Negative Implications of a Power-Law Distribution: A Study on Networks of Scientific Reviewers

Song Qin, Marius C. Silaghi, Ronaldo Menezes and William Cheung

Abstract Traditional peer-reviewing is a process whereby submissions by various scientists are selected based on certain criteria passed on to reviewers by organizers of conferences or editors of journals. This process has been used to maintain the quality of the works being presented and also to help grouping reports relevant to a given community (or topic). However, certain scientific opinions and theories compete and have partisans. Common examples of such competitions appear when deciding the most important metric in classification algorithms, what to use as a basis for recommendation algorithms, the best predicting models for a known phenomena, to name a few. The common assumption is that the community will be equally informed about the arguments of all involved studies, in order to come out with objective conclusions. This assumption is reasonable when partisans of each competing opinion can eventually review and recommend for publication the studies that agree with their perspective. In its turn, this can be expected to eventually happen whenever expert reviewers are randomly assigned to corresponding papers. However in recent years we have seen that power-law distributions instead of randomness are present in many social relationships. In this study we investigate what happens in the world of peer-reviewing, more specifically in a network of reviewing relations for an open review journal. We found that a power-law distribution is indeed present, as a small group of reviewers evaluates a significant fraction of all sub-

Song Qin
HDSS Lab, Department of Computer Sciences, Florida Institute of Technology, Melbourne, Florida, USA e-mail: qsong2008@my.fit.edu

Marius C. Silaghi
HDSS Lab, Department of Computer Sciences, Florida Institute of Technology, Melbourne, Florida, USA e-mail: msilaghi@cs.fit.edu

William Cheung
Hong Kong Baptist University, Hong Kong e-mail: william@comp.hkbu.edu.hk

Ronaldo Menezes
BioComplex Laboratory, Department of Computer Sciences, Florida Institute of Technology, Melbourne, Florida, USA e-mail: rmenezes@cs.fit.edu

missions. The problem however is that this is undesirable since these “hubs” have an unmatched influence on what gets published. This experiment presents a first case where arguably the power-law structure of the social network can be considered as an overall negative factor. It also supports an argument for employing the social graph of reviewers as an additional metric of the quality of a journal/conference.

Key words: Social Networks, Peer-Reviewing Process, Power-Law Distributions

1 Introduction

Peer reviewing is an essential mechanism of the modern research process. Traditionally, journals have used the peer-review process as a filter to decide which submissions should be selected for publication, potentially based on criteria such as relevance to the main topic of the journal and technical quality. In the past it was difficult to study the properties of reviewing processes due to requirements of anonymity of reviewers. It is generally believed that this process is fair and it is assumed that it avoids bias given the fact that most journals and conferences employ at least 3 individuals to evaluate each publication.

Recently however, several journals and workshops have adopted new models of peer reviewing, where the names of the reviewers are made public [8, 17, 13, 14]. Among them the journal of Biology Direct, which has used open peer-review for several years, yields a large amount of data for verifying assumptions. This allows us to analyze the process and to try to understand peer-reviewing a little better.

While the general public is familiar with “reviewing” for items sold on eBay and Amazon, that type of reviews does not have as purpose the forbidding of the sale of poor quality items. Instead, the poor quality items are still left for sale but they are attached with the relevant information for warning potential buyers.

Traditional peer-reviewing for scientific articles is a very different concept. Reviews of scientific articles are not commonly published with these articles and are not commonly used for warning potential readers of the failures and qualities of the article. Rather, these reviews are meant for helping a chair of conference or journal editors to decide which submissions should be filtered out. The general public (scientists who read these papers) are not aware of any concerns or discussion that took place during the review process.

1.1 Problem

While reviewing aims to be impartial, it is difficult for introspection alone to remove certain biases stemming from the school of thought where the reviewer was herself educated as a scientist. There exist competing schools of thought in economics, science, engineering, and many other fields [15, 10, 7]. A factor that is arguably desirable for the objectivity of a journal, is how well it gives equal chances to researchers from different currents to state their arguments. Based on the assumption of potential bias, it is desirable to have a diverse range of reviewers being able to review and participate in filtering, balancing the chances of the competing schools of thought. While not all schools of thought will review each given submission, each of them should at least get a fair chance of reviewing relevant submissions.

There are four major types of peer-review mechanisms used in reviewing of scientific articles, as well as in (medical) experiments on human subjects. Common *simple blind review* mechanisms are those where reviewers know the name of the authors but authors do not know the names of their reviewers. The motivation is to avoid that reviewers would fear retribution for their reviews. *Double-blind reviews* are those made in such a way that reviewers do not know the name of the paper authors and paper authors do not know the name of their reviewers; they stem from concern of bias for reviewers (e.g. when reviewing an influential person). *Open peer reviews* are mechanisms where reviewers know the names of the authors and authors know the names of their reviewers, while reviews and reviewer names are published with the articles. The motivation is to encourage reviewers to write responsible reviews. *Blinded review* is another mechanism where reviewers do not have access to the names of authors but the names of reviewers are published together with their reviews. The idea is to encourage reviewers to be responsible, by accountability, while helping them to avoid bias (as in the case of the double blind review).

In the aforementioned processes there exist relations between authors and reviewers (even if these relations are sometimes hidden from both parts). According to theories in social network analysis and complex networks, the structure of these relations should tell us something about the process itself. This study looks at the characteristics of the social network of reviewers and its relation to the fairness of the review process.

The online availability of the reviewing information from the Biology Direct Journal gives us a window into the scientific review process. It allows for verifying whether submissions have fair chances to be reviewed by a varied number of researchers, potentially covering multiple schools of thought. What we noticed is that a couple of reviewers eventually reviewed a large fraction of the accepted articles. This took place despite the editor's effort of finding different people to review each work. The issue however is that each submission could have different reviewers that appear to be chosen at random but at the global level the picture is quite different and what emerges when we combine the seemly individual random choices for each paper is a network in which few reviewers are hubs and review quite a lot of papers, a typical social network with hubs and long-tail distributions [1]. Note that this would be less prominent in a conference if we analyze only one edition. Normally conferences enforce a maximum number of reviews per person. For a conference one would have to look at the global picture across many editions (years) of the event.

From the perspective of review processes, we claim that the power-law distribution is detrimental to the process itself because it points to an unbalanced influence for any potential bias that these hub reviewers may have. While a power-law distribution for networks was so far considered as a positive characteristic signaling the stability of the social network, we have identified here an argument for considering the power-law distribution in peer-reviewers networks to be a negative trait. Having journals and conferences publish the networks of reviewers (even if just the structure without actual names) can help developing a new metric for their objectivity.

We discuss the related work in Section 2. Section 3 describes the dataset and how it was modeled as a network. We then look at these networks formed from author-reviewer relations and discuss our findings in Section 4. Finally we conclude with suggestions of how these findings can be used to improve peer-review processes.

2 Related Work

2.1 Social Network Analysis

A social network is a structure that represents social interactions and personal relationships. Examples of social networks include: friendship [6], collaboration [16, 5] and email networks [3]. In general a social network can be abstracted as a structure in which the entities are people and the links between these people are extracted from some social relationship. In this paper we address the social network of reviewers. This network is obtained by projecting the bipartite network of reviewers and papers shown in Figure 1 unto the set of reviewers. Two reviewers are connected if they have reviewed the same paper. The strength of their link is given by the number of submission that they have reviewed together.

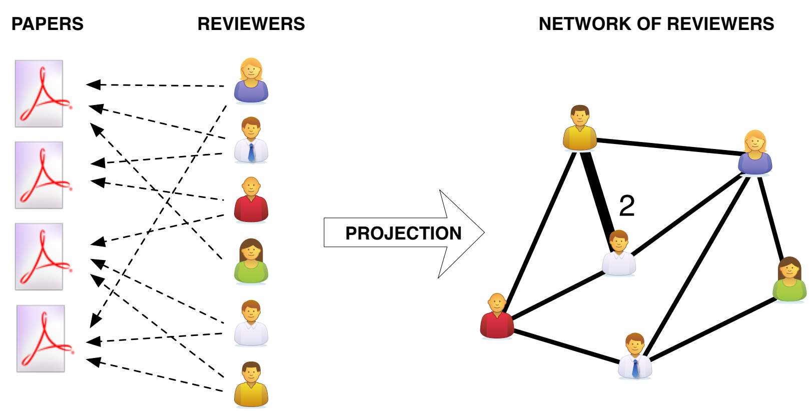

Fig. 1 From a bipartite network in which reviewers are linked to the papers they review, we can project a network of reviewers where reviewers are linked directly if they reviewed the same paper.

When looking at networks one can use several metrics to understand the represented phenomena. The metrics we use are the degree distribution, betweenness centrality, closeness centrality and clustering coefficient, explained below:

Degree Distribution: This distribution expresses the probability, $p(k)$, that a node in the network will have k connections. It has been observed that in many real networks [12] their degree distribution roughly follows a power law as given by Equation 1,

$$p(k) = ck^{-\lambda}, \tag{1}$$

where, c and λ are constants. For most of the real networks $2 \leq \lambda \leq 3$. Nodes that have more ties to other nodes may be in advantageous positions.

Betweenness Centrality: The betweenness is a measure of the centrality of a vertex in a network. Betweenness is calculated as the fraction of the shortest paths between vertex pairs, that pass through the vertex of interest. For a network $N = (V, E)$ with n vertices, the betweenness $C_B(v)$ for a vertex v is:

$$C_B(v) = \sum_{s \neq t \neq v \in V} \frac{\alpha_{st}(v)}{\alpha_{st}} , \tag{2}$$

where α_{st} is the number of shortest paths from s to t, and $\alpha_{st}(v)$ is the number of shortest paths from s to t that pass through vertex v. Vertices exhibiting a high level of betweenness are in a position to control information flow in the network.

Closeness Centrality: Centrality is defined as the inverse of the distance from a vertex to all other vertices in the network

$$C_c(i) = \frac{1}{\sum_k d(i,k)} , \tag{3}$$

where $d(i, k)$ is the shortest path between vertices i and k. This metric gives low values for the central nodes and high values for the less central ones. Nodes with high closeness are generally in a position to influence other nodes because they can reach them very quickly.

Clustering Coefficient: This coefficient is a measure of the ratio in which nodes in a graph tend to cluster together. The clustering coefficient, C_i, of node i is given by Equation 4, where, m_i is the number of links between the k_i neighbors of i; the clustering coefficient of the entire network is just the average of all C_i over the number of nodes in the network n. Clustering is relevant to social networks. It can be used to identify small-world networks [19] which are expected to have high clustering and short average path lengths.

$$C_i = \frac{2m_i}{k_i(k_i - 1)}. \tag{4}$$

In this paper we study the aforementioned metrics in relation to a reviewers' network from a peer-review process. However our main discussion in this paper focuses on the drawbacks of the degree distribution to the peer-review process itself.

2.2 Open Peer-Review Process

One of the most common reviewing processes is the *double-blind review.* Under this scheme one publishes only accepted articles and the names of the organizing committee members (names which are supposed to witness to the quality of the reviewing process). Reviewers are not expected to know the names of the authors at the time of the review, and authors will never find out the names of the reviewers of their submission. Reviews are not published and are not digitally signed, and authors have no way to prove that they have received any given review, or even that they have submitted any given article. The camera-ready article eventually published in the proceedings can be completely different from the corresponding article actually evaluated by reviewers (and even the title can be completely changed). As example of venue using double blind review we mention the *International Joint Conference on Artificial Intelligence (IJCAI).*

Multiple models of open peer-review processes have been proposed and experimented with. The openness varies in terms of *what is revealed*, in terms of *the degree of the revelations*, and in terms of *the articles to which the revelation applies* (e.g., only to accepted articles or to all submitted articles). A classification from the perspective of the object of the revelation contains the following dimensions:

Open/Closed Author Names: This dimension tells whether reviewers are informed about the names of the authors at review time. This information is supposed to help reviewers correctly assess the originality of the submission. An example of open author names reviewing is employed by the *Conference on Principles and Practice of Constraint Programming (CP).*

Open/Closed Article: Venues may either publish articles, or keep them as part of a closed meeting. Certain workshops do not produce public proceedings and the articles accepted and presented in their forum are not considered published.

Open/Closed Submission: Some venues do publish the submission actually evaluated by reviewers, while others only use the submission as an acceptance criteria for publishing a different article. The actually published (aka *camera-ready*) article is typically assumed to be an improvement of the submitted article based on feedback from reviewers. However, only few journals have mechanisms to ensure that the actually published article has any relation whatsoever with the actually reviewed submission. In practice submitters can change even the title and the list of author names of a submission. An example of venue that publishes both the original submission and the camera ready version is the *Workshop on Decentralized Coordination*.

A limited degree of openness of submissions is offered by some cryptology conferences that give authors digital signature certificates for their submissions, helping them to prove that they have submitted the corresponding articles.

Open/Closed Reviews Summary: At various conferences, a senior committee member is charged with writing a short summary of the reviews, to be privately communicated to the author and to the editor making the acceptance decision. Under schemes with open reviews summary, an anonymous committee member is charged with writing a public summary of the anonymous reviews for an article (Example: *IEEE Conference on Peer to Peer Computing (P2P)*).

Open/Closed Reviews: With open reviews, the actual reviews received by the article are made publicly available (not necessarily publishing the names of the individuals). This procedure is more common in online journals such as *Philica.com*.

Open/Closed Reviewing Activity: This revelation dimension quantifies whether one publishes the number of articles reviewed by each reviewer (helping to quantify their impact). Most conferences do publish an average of the number of articles reviewed by its reviewers, but note that an average is not significant if the distribution of reviews per reviewer follows a power-law.

Open/Closed Article Reviewers: The names of the reviewers of each article may or may not be published in association with that article. Certain journals publish articles labeled with names of researchers recommending them.

Open/Closed Review Authorship: The dimension of review authorship is used to specify whether the name of each reviewer is published in association with the corresponding review. With open review authorship the reviewers assume the responsibility of their reviews, and they also get credit for improvements suggested in these reviews. Sample venues presenting this feature are the *2013 Workshop on Decentralized Coordination* and the online journal *Biology Direct*.

From the perspective of the degree of revelation, each of the aforementioned items of information can be revealed to any subset of the following groups:

- conference chairs
- reviewers
- authors
- conference audience
- general public

For example, the *2007 Workshop on Material Thinking Design* revealed reviewer names to authors but did not publish them.

Here we focus on the effects of using an *open reviewing activity*. These effects appear also in the case of a stronger revelation that implies the opening of the reviewing activity, such as using *open article reviewers* or *open review authorship*.

There are multiple (more or less logic) qualitative arguments for and against each of these degrees of openness, and there are at least two series of conferences dedicated to the practice of Peer Reviewing. Nevertheless it is a remarkable scientific challenge to design quantitative metrics that can be used for founding a systematic study of this area [13]. We believe that the use of the social graph can move us one step closer to a mechanism in which the general audience can have a better idea of the quality of the peer-reviewing process.

3 Network of Reviewers

Biology Direct[1] is an open peer-reviewed journal where publications and their reviews (also the reviewers' name) are publicly accessible through unique URLs. We downloaded information for about 314 papers (titles) and the corresponding reviewers up to March 2013. The dataset contained 843 reviewers. Once the projection was done as explained earlier, the network contained 843 nodes and 2,512 relations.

We first performed an analysis of the degree distribution of the network and found that two nodes dominated the reviewing process with 192 and 125 reviews each. This is respectively 3 and 2 times as much as the third reviewer who reviewed 65 submissions. The obtained degree distribution is shown in Figure 2.

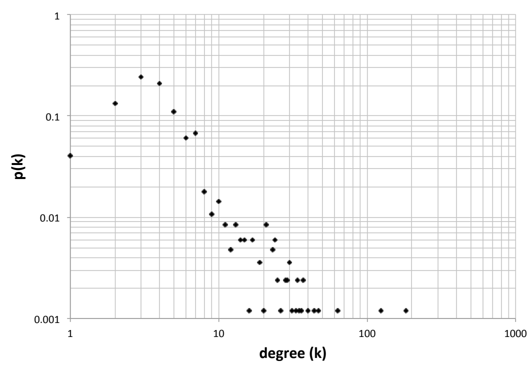

Fig. 2 The degree distribution of reviewers shows a scale-free network in which a couple of nodes are hubs.

Next we performed a community analysis to understand whether the hubs are part of the same group of people. Fortunately, in this case they belong to different communities which we believe indicate that they belong to perhaps two different groups of individuals interested in reviewing papers. The community detection algorithm that we used here was proposed by Blondel et al. [4], and it identifies 14 communities. In Figure 3 we can observe the network of reviewers where the size of the node represents the degree and the colors highlight the different communities. Diversity in the community can be used to positively assess the quality of the review process.

As mentioned before, the network of reviewers is built by using nodes to represent the reviewers and arcs to illustrate whether they are connected by reviewing together the same paper. The tool we used for visualization is Gephi [2]. Note that the size of a node is proportional to its degree. Nodes are colored based on the 14 different communities they are in. The employed community detection algorithm is proposed in [4, 9].

[1] http://www.biology-direct.com/

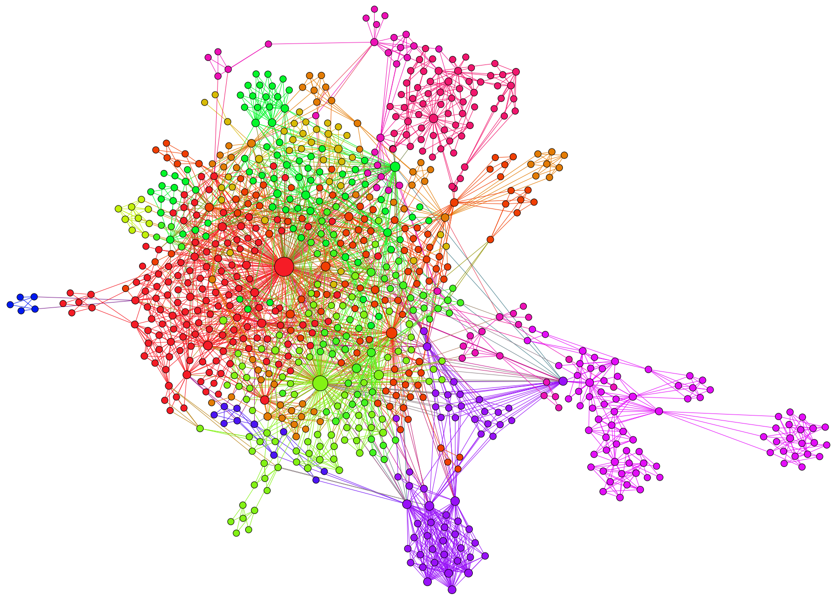

Fig. 3 Network of reviewers highlighting the communities they belong to. 14 communities can be observed in this network. The existence of many communities is a positive indicator of the review quality. Communities can be seen as groups of reviewers interested in similar subjects.

In this network we have a high correlation between the degree of a node and its betweenness and closeness so we decided not to show the visualizations for these centralities due to lack of space in the article. The high correlation is due to the fact that the people doing a lot of reviews end up linking different groups and being close to most of the other reviewers. What these high centralities mean is that highly-connected reviewers have an unhealthy chance to exert a strong influence on other reviewers and consequently on the review process. Not only that their reviewing leads to a collection of accepted articles that fits the scientific view of these hubs, but the other reviewers are likely to be indirectly influenced by the reviews of the hubs. In a reviewing process it is common to have reviewers modify their review after they see the reviews of others or participate in discussions. A hub acts as an authority in the process because she has participated in several other discussions.

Last we looked at the clustering coefficient of the nodes in the network and at the average clustering coefficient. Figure 4 shows a network in which the color of the node represents its clustering coefficient; the stronger the color the more clustered the node is.

Note that Figure 4 indicates the existence of a few highly connected groups. These groups are not good for the peer-review process and can indicate the existence of some “mob” phenomena in which a group of individuals may work together to achieve a certain goal—in the case here the goal could be to influence the accep-

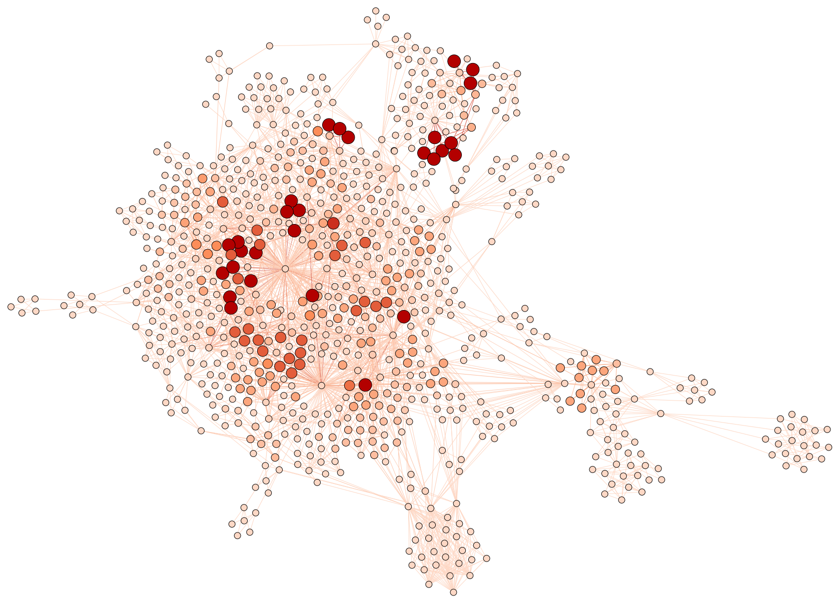

Fig. 4 Network of reviewers by clustering coefficient. Darker colors represent higher clustering coefficient.

tance/rejection of a paper. Fortunately however this is not generalized throughout the network.

4 Discussion

While each given submission to a venue may be reviewed by a varied set of researchers, the aggregated distribution of reviewers may be less uniform. The aggregated distribution of reviewers can end up with a few reviewers having disproportionate influence. The danger of this phenomenon is increased if one cannot always guarantee the relation between expertise and influence.

The existence of communities is more complicated to understand. Although the existence of communities can be positive for showing a variety of individuals with different interests (as we said before), one has to also be careful because the communities may also mean that we have individuals who review papers together but rarely mix with other groups. When one uses the social graph to assess the quality of a conference, the community structure should be carefully analyzed to help one understand its benefits and drawbacks. Moreover on the drawback side, communities show that the peer-review process is not being vetted by reviewers with diverse backgrounds and interests. We believe a good reviewing network will not have a good resolution on the division of communities, having at most a small number of communities defined (the exact number really depends on the size of the network and on the field of research).

One can also look deeper into the structure of the neighbors of each researcher. Ideally each submission would be reviewed by reviewers that are as disconnected as possible, reducing social contagion (while still being somewhat connected due to their expertise), such as to obtain diverse points of view [18].

A parallel can be drawn between peer-reviewing and genetics. Multiple sets of genes come together to generate a more robust set of genes where the union overshadows individual damaging mutations from each independent set. Similarly, the opinion of multiples reviewers come together to select and influence an article.

With current genome donors one has raised the issue of dangers coming from the disproportionate usage of certain sources [11]. In particular, offspring of the same donors will inherit mutations that can dangerously surface in subsequent generations.

Similarly, there are dangers from disproportionate usage of certain reviewers. Reviewers (as any human) can have preconceived ideas and can subjectively favor certain metrics or scientific views. The relevance of peer-reviewing as a pillar of the modern science is also due to the theory that it can mitigate such subjectivity by joining diverse opinions. Disproportionate involvement of certain reviewers can endanger this property, as their subjectivity will disproportionately impact on the venue and thereby on subsequent generations of researchers.

5 Conclusions

We raise the issue that the power-law distribution generally seen as a positive factor for the stability of social networks can also have negative undesirable connotations. We raise this issue in the context of peer-review social networks, where the fact that certain reviewers are found to be involved in a disproportionate number of articles, conflicts with the objectivity expected from scientific reviewing. Human individuals are intrinsically subject to preconceived ideas, errors and subjective reasoning. Diversity of reviewers is therefore the basis that makes from peer-reviewing the pillar of modern science, where the mixture of multiple views can lead to sounder aggregated result (just as the combination of genes can help defend against damaging mutations in each individual contribution). Just as the defects of a disproportionately frequent donor of genes risk to appear more frequently in subsequent generations, the subjectivity of one disproportionately involved reviewer can impact negatively on generations of researchers.

The availability of data for our study was made possible by the recent trend of openness in journal and conference peer-reviewing. In particular we made an extensive usage of the information on reviews and reviewers made available by the Biology Direct online journal. In the proceedings of this journal we find that a couple of reviewers were involved in reviewing a significant fraction of the articles being published in the venue. We conclude that such openness (at least *openness* *of reviewer activity*) can be recommended to journals and conferences that want to convince the public about the soundness of their reviewing procedures. Further openness concerning the structure of reviewer neighborhoods (as offered by *openness of article reviewers*) can also be recommended as a way to detect risks of social contagion and detection of undesirable segregation into communities.

While significantly more research is required for establishing a sound scientific foundation to the peer-reviewing procedures found at the foundations of modern science, this study brings a small but clear and objective contribution.

References

1. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
2. M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. In *Intl. AAAI Conference on Weblogs and Social Media*, 2009.
3. C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories*, pages 137–143. ACM, 2006.
4. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
5. P. Divakarmurthy, P. Biswas, and R. Menezes. A temporal analysis of geographical distances in computer science collaborations. In *Privacy, security, risk and trust (PASSAT), 2011 ieee 3rd international conference on social computing (SocialCom)*, pages 657–660. IEEE, 2011.
6. N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.
7. D. Hongwen. A new school of thought in sequence stratigraphic studies in us: High-resolution sequence stratigraphy [j]. *Oil & Gas Geology*, 2, 1995.
8. E. Koonin, L. Landweber, and D. Lipman. Biology direct. http://biologydirect.com/, 2013.
9. R. Lambiotte, J.-C. Delvenne, and M. Barahona. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*, 2008.
10. B. Moingeon and B. Ramanantsoa. Understanding corporate identity: the french school of thought. *European Journal of Marketing*, 31(5/6):383–395, 1997.
11. J. MROZ. One sperm donor, 150 offspring. nytimes.com/2011/09/06/health/06donor.html, 2011.
12. M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
13. M. J. Peterson, M. C. Silaghi, and M. Yokoo. Game theoretical modeling and studies of peer-reviewing methods. In *Intl. Symposium on Peer Reviewing*, volume I, pages 267–272, 2009.
14. M. Silaghi, S. Qin, and W. Cheung. Open peer-review experiment in the decentralized coordination workshop. *IEEE Intelligent Informatics Bulletin*, 13(1), 2013.
15. H. W. Spiegel. *The growth of economic thought*. Duke Univ Press, 1971.
16. B. A. . V. T. Evolution of the social network of scientific collaborations. *Physica A*, 311:590–614, 2002.
17. C. Tonkinwise. Material thinking design reaearch workshop: An experiment in open peer review process at connected: International conference on design education. materialthinking.org, 2007.
18. J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *PNAS*, 109(16), 2012.
19. D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.