Six degrees of separation: the small world of medical education

Wolf E Hautz,[1] Gert Krummrey,[1] Aristomenis Exadaktylos,[1] & Stefanie C Hautz[2]

[1] Department of Emergency Medicine, Inselspital University Hospital, Bern, Switzerland
[2] Institute of Medical Education, Department of Assessment and Evaluation, Medical Faculty, University of Bern, Bern, Switzerland

Correspondence
Wolf E Hautz, MD, MME
Department of Emergency Medicine
Inselspital University Hospital
Freiburgstrasse 16c
3010 Bern, Switzerland
Tel: ++41 31 632 21 11
E-mail: wolf.hautz@insel.ch

Abstract

Context

Conventional wisdom has it that everyone on earth is on average only six steps away from knowing any other person through 'a friend of a friend'. On a local level, however, many people experience that most of their acquaintances know each other. It is thus hard to imagine how such a highly clustered group could be so well connected to the rest of the world. In this paper, we investigate how co-authorship connects scholars in medical education and whether the six degrees of separation hypothesis also applies to the network of authors in the field.

Methods

We constructed a mathematical graph from publication data obtained on the top three journals in the field and analysed it using social network analysis methods. We found Lorelei Lingard to be one centre of the network of co-authors and determined the numbers of authors who were one, two or more steps away from her. We further created a website that makes it possible to identify the shortest path from any author in the field to any other, including links to the connecting papers.

Results

The analysis covered 16 653 papers by a total of 24 258 different authors. Co-authorship connected authors into 68 663 unique pairs, of which 61 937 had co-authored only one article. 67.43% of all authors were linked to each other through a 'co-author of a co-author'. The average shortest path between any two authors in this network was 5.98 (min 1, max 17); the average distance to Lorelei Lingard was 4.17 (min 1, max 10).

Conclusion

The field of medical education represents what social network analysts term 'a small world network'. Making the connections between its actors visible may provide a new perspective on social phenomena that occur in this world, including peer review, citation and conference invitations.

Introduction

Getting to know the field of medical education (ME) is a challenge for many newly arriving researchers. Fortunately, existing papers provide insights into the field's predominant themes, institutions and journals,[1,2] and numerous books introduce commonly used methods[3] and frequently encountered theories.[4] Getting to know the people behind these theories and finding out who sets the themes is much harder but could benefit authors new to the field.[5]

According to social network theory,[6] the spread of ideas and the propagation of information critically depends on the people who provide the greatest connectivity within a group. These key figures are not necessarily those who write the most articles but those with the most connections to other scholars in the field. We thus applied social network analysis (SNA) methods[7] to the field of medical education to identify the best connected researchers in the field.

In a highly popular application of SNA, Kevin Bacon is the centre of the acting universe: Actors who have been in a film with Kevin Bacon have a Bacon Index of 1. Those who have not, but who have been in the same film as someone with a Bacon Index of 1, have a Bacon Index of 2. And so on. The average Bacon Index of all linkable 1 911 760 actors currently listed in the internet movie database (IMDb) is around 3; the maximum finite index is 10.[8] Around 60% of all actors in the IMDb can be linked to Kevin Bacon. This large connected network within the group of all actors is termed a *giant component*[6] (see Table 1 for an explanation of technical terms in SNA).

More generally, the six degrees of separation concept suggests that everyone on earth is on average just six handshakes away from every other person.[6] On the local level, however, we find that most of our friends not only know us but also know each other. Our acquaintances tend to cluster in groups, where everybody is connected to almost everybody else. The daily experience of such local clusters leads most people to feel much less connected than the six degrees hypothesis suggests.

When we meet someone new outside our daily routine and unexpectedly discover that we are connected through a mutual acquaintance, we often agree that it's a *small world*. Networks in which such phenomena occur are therefore called *small world networks*[6] and have three key characteristics. First, small world networks are usually giant components, as they typically connect most people within the group. What distinguishes them from other giant components is that, second, the clustering coefficient (the likelihood that two persons, A and B, who both have links AC and BC to a person C, will have a common link AB between themselves) is high. Third, every member of a small world network is linked to every other member via a path that is very short relative to the overall number of people in the network.[6] A small world network structure has profound effects on the spreading of information, ideas or any other entity that is transmitted between vertices.[9]

In this paper, we examine whether the world of ME is also small in this sense. Are ME scholars connected within a giant component containing most persons in the network? What is the average distance between any two scholars in the network? Does this network have a centre? And do authors cluster in small workgroups or are patterns of co-authorship more promiscuous?

Methods

To determine the characteristics of the ME world, we constructed a mathematical graph from publicly available publication data. We first determined the top three ME journals from their impact factors in the ISI Web of Knowledge[10] in July 2015. We then retrieved bibliometric data for each article published in these journals through Pubmed. A custom Perl script[11] converted these data into graphml format.[12] The strength of a connection between two authors was determined by the number of papers they had co-authored. If authors with the same last name and first initial had more than four papers in the dataset, we manually checked whether they were indeed the same person.

SNA[7] was conducted on the resulting graphml file using gephi 0.8.2.[13] We determined the largest connected component within the data and derived the degree, clustering coefficient, centrality and eccentricity for each author in that component. We furthermore computed the average length of the shortest path between any two authors in the component and the diameter of the component (see Table 1).

A number of approaches appeared plausible for determining the centre of the ME world. One would be to identify the author through which the most paths between any other two authors pass (equivalent to the highest betweenness centrality); others would be to choose the author with the shortest average path to anyone else, the lowest eccentricity or the highest number of unique co-authors. The centre of a network may change depending on the measure chosen,[6] suggesting that there is no one unique centre. However, identifying one rather central author helps to illustrate the connectivity within a network, makes it easily comparable to other known networks and may ease understanding of the network's features for people less familiar with SNA. We thus aimed to identify one author who all the above measures placed near the centre, and who was preferably female (as women outnumber men in medicine and ME[14]) with a unique and memorable name (for illustrative purposes).

Results

At the time of the analysis, the three top journals in ME—*Med Educ* (Impact Factor IF = 3.196; including its predecessor *Br J Med Educ*), *Acad Med* (IF = 3.06) and *Adv Health Sci Educ* (IF = 2.124)—had published a total of 16 653 articles by a total of 24 258 different authors. The average number of different co-authors per author was 2.62 (min 0, max 325 for Cees PM van der Vleuten). Co-authorship connected authors into 68 663 unique pairs, of which the majority (61 937) had co-authored only one article, while the ten most productive pairs had each co-authored at least 20 articles (usually together with further co-authors). While some of the highly productive authors tended to repeatedly write in the same pairs, others successfully collaborated with a larger number of scholars (see Figure 1).

The authors were connected into 2209 different components. 67.43% of all authors were connected within one giant component; the second largest component contained only 0.18% of all authors. When the analysis was restricted to the 6904 authors who had contributed to more than one article, 84.92% of all authors were connected within one giant component (and the second largest component contained 0.33% of all authors). This suggests that the many smaller components in the overall dataset mainly contained authors who appeared on just one article.

The number of articles written by an author correlated strongly with the number of his or her unique co-authors ($r = 0.82$, $p < 0.001$). The likelihood that any two co-authors of a given author had written an article together was around 82%. However, this clustering coefficient dropped sharply with the number of articles an author had written (see Figure 2). The correlation between the log/log-transformed clustering coefficient of any author and the number of his or her articles was in fact $r = –0.89$ ($p < 0.001$). Thus, the more productive an author in ME is, the less likely it is that his or her co-authors will have written an article together (even together with the productive author).

However, all of the highly productive authors (and thus their co-authors) were part of the giant component on which the following network-level analysis was based. Each author within this component had, on average, 6.97 co-authors. No author in the giant component was more than nine steps away from Mark Albanese, Stephen Durning, Richard Hays, Eric Holmboe, Maxine Papadakis, Olle Ten Cate or Cees van der Vleuten (the authors with the lowest eccentricity). The average shortest path between any two of the 16 358 authors in the giant component was 5.98 (min 1, max 17). The authors best reachable by a notably shorter path were John Norcini (average shortest path = 3.79), Stephen Durning (3.83) and Glen Regehr (3.84). The best connected female by this measure was Lorelei Lingard (4.17), who was also fairly central in terms of degree (92 unique co-authors) and eccentricity: No author who can be linked to Lorelei Lingard was more than 10 handshakes away from her. Table 2 provides an overview of the distribution of finite 'Lorelei Links' relative to Bacon Indices and Erdös Numbers (that give a mathematician's distance to Paul Erdös[15]). We have created a website that allows every author in the field to identify their distance (and route) to Lorelei Lingard (or any other author in the field): www.semedis.eu/oracle/index.aspx.

Discussion

Consistent with the impression one may well get from attending conferences and reading journals in the field, ME is indeed a small world. Although the number of people contributing to articles in the field is quite high,[16] at least two thirds of the authors in the field are connected to everyone else by very short chains of colleagues—on average, six. The hypothesis that most people are only about six degrees of separation apart thus holds in ME as well as among actors[8] and mathematicians.[16]

Most authors new to the journals included in our analysis tend to cluster in groups, where everyone they have written articles with has also worked with almost every one of their co-authors. This may result from the common observation and advice that working in groups fosters productivity.[17] Yet our analysis identified a contrary pattern: the more articles an author contributed, the less likely it is that his or her co-authors will have written an article together. Academic 'promiscuity' thus seems to be a key feature of highly productive scholars in ME. One possible explanation is that many of these productive scholars separately co-author papers with the students they supervise, but that these students rarely write papers together.

Previous work has focused on themes, institutions or journals in ME[1,2,16] and on the geographic location of authors.[18] Our study extends these studies by focusing on patterns of collaboration among researchers in ME. One interesting result is that, of the more than 24 000 authors who have published in the three journals considered, less than a quarter have contributed more than one article. This finding seems to confirm that publishing in ME is rather difficult.[16] The website we have created may help new researchers in the field to identify colleagues socially close to them, who can introduce them to more experienced scholars with a common interest.

Some limitations of our approach warrant consideration: First, our analysis was restricted to the three journals with the highest impact in the field. Second, our design neglects relationships other than co-authorship (two of the authors of this article share an obvious relationship that does not result from writing together). Third, for feasibility reasons, we only checked whether authors with the same last name and first initial were identical if they had contributed more than four articles. This could result in a bias towards underestimation of the overall network connectivity. Any such bias should be reasonably small, however, because connectivity tends to depend on well-connected people.[6] Finally, although centring on Lorelei Lingard might look like a rather random choice, it has the desirable outcome that medical educationalists—as opposed to mathematicians and actors—can locate a woman at the centre of their field.

Conclusion

ME is a small world. Making the connections between its actors visible may provide a new perspective on social phenomena (including peer review, citation and conference invitations) that occur in this world. It may further help scholars new to the field to consciously develop their personal academic network.

References

1. Rotgans J. The themes, institutions, and people of medical education research 1988–2010: Content analysis of abstracts from six journals. Adv Health Sci Educ Theory Pract 2011; 17:515–27.

2. Lee K, Whelan JS, Tannery NH, Kanter SL, Peters AS. 50 years of publication in the field of medical education. Med Teach. 2013;35(7):591-8.

3. Norman GR, van der Vleuten CPM, Newble DI (Eds.). International Handbook of Research in Medical Education. Dordrecht: Kluwe Academic Publishers 2002

4. Cleland J, Durning SJ (Eds.). Researching Medical Education. Oxford, UK: Wiley Blackwell 2015

5. Lingard L, Driessen E. How to tell compelling scientific stories: tips for artful use of research manuscript and presentation genres. In: Cleland J, Durning SJ (Eds.). Researching Medical Education. Oxford, UK: Wiley Blackwell 2015. p. 259-68

6. Watts DC. Six Degrees: The Science of a Connected Age. New York, NY: W.W. Norton & Company 2003

7. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. Science 2009;323(5916):892-5.

8. Reynolds P. The Oracle of Bacon. [Internet] [Place unknown] [Publisher unknown] Updated 2014; cited 2015 Sept 13. Available from http://oracleofbacon.org

9. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998;393(6684):440-2.

10. Thomson Reuters. Web of Science [Internet] London, UK: Thomson Reuters. Updated 2015; cited 2015 Sept 13. Available from http://wokinfo.com/

11. The PERL Foundation. The PERL Programming language. [Internet] Walnut, CA: The

PERL Foundation. Updated 2015; cited 2015 Sept 13. Available from http://www.perl.org/

12. The GraphML Working Group. The GraphML File Format. [Internet] [Place unknown] [Publisher unknown] Updated 2015 Jun 20; cited 2015 Sept13. Available from http://graphml.graphdrawing.org/

13. The Gephi Consortium. Gephi. [Internet] Paris, France: The Gephi Consortium. Updated 2015; cited 2015 Sept 13. Available from http://gephi.github.io/

14. Bleakley A. Gender matters in medical education. Med Educ 2013;47 (1):59–70.

15. Erdös Number Project: The Erdös Number Project. [Internet] Rochester, MI: University of Oakland. Updated 2015 Aug 10; cited 2015 Sept 13. Available from http://wwwp.oakland.edu/enp/

16. Azer SA. The top-cited articles in medical education: a bibliometric analysis. Acad Med 2015;90(8):1147-61.

17. Bland CJ, Center BA, Finstad DA, Risbey KR, Staples JG. A theoretical, practical, predictive model of faculty and department research productivity. Acad Med. 2005;80(3):225-37.

18. Tutarel O. Geographical distribution of publications in the field of medical education. BMC Med Educ 2002;2:3. http://www.biomedcentral.com/1472-6920/2/3

Contributions

WH and SH designed the study, WH and GK analysed the data, WH wrote the first draft of the manuscript. All authors contributed to the interpretation of results, revised the first draft for important intellectual content and provided final approval of the version to be published. All authors agree to be held accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

The authors would like to thank Lorelei Lingard, University of Western Ontario, for agreeing with grace and good humour to her prominent role in the paper. We further thank Georges Bordages, University of Illinois at Chicago, for pointing us to some publications relevant to this paper and three reviewers at *Medical Education* for their constructive critique of the manuscript. We further thank Susannah Goss, Berlin, for editing the language of the manuscript.

Funding none.

Conflicts of interest none.

Ethical approval not applicable.

Term in Social Network Analysis (SNA)	Explanation
Global properties of the network	
Vertex	A connected entity within a network; equivalent to an author in the network of ME scholars.
Edge	A connection between two authors; in the network of ME scholars: common authorship.
(Giant) component	A group of authors within a network who are connected (directly or through others). A giant component is by far the largest component in the network (if a considerable difference in size exists between components).
Path	A chain of edges and vertices between two given authors in a component.
Diameter	The longest path within a component.
Small world network	A (giant) component with a high average clustering coefficient and average shortest paths between any two authors that are short relative to the overall number of authors in the component.
Local properties of each author	
Degree	The number of different co-authors of a given author.
Clustering coefficient	The probability that any two co-authors of a given author have authored a paper together.
Closeness centrality	The average distance from a given author to all other authors in a component.
Betweenness centrality	How often an author is in the shortest path between any other two authors in a component.
Eccentricity	The distance from a given author to the author furthest from him or her in a component.

Table 1: Terms used in social network analysis (SNA) applied to the network of co-authorship in medical education (ME). Adopted from [6,7].

Field	**Medical Education**	**Actors**	**Mathematicians**
Distance to	**Lorelei Lingard**	**Kevin Bacon**	**Paul Erdös**
1	92	3018	504
2	1065	348805	6593
3	3704	1225913	33605
4	5498	305777	83642
5	3839	24890	87760
6	1568	2945	40014
7	433	361	11591
8	138	32	3146
9	14	16	819
10	7	2	244
11	-	-	68
12	-	-	23
13	-	-	5
Average	4.17	3.006	4.65
Number of linkable persons	16 358*	1 911 760	268 014

Table 2: Distribution of 'Lorelei Links' among authors in medical education relative to the distribution of Bacon Indices among actors (data from [8]) and Erdös Numbers among mathematicians (data from [15]). Number of hops required to the target person. *Probably more if the analysis were extended to more than three journals.

Figure 1: Co-authorship network between the most productive authors in the dataset (50 or more articles). Circle size proportional to the number of articles of the given author; colour proportional to the authors betweenness centrality (darker is higher); connection strength proportional to the number of articles two authors have co-authored.

Figure 2: Correlation between the log-transformed number of different articles by a given author and his or her log-transformed clustering coefficient (i.e. the likelihood that any two of his or her co-authors have written an article together).