Available online at www.sciencedirect.com

ScienceDirect

Health Professions Education 5 (2019) 218–224

www.elsevier.com/locate/hpe

Social Networks as an Approach to Systematic review

Jimmie Leppink[a,*], Patricia Pérez-Fuster[b]

[a]*School of Health Professions Education, Maastricht University, the Netherlands*

[b]*Faculty of Psychology, University of Valencia, Spain*

Received 14 April 2018; received in revised form 28 August 2018; accepted 2 September 2018

Available online 5 September 2018

Abstract

Whether we are in the process of designing a new empirical study or our interest lies in conducting a review study, a solid literature review is needed to acquire an accurate idea of the current state of affairs with regard to a phenomenon of interest. Even if we can find contributions to the literature by entering keywords in search engines, we need tools that can help us to structure all the contributions encountered in terms of their interrelations and impact. This article presents social network analysis as such a tool. Although social network analysis is commonly thought of as a method in a particular empirical study, where individuals and groups of participants are studied, we can view writing and citation behavior in a field as an empirical study as well. In that context, participants can be individual authors and author teams as well as publications. Social network analysis can provide indicators that can help to qualify and quantify impact of contributions to a field across time.

Keywords: Social network analysis; Systematic review; Citation; Coauthoring; Impact

1. Introduction

Whether we are in the process of designing a new empirical study or our interest lies in conducting a review study, a solid literature review is needed to acquire an accurate idea of the current state of affairs with regard to a phenomenon of interest. Unfortunately, partial reviews that are biased towards researchers' hypotheses are not uncommon[1] yet these and other flawed practices can have grave effects on the quality of our research.[2] Moreover, even if there are no interests in selective reporting for time constraints or to support particular hypotheses and we can find all relatively meaningful contributions to the literature by entering the right keywords in search engines, we need tools that can help us to structure all the contributions encountered in terms of their interrelations and impact. This article presents social network analysis[3–5] as such a tool. Although social network analysis is commonly thought of as a method in a particular empirical study, where individuals and groups of participants are studied, we can view writing and citation behavior in a field as an empirical study as well. In that context, participants can be individual authors and author teams as well as publications. Social network analysis can provide indicators that can help to qualify and quantify impact of contributions to a field across time.

Succinctly put, social network analysis is the study of *social structures*: networks of relations between individuals and/or groups such as organizations. To model

*Correspondence to: School of Health Professions Education, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands.

E-mail address: jimmie.leppink@maastrichtuniversity.nl (J. Leppink).

Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region

https://doi.org/10.1016/j.hpe.2018.09.002

2452-3011/

these relations, *graph theory* is used: mathematical structures or *graphs* to model pairwise relations such as between two individuals or between an individual and an organization. Depending on the context, these relations may be *symmetric* (i.e., A and B are related: communication may occur in both directions) or *asymmetric* (i.e., there is communication from A to B or from B to A but not both). In an author team, all relations between individual authors are supposed to be *symmetric*: ideas expressed on paper ought to result from constructive dialogue and consensus not from one individual dictating others what to write or what to agree on. However, in the study of citations of published work we will find many more *asymmetric* (e.g., a paper published in 2015 citing work published earlier that year but *not* vice versa) than symmetric relations (i.e., two papers citing each other). Social network analysis can deal with both symmetric and asymmetric relations; they may consist of only symmetric relations, only asymmetric relations, or some combination thereof.

In the context of publications and citations, making connections between actors such as authors and author teams in a field visible can enrich our perspectives on social phenomena in the field (e.g., peer review, citations, and conference invitations).[6] For example, in a large-scale social network analytic study involving 16,653 papers by a total of 24,258 different authors, Hautz and colleagues[6] found that coauthorship connected authors into 68,663 unique pairs of which 61,937 had coauthored only one article, and 67.43% of all authors were linked to each other through a coauthor of a coauthor. The most productive and most connected authors in the field were easily identified as key scholars in the field of medical education, including: Cees van der Vleuten, Geoffrey Norman, Kevin Eva, Albert Scherpbier, Lambert Schuwirth, Henk Schmidt, Glenn Regehr, and Lorelei Lingard (see Fig. 1 in 6 for the full coauthorship network of the twenty-seven most productive authors at the time). Studies such as the one by Hautz and colleagues provide very detailed insights into coauthorship, citations and other social phenomena that occur in a field. Hautz et al.[6] concluded (p. 1274) that the "field of medical education represents what social network analysts term 'a small world network'." Although more stars have risen in the field of medical education since the publication of this study about two years ago, the work by Hautz et al. helps scholars in and newcomers to the field to understand how different medical education scholars are connected, who may be working on common topics based on shared ontological and epistemological views and built on the same educational theories.

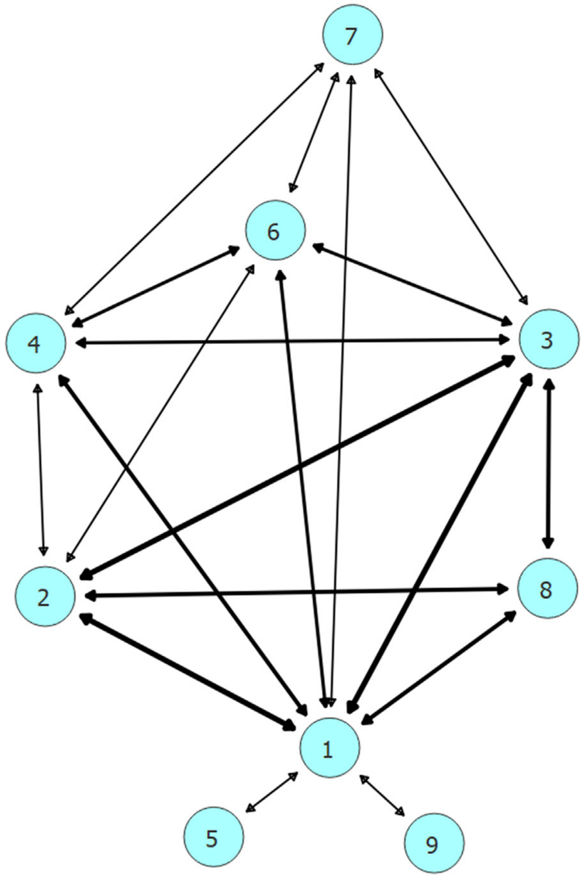

Fig. 1. Social network of the nine authors.

When our interest lies in *how* individual actors (e.g., researchers) in a network are connected or *which* individual actors tend to form sub-networks aka *cliques* (e.g., research groups or author teams), one individual can be a member of several cliques, but cliques can be distinguished based on unique collaborations (e.g., 68,663 unique pairs of coauthors of peer-reviewed publications in 6). Graphical visualizations and numerical indicators can help us qualify and quantify collaboration and impact. Although a full discussion and presentation of all possible uses of social network analysis would require a series of articles or eventually a book, through a hypothetical example, we explain in this paper how social network analysis can help researchers to study amongst others which researchers tend to form cliques and how they cite each other. Although the reader might wonder why take a hypothetical example and not a social network analysis on an actual topic, the latter may easily result in numbers of papers and author teams that make it more difficult for readers to follow the explanations throughout the example. Moreover, the discussion of concepts related to the actual topic would in that case also distract from the concepts of social network analysis. Using a hypothetical example with a limited number of publications and author teams facilitates the explanation and understanding of the concepts of social network analysis. Since the concepts and indicators discussed in this paper equally apply to larger networks (e.g.,6), this paper can serve as a worked example for social network analyses on actual topics in or relevant to (health professions) education.

2. Method

We have nine researchers who altogether (co)authored eleven publications on the same specific *Topic X* in a given field (e.g., health professions education) or context (e.g., postgraduate education). We first explore which authors are involved in which publications and which publications are cited in which other publications. Based on that information, we can compose a so-called *adjacency matrix* for authors and for publications, respectively. An adjacency matrix is a k rows by k columns square matrix, where k is the number of authors or the number of publications. Given nine researchers or authors ($k = 9$), the adjacency matrix for authors has nine rows and nine columns. Given eleven publications ($k = 11$), the adjacency matrix for publications has eleven rows and eleven columns. For k, Hautz et al.[6] use the term *vertex*: "A connected entity within a network" (p. 1275). In the case of k authors, a vertex is an author in the network; in the case of k publications, a vertex is a publication in the network.

At the *author level*, the adjacency matrix provides information with regard to which authors have appeared together on one or more publications. Therefore, the numbers on the diagonal are zero: no single author is listed twice as an author on the same publication. In its simplest form, the off-diagonal numbers of the adjacency matrix for authors are 0s and 1s: they indicate which two authors have been listed as authors on the same publication *at least once* ('1') and which two authors have thus far never been listed as authors on the same publication ('0'). Every '1' means a direct connection between two authors (i.e., an "edge" in terminology of Hautz et al.[6]), whereas every '0' means absence of such a direct connection. Any two authors who have no direct connection may be connected indirectly, through a chain of edges and vertices aka *path* (e.g., coauthor of a coauthor). When the question is not *if* two authors have (co)authored at least one publication together but *how many* publications they have (co)authored together, the off-diagonal numbers represent the number of publications each pair of authors was involved in. For example, if authors A and B have published six publications together, the two off-diagonal cells for this pair of authors – cell 1: row A and column B; cell 2: row B and column A – have the value '6'. The reason that *both* cells for this pair of authors have the same value is that we assume that ideas expressed on paper ought to result from constructive dialogue and consensus and hence the adjacency matrix for authors is symmetric. The numbers in the adjacency matrix provide the input for hierarchical cluster analysis[7] to identify cliques (i.e., subnetworks, here: author teams) and for statistics such as the *standardized information centrality*[8] as a measure of the proportion (i.e., 0–1) of information flow associated with authors (i.e., positions of individual authors in the collection of author teams), in this case based on the number of publications (co)authored together.

Given the typically asymmetric (i.e., one-way) nature of citations, at the *publication level*, the adjacency matrix is typically asymmetric. The numbers on the diagonal are zero (no publication cites itself), whereas the off-diagonal numbers are zeros and ones depending on whether or not a publication in question has been cited in another publication. The numbers in the adjacency matrix provide the input for the so-called *co-citation matrix*: a matrix that provides us with information about citations of individual publications (on the diagonal) and of pairs of publications (off the diagonal) in other publications. That is, the numbers on the diagonal of this matrix indicate the number of citations each publication has received from the other publications in this list (i.e., the number of publications selected that refer to a given publication). Where the diagonal focuses on *individual* publications, the off-diagonal numbers represent the number of times each *pair* of publications has been referred to in the same other publications in the list. Thus, the on-diagonal and off-diagonal numbers shed light on two aspects of potential topical connectedness of the different publications selected. The numbers on the diagonal are also called the *degree prestige*[9] and, when dividing them by the number of possible citations, form a *standardized degree prestige*[9] as a measure of proportion of citations of a given publication by other publications in a given set. For instance, in a list of eleven publications, any publication can be cited in at most ten other publications; if all ten other publications cite that publication, the degree prestige equals 10 and the standardized degree prestige equals 10/10 or 1 (100%).

All analyses were done in *SocNetV* version 2.4,[7] which is a freely downloadable (i.e., zero cost), Open Source software program that can provide graphical visualizations of social networks along with statistics

Table 1
Publications, (order of) authors, citations and type of work in the context of Topic X.

Pub.No.	Auth. IDs& order	Cited inPubl. No.	Totalcites	Type ofwork
1	1, 2, 3	2-11	10	Theoretical foundation
2	1, 4	5, 7, 9, 11	4	Empirical study 1
3	3, 2	6, 7, 9, 11	4	Empirical study 2
4	5, 1	5-11	7	Special Issue paper on theory development
5	1, 2, 4, 3, 6	6, 8, 10	3	Empirical study 3
6	3, 1, 4, 6, 7	7-11	5	Empirical study 4
7	8, 2, 3, 1	8, 9, 11	3	Empirical study 5 (Project X, PhD candidate A)
8	9, 1	11	1	Empirical study 6 (Project X, PhD candidate B)
9	8, 2, 3, 1	11	1	Empirical study 7 (Project X, PhD candidate A)
10	2, 1	–	0	Side kick: Application in new area
11	8, 2, 3, 1	–	0	Empirical study 8 (Project X, PhD candidate A)

Table 2
Adjacency matrix of the nine authors.

Author	1	2	3	4	5	6	7	8	9
1	0								
2	6	0							
3	6	6	0						
4	3	1	2	0					
5	1	0	0	0	0				
6	2	1	2	2	0	0			
7	1	0	1	1	0	1	0		
8	3	3	3	0	0	0	0	0	
9	1	0	0	0	0	0	0	0	0

such as measures of social cohesion and prominence (power).

3. Results

Table 1 outlines which of the researchers (co) authored which of the publications along with some other information.

The first (left) column in Table 1 represents publication 'number' or the order in which the publications appeared online. Some publications may have appeared online in the same year (e.g., publications 1–3 in 2015, publications 4–7 in 2016, publications 8–9 in 2017, and publications 10–11 in 2018), but 1–11 represents the chronological order in which they appeared online. The second column lists authors (IDs) 1–9 in the order in which their names appeared on the respective publication. The third and fourth column indicate in which of the other works the given publication was cited and the implied number of citations, respectively. Finally, the fifth (right) column could in a review study result from a preliminary analysis of the type of work presented in each of the publications. For the sake of the example, suppose that authors 1–3 are the founders of Topic X, authors 4–7 are 'occasional' researchers in that at some point one or more opportunities emerged for them to do or join a study in the context of Topic X, and authors 8–9 are two PhD candidates who are part of Project X and focus on related subtopics within Topic X.

3.1. Authors

Table 2 presents the adjacency matrix of the nine authors (i.e., the presumably symmetric relations between authors 1–9 of publications 1–11 on Topic X) and Fig. 1 provides a graphical representation of the numbers in that adjacency matrix.

In line with Table 1, author 1 is connected with all other authors because s/he (co)authored at least one publication with each of the other authors. Also note that the connections between author 1 and authors 2–3 are stronger (i.e., the arrows are thicker) than the connections between author 1 and other authors, because author 1 published much more with authors 2–3 than with any other authors. We can also conclude that authors 5 and 9 are the least connected in this community of authors; each of them was involved in only one publication that had only one other author (i.e., author 1).

Subjecting the numbers from the adjacency matrix to hierarchical cluster analysis, we find five cliques: (1) *authors 1 and 5*: responsible for publication 4; (2) *authors 1 and 9*: responsible for publication 8; (3) *authors 1–3 and 8*: as team responsible for publications 7, 9 and 11, and authors 1–3 are in turns engaged in publications 1, 3 and 10 two of which are cited in work with author 8; (4) *authors 1, 3, 4, 6 and 7*: publication 6; and (5) *authors 1–4 and 6*: as team responsible for publication 5, and authors 1 and 4 are also in publication 2 which is cited in publication 5.

For this network of authors, standardized information centrality ranges from 0.070 (7.0%) for each of authors 5 and 9 (i.e., both have only 1 publication, which in

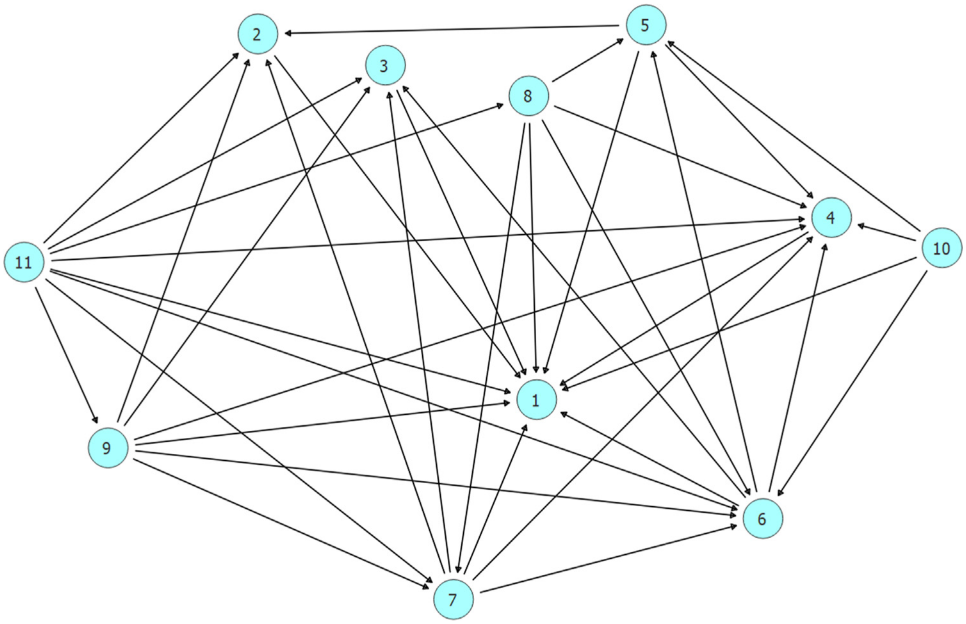

Fig. 2. Social network of the eleven publications.

Table 3
Co-citation matrix of the eleven publications.

Pub.	1	2	3	4	5	6	7	8	9	10	11
1	10										
2	4	4									
3	4	3	4								
4	7	4	4	7							
5	3	0	1	3	3						
6	5	3	3	5	2	5					
7	3	2	2	3	1	3	3				
8	1	1	1	1	0	1	1	1			
9	1	1	1	1	0	1	1	1	1		
10	0	0	0	0	0	0	0	0	0	0	
11	0	0	0	0	0	0	0	0	0	0	0

both cases is coauthored by author 1) to 0.146 (14.6%) for author 1.

3.2. Publications

Fig. 2 provides a graphical representation of the (adjacency matrix of the) links between publications 1–11 obtained through citations.

In this example, there are no publications that mutually refer to one another and hence all links between publications are asymmetric. For instance, even though publications 6 and 7 both appeared online in 2016, publication 6 is cited in publication 7 but publication 7 is not cited in publication 6. This explains why all connections between any two publications are one-way (e.g., the arrow going from publication 4 to publication 1 indicates that publication 1 was cited in publication 4). Note further that all arrows in Fig. 2 are equal in size since they indicate whether or not instead of how many times publication A was cited in publication B.

Table 3 presents the so-called *co-citation matrix* of the eleven publications.

The numbers on the diagonal of this matrix indicate the number of citations each publication has received from the other publications in this list, whereas the off-diagonal numbers represent the number of times each pair of publications has been referred to in other publications in the list. For instance, publication 1 received 10 citations (out of 10 possible citations, hence a degree prestige of 10 and a standardized degree prestige of 10/10 or 1) and was cited 7 times in a publication where also publication 4 was cited. Publication 6, although more recent than publications 1–5, already received 5 citations. Although the standardized degree prestige based on all publications equals 0.5 (i.e., 5/10 or 50%), as illustrated in Fig. 2 all publications appearing after publication 6 refer to publication 6 (i.e., a standardized degree prestige of 1 or 100% based on that subset). Hence, especially publications 1 (standardized degree prestige of 1) and 4 (standardized degree prestige of 0.7) may have provided the theoretical foundation – or part of it – for many of the later publications including publication 6, and (some of) the empirical findings and/or implications of these findings reported in publication 6 (standardized degree prestige of 0.5) apparently informed publications 7–11 in one way or another. Finally, since publications 10 and 11 are not cited anywhere in this set of publications, all on- and off-diagonal numbers for these two publications are zero based on this analysis. If at some point publication 12 appears online and cites publication 10 or 11, some of the zeros for that publication cited will disappear.

4. Discussion

This article demonstrates that coauthoring and citation information can be summarized in an author-level and publication-level adjacency matrix, which can serve as input for social network analysis. Assuming symmetry of author interrelations, cliques and author-level standardized information centrality provide insights into collaborative subgroups and influence of individuals in a given network of collaborative subgroups. Citation behavior across publications is typically asymmetric, and the co-citation matrix and the standardized degree prestige provide useful indicators of citation behavior.

4.1. What social network analysis adds to established citation indices

Through the aforementioned matrices, graphs, and statistics, social network analysis provides a number of insights that cannot really be derived from common metrics such as the Hirsch index aka *h-index*[10] or i10-index which indicate how many articles *h* have been cited at least *h* times and how many articles have been cited at least 10 times, respectively. Although these established indices commonly used by for instance Google Scholar have useful features and may even be of help in social network analysis, they are limited to *overall* numbers of citations and do not indicate which citations are made on which topic, in which context, and by which author teams. In the context of the latter, the *h*-index and the i10-index do not distinguish between citations by others and self-citations. In the example discussed in this article, author 1 has many collaborations with many other authors. Had the collaborations of author 1 been limited to just a few other authors and that group would mainly cite their own work, that clique or subgroup feature would be revealed by social network analysis. In other words, social network analysis may identify ‘islands’ in a particular topic where the h-index and i10-index just see citations.

Besides islands in a topic, social network analysis can also help to understand to what extent work from outside a field is cited and used to inform research and practice in medical education. For instance, at the International Association for Medical Education Europe (AMEE) 2018 Meeting in Basel, Switzerland, during the Symposium entitled “*Intersections, Introspections and Divergences: Sustaining the Growth of Medical Education Research and Training*” (i.e., Symposium 4B), Mathieu Albert delivered a presentation entitled “*Talking to Ourselves*” about a study that reveals that about 69% of citations in medical education articles come from medical and health professions education journals. Statistics like this 69% and percentages of where the other citations come from provide important input for social network analysis and help us understand which fields have inspired medical education and from which other fields medical education scholars might find new inspiration.

4.2. Social network analysis based on content

Although the matrices, graphs, and statistics discussed in this paper can help us to shed light on aspects of citation behavior that cannot really be studied through common metrics such as the aforementioned *h*-index or i10-index, social network analysis cannot replace content analysis. To acquire a deeper understanding of topical interrelations in a series of selected publications on a given topic, such as in a systematic review study, social network analysis ought to be used in combination with methods such as *co-word analysis*.[11] Co-word analysis uses frequencies of words across publications to reveal the major themes in a given area of interest. Just like authors can be linked in terms of publications, social network analysis based on co-word analysis may provide insight into how different publications are linked content-wise. Although various factors may influence authors' decisions with regard to what work to cite and what not, some kind of topical link should be present for work to be cited and, as such, work that is considered of key importance by a community ought to be cited more frequently (i.e., have more links in the network) than work that is given a somewhat lower importance or is considered relevant mainly in a particular context or facet of a topic. When integrated in the methodology for a review study, social network analysis may reveal insights in commonly perceived links between content covered in different publications on the same topic or may help to position a particular publication in the topic. For instance, in the context of Topic X, publication 10 focuses on the application of the theory covered in publications 1 and 4 in an area where this theory has not yet been applied (at least not in published work) but builds on empirical findings reported in publications 5 and 6 that have implications for the research questions and/or study setup in the new area.

4.3. Questions inform the exact use of social network analysis

For the simplicity of the example, the numbers of authors and publications subjected to social network analysis in this article are small and probably much smaller than when we are going to use this method in an actual study. Social network analysis is a method that many researchers in (health professions) education are not familiar with, and a (for researchers) new method is more easily explained with a fairly simple small-number example than with a large review study on a given topic. However, the matrices, graphs, and statistics presented in this article can also be used when the numbers of authors and publications are much larger (i.e., in the hundreds or even more). As such, this paper can serve as a worked example for social network analyses on actual topics in or relevant to (health professions) education.

Social network analysis can help us to study complex social networks and their dynamics, and this strength comes with a challenge as well: the larger our numbers of authors and publications, the more complex things become. The broader our questions, the more keywords we may need to include in our search, and the larger the numbers of authors and publications we may need to include. Therefore, it is important to carefully define our questions and keywords before we proceed with social network analysis. The numbers of authors and publications under study result from the questions that drive our systematic review. We will rarely if ever do a systematic review on an entire field such as the whole medical education, statistics education or language education. Our interest typically lies in a particular phenomenon such as managing cognitive load in emergency medicine, visual expertise in radiology, fostering clinical reasoning through physical examination with simulated patients, enhancing education through assistive technologies, and the like. Besides, in not so few cases, there are even time constraints (e.g., only considering work published between [one year] and [another year]). Such interests generally narrow down our search to such an extent that we end up dealing with 50–200 authors rather than 1,000 authors and that we have a set of 100–300 rather than 1,000 or more publications. Presenting a full adjacency matrix for such numbers may become somewhat inefficient. However, the graphical visualizations and statistics discussed in this article will still be useful (e.g., Fig. 1 in 6), the statistics will still provide the same information that cannot be captured by common metrics such as the *h*-index or i10-index, and it can still be combined with content methods such as co-word analysis to acquire a deeper understanding of topical interrelations in a series of selected publications on a given topic.

4.4. Social network analysis at different levels

From the example discussed in this article, it becomes clear that social network analysis can help researchers to study a topic of interest through different lenses and with units from different levels as subjects of study. Although authors and publications constitute the subjects or levels of study in this article, other entities such as *research groups* or situations such as *research programs*, *research grant recipients* or *conference sessions* also constitute possible subjects or levels of study. Individuals and groups of individuals are study subjects we are all familiar with, but publishing – just like (any other form of) learning – is a social phenomenon and that invites for studying *situations* as well. Social network analysis enables researchers to have it all. The methods and the software are out there and ready to be used.

References

1. Picho K, Artino AR. 7 deadly sins in educational research *J Grad Med Educ* 2016;8:483–487.
2. Maggio LA, Sewell JL, Artino AR. The literature review: a foundation for high-quality medical educational research. *J Grad Med Educ* 2016;8:297–303.
3. Scott J. Social network analysis. *Sociology* 1988;22:109–127.
4. Scott J. *Social Network Analysis*, 4th ed., London: Sage; 2017.
5. Wasserman S, Faust K. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press; 1994.
6. Hautz WE, Krummrey G, Exadaktylos A, Hautz SC. Six degrees of separation: the small world of medical education. *Med Educ* 2016;50:1274–1279.
7. Bridges CC. Hierarchical cluster analysis. *Psychol Rep* 1966;17: 851–854.
8. Stephenson K, Zelen M. Rethinking centrality: methods and examples. *Soc Net* 1989;11:1–37.
9. Kalamaras DV. Social network visualizer (version 2.4). [socnetv.org] Accessed 14 August 2018; 2018.
10. Hirsch JE. An index to quantify an individual's scientific research output. *Arxiv* 2005. 〈[https://arxiv.org/PS_cache/physics/pdf/0508/0508025v5.pdf]〉. Accessed 14 August 2018.
11. Kostoff RN. Co-word analysis. In: Bozeman B, Melkers J, editors. *Evaluating R&D Impacts: Methods and Practice*. Boston: Springer; 1993. pp. 63–78.