not so clearly defined as in the case of citations (where each unit corresponds to a paper), it is necessary to modify slightly the definition of impact factor. Simply put, the core number and the average degree are calculated over the whole network which means that, if we disregard all citations except the ones of a specific window, we will not capture the network structure nor the essence of the RCG which requires publications possibly older than the range of a specific time window. If we dismiss those (older) publications, we cannot tell much about the "'growth" of the RCG. Since we cannot compute the average core influence that articles published in the journal over a specific window in time, we instead compute the average increase of their core influence over said window. The same applies to the case of average degree. Note, however, that average degree may decrease from one year to another, hence, impact factors based on the average degree may take negative values. The above definition of impact factor does not violate any properties of the traditional definition. In fact, the impact factor itself is essentially the average increase of the citation counts occurred this year of articles published in the journal in the past two years.

Figure 6 illustrates the five-year impact factors of all physics journals according to the the three measures and compares each pair of them. The impact factors were computed for the year 2011. Therefore, the papers that participate in the calculation of the impact factors were all published between 2005 and 2010. As can be seen from the Figure, in general, the impact factors according to the three measures agree with each other for journals attaining small and medium values. However, while impact factors based on core influence and average degree also provide quite similar evaluations for the rest of the journals, this is not true when impact factor based on citation counts is compared to them. Journals that have a high impact factor according to the number of citations may have a low impact factor according to core influence and average degree, and vice versa. The plots of Figure 7 are similar to those of Figure 6. However, in this case, not all physics journals are included in the plots, but only a subset of them. More specifically, for each pair of impact factors, only the journals that were consistently in the top 30 journals according to each impact factor from 2005 to 2011 are taken into account. The size of the marker representing each journal is proportional to the number of papers over which impact factors are computed, that is, equal to the number of papers published by that journal from 2005 to 2010. The colour of the marker indicates from which impact factor the corresponding journal emerged. For example, a green marker indicates that the journal was every year present in the top 30 journals according to both impact factors that are compared to each other. Interestingly, all the top journals selected using the core-based influence impact factor were also selected when using the citation-based impact factor. Anecdotally, Icarus is one of the highest impact journals according to all versions of impact factors; Nature Materials, although it is one of the most important journals according to the traditional impact factor, its core influence and average degree impact factors are at relatively low values. Furthermore, Reviews of Modern Physics is the top-rated journal of the citation-based impact factor, but it has slightly above average impact with respect to the two other impact factors. Finally, Physics Reports while being highly-rated by the core influence impact factor, it is not ranked that high when utilizing the two other impact factors.

Discussion

The RCG representation opens up new avenues of research that can potentially lead to very interesting measures of influence of research papers or even other domains like social network analysis. As mentioned above, the total number of citations, which has been traditionally used as a measure of the impact of a paper, only takes into consideration papers that are directly influenced by a given paper. However, it ignores papers that may have been implicitly influenced by that paper. Undoubtedly, the RCG of a paper contains all the papers that may have been potentially influenced by that paper. This representation is thus all we need for quantifying the true impact of a research manuscript. Based on the RCG representation, we can define measures that aim at discerning the genuine number of publications the paper has had an impact upon. In this article, we defined two such measures, but there are several more that can be defined (e. g. centrality measures). The evaluation of all candidate methods is out of the scope of this article. However, we hope that this article will serve as a stimulus for further work on designing new measures of scientific impact based on the RCG representation.

It is interesting to note that in this article, we have presented the simplest form of the RCG representation. In fact, both edges and vertices of RCGs can be enriched by additional attributes. For example, we may add edge attributes that denote the absolute time difference between the publication dates of the two endpoint papers of each edge. Similarly, vertex attributes such as those representing the research field of each paper or its set of authors may prove quite beneficial when computing a paper's impact. However, this enriched RCG representation requires more sophisticated measures which take these attributes into account, and which may result in additional computational complexity. Nevertheless, it is clear that the flexibility of the RCG representation allows us to build a plethora of measures quantifying a paper's influence, some of them yielding higher descriptive power than others.

Regarding the proposed core influence measure, it seems to be capable of distinguishing those papers that have a great influence in their corresponding field. In contrast to the in-degree which is a local in nature measure, k-core is a more global measure since its calculation requires the complete (sub)network structure. Hence, it is very likely that it can identify structural information that in-degree could never detect. The number of citations a paper has received is an upper bound of its core influence. Specifically, let v be a paper and $\mathcal{N}(v)$ the set of papers that cite v. If for each paper $u \in \mathcal{N}(v)$, it holds that, besides v, u also cites all the papers in $\mathcal{N}(v)$ that were published earlier than itself, the core influence of v will be equal to its in-degree, that is, equal to its total number of citations. Conversely, if all the papers in $\mathcal{N}(v)$ do not cite others from $\mathcal{N}(v)$, the core number of the root will be equal to 1. It is thus clear that a highly-cited paper v can have very low core influence if the papers in $\mathcal{N}(v)$ do not cite other papers in that set. Furthermore, since core influence is a monotonically increasing function of time, its computation over a time period could give an indication of the diffusion of the paper in its field, and reveal much information regarding the reception of the paper by the scientific community.

We should note that the two proposed measures of influence suffer from several limitations which are also encountered in the case of citation counts. For example, core influence can be easily manipulated by an individual through self-citation to increase her own value of influence. However, strategies for dealing with such problems that have already been developed for citation counts can in most cases be adapted to the case of the proposed measures. For example, we may remove from the RCG of a paper all those papers that are written by the same authors, or we can assign lower weights to the edges adjacent to those papers and then, use a weighted k-core decomposition algorithm to determine core influence.

It is quite encouraging that the proposed measures can directly serve as the building blocks of well-established methods that quantify the impact of researchers, teams, institutions and journals. It is our belief that the field of scientific impact evaluation is now mature enough such that "destroy and build from scratch" approaches are not well-suited. This is why we capitalize on measures that have survived through time to quantify the impact of higher-lever entities such as authors and journals. Of course, as mentioned above, these measures also suffer from several limitations. To this point, it is important to stress out that by defining measures that remain consistent with existing ones (like the ones we proposed which are in line with citation counts), a whole toolbox of measures of scientific influence at different granularity levels becomes available.

The landscape of scientific impact evaluation has undergone significant changes in the last decades, and has reached a steady state and we do not expect dramatic changes on that landscape. The scientific impacts of papers, authors and journals are dominated by the number of citations, the *h*-index and the journal impact factor, respectively. However, since all measures of scientific impact suffer from some limitations, it is not fair to use a single measure. Instead, it would be beneficial to use a variety of them. The suggested RCG representation and methodology to measure impact goes against the narrow nature of citation counts, putting forward an alternative ideology to quantify influence. We thus believe that the proposed measures can serve as useful complementaries of the long-established measures of scientific influence.

Methods

k-core. Let G be a graph and G' a subgraph of G induced by a set of vertices S. Then, G' is defined to be a k-core of G, denoted by C_k, if it is a maximal subgraph of G in which all vertices have degree at least k. Hence, if G' is a k-core of G, then $\forall v \in S$, $d_{G'}(v) \geq k$. Each k-core is a unique subgraph of G, and it is not necessarily connected. The core number $c(v)$ of a vertex v is equal to the highest-order core that v belongs to. In other words, v has core number $c(v) = k$, if it belongs to the k-core but not to any $(k+1)$-core. The degeneracy $\delta^*(G)$ of a graph G is defined as the maximum k for which graph G contains a non-empty k-core subgraph, $\delta^*(G) = \max_{v \in V} c(v)$. Furthermore, assuming that $\mathscr{C} = \{C_0, C_1, \ldots, C_{\delta^*(G)}\}$ is the set of all k-cores, then $\mathscr{C}$ forms a nested chain: $C_{\delta^*(G)} \subseteq \ldots \subseteq C_1 \subseteq C_0 = G$. Since the k-cores of a graph form a nested chain of subgraphs, the k-core decomposition is a very useful tool for discovering the hierarchical structure of graphs. The popularity of the k-core decomposition stems mainly from the fact that it can be computed in linear time[27,28]. The algorithm runs in $\mathscr{O}(max(n,m))$ time. The underlying intuition is that we can obtain the i-core of a graph if we recursively remove all vertices with degree less than i and their incident edges from the graph until no other vertex can be removed. Since higher-order cores are nested within lower-order cores, we compute k-cores sequentially from $k = 0$ to $k = \delta^*(G)$. Therefore, at each iteration, we remove the lowest degree vertex and sets its core number accordingly.

Properties of Rooted Citation Networks. A citation network is a social network whose vertices correspond to papers and are linked to each other by citation relationships[20]. Given a single paper, we defined its rooted citation graph (RCG) as follows: Given a paper v of a citation network $G = (V, E)$, the RCG of v is the subgraph of G induced by a set of vertices S which contains v and all the vertices which can reach v by a directed path. Set S is formally defined as $S = \{v\} \cup \{u : (u, e_1), (e_1, e_2), \ldots, (e_n, v) \in E\}$. From a different perspective, S is a set that contains v and has the property that the children of all its elements are also in the set. We next give some results and properties related to Rooted Citation networks.

Lemma 1 *Let $d_{in}(v)$ be the in-degree of the root v of a rooted citation network. Then, $d_{in}(v) \geq c(v)$.*

Lemma 2 *Let v be the root of a rooted citation network and $C(v)$ denote the set of vertices belonging to the $c(v)$-core and from which there is a path to the root v. Then, the out-degree of all the vertices except v in the subgraph induced by $C(v)$ is at least* 1.

Lemma 3 *Let v be the root of a rooted citation network and $C(v)$ denote the set of vertices belonging to the $c(v)$-core and from which there is a path to the root v. Then, there is a vertex whose out-degree in the subgraph induced by $C(v)$ is at least $c(v)$.*

Lemma 4 *Let $C(v)$ denote the set of vertices belonging to the $c(v)$-core of a rooted citation network and from which there is a path to the root v. Then, every cycle in the undirected subgraph induced by $C(v)$ corresponds to two directed paths from exactly one vertex in the cycle to another vertex in the cycle in the directed subgraph induced by $C(v)$.*

References

1. Gross, P. & Gross, E. College libraries and chemical education. *Sci.* **66**, 385–389 (1927).
2. Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl. academy Sci. United States Am.* **102**, 16569–16572 (2005).
3. Egghe, L. Theory and practise of the g-index. *Sci.* **69**, 131–152 (2006).
4. Garfield, E. Journal impact factor: a brief review. *Can. Med. Assoc. J.* **161**, 979–980 (1999).
5. Waltman, L. & Van Eck, N. J. The inconsistency of the h-index. *J. Assoc. for Inf. Sci. Technol.* **63**, 406–415 (2012).
6. Editors, P. M. *et al.* The impact factor game. *PLoS medicine* **3**, e291 (2006).
7. MacRoberts, M. H. & MacRoberts, B. R. Problems of citation analysis: A critical review. *J. Am. Soc. for information Sci.* **40**, 342 (1989).
8. Adler, R., Ewing, J., Taylor, P. *et al.* Citation statistics. *Stat. Sci.* **24**, 1 (2009).

9. Thorne, F. C. The citation index: Another case of spurious validity. *J. Clin. Psychol.* **33**, 1157–1161 (1977).

10. Seglen, P. O. The skewness of science. *J. Am. Soc. for Inf. Sci.* **43**, 628 (1992).

11. Zhu, X., Turney, P., Lemire, D. & Vellino, A. Measuring academic influence: Not all citations are equal. *J. Assoc. for Inf. Sci. Technol.* **66**, 408–427 (2015).

12. Waltman, L. & van Eck, N. J. A systematic empirical comparison of different approaches for normalizing citation impact indicators. *J. Informetrics* **7**, 833–849 (2013).

13. Zitt, M. & Small, H. Modifying the journal impact factor by fractional citation weighting: The audience factor. *J. Assoc. for Inf. Sci. Technol.* **59**, 1856–1860 (2008).

14. Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S. & van Raan, A. F. Towards a new crown indicator: Some theoretical considerations. *J. informetrics* **5**, 37–47 (2011).

15. Hutchins, B. I., Yuan, X., Anderson, J. M. & Santangelo, G. M. Relative citation ratio (rcr): A new metric that uses citation rates to measure influence at the article level. *PLoS biology* **14**, e1002541 (2016).

16. Bergstrom, C. T., West, J. D. & Wiseman, M. A. The eigenfactor^{TM} metrics. *J. Neurosci.* **28**, 11433–11434 (2008).

17. Aragón, A. M. A measure for the impact of research. *Sci. Reports* **3**, 1649 (2013).

18. Martin, B. R. The use of multiple indicators in the assessment of basic research. *Sci.* **36**, 343–362 (1996).

19. Price, D. J. D. S. Networks of scientific papers. *Sci.* 510–515 (1965).

20. Egghe, L. & Rousseau, R. *Introduction to informetrics: Quantitative methods in library, documentation and information science* (Elsevier Science Publishers, 1990).

21. Redner, S. How popular is your paper? an empirical study of the citation distribution. *The Eur. Phys. J. B-Condensed Matter Complex Syst.* **4**, 131–134 (1998).

22. Lehmann, S., Lautrup, B. & Jackson, A. Citation networks in high energy physics. *Phys. Rev. E* **68**, 026113 (2003).

23. Qazvinian, V. & Radev, D. R. Scientific Paper Summarization Using Citation Summary Networks. In *Proceedings of the 22nd International Conference on Computational Linguistics*, 689–696 (2008).

24. Pei, S., Muchnik, L., Andrade Jr, J. S., Zheng, Z. & Makse, H. A. Searching for superspreaders of information in real-world social media. *Sci. Reports* **4**, 5547 (2014).

25. Seidman, S. Network Structure and Minimum Degree. *Soc. networks* **5**, 269–287 (1983).

26. Fagin, R., Kumar, R. & Sivakumar, D. Comparing top k lists. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '03, 28–36 (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2003).

27. Matula, D. & Beck, L. Smallest-last Ordering and Clustering and Graph Coloring Algorithms. *J. ACM* **30**, 417–427 (1983).

28. Batagelj, V. & Zaveršnik, M. Fast algorithms for determining (generalized) core groups in social networks. *Adv. Data Analysis Classif.* **5**, 129–145 (2011).