Identifying conflicts of interest of reviewers using authors bibliometric networks

V. Duarte-Martínez
Fac. de Ing. en Electricidad y Computación
Escuela Superior Politécnica del Litoral
Guayaquil, Ecuador
vealduar@espol.edu.ec

A.G. López Herrera
Dept. Computer Science and Artificial Intelligence
University of Granada, Granada, Spain
lopez-herrera@decsai.ugr.es

M.J. Cobo
Dept. Computer Science and Engineering
University of Cádiz, Cádiz, Spain
manueljesus.cobo@uca.es

***Abstract*—The process of peer review is a very important task that must be followed by a scientific document to be approved and then published in a scientific journal. Being such a delicate process, it is required that the elected reviewers guarantee the validity and quality of the submitted work, without influencing in its decision any type of link that may exist between the authors of a scientific document and the reviewers. However, many times this goes unnoticed and the reviewers are selected only according to their experience in a given field.**

This contribution proposes an approach to do recommendations based on the design of bibliometric networks such as: co-authorship networks, author co-citation networks and direct link-citation networks. In this way, we seek to discover relationships between reviewers and authors of scientific documents in order to suggest appropriate reviewers, thus avoiding the recurring problem of conflicts of interest.

***Index Terms*—Reviewers selection, bibliographic networks, co-author, author co-citation, direct link-citation**

I. Introduction

Much of the scientific job is based on the search for answers. To accomplish this, it must be performed many researches and experiments and their results are analyzed, discussed and disseminated, generally made through publications. However, in order that scientific documents become published must go through an assessment process, which includes peer review, to ensure the quality of the document and above all the validity of what has been investigated and is exposed in the document.

Scholarly peer review can be viewed as the central part of the publishing workflow, being considered necessary to ensure the quality of research journals. In fact, it must determine whether a manuscript should be accepted or rejected. This method leaves the work open to scrutiny, and often to the annotation or modification, by authors of similar or superior to the author's range. Since peer review is based in the criteria of experts, it needs of a set of experts in a particular research field (i.e. the research field of the manuscript to be reviewed) who are qualified and able to carry out an impartial review.

During this process, the role of referees is advisory. Moreover, in the scientific literature, the referees do not act as a group; not communicate with each other. Generally, they are not aware of the identity or the reports of their colleagues. In general, it is not necessary to achieve consensus. Therefore, the group dynamics are very different to a jury. Sometimes their opinion is not unanimous. In such cases can be applied different options to make a decision.

Usually, it is desired to choose reviewers who are not close to the authors and also have no links with them. These are expected to inform publishers about potential conflicts of interest for evaluation. Some publishers or publications ask authors a list of potential arbitrators, as well as people they deem inappropriate to arbitrate their work. This is necessary, especially, when the subject of a work is so specialized that publishers cannot locate specialists in the field by themselves.

Therefore, the selection of impartial reviewers is a difficult and daunting task, since reviewers could have some conflicts of interest with the authors of the manuscript. For example, a reviewer could be a recent co-author, could be cited by the manuscript or could be usually cited by the author, which would mean, in a way that the reviewer might be interested in the publication of this manuscript. This and other issues could bias the reviewer criteria.

Usually, the selection of reviewers is based on the field of expertise of the reviewers, and the inherent and hidden social network of them is not taken into account in the process. So, in this contribution a novel approach based on bibliometric networks [1], to uncover hidden relationships between reviewers and authors, is presented. In particular, we focus on bibliometric networks based on the social aspects, i.e., authors. To do that, co-author, author co-citation and author direct link citation networks are used to build a global social bibliometric network.

II. Bibliographic networks

Networks provide an interesting abstraction of a variety of complex systems [2], such as, social networks or biological networks. Moreover, scientific knowledge could be understood as a complex system, where the network structure is frequently used to model the interaction of scientific actors (authors, journals, keywords, references, etc.).

On the other hand, bibliometric is dedicated to the analysis of the scientific output. Formally, Bibliometrics are a set of methods, which can be used to analyze academic literature quantitatively and its changes over time [3]. It is an important tool for assessing and analyzing the academic research output contributing to the progress of science in many different ways.

There are two main bibliometric methods for exploring a research field: performance analysis and science mapping [4]. While performance analysis aims to evaluate the citation impact of the scientific production of different scientific actors [5], science mapping aims at displaying the conceptual, social or intellectual structure of scientific research and its evolution and dynamical aspects [6] [7].

As above mentioned, the scientific knowledge can be represented as a graph. In this sense, science mapping analysis uses bibliographic networks [1] in order to represent the different relationships among scientific actors. Depending on the kind of aspects that will be represented, three kind of networks could be identified:

- Collaboration networks are used to show how authors or institutions relate to others in the field of scientific research. The most common kind of collaboration network are co-author networks. With this type of network can be discovered, for example, groups of regular authors, influence authors, hidden communities of authors, relevant institutions in a specific research field, etc. [8]
- Conceptual networks [9] represent relations between concepts or words in a set of publications. That means that, for instance, words which appear together in a document, will be related in a network. It is also known as co-word network. This type could be used to understand the topics covered by a research field [10], to define which are the most important and the most recent issues. It could also help in the study of the evolution of subjects over time and it could give good impression of cognitive relations between different research groups.
- Publication Citation networks [11] [12] show relationships between nodes which represent publications, while the edges can have different interpretations depending on the network type (co-citation, bibliographic coupling or direct link).

In this contribution we focus on the social bibliographic network which have become a very important topic of study that has attracted the attention of researchers because these kinds of networks represent a good prototype of complex evolving system to explore, where the network is constantly expanding with the occurrence of new authors of scientific documents, or new links between existing authors.

These networks are studied with two main aims. Mainly, to evaluate the collaboration status of scientific disciplines, that is related with bibliometric mapping, assisting in discovering of structural and dynamic aspects of scientific research, to better understand the organization of scientific fields and their evolution [6]. And also, to examine the performance of social structure of collaboration networks, using techniques of social networks analysis [13].

In Bibliometrics, social bibliographic networks are analyzed using indicators to measure and extract the inherent structure to a set of publications; but also these are studied as social networks to explore interactions between actors who may be researchers, groups, institutions, etc. that collaborate in the development of scientific papers on research topics. These are usually weighted to represent the strength of different relations between actors.

Social bibliographic networks might reveal substantial knowledge such as: researcher communities, professional interactions between scientists, central nodes that act as hubs, leaders or gatekeepers; highly connected groups; and patterns of interactions between groups [14]. Very large networks can be assembled in this way.

III. Methodology

As above mentioned, different bibliographic networks have been proposed based on co-occurrence, coupling and direct link. Moreover, depending of the selected unit of analysis, different aspects could be uncovered. The approach presented in this contribution focuses on the social aspects.

Particularly, in a research document the unit of analysis author can be found in two different parts. The former is the authors list, being the authors who write the research document. The latter is the author-references list, which are the authors cited by the research document (the intellectual social base). Using these two types of author lists, three kinds of bibliometric networks can be built: co-author, author co-citation and author direct link citation.

- Co-author networks, are those containing all authors of a document and related to each other, i.e. two scientists are considered connected if they have coauthored one or more papers together. Such networks reflect groups of researchers in scientific fields.
- Author co-citation networks, are those which relate all authors referenced in a document to each other. This kind of networks shows author groups which are referenced together in the same papers.
- Author direct citation networks, are those that express the direct relationship between the author of a document and the cited author, that is, it shows the authors cited by other authors. This information shows a very important social factor that may allow discover which are the inspirations of a creator author, and the bases that allows the author to develop his/her research.

These networks show us the connections that may arise between two researchers, according to the nature of their link, and represent the main study object in this document, for example, let suppose a set of three research documents where each document is composed of a set of authors, and a set of authors in the references list as it is shown in Table I. Thus, the three social bibliometric networks built from this example corpus are shown in Figure 1.

TABLE I
Example corpus.

Document	Authors	Authors in reference
Document 1	a1, a2, a3	a2, a4
Document 2	a2, a3, a4	a1, a3, a5
Document 3	a1, a3, a5	a2, a4, a6

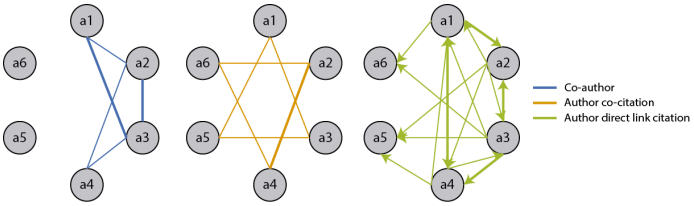

Fig. 1. Example of bibliometric network.

We should point out that an arc between two nodes represents a bibliometric relation between these authors. Also, the thickness of the arcs represents the raw co-occurrence or number of direct links depending on the kind of relationships.

Once the three kinds of bibliometric networks have been built, three levels of social relationships could be assessed:

- Level 1: authors who have only one kind of social relationship. At this level we just focus in one kind of relationship, and the remaining ones are not taken into account.
- Level 2: authors who have two kinds of social relationships. In this level, two authors could 1) collaborate together in a document and be co-cited, or 2) collaborate together in a document and have a direct link citation relationship between them, or 3) be co-cited and have a direct link citation between them.
- Level 3: authors who have the three kinds of social relationships. In this case, two authors collaborate together in (at least) a paper, are co-cited and also one of them cites to the other.

The above mentioned levels of relationships can be used to achieve the grade of conflict of interests between researchers. Therefore, from a set of bibliographic records downloaded from a bibliographic database (e.g. WoS or Scopus), a whole network containing the three types of relationship can be built.

Thus, from a set of documents related with a research field, a set of reviewers candidates could be achieved based on their bibliographic social relationships in order to avoid possible biases. To do that, the journal office should look at the social bibliometric relationships between the authors of the submitted manuscript and the reviewer community / panel in order to avoid biased reviewers. Furthermore, based on the three kinds of levels of social relationships, we propose an algorithm to identify different levels of bias prevention that could be configured. It is described as follows:

Having a graph G with a set of nodes V and edges between nodes V_i, $V_j \in G$. It is assumed that V_i is a document author while V_j is a reviewer candidate. Therefore, we could asses levels using the following inclusion criterion: **For each** edge(V_i, V_j) where $V_i \neq V_j$, level of bias prevention will be **Level 1 if** V_i co-cited V_j **or** V_i co-authored V_j **or** V_i direct link cited V_j. Otherwise, level of bias prevention will be **Level 2 if** (V_i !co-cited V_j **and** V_i !co-authored V_j) **or** (V_i !co-authored V_j **and** V_i !direct link cited V_j) **or** (V_i !co-cited V_j **and** V_i !direct link cited V_j). In any another way, level of bias prevention will be **Level 3 if** V_i !co-cited V_j **and** V_i !co-authored V_j **and** V_i !direct link cited V_j.

The Level 3 is the most restrictive one. That means that the authors of the submitted manuscript and the reviewer candidate never appear together (neither in the author list, neither in the author-reference list) in any paper, in spite of, they investigate in the same research area. The Level 3 suggests no bias or conflicts of interest among authors and reviewers candidates, those last are supposed impartial. In Figure 2, it is shown a Venn diagram to explain the intersection of these three levels where each type of relationship is represented as a set and the levels are: The level 1, the less least restrictive, that covers every relationship; the Level 2, which allows the existence of two types of relationships at the same time; and the Level 3, which does not allow any kind of relationship.

Fig. 2. Different levels of bias prevention.

Thus, the level of rigor to be established depends on the editor's criteria; however, here is proposed this methodology that aims to deal with the problem of conflict of interest in scientific collaboration.

IV. Case Study

With the aim of showing the applicability of this approach, a case of study have been built. It is described with details as follows.

Initially, we have collected data from Scopus database. For this, we have chosen 3891 research papers about Cloud Computing thematic only because this is a hot topic in Computer Sciences field.

Then, we have had used preprocessing techniques to decompose bibliographic data in order to extract specific data about principal authors and referenced authors. Afterwards, we have been able to build different bibliometric networks mentioned previously, that is, co-author, author co-citation and direct link citation networks. In these networks, we can explore

through their relationships and we can realize, as is logical, that an author related to another directly would not be a good candidate to be selected as a reviewer of a scientific document of him, but through these relationships you can navigate to find other researchers who have worked in the same subject but not related to the main author. The more levels of relationship are excluded, the more reliable it can be assured that there will be no conflicts of interest, obviously this can be assured from the point of view that encompasses scientific work.

In this analysis, the objective has been to find relationships of the type: *Friend of my friend who is not my friend*, taken to the field of research, Understanding that, if two researchers A and B have worked together, as collaborators of a scientific document, then from A and from B you can find possible reviewers candidates that are not related to A and/or B but with other researchers who have worked with authors A and B. In this way it is sought to avoid choosing reviewers of scientific articles that could have conflicts of interest with the authors of an article sent to a specific journal. The journal office, after receive the original manuscript, could use this analysis in order to select and recommend the best reviewers within the research area of the article, along with avoiding of conflict of interest, that is, authors which have no direct relationship with the authors of the submitted document (not cited author, or not collaborator).

To demonstrate this approach, we study each type of network that is generated from an author. For example, we have chosen, at random, an author of scientific documents dealing with the subject of this case study (Cloud Computing), the researcher Zheng Q. (Zheng, Qinghua). From this, co-author network, co-citation network and direct link citation network have been generated.

In Figure 3, we can see the co-author network built from one author (Zheng Q.) where the principal researcher is in purple node. In addition, the authors with whom he has shared co-authorship in scientific documents (Li J., Zhang J., Li X. and Li R.) are represented with orange nodes. From each author that have worked with Zheng Q. we have gotten all the recommended authors which are represented with green nodes. Recommended researchers are those who have worked with others who have co-authored scientific documents with Zheng Q., but have not worked directly with him. This would be the minimum level of connection that should be considered, although we could explore more in networks until we can find even more indirect relationships. In Table II, the list of recommended authors can be seen. We have pruned the networks built to show only 15 recommended authors thinking of simplifying the visualization of the results in this example.

We have also generated the co-citation network for the author, this means that the network has been built with all the relationships defined by the authors referenced in the scientific documents written by Zheng Q. This network can be seen in Figure 4, where node in purple color represents the principal author, orange nodes represents authors referenced in papers written by Zheng Q. and green nodes represents recommended authors. Summarized data can be seen in Table III.

TABLE II
TABLE OF RECOMMENDATIONS FROM CO-AUTHOR NETWORKS.

Recommended Author
Cao Y.
Chen L.-X.
He T.-Q.
Lee S.-Y.
Li B.
Li C.
Liao S.
Su D.
Su G.
Wong D.W.S.
Xu P.
Xue W.
Yang Y.
Yu N.
Zhou F.-F.

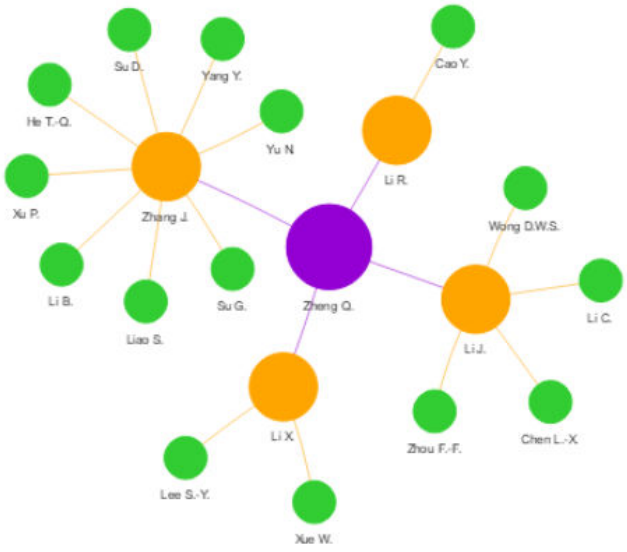

Fig. 3. Recommendations from co-author graph.

And finally, we have built an author direct link citation network. It models the relationships between the author and all cited authors in reference section from a scientific document. The Figure 5, shows the built graph where principal author, Zheng Q. is shown in purple node, referenced authors are shown in orange nodes and recommended authors are shown in green color. The list of recommended author is shown in the Table IV.

With all these recommended authors, we can identify who are those that fit in each level of bias prevention, according to what is established in section III. However, since the complete networks are very extensive (especially the co-citation networks that could contain thousands of relationships) we have summarized the results obtained. In Table V, it is shown some scientific authors that belong to level 1, level 2 and level 3 respectively. This means that, for example. the author Cheng

Fig. 4. Recommendations from author co-citation graph.

Fig. 5. Recommendations from direct link citation graph.

TABLE III
TABLE OF RECOMMENDATIONS BASED ON AUTHOR CO-CITATION NETWORKS.

Recommended Author
Dunagan J.
Ghaffarkhah A.
Hara T.
Letier E.
Ma X.
Megiddo N.
Mignot J.-C.
Nagin K.
Ota K.
Rai A.
Talbi E.-G.
Tam A.S.W.
Tan K.
Wakelam V.
Zhu M.

TABLE IV
TABLE OF RECOMMENDATIONS BASED ON DIRECT LINK CITATION NETWORKS.

Recommended Author
Campbell R.H.
Cardellini V.
Chao K.-M.
Chen C.
Jararweh Y.
Paraiso F.
Pernici B.
Rong H.
Safieddine I.
Sharma A.K.
Sivic J.
ValiKardan S.
Williams P. Yan Z.
Yang K.

C. who belongs to level 3, has not collaborated with Zheng Q. as co-author, nor has he quoted by him, nor does there exist direct link citation between them. Therefore, Cheng C. would be a good reviewer candidate because no relationship has been identified that could incur a problem of conflicts of interest when reviewing the scientific document written by Zheng Q. Depending on the level of restriction that you wish to apply, the authors that are pigeonholed in levels 1 and 2, can also be considered.

V. Conclusions

Per review plays a central role in the research life, since scientific publications must pass the judge of anonymous colleagues in order to determine the quality of the manuscript. Also, per review is not only restricted to publications in journals, conferences or books. It is commonly used to determine research project proposals, grants and scholarship positions.

TABLE V
TABLE OF BIAS PREVENTION LEVELS.

Level 1	Level 2	Level 3
Chao K.-M.	Campbell R.H.	Chen C.
Chen L.-X.	Cao Y.	Lee S.-Y.
Dunagan J.	Cardellini V.	Su G.
Ghaffarkhah A.	Hara T.	Xue W.
He T.-Q.	Jararweh Y.	Yan Z.
Letier E.	Li C.	
Ma X.	Liao S.	
Megiddo N.	Pernici B.	
Mignot J.-C.	Rong H.	
Nagin K.	Su D.	
Ota K.	Wong D.W.S.	
Paraiso F.	Yu N.	
Rai A.	Zhou F.-F.	
Safieddine I.		
Sharma A.K.		
Sivic J.		
Talbi E.-G.		
Tam A.S.W.		
Tan K.		
ValiKardan S.		
Wakelam V.		
Williams P. Xu P.		
Yang K.		
Yang Y.		
Zhu M.		

Reviewers must act following an ethic rules and conflict of interests must be avoided. The criteria of the researchers could be biased by several aspects: i) collaboration (e.g. co-authorship), ii) social relationship (e.g. belonging to the same department), and iii) intellectual (e.g. reviewer could be cited several time). In fact, the biases could be positive or negative. That is, a reviewer could favor or work against the author. Some of these effects could be avoided using a blind review, where the reviewers do not know the identity of the authors.

In this sense, the selection of reviewer without conflict of interests with the author could be a difficult task, since a reviewer should not have interacted socially with the author.

As above mentioned, researchers have many kinds of relationships with their colleagues in their academic life. These interactions could be modeled as a bibliographic network, where the nodes are the researchers and the edges are the interactions among them.

In this contribution, a novel approach to identify potential reviewers without conflict of interest based on social bibliographic networks is presented. Particularly, three kinds of social network are used: co-author, author co-citation and author direct link citation.

Using this variety of scientific social networks, three levels of social relationships can be determined. Therefore, three levels of bias prevention could be used, being the level 3, the most restrictive, where a researcher could be a reviewer if he/she never appear as co-author, neither in the author reference list.

For this, the transitivity property of networks have been studied in order to find possible candidates for reviewers using an undirected relationship between authors that we have called: *Friend of my friend who is not my friend*. This is an application of the transitivity property that indicates the fact that a node A knows a node$B$ and a node B knows a node C does not guarantee that A knows C. Therefore, with these candidates we can apply level 3 of bias prevention in order to find real good possible reviewers.

The presented approach could be used by journal officers or panel experts in order to select the most suitable reviewers to evaluate different kind of research outputs.

As future work, we propose to include in the preprocessing process a disambiguation task in order to avoid the problem of finding two or more nodes that actually represent the same person. For instance, we have found {DÍAZ BRUGERA, JAVIER} and {DÍAZ BRUGUERA, JAVIER} which obviously refers to the same person but his/her name is misspelled.

ACKNOWLEDGMENT

This study was funded by FEDER funds in the Spanish Department for Economy and Competitiveness project (TIN2016-75850-R).

REFERENCES

[1] V. Batagelj and M. Cerinšek, "On bibliographic networks," *Scientometrics*, vol. 96, no. 3, pp. 845–864, 2013.
[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: Structure and dynamics," pp. 175–308, 2006.
[3] M. Cobo, A. López-Herrera, E. Herrera-Viedma, and F. Herrera, "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field," *Journal of Informetrics*, vol. 5, no. 1, pp. 146–166, 2011.
[4] E. Noyons, H. Moed, and M. Luwel, "Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study," *Journal of the American Society for Information Science*, vol. 50, no. 2, pp. 115–131, 1999.
[5] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16 569–16 572, 11 2005.
[6] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 7, pp. 1382–1402, 2011.
[7] M. J. Cobo, A. G. Lõpez-Herrera, E. Herrera-Viedma, and F. Herrera, "SciMAT: A new science mapping analysis software tool," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 8, pp. 1609–1630, 2012.
[8] W. Glanzel, "National characteristics in international scientific co-authorship relations," *Scientometrics*, vol. 51, no. 1, pp. 69–115, 2001.
[9] M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Social Science Information*, vol. 22, no. 2, pp. 191–235, 3 1983.
[10] M. Cobo, M. Martínez, M. Gutiérrez-Salcedo, H. Fujita, and E. Herrera-Viedma, "25years at Knowledge-Based Systems: A bibliometric analysis," *Knowledge-Based Systems*, vol. 80, pp. 3–13, 5 2015.
[11] M. M. Kessler, "Bibliographic coupling between scientific papers," *American Documentation*, vol. 14, no. 1, pp. 10–25, 1963.
[12] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *Journal of the American Society for Information Science*, vol. 24, no. 4, pp. 265–269, 7 1973.
[13] C. C. Aggarwal, Ed., *Social Network Data Analytics*. Boston, MA: Springer US, 2011.
[14] S. Wasserman and K. Faust, *Social network analysis : methods and applications*. Cambridge University Press, 1994.