

Looking deeper into academic citations through network analysis: popularity, influence and impact

Salvatore F. Pileggi¹

Published online: 10 July 2017
© Springer-Verlag GmbH Germany 2017

Abstract Google Scholar (GS) has progressively emerged as a tool which “provides a simple way to broadly search for scholarly literature across many disciplines and sources.” As a free tool that provides citation metrics, GS has opened the academic word to a much larger audience, according to an open information philosophy. GS’ profiles are largely used not only to have a quick look at the authors and their works but, more and more often, as a “de facto” metric to quickly evaluate the research impact. This process looks unstoppable and discussing about its fairness, advantages and disadvantages, as well as about social implications is out of the scope of this paper. We rather prefer to (1) briefly discuss the changes and the innovation that GS has introduced and to (2) propose possible improvements for analysis on academic citations. Our methods are aimed at considering a GS profile in its proper context, providing a social perspective on academic citations: Although maintaining a fundamentally quantitative focus, novel approaches, based on complex network analysis, distinguish between a research impact on the authors’ research network and a more general impact on the scientific community.

Keywords Open data · Network analysis · Analytics · Social network · Research impact

1 Introduction

Google Scholar (GS) [3] has progressively emerged as an open tool which “provides a simple way to broadly search for scholarly literature across many disciplines and sources.” GS actually proposes an enriched environment including citation count and metrics.

GS is definitely not the first tool which provides that kind of information [12]; while other databases with a similar purpose (e.g., Web of Science [9], SCOPUS [7], IEEEExplore [4], ACM Digital Library [1]) are restricted to subscribers, GS has quickly gained a relevant (if not dominant) position, mostly thanks to the fact that it is freely available to everyone. Something of similar has happened also in the context of online social networks where ResearchGate [6], the largest academic social network in terms of active users, has integrated common networking features with citation count and other metrics.

Indeed, in the Information Society the accessibility of information is a primary and key issue [24]: regardless of the target profile (student, researcher, common user) and of its expertise, an ideal learning process assumes (or should assume) an environment where knowledge is completely available and accessible to everyone. Despite the persistent and increasing support of technology in a de facto completely digital world [36], the accessibility skill seems to be far away from the optimal one: more and more people are effectively getting that skill through the Web; but that skill is limited to freely available contents, while the best sources are still on payment; searching and discovering content is not always easy and, anyway, depends on the policy or strategy implemented by common brokers (e.g., research engines or online social networks); such mechanisms are often modeled according to business criteria; last but not the least, there is a serious risk to feel lost in a

✉ Salvatore F. Pileggi
s.pileggi@uq.edu.au;
<http://www.flaviopileggi.net>

¹ School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia

digital world where tracking the reliability and the quality of sources is getting harder everyday.

In this context, GS is probably a step forward with respect to its main competitors. As a free tool, GS has opened the academic word to a much larger audience [21]. Indeed GS' public profiles have already entered the life of most scientists and researchers. They provide a quick overview of authors' contribution, as well as of its impact on the community through citations count, metrics, and the consequent documents ranking [34].

GS' profiles are largely used not only to have a quick look at the authors and their works but, more and more often, as a de facto compact metric to quickly evaluate research "quality" with an important impact on researchers' careers. The universally accepted assumption is the questionable equivalence between popularity and impact/quality. That is a controversial concept [11] which leads to an intrinsically ambiguous evaluation model [23]. This process looks unstoppable and discussing about its fairness, advantages and disadvantages, as well as social implications is out of the scope of this paper.

Like any other product, GS has its pros and cons [25]. From a critical perspective, apart from the already mentioned open and free approach, the most significant difference between GS and its main competitors [15, 28] is the fact that GS considers *information sources extended to non-academic documents* (including gray literature [14]): Whichever document published on the Web can be potentially aggregated to the GS's database. In practice, this approach is in a strong contrast with similar tools that only consider academic documents or, more realistically, a part of them. On the one hand, this approach could open to a more exhaustive understanding of the real impacts; however, on the other, it addresses evident concerns on reliability and accuracy in the case of non-supervised analysis [33]. Despite the fact that it is hard to completely rely on documents from uncertain sources, the attempt to consider the research impact in a context wider than the simple well-known academic environment is valuable: It looks more convincing than other approaches, such as volatile metrics (e.g., read and download count) recently adopted within reputable portals.

Evident limitations common to all platforms dealing with citation metrics can be summarized as follows:

- *the quantitative nature of the approach* [30], as all documents have the same weight/importance, meaning they contribute atomically and homogeneously to measure impacts as pieces of knowledge. That is somehow in line with the current trends which assume significant approximations on Big Data based on quantitative analysis. However, it appears a severe barrier for a qualitative analysis [20]. Recent studies

(e.g., [28]) clearly show the sensitivity of the impact with respect to the sources considered: Varying the sources, even in a reputable context, has a huge impact on the estimation of citations counting and on consequent rankings.

- *the popularity of a research as the measure of its impact* [37] according to the current technological climate (Web 2.0 [31, 32]). That is at least questionable in a generic meaning but largely accepted in a context that assumes reputable sources of information and quantitative analysis. Considering contextual analysis and/or a less radical relation between popularity and impact could really provide strong benefits in terms of analysis capabilities.

This first part of the paper is completed by the two next subsections aimed, respectively, at deeply understanding the use of GS in practice from the researchers' perspective and at discussing the goal of this work. The second part of the paper deals with details on possible techniques of contextual analysis to correctly interpret GS profiles. The paper ends with a section of conclusions and future work.

1.1 Use or misuse?

The scope of GS is stated as follows:

Google Scholar aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature [3].

As researchers are often looking at the world in a different way, how are they really using the information GS provides? Namely, what is the perspective of a researcher?

In order to provide adequate and effective techniques of analysis, it is important to fully understand that perspective. Therefore, we have performed a small-scale survey among active researchers, meaning those persons who are regularly involved in collaborative research projects and are usual to publish their works in internationally recognized venues (e.g., ranked conferences [2] and JCR indexed journals [5]). That should be a guarantee of belonging to a significant research network. We have focused on experts from academia, and, due to the purpose of the study, we preferred unconnected (at least in theory) researchers: We selected the participants from different countries and institutions, from different areas of expertise and with no (known) past or ongoing collaborations (e.g., common papers or projects). The common pattern that connects the participants is their interest, direct or indirect, in IT and its applications. Basically we have considered researchers working in technology or in strong touch with it. Generally speaking, designing a comprehensive,

context-less, unambiguous survey would have been quite complex. We preferred to pragmatically focus on the practical behavior of researchers, trying to escape, in the limit of the possible, their personal opinions.

The featured part of our survey includes the questions below:

Do you look at the number of citations of a paper before citing it? Most interviewed researchers (80%) have admitted to look at the number of citations of a paper before citing it. While 40% of them does it only occasionally, the remaining part considers it a part of its normal behavior.

Do you prioritize the citations of works from colleagues? A half of interviewed answered positively, with a part (20%) that considers it not intentional.

Do you associate the quality mostly with the venue or with the citations? Most researchers (60%) consider the number of citations like a parameter of quality, even though only 10% considers it the only parameter of quality.

A summary of the survey is shown in Fig. 1. Looking at the results as a whole, regardless of the multiple possible interpretations, our doubts and concerns on the use and interpretation of the information provided by GS and similar tools are far away to be solved or to disappear.

Focusing on academia, we extended our investigation to CV assessment/selection and grant evaluation. As such evaluations normally concern only senior academics, it has been much harder to perform a significant number of interviews. Therefore, more than a proper survey, we have undertaken an informal study based on the interview of few selected senior academics. The purpose of a CV assessment is the evaluation of a given profile to match the requirements of a certain position, role or task. As an academic CV is normally composed of two main parts, teaching and research, the assessment of the impact and quality of

research is usual to play a major role. In the context of grant evaluation, that factor may have a variable weight. For example, in Australia, the Australian Research Council (ARC) is usual to consider the CV of the proposer up to 40% of the overall score for the proposal. All interviewed academics have pointed out that they are usual to undertake some search online about the candidate, in either cases of CV assessment and grant evaluation. They consider GS the primary source of information but not the only one as, depending on the case, they are usual to compare and contrast the information retrieved looking also at other sources. They always rely on common metrics.

1.2 Objective

This work is aimed at overcoming the strictly numeric analysis of citations (normally limited to ambiguous bibliometrics [19]). Even though maintaining a quantitative character, the model proposed in the following sections of the paper extends in fact the common data processing model taking into account the research context. That is in line with most novel approaches which push a semantic enforcement of the scholarly ecosystem (e.g., Semantic Scholar Project [8]).

Those extensions for the analysis capabilities are modeled on graphs, and therefore, their processing can be performed according to common graph analysis techniques [10]. Thus, this paper focuses on the specification of the model and its semantics. Within the model, the research context plays a key and central role. Indeed, it is semantically equivalent to a social context in a social network as it defines a contextual structure that reflects the relationships (e.g., co-authorship) and the interactions (e.g., citation) among authors. Once the research context is modeled, advanced techniques of analysis can be applied in a non-exclusively numeric environment.

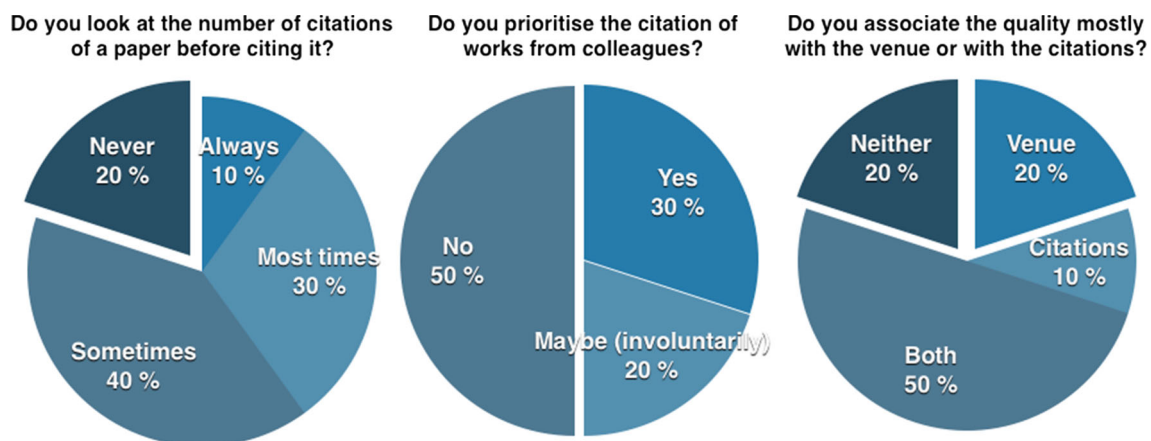


Fig. 1 A small-scale survey performed by interviewing active researchers

Although the paper explicitly refers to GS, the proposed model applies to any other system with an equivalent scope. However, the increased capabilities in terms of analysis might have a much more sensitive role considering open environments and, more in general, a model of knowledge which is not restricted to a number of controlled sources.

2 Social perspective on academic citations

Moving from a context-less to a contextual analysis model is probably the most effective path to overcome the limitations of the existing analysis methods. Indeed, if properly modeled, research networks may play a key role in the analysis of academic citations, as such structures represent the context in which the research initiatives are designed and developed. There are different approaches to define a research network. In the context of this work, we have chosen a social perspective that assumes the research network directly inferred from the collaborations among authors. This approach implicitly assumes a research network conceptually equivalent to a social network.

In order to provide a social perspective on academic citations analysis, we first formally define the generic research network model; then, an overlay application-specific network, built on the former, is proposed; finally, the potential impact of the provided extensions on the analysis capabilities is briefly discussed.

2.1 Modeling research network

A *research network* [16] is a social structure composed of a subset of members within the research community identified by a set of relations. Like any other social structure, a research network is defined by the nature of the relations that connect its members. Common research networks can be modeled according to *explicit linking*, meaning that members build their network by explicitly defining their connections on the model of commercial social networks. Those dynamic links are constantly evolving because of the members' activity [35]. Another common approach adopts an *interest-based* connection [26], where people converge as the function of their research interests. The previously mentioned approaches are often integrated with profile-based networks, where rich user profiles play a relevant role [13], driving the interaction among members and, consequently, the building of the network itself.

In the context of this work, the research network has a limited though key scope at both a modeling and a processing level: It is used to perform a contextualized analysis of the research impact as an extension of common context-less approaches. Indeed, a simplified network

model based *exclusively on the information appearing in the documents and in their citations* is adopted as a driver factor. Basically, we are building the research network upon the information provided by GS.

Two members of the community, a and b , are *strongly (or directly) connected* if they are co-authors of at least one paper p (Eq. 1a). Indeed, members are *indirectly (or eventually) connected* if a path of i direct connections that links them exists. Indirect connections are recursively defined as in Eq. 1b according to a Prolog-like notation (X is a generic variable that matches direct connections among authors). Only minimum paths are considered so $Connected_i(a, b)$ assumes a minimum path between a and b of i steps (Eq. 1c). Furthermore, as the resulting graph is not oriented, $Connected$ is a symmetric relation (Eq. 1d).

$$Connected_0(a, b) \leftarrow \exists p : \text{author}(a, p), \text{author}(b, p) \quad (1a)$$

$$Connected_i(a, b) \leftarrow Connected_0(a, X), Connected_{i-1}(X, b) \quad (1b)$$

$$Connected_i(a, b) \rightarrow \nexists Connected_j(a, b), \quad j < i \quad (1c)$$

$$Connected_i(a, b) \leftrightarrow Connected_i(b, a) \quad (1d)$$

An example of research network as the function of the *path length* is represented in Fig. 2. Member i is co-author of members a, b, c and d . i is indirectly related to e and g according to a factor 1 (a and e are co-authors, as well as d and g). He is also related to f with a factor 2 (e and f are directly related).

This simple model is integrated with the concept of *citation* in order to define a number of overlay networks [38]. As in the common semantic, a citation of a paper p (Cit^p) implies the existence of another paper c citing p (Eq. 2a). The overall number of citations for a paper is obtained by summing single citations (Eq. 2b). Finally, as this work focuses on individual-centric analysis, the citations per author are defined by the sum of the citations of single papers authored (Eq. 2c).

$$Cit^p \leftarrow \exists c : \text{citation}(p, c) \quad (2a)$$

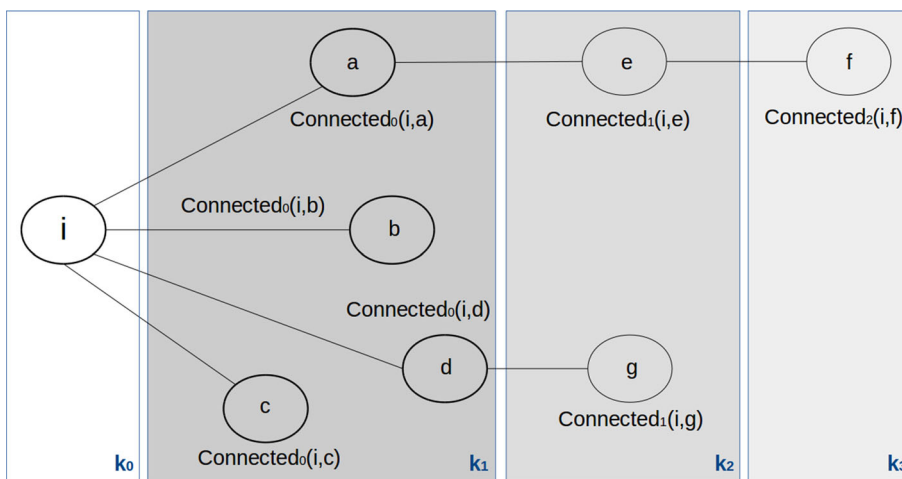
$$CIT^p = \sum_{\exists c : \text{citation}(p, c)} Cit^p \quad (2b)$$

$$CIT(a) = \sum_{\forall p : \text{citation}(a, p)} CIT^p \quad (2c)$$

2.2 k-vector

In the previous subsection, a simple research network based on authorship has been defined, as well as the straightforward concept of citation has been formalized. As already mentioned, citations can be used in a way similar to authorship to define overlay networks on the top of the main research network inside the research community. On

Fig. 2 An example of research network



the other hand, authorships and citations can be merged and orthogonally analyzed to provide a more solid social perspective, which is the purpose of the *k*-vector.

An overlay network based on citations can be recursively defined as in Eq. 3. A *self-citation* is intuitively defined as a citation of a paper *p* by another paper *c* authored/co-authored by at least one of the authors of *p*.

Two members are *directly (or strongly) connected* if Cit_1^p (as in Eq. 3) exists. They are *indirectly (or eventually) connected* if Cit_i^p exists, where *i* is higher than 1 (as in Eq. 3).

$$\begin{aligned}
 Cit_0^p &\rightarrow \exists c : \text{author}(a, p), \text{author}(a, c), \text{citation}(p, c) \\
 Cit_1^p &\rightarrow \exists c, b : \text{author}(a, p), \text{author}(b, c), \text{citation}(p, c), \\
 &\quad \text{Connected}_0(a, b) \\
 Cit_2^p &\rightarrow \exists c, b : \text{author}(a, p), \text{author}(b, c), \text{citation}(p, c), \\
 &\quad \text{Connected}_1(a, b) \\
 Cit_3^p &\rightarrow \exists c, b : \text{author}(a, p), \text{author}(b, c), \text{citation}(p, c), \\
 &\quad \text{Connected}_2(a, b) \\
 &\quad \dots \\
 Cit_i^p &\rightarrow \begin{cases} \exists c : \text{author}(a, p), \text{author}(a, c), \text{citation}(p, c) \\ \quad \text{if } i = 0 \\ \exists c, b : \text{author}(a, p), \text{author}(b, c), \text{citation}(p, c), \\ \quad \text{Connected}_{i-1}(a, b) \quad \text{if } i > 0 \end{cases}
 \end{aligned} \tag{3}$$

The overall number of citations for a given factor *i* is obtained as in Eq. 4a and the correspondent view per author as in Eq. 4b.

$$CIT_i^p = \sum_c Cit_i^p \tag{4a}$$

$$CIT_i(a) = \sum_{\forall p:\text{author}(a,p)} CIT_i^p \tag{4b}$$

Each of the elements of the *k*-vector (*K*(*a*)) is the number of citations per author with the factor correspondent to the index of the vector (Eq. 5).

$$\begin{aligned}
 K(a) &= [k_0(a), k_1(a), \dots, k_{i-1}(a), k_i(a), k_{i+1}(a), \dots, k_n(a)] \\
 k_m(a) &= CIT_m(a) \\
 \hat{K}(a) &= [\hat{k}_0(a), \hat{k}_1(a), \dots, \hat{k}_{i-1}(a), \hat{k}_i(a), \hat{k}_{i+1}(a), \dots, \hat{k}_n(a)] \\
 \hat{k}_m(a) &= CIT_{tot} - CIT_i(a)
 \end{aligned} \tag{5}$$

2.3 Contextual analysis of citations and its social impact

The *k*-vector as previously described provides a formalized social perspective for a simple yet effective contextual citations analysis. More concretely, the major extensions to the common analysis techniques that the model allows are:

- *Basic filtering* detecting self-citations, as well as citations from closely related colleagues, is the very first obvious step for a deeper and more sophisticated analysis.
- *Estimation of the influence inside a given research network*; an individual-centric analysis can be performed in the context of a well-defined data-centric model [17] that allows to distinguish very close collaborators inside a wider research network, as well as indirect influence propagation.
- *Impact outside the contributor’s research network* as a complementary analysis, it is possible to clearly distinguish between the influence inside a concrete research network a considered researcher belongs to, and its impact outside that given sub-network. Although without clearly distinguishing between popularity and impact, this is a straightforward path to

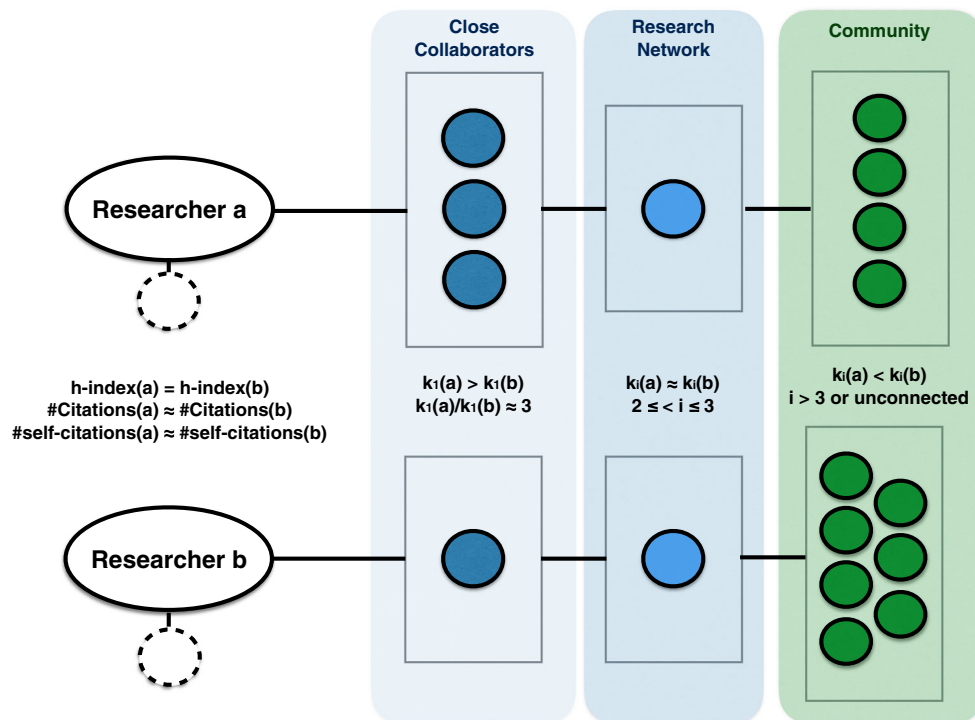


Fig. 3 An example of extended analysis

understand how the scientific community is affected or influenced by a certain contributor or work.

- *Connection among different research networks* individual-centric analysis can easily evolve toward a cluster-oriented study of the network [18], where clusters can be isolated and the relations existing among them can be detected and analyzed accordingly.

One of the questions we often received about this work is the following:

Is that just a different, hopefully improved, way to present academic citations?

It is a key question actually. A better presentation would be useful though, definitely, it does not provide any further knowledge. A simple improvement of the presentation is not the goal of this work, which aims to provide a more sophisticated, rather simple, context-based method of analysis.

In order to support that statement, we have provided some analysis on real profiles, as a preliminary evaluation of our work. We have detected researchers with a strong similarity according to the most common metrics: a similar number of citations and the same h-index. We have not considered the number of documents published. Moreover, as we do not have a direct systematic access to large academic datasets, we have focused on young scientists with a relatively low number of citations. That is because the accurate collection of data is an expensive and time-

consuming process. This very first experiment has allowed some evaluations at a very low scale. Furthermore, we have addressed a further level of similarity by considering multiple data sources, meaning we have tried to identify researchers with very similar performance according to more than one source (GS, ResearchGate and Scopus).

In Fig. 3 we compare the performance of two researchers, *a* and *b*. As previously said, their performance are very similar according to common metrics.

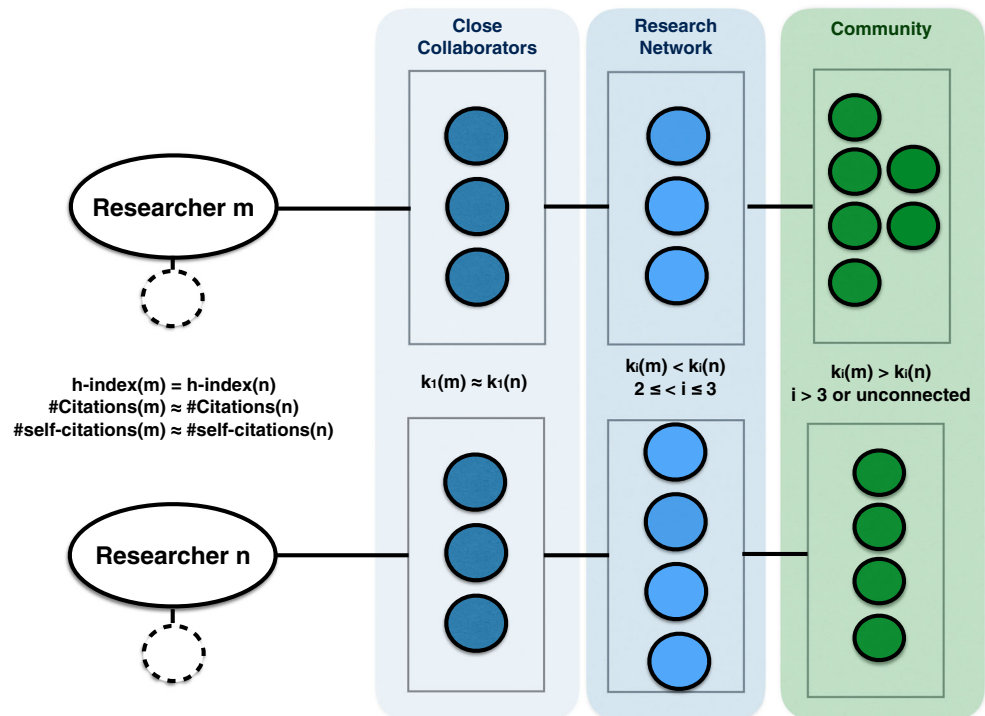
However, our analysis method points out two well different research profiles. Indeed, *a* seems to have a direct influence on closer collaborators and, at least numerically, an equivalent impact on the community. Contrariwise, *b* has a much stronger impact on the community. Neither *a* nor *b* seems to be part of a big research network. The natural conclusion according to this method is that *b* is performing much better than *a*.

A further example is depicted in Fig. 4. Researchers *m* and *n* propose a similar influence on their closer members of their respective networks. *m* has a lower impact within the research network but a much higher impact on the community.

3 Conclusions and future work

Google Scholar is a powerful and well-designed tool with the capability to address Big Data [27] according to an open perspective. It can widely extend its scope to include,

Fig. 4 A further example of analysis



potentially, also alternative kinds of contents and data sources (e.g., blogs and content from social networks [29]).

On the other hand, the capabilities in terms of analysis on this enormous scholarly ecosystem are currently very limited. Novel approaches are still largely unexplored. A contextual analysis of the information which still considers common metrics (e.g., the well-known *h-index* [22]), resulting by processing the information in an appropriate context (e.g., a research network), can provide an added value and enforces more consistent semantics to the target knowledge.

The model for contextual processing and analysis proposed in this paper is relatively simple because it does not assume additional external data sources and focuses exclusively on making as explicit as possible the knowledge already present inside the GS database. As mentioned, the proposed approach applies to any other system different from GS with a similar scope. By adopting such techniques of analysis, typical advice looking at common metrics (e.g., “indices are comparable only within the same research area”) could be replaced by objective, effectively comparable and unambiguous parameters supporting the main indices. In other words, we do not want to provide new metrics; we want to correctly interpret the existing ones.

The model proposed in this paper is not just another way to present academic citations but, rather, an analysis technique to give a further and more reliable meaning to the current metrics for the measure of the research impact.

Future steps will be oriented to the validation of the model proposed. The ideal environment should include real data, meaning a cluster of the GS network. At the moment, we are testing the proposed method to produce contextual metrics on synthetic data sets. The validation process aims at the critical comparison between existent and emerging techniques as the function of their cost and complexity. Our experience so far has pointed out that the methods based on network analysis may be expensive; at the same time, it has clearly shown a further critical step forward toward the differentiation among research popularity, research influence and research impact.

Acknowledgements We would like to thank the survey participants who have provided insight and expertise that greatly assisted this research.

References

1. Acm digital library. <http://dl.acm.org>. Accessed 21 Oct 2015
2. Core—conference portal. <http://portal.core.edu.au/conf-ranks/>. Accessed 01 Feb 2017
3. Gogglescholar. <http://scholar.google.com>. Accessed 21 Oct 2015
4. Ieexplore. <http://ieeexplore.ieee.org>. Accessed 21 Oct 2015
5. Journal Citation Report (JCR). <http://about.jcr.incites.thomsonreuters.com>. Accessed 01 Feb 2017
6. Researchgate. <https://www.researchgate.net>. Accessed 21 Oct 2015
7. Scopus. <https://www.scopus.com>. Accessed 21 Oct 2015
8. Semantic Scholar Project. <https://www.semanticscholar.org/>. Accessed 01 Feb 2017

9. Thomson ISI Web of Science. <http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science.html>. Accessed 21 Oct 2015
10. Aggarwal, C.C., Wang, H., et al.: *Managing and Mining Graph Data*, vol. 40. Springer, Berlin (2010)
11. Antelman, K.: Do open-access articles have a greater research impact? *Coll. Res. Libr.* **65**(5), 372–382 (2004)
12. Bakkalbasi, N., Bauer, K., Glover, J., Wang, L.: Three options for citation tracking: Google scholar, scopus and web of science. *Biomed. Digit. Libr.* **3**(1), 7 (2006)
13. Balduzzi, M., Platzer, C., Holz, T., Kirda, E., Balzarotti, D., Kruegel, C.: Abusing social networks for automated user profiling. In: Jha, S., Sommer, R., Kreibich, C. (eds.) *Recent Advances in Intrusion Detection*, pp. 422–441. Springer, Berlin (2010)
14. Banks, M.: Connections between open access publishing and access to gray literature. *J. Med. Libr. Assoc.* **92**(2), 164 (2004)
15. Bar-Ilan, J.: Which h-index? A comparison of wos, Scopus and Google Scholar. *Scientometrics* **74**(2), 257–271 (2008)
16. Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. *Phys. A Stat. Mech. Appl.* **311**(3), 590–614 (2002)
17. Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., Fraternali, P., Quarteroni, S.: *Web information retrieval. Data-centric systems and applications*, Springer, Berlin (2013)
18. Cook, T.D.: Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. *Ann. Am. Acad. Political Soc. Sci.* **599**(1), 176–198 (2005)
19. Glänzel, W., Moed, H.F.: Journal impact measures in bibliometric research. *Scientometrics* **53**(2), 171–193 (2002)
20. Golafshani, N.: Understanding reliability and validity in qualitative research. *Qual. Rep.* **8**(4), 597–606 (2003)
21. Harzing, A.-W.: *Google Scholar-A New Data Source for Citation Analysis*. University of Melbourne, Parkville (2008)
22. Hirsch, J.E.: An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **102**(46), 16569–16572 (2005)
23. Holbrook, J.B., Barr, K.R., Brown, K.W.: Research impact: we need negative metrics too. *Nature* **497**(7450), 439–439 (2013)
24. Holzinger, A.: Universal access to technology-enhanced learning. *Univ. Access Inf. Soc.* **7**(4), 195–197 (2008)
25. Jacsó, P.: Google scholar: the pros and the cons. *Online Inf. Rev.* **29**(2), 208–214 (2005)
26. Khambatti, M., Ryu, K.D., Dasgupta, P.: Structuring peer-to-peer networks using interest-based communities. In: Aberer, K., Koubarakis, M., Kalogeraki, V. (eds.) *Databases, Information Systems, and Peer-to-Peer Computing*, pp. 48–63. Springer, Berlin (2004)
27. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., Barton, D.: Big data. The management revolution. *Harvard Bus Rev* **90**(10), 61–67 (2012)
28. Meho, L.I., Yang, K.: Impact of data sources on citation counts and rankings of lis faculty: web of science versus scopus and google scholar. *J. Am. Soc. Inf. Sci. Technol.* **58**(13), 2105–2125 (2007)
29. Murugesan, S.: Understanding web 2.0. *IT Prof.* **9**(4), 34–41 (2007)
30. Neuman, W.L.: *Social Research Methods: Quantitative and Qualitative Approaches*, vol. 13. Allyn and Bacon, Boston (2005)
31. O'reilly, T.: What is web 2.0: design patterns and business models for the next generation of software. *Commun. Strateg.* **65**(1), 17 (2007)
32. Pileggi, S.F., Fernandez-Llatas, C., Traver, V.: When the social meets the semantic: social semantic web or web 2.5. *Future Internet* **4**(3), 852–864 (2012)
33. Sanchez, C.A., Wiley, J., Goldman, S.R.: Teaching students to evaluate source reliability during internet research tasks. In: *Proceedings of the 7th International Conference on Learning Sciences*, pp. 662–666. International Society of the Learning Sciences (2006)
34. Serenko, A., Bontis, N.: Meta-review of knowledge management and intellectual capital literature: citation impact and research productivity rankings. *Knowl. Process Manag.* **11**(3), 185–198 (2004)
35. Skyrms, B., Pemantle, R.: A dynamic model of social network formation. In: Gross, T., Sayama, H. (eds.) *Adaptive Networks*, pp. 231–251. Springer, Berlin (2009)
36. Stephanidis, C., Savidis, A.: Universal access in the information society: methods, tools, and interaction technologies. *Univ. Access Inf. Soc.* **1**(1), 40–55 (2001)
37. Ullrich, C., Borau, K., Luo, H., Tan, X., Shen, L., Shen, R.: Why web 2.0 is good for learning and for research: principles and prototypes. In: *Proceedings of the 17th International Conference on World Wide Web*, pp. 705–714. ACM (2008)
38. Xiang, R., Neville, J., Rogati, M.: Modeling relationship strength in online social networks. In: *Proceedings of the 19th International Conference on World wide web*, pp. 981–990. ACM (2010)