Network Analysis of Academic Journals: Promoting Influential Research Through Collaboration

Scholarly Report submitted in partial fulfillment of the MD Degree at Harvard Medical School

Date: 26 February 2020

Student Name: Alexander J. Ordoobadi, BA

Scholarly Report Title: Network Analysis of Academic Journals: Promoting Influential Research Through Collaboration

Mentor Name and Affiliation: David C. Chang, PhD MPH MBA, Department of Surgery, Massachusetts General Hospital.

Collaborators, with Affiliations: Numa P. Perez, MD[1,2]; Maggie L. Westfal, MD MPH[1,2]; Cassandra M. Kelleher, MD[1,2]

1. Harvard Medical School, Boston, MA.
2. Department of Surgery, Massachusetts General Hospital, Boston, MA.

Table of Contents

Abstract

Title: Network Analysis of Academic Journals: Promoting Influential Research Through Collaboration

Purpose: The collaboration patterns of scientists may affect the impact of their research. In particular, clustering among authors in scientific journals may stifle innovation. The purpose of this paper was to explore the relationship between author clustering and journal impact factor.

Methods: Coauthor networks were generated for articles published in 2010-2015 in 31 journals within the fields of surgery and internal medicine. The overall degree of clustering within each journal was assessed by using network analysis techniques to calculate the average clustering coefficient (ACC). ACC values were compared between broad-interest and specialty-specific journals within surgery and internal medicine. Spearman's correlation coefficient was calculated between each journal's ACC and its impact factor, an established metric of a journal's influence.

Results: ACC was lowest in broad interest journals, like Science (0.014) and Nature (0.015), and clinical journals with a broad scope, like JAMA (0.025) and NEJM (0.026). In surgery and internal medicine, ACC increased as the journal became more specialized. There was a negative correlation between a journal's ACC and its impact factor (Spearman's r_s=–0.49, p=0.005).

Conclusions: Author clustering is higher in specialty-specific journals and is negatively correlated with a journal's impact factor. Open collaboration networks may promote influential research.

Glossary of Abbreviations

ACC: Average Clustering Coefficient

Description of Scholarly Work

Statement of Scholarly Project Question

It is commonly held that diversity among teams of researchers leads to higher quality research, and there is a growing body of evidence highlighting the importance of ethnic diversity within research groups (Freeman & Huang 2015; AlShebli et al. 2018). Other types of diversity are likely important too, including bringing together researchers of different scientific professional backgrounds (Fortunato 2018). Together, diverse perspectives among research team members likely result in novel and innovative ideas that may result in higher impact research.

Underpinning the benefit of diversity is collaboration. If researchers work as individuals rather than collaborate, they are unable to harness the benefits of diverse perspectives. This central benefit of collaboration may explain why scientific research has increasingly shifted from an individual pursuit to the work of teams (Wuchty et al. 2007).

We sought to harness the technique of network analysis to study collaboration patterns of academic researchers. In particular, we aimed to determine whether working with the same group of researchers on multiple projects would limit the impact of subsequent research projects because team members would not be exposed to new and diverse perspectives. In network analysis, clustering is a metric used to describe this interconnected model of collaboration where researchers work repeatedly with prior collaborators. We hypothesized that higher levels of clustering would be associated with lower impact research.

To study this hypothesis, we studied collaboration among authors who published in the same academic journal over multiple years. Journals are a useful environment to study the outcome of collaboration among a cohort of researchers because the average impact of all of the studies published in a journal is quantified as the journal's impact factor. In addition, co-authorship on a publication serves as objective evidence of collaboration. We had several specific aims for our research:

- <u>Aim 1</u>: Describe current collaboration patterns among authors who have published in top surgical and internal medicine journals.

- Aim 2: Compare overall levels of clustering between broad-interest and specialty-specific journals in both internal medicine and surgery.
- Aim 3: Determine whether there is an association between a journal's overall level of clustering (as quantified by the average clustering coefficient [ACC]) and the journal's impact factor.

Description of Student Contribution to Work

I am the lead author on this study and contributed extensively to all aspects of the project. All coauthors worked together to design the study over the course of several lab meetings. I collected the data from online databases and wrote the custom scripts in the R programming language. Because the datasets were so large, I worked with Saam Aghevli, a computer science engineer at Google, to optimize the efficiency of the custom scripts. I performed the statistical analysis using Gephi, an open source network analysis software program, and the statistical software Stata.

All members of the research team discussed the results and determined how to present the data in the manuscript over the course of multiple lab meetings. I was the primary author of all portions of the manuscript and created all figures. Dr. Perez provided substantial edits to multiple portions of the manuscript, particularly the first paragraph of the introduction and portions of the discussion. Dr. Westfal also provided thoughtful edits. Dr. Kelleher and Dr. Chang provided oversight of the entire project and also contributed edits to multiple portions of the manuscript.

Dr. Chang was my primary mentor on the project. He was a very generous mentor, always available to discuss my project even if it meant staying late after work. He taught me his approach to every step of the research process, from formulating a question to performing the statistical analysis and drafting the manuscript. I learned from him how to code in Stata and how to apply statistical methods to answer research questions. But more importantly, he encouraged me to follow my interests and supported me whenever I had new ideas for projects or a new dataset I wanted to analyze. He encouraged me to develop a personal vision for my research and clinical career goals, and he helped me to pursue my interest in academic surgery.

When it came time to apply to residency, he advised me about programs to consider, reviewed my personal statement, and helped me to prepare for interviews. I am very grateful for Dr. Chang's mentorship.

Manuscript

Introduction

The goal of research in academic medicine is to produce novel and impactful findings that can improve patient outcomes. Over the past 50 years, the production of research has shifted from the domain of individual researchers to the work of teams, and it is now clear that teams outperform individuals in generating high-impact research.[1] However, the ideal structure and composition of research teams remains unclear. Collaborating frequently with the same group of individuals may be an efficient way to produce research because familiar individuals are likely to work together effectively. Conversely, collaboration among academic physicians of diverse backgrounds (i.e. ethnic, scientific, professional, etc.) may provide a range of perspectives that leads to novel ideas and research of a higher impact.[2–5]

Studying the collaboration networks of academic physicians may help to uncover collaboration patterns that promote exposure to a diversity of perspectives.[6] Collaboration networks can be generated using co-authorship on scientific publications as objective evidence of collaboration.[7–10] In these coauthor networks, researchers are connected to each other if they are both listed as an author on the same publication. A variety of statistical techniques can then be applied to the network to quantify the manner in which individuals collaborate to produce research.[11] One important metric used to describe collaboration networks is clustering, which occurs when individuals form highly interconnected groups. In the case of coauthor networks, clustering occurs when an author's coauthors have also been coauthors with each other on another article.[12] Because high levels of clustering suggest that academic physicians are collaborating frequently with the same individuals, higher levels of clustering could have the unintended result that researchers are not exposed to a diversity of views.

To explore the relationship between clustering and research impact, we evaluated the degree of clustering within the cohort of authors who published in the same academic journal. We chose to study the collaboration networks of academic journals because their overall degree of influence is quantified by the journal's impact factor. We hypothesized that higher degrees of author clustering would be found in more specialized journals and would have an inverse association with journal impact factor.

Methods

Selection of Journals

We selected 31 scientific journals indexed by Medline for inclusion in the study. We categorized journals based on the scope of articles that they accept. We used three main groupings of journals: general interest, internal medicine, and surgery. Each group was divided into two sub-groups. General interest journals were divided into general science journals (e.g., Science and Nature) and general clinical journals (e.g., JAMA and New England Journal of Medicine). Internal medicine and surgery journals were categorized as broad-interest if they accepted articles from across the entire field or specialty-specific if they accepted articles from a sub-specialty of internal medicine or surgery. To appropriately categorize the journals, we reviewed “About the Journal” and “Instructions for Authors” pages on each journal’s website. For example, Annals of Surgery was categorized as a broad-interest surgical journal because it accepts articles that are from a range of surgical specialties,[13] while Annals of Thoracic Surgery was classified as a specialty-specific journal because it publishes articles on cardiothoracic surgery,[14] a surgical specialty. Journal names presented in tables and figures were abbreviated using the ISO 4 standard.[15]

Data Sources

We used publicly available authorship data from PubMed (U.S. National Library of Medicine, Bethesda, MD) and Web of Science (Clarivate Analytics, Philadelphia, PA) to construct the coauthor networks. PubMed does not provide authors’ full first and/or middle names, making the process identifying unique authors difficult because many individuals could share the same last name and initials. Web of Science, on the other hand, does provide authors’ full first names and middle initials. Therefore, we primarily used data obtained from Web of Science to construct the coauthor networks, with PubMed data used to verify the accuracy of the Web of Science data. We included articles that were published in the selected journals from 2010 through 2015. We filtered article types within Web of Science and only included research articles, letters, review articles, and editorial materials within the analysis. We excluded articles with more than 22 authors as the PubMed database does not include complete author lists for these articles.

We obtained the impact factor for each journal included in the study from Journal Citation Reports (Clarivate Analytics, Philadelphia, PA). The journal's yearly impact factor was averaged across the 2010-2015 study period.

Statistical Analysis

We employed network analysis techniques to generate coauthor networks for each journal. In the coauthor networks, individual authors are represented as nodes. When two nodes are both listed as authors on the same article, an edge is drawn between them, signifying co-authorship.[7] In this study, we generated each journal's coauthor network by connecting the last author of each article to all other coauthors in that article, with all connections directed from the last author to the other authors (i.e. a "directed" network). Constructing the network in this manner prevents the connections between authors on a single article from contributing to the overall level of clustering within a journal.

The networks were constructed using custom scripts in the R programming language (R Foundation for Statistical Computing, Vienna, Austria). We first reformatted the authorship data obtained from Web of Science to remove capitalization and special characters in an effort to prevent the same author from being split into two by the network analysis software due to minor differences in punctuation or accents. We then generated an edge list (i.e. a list of connections between authors) for each journal to describe the network. The edge list was exported from R for analysis in network analysis software.

We graphed and analyzed the coauthor networks using the open-source network analysis program Gephi, version 0.9.2 (Gephi Consortium, Paris, France). Graphical representations of the coauthor networks were constructed using the OpenOrd algorithm within Gephi. To quantify the degree of clustering within each journal, we calculated the directed average clustering coefficient (ACC), which measures clustering based on the proportion of an author's coauthors who have also been coauthors on other articles in the same journal.[9,12] ACC is normalized on a scale from 0 to 1, with higher ACC indicating more clustering among authors.

We calculated the Spearman correlation to assess the association between journals' ACC and impact factor. We used two-tailed t-tests to compare ACC between broad-interest and specialty-specific journals. Significance level was set at $\alpha \leq 0.05$. These statistical analyses were performed using Stata, version 15.1 (StataCorp, College Station, TX).

Results

We performed coauthor network analysis on a total of 159,895 articles published in 31 journals across a range of disciplines, including 12 surgical journals, 12 internal medicine journals, and 7 general interest journals. Table 1 lists the journals included in the study and their categorization as either broad-interest or specialty-specific journals as well as descriptive statistics for each journal's coauthor network. Coauthor networks and associated ACC for 4 representative journals are depicted in Figure 1.

ACC was lowest in broad interest scientific journals like Science (0.014) and Nature (0.015). Clinical journals with a broad scope, like JAMA and New England Journal of Medicine (NEJM), had a slightly higher ACC (0.025 and 0.026, respectively). ACC was highest in specialty-specific journals like Annals of Surgical Oncology (0.052) and Annals of the Rheumatic Diseases (0.053). Figure 2a demonstrates the distribution of ACC scores across all journals included in the study. There was a moderate negative correlation between a journal's ACC and its impact factor (Spearman's $r_s = -0.49$, p = 0.005) (Figure 2b).

Among surgical journals, broad interest journals like JAMA Surgery (0.031), Surgery (0.032), and Journal of the American College of Surgeons (0.036) had lower ACC than subspecialty surgical journals like Annals of Surgical Oncology (0.052) and Journal of Pediatric Surgery (0.047). The mean ACC of broad-interest surgical journals was significantly lower than that of specialty-specific surgical journals (0.034 vs. 0.045, respectively; p = 0.01) (Figure 3). This trend toward higher ACCs among specialty-specific journals was also observed in the field of internal medicine, with broad-interest internal medicine journals having lower mean ACC than specialty-specific journals (0.026 vs. 0.040, respectively; p = 0.03).

Discussion

Our study demonstrates that author clustering is higher in specialty-specific journals than in broad-interest journals and is negatively correlated with a journal's impact factor. Together, these findings suggest that clustering is more common in sub-specialty research and is associated with publication in journals with lower impact factors.

The negative association between author clustering and journal impact factor suggests that high degrees of clustering could impede the generation of influential research. Our finding that clustering is associated with research of lower impact is supported by a previous study that found that clustering among surgical researchers is associated with lower levels of innovation.[16] Clustering likely limits work on novel topics by decreasing the diversity of ideas within research groups. When collaboration among researchers is modeled over time, teams that include newcomers (i.e. individuals with whom the team has not previously collaborated) produce research of higher impact than teams comprised of members who have worked together previously.[6] This observation has been attributed to the ability of new team members to add to the diversity of perspectives of the research group.[6] Adding new team members to a research group is one potential mechanism for researchers to decrease clustering, increase the diversity of ideas, and produce research of greater impact.

Clustering may also be negatively associated with journal impact factor because it decreases the breadth of the audience that the research attracts. Because a journal's impact factor is calculated based on the number of citations each article receives divided by the total number of articles in the journal, journals with articles that are seen by a wider audience are more likely to receive citations and generate a higher impact factor. Importantly, the diversity of an article's authors may contribute to the subsequent breadth of its readership.[17] Therefore, the effect of clustering on a journal's impact factor may be at least partially modulated by lower levels of clustering attracting a broader audience.

Although clustering was negatively associated with impact factor, some degree of clustering is expected whenever individuals collaborate and is necessary for ideas to be shared among research groups. While the implications of clustering have not been extensively studied in the domain of scientific research, the effects of clustering have been studied on the critical and economic success of Broadway musicals.[18] In this setting, both low and high levels of clustering were shown to be associated with decreased financial and critical success of Broadway shows.[18] Low levels of clustering likely limited the success of a production because it meant that production staff were not collaborating, while excessive clustering prevented diverse and creative ideas from forming. Low levels of clustering were not shown to be negatively associated with impact factor in this study, likely because the sample size of included journals was relatively small.

There are several possible explanations for the finding that specialty-specific journals have higher levels of clustering than broad-interest journals. Coauthors in specialty-specific journals are more likely to share similar research interests, leading to more collaboration among coauthors within these journals over time and therefore increased clustering. An interesting alternative interpretation is that ACC might suggest the degree of sponsorship among a journal's senior authors. Because we constructed the coauthor networks by linking the last author on each article to the other authors, our analysis may highlight the role of the senior author in forging connections between coauthors that lead to future publications within the same journal, with higher levels of sponsorship by senior authors leading to higher ACC. Specialty specific journals may therefore have higher degrees of sponsorship by senior authors.

A key strength of this study is our unique approach for constructing authorship networks by connecting the last author on each article to the article's other others. This method prevents connections between individual authors on a single paper from artificially increasing the overall degree of clustering. Our study is limited by problems with author name disambiguation, a known challenge when developing authorship networks that occurs when network analysis software inadvertently splits a single author into two or vice-versa.[19] Our use of data obtained from Web of Science, which contains authors' full first name and middle initial, greatly decreased the level of disambiguation errors compared to using PubMed data. Moreover, any residual uncertainty surrounding author names is likely to have occurred in equal rates across all journals included in the study and is therefore unlikely to have significantly affected our overall results. Our study is also limited by the relatively small number of journals that were included. Moreover, the included journals all had impact factors within the top 50 percent of all journals within their field, limiting the ability of our study to describe the association between clustering and impact factor in lower tier journals.

ACC is a novel way to quantify collaboration among a journal's authors that provides new insights into collaboration patterns and how they generate influential research. ACC provides the research community with a new metric to evaluate a journal's academic mission and track changes in collaboration over time that may ultimately influence the journal's impact factor. For individual authors seeking to publish in a broad-interest journal, this study supports taking on new team members and engaging with authors of diverse specialties and backgrounds to help make their research more influential.

Acknowledgements: The authors would like to thank Saam Aghevli for his programming advice.

Funding/Support: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Approval: Not applicable.

Previous Presentations: This work was presented at the American College of Surgeons Clinical Congress 2019 on October 29, 2019 in San Francisco, CA.

References

1. Wuchty S, Jones BF, Uzzi B. The Increasing Dominance of Teams in Production of Knowledge. *Science*. 2007;316(5827):1036-1039. doi:10.1126/science.1136099

2. Freeman RB, Huang W. Collaborating with People Like Me: Ethnic Coauthorship within the United States. *Journal of Labor Economics*. 2015;33(S1):S289-S318. doi:10.1086/678973

3. AlShebli BK, Rahwan T, Woon WL. The preeminence of ethnic diversity in scientific collaboration. *Nature Communications*. 2018;9(1):5163. doi:10.1038/s41467-018-07634-8

4. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. *Facilitating Interdisciplinary Research*. Washington, DC: The National Academies Press; 2004. doi:10.17226/11153

5. Fortunato S, Bergstrom CT, Börner K, et al. Science of science. *Science*. 2018;359(6379):eaao0185. doi:10.1126/science.aao0185

6. Guimerà R, Uzzi B, Spiro J, Amaral LAN. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science (New York, NY)*. 2005;308(5722):697. doi:10.1126/science.1106340

7. Fonseca B de PFE, Sampaio RB, Fonseca MV de A, Zicker F. Co-authorship network analysis in health research: method and potential use. *Health Res Policy Syst*. 2016;14(1):34. doi:10.1186/s12961-016-0104-5

8. Melin G, Persson O. Studying research collaboration using co-authorships. *Scientometrics*. 1996;36(3):363-377. doi:10.1007/BF02129600

9. Newman ME. The structure of scientific collaboration networks. *Proc Natl Acad Sci USA*. 2001;98(2):404-409. doi:10.1073/pnas.021544898

10. Newman MEJ. Coauthorship networks and patterns of scientific collaboration. *PNAS*. 2004;101(suppl 1):5200-5205. doi:10.1073/pnas.0307545100

11. Newman M. *Networks: An Introduction*. 2nd ed. Oxford University Press; 2010. doi: 10.1093/acprof:oso/9780199206650.001.0001.

12. Newman ME. Clustering and preferential attachment in growing networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2001;64(2 Pt 2):025102. doi:10.1103/PhysRevE.64.025102

13. Editorial Manager - Annals of Surgery. http://edmgr.ovid.com/annsurg/accounts/ifauth.htm. Accessed February 20, 2020.

14. Aims and Scope: The Annals of Thoracic Surgery. https://www.annalsthoracicsurgery.org/content/aims. Accessed February 20, 2020.

15. International Standard Serial Number International Centre. ISO 4 List of Title Word Abbreviations. https://www.issn.org/services/online-services/access-to-the-ltwa/. Accessed February 20, 2020.

16. Garas G, Cingolani I, Patel V, et al. Surgical Innovation in the Era of Global Surgery: A Network Analysis. *Ann Surg*. December 2018. doi:10.1097/SLA.0000000000003164

17. Freeman RB, Huang W. Collaboration: Strength in diversity. *Nature News*. 2014;513(7518):305. doi:10.1038/513305a

18. Uzzi B. A social network’s changing statistical properties and the quality of human innovation. *J Phys A: Math Theor*. 2008;41(22):224023. doi:10.1088/1751-8113/41/22/224023

19. Fegley BD, Torvik VI. Has Large-Scale Named-Entity Network Analysis Been Resting on a Flawed Assumption? *PLoS One*. 2013;8(7). doi:10.1371/journal.pone.0070299

Tables and Figures

Table 1. Categorization and network properties of journals included in the study.

Journal Categorization		Journal Name	Number of Articles	Number of Nodes	Mean Degree
General Interest (n = 7 journals)	Science	Science	10023	35707	0.949
		Nature	12138	35698	0.972
		Proc Natl Acad Sci USA (PNAS)	24572	114541	1.082
	Clinical	JAMA	6021	13772	0.971
		New Engl J Med (NEJM)	9338	21012	0.952
		Lancet	9607	20845	1.011
		BMJ	10872	16322	0.967
Internal Medicine (n = 12 journals)	Broad	JAMA Intern Med	3587	8911	0.986
		Ann Intern Med	3522	7656	0.989
		J Gen Intern Med	2237	6059	1.065
		Am J Med	2434	7213	0.936
	Specialty	J Amer Coll Cardiol	5885	18541	1.302
		Circulation	5972	22322	1.232
		Arthritis Rheumatol	2922	13670	1.137
		Ann Rheum Dis	2543	10774	1.310
		Clin J Am Soc Nephrol	1917	7644	1.190
		J Am Soc Nephrol	1707	8406	1.132
		Gastroenterology	3314	14408	1.071
		Am J Gastroenterol	2203	8357	1.056
Surgery (n = 12 journals)	Broad	JAMA Surg	1849	5528	1.010
		Surgery	2335	9305	1.107
		J Am Coll Surg	2812	7694	1.156
		Ann Surg	2595	10491	1.101
		World J Surg	2841	10559	1.002
	Specialty	Ped Surg Int	1264	4513	1.040
		J Ped Surg	6064	10042	1.158
		Surg Endosc	3473	13826	1.082
		J Trauma Acute Care Surg	3978	12123	1.256
		Dis Colon Rectum	1428	4669	1.052
		Ann Surg Oncol	3989	17490	1.187
		Ann Thorac Surg	6453	19933	1.258

Figure 1. Coauthor social networks and associated average clustering coefficient (ACC) for 4 selected journals.

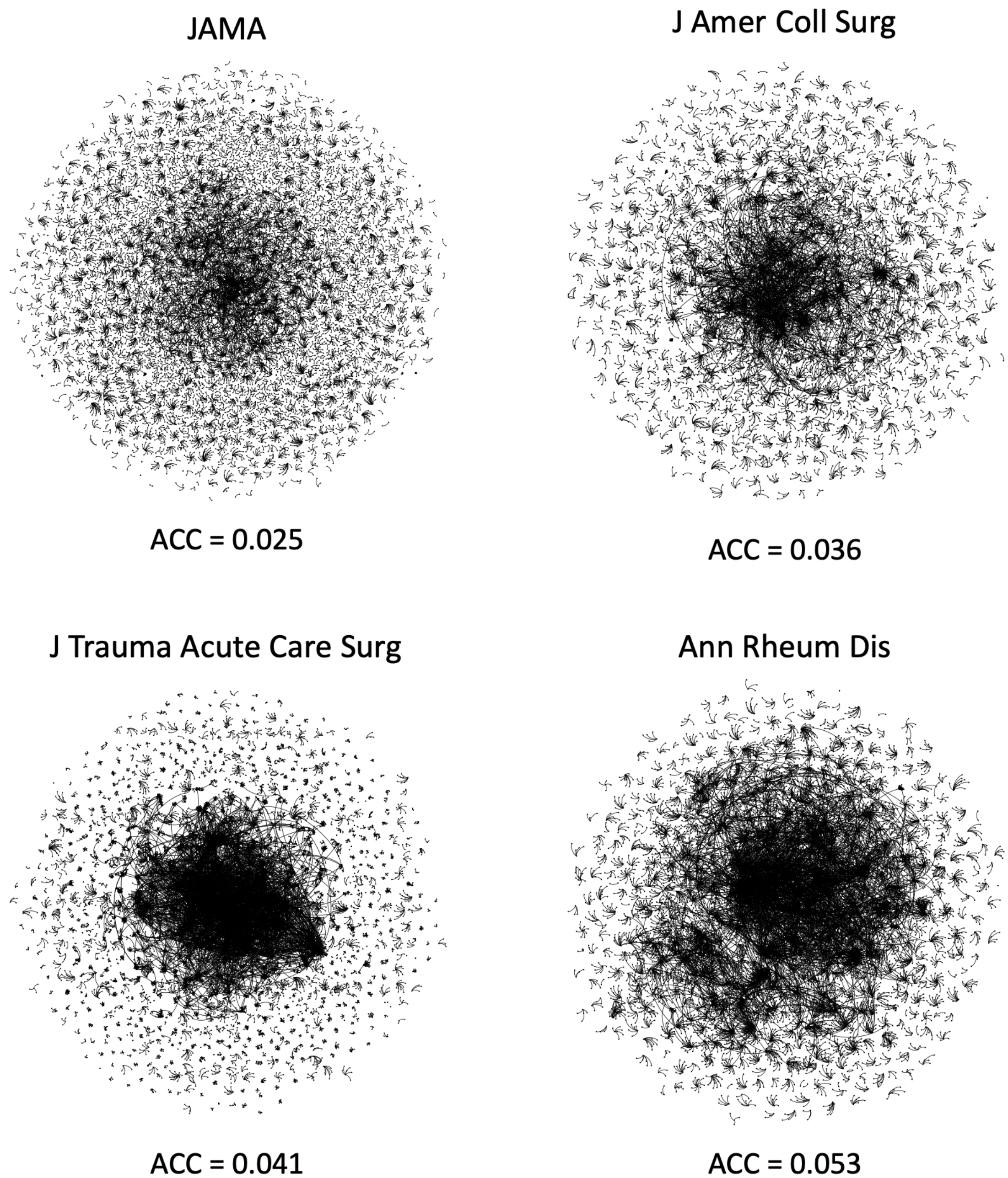

Figure 2. a) Distribution of scientific journals by average clustering coefficient (ACC). **b)** Relationship between ACC and impact factor.

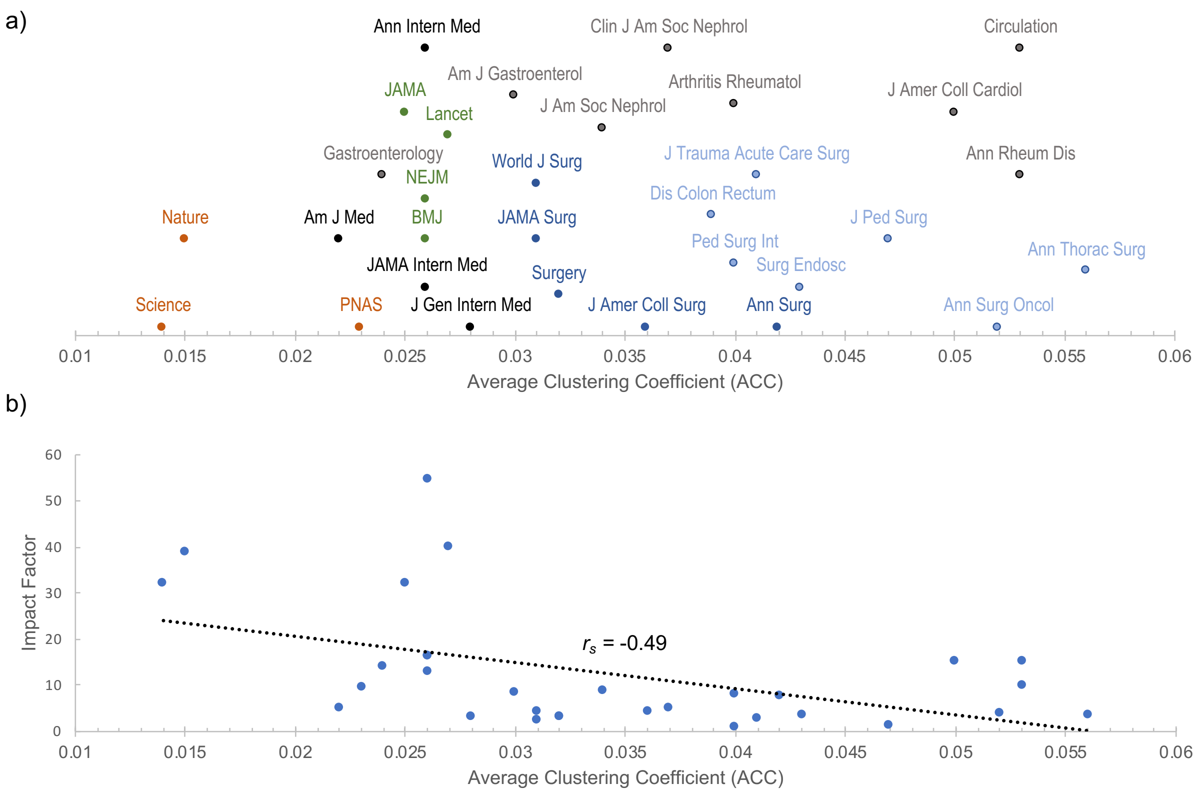

Figure 3. Comparisons of average clustering coefficient between broad-interest and specialty-specific journals in the fields of medicine and surgery.