Journal of Biomedical Informatics 52 (2014) 311–318

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Associating co-authorship patterns with publications in high-impact journals

Michael E. Bales [a,b], Daniel C. Dine [b], Jacqueline A. Merrill [a,c], Stephen B. Johnson [d], Suzanne Bakken [a,b,c], Chunhua Weng [a,b,*]

[a] Department of Biomedical Informatics, Columbia University, New York City, United States
[b] The Irving Institute for Clinical and Translational Research, Columbia University, New York City, United States
[c] School of Nursing, Columbia University, New York City, United States
[d] Weill Cornell Medical College, New York City, United States

ARTICLE INFO

Article history:
Received 2 December 2013
Accepted 11 July 2014
Available online 19 July 2014

Keywords:
Co-authorship patterns
Social networking
MEDLINE
Journal impact factor
Evaluation

ABSTRACT

Objectives: To develop a method for investigating co-authorship patterns and author team characteristics associated with the publications in high-impact journals through the integration of public MEDLINE data and institutional scientific profile data.

Methods: For all current researchers at Columbia University Medical Center, we extracted their publications from MEDLINE authored between years 2007 and 2011 and associated journal impact factors, along with author academic ranks and departmental affiliations obtained from Columbia University Scientific Profiles (CUSP). Chi-square tests were performed on co-authorship patterns, with Bonferroni correction for multiple comparisons, to identify team composition characteristics associated with publication impact factors. We also developed co-authorship networks for the 25 most prolific departments between years 2002 and 2011 and counted the internal and external authors, inter-connectivity, and centrality of each department.

Results: Papers with at least one author from a basic science department are significantly more likely to appear in high-impact journals than papers authored by those from clinical departments alone. Inclusion of at least one professor on the author list is strongly associated with publication in high-impact journals, as is inclusion of at least one research scientist. Departmental and disciplinary differences in the ratios of within- to outside-department collaboration and overall network cohesion are also observed.

Conclusions: Enrichment of co-authorship patterns with author scientific profiles helps uncover associations between author team characteristics and appearance in high-impact journals. These results may offer implications for mentoring junior biomedical researchers to publish on high-impact journals, as well as for evaluating academic progress across disciplines in modern academic medical centers.

1. Introduction

Biomedical research is becoming increasingly interdisciplinary [1]. Numerous organizational factors have been recognized as barriers or facilitators of interdisciplinary research [2]. Although there are significant challenges in projects spanning multiple departments or disciplines [3], interdisciplinary research has been shown to be important for accelerating innovation [4].

A variety of analytical approaches, such as social–ecological models, systems thinking and complexity theories, social-determinants paradigms, and hierarchical analytic frameworks [5], have been employed to understand patterns of scientific collaboration. A prior bibliometric study has shown differences in co-authorship patterns across disciplines [6]. However, factors associated with the differences in scientific productivity have not been systematically quantified.

Given the central importance of scholarly publications and team-based scientific work, in this study we sought to understand scientific collaborations in biomedical research by investigating co-authorship patterns. Specifically, we sought to identify associations between co-authorship patterns and the impact factors of the journals of the publications. We leveraged the open-access Columbia University Scientific Profiles (CUSP) (http://irvinginstitute.columbia.edu/cusp) to obtain information about published researchers at our institution. Using CUSP, we enriched publication

* Corresponding author at: Department of Biomedical Informatics, Columbia University, New York City, United States.
E-mail address: cw2384@columbia.edu (C. Weng).

http://dx.doi.org/10.1016/j.jbi.2014.07.015
1532-0464/

data with institution-internal human resources data, including author academic rank and departmental affiliation. We employed two methodological approaches: analysis of authorship patterns and co-authorship networks. We then compared departments with respect to the ratio of within- to outside-department collaboration, as well as the overall levels of structural integration, all within our institution.

2. Materials and methods

2.1. Data sources and sample selection

Data were retrieved from our institution's research networking system, CUSP. CUSP was funded by Columbia University's Clinical and Translational Science Award (CTSA) to facilitate research networking and to help researchers identify experts and potential collaborators at CUMC. CUSP includes grants from institutional financial databases and publications from MEDLINE, along with job title, highest degree completed, and departmental affiliation from institutional human resource data. A core feature of CUSP is ReCiter [8], a method developed by the Columbia University CTSA for author name disambiguation for publications in scholarly databases. Researcher profiling systems often require investigators to populate their own publications manually. ReCiter keeps publications up to date by populating author publication lists automatically in CUSP through monthly feeds from MEDLINE. CUSP is interoperable with the open-source semantic web application VIVO, which enables the discovery of researchers across institutions [7].

When determining a time frame appropriate for article selection we sought to include enough articles to provide sufficient statistical power to address our research questions, while also minimizing the effects of missing data in older years. As CUSP only provides a snapshot of researchers currently employed at the university, historical data on academic rank and departmental affiliation were not available for current authors and no data is available for those who have left the university. Since personnel fluctuation is frequent in our university, it is appropriate to use a time period, e.g., a 5-year time period, that is shorter than our standard promotion time window (i.e., 7–11 years) for the analysis of patterns of authorship based on academic rank and departmental affiliations (Section 2.2) so that we can assume such information is less likely to have substantially changed during the short time frame.

Moreover, due to a recent major upgrade of administrative systems that provides departmental affiliation and rank to CUSP at our institution, year 2011 provided the most complete data at the time of the analysis. Therefore, we retrieved 7997 MEDLINE articles from 2007 to 2011 that included at least one author in CUSP. From this list of articles we identified 182 journals with an impact factor record, in which 10 or more articles were published during the time period. From these 182 journals we identified 3996 articles involving 2001 unique authors for this analysis. In contrast, for the co-authorship network analysis (Section 2.3), as social connections among researchers take time to develop, we sought to ensure that sufficient data on social links would be included. We therefore selected a 10-year period, 2002–2011, corresponding to a data set with 13609 articles, 2893 unique authors, and 2072 journals, from which individual co-authorship networks were generated for each of the top 25 departments in publishing volume (i.e., the 25 most prolific departments).

2.2. Co-authorship impact analysis

The first goal of this research was to characterize associations of author academic rank and departmental affiliation with publication in high-impact journals. After preliminary descriptive analysis we formulated our research questions as (1) what are the typical co-authorship patterns with respect to five specific author team properties (i.e., total number of authors, mixing of academic rank, inclusion of senior researchers, inclusion of junior scientists, and inclusion of basic or clinician scientists); and (2) which co-authorship patterns are associated with publications in high-impact journals?

To assign each article to a distinct journal impact tier we first ranked the articles based on journal impact factor for the year 2012, as reported in the ISI Journal Citation Reports [9]. We then divided the journals into three tiers based on journal impact rank: (1) High: [5.704, 51.658, $n = 60$]; (2) Medium: [3.371, 5.635, $n = 61$]; and (3) Low: [0.871, 3.320, $n = 61$]. We further labeled each article with one of these three journal impact categories.

We extracted academic rank and departmental affiliation for all authors having profiles in the CUSP system. For the analyses involving academic rank we included only investigators with an academic rank of postdoc, research scientist, assistant professor, associate professor, or professor (in our author academic rank notation, the term *professor* is used to denote full professors). Authors for whom academic rank was unavailable (e.g., authors at other institutions and researchers no longer employed at our university) were excluded. Authors were labeled according to their primary department affiliation.

In this context, authorship patterns are based on academic rank and on departmental affiliation. Possible combinations based on academic rank might be *one professor and one assistant professor*, or *one associate professor and two postdocs*. Similarly, possible combinations based on department type might be *one researcher from a clinical department and one researcher from dental medicine*, or *one researcher from public health and two researchers from basic science departments*. We enumerated author patterns for each paper as follows. First, we enumerated distinctive combinations of co-authors based on academic rank irrespective of author order. For example, if one paper had a professor as its first author, an assistant professor as its second author, and another professor as its third author, its academic rank pattern was PPI, representing *two professors* (*P*) *and one assistant professor* (I). More example patterns are provided below: (1) IP = one *assistant professor* and one *professor*; (2) OP = one *associate professor* and one *professor*; (3) IIP = two *assistant professors* and one *professor*; and (4) DP = one postdoc and one *professor*.

Second, we enumerated combinations of co-authors based on department type. A paper was considered to belong to a department type if at least one author on the paper was from the department type; as such, some papers included multiple department types. We calculated the number of departments involved on each paper. In this research we used the term *department* to refer to major organizational entities at our university, including departments within the school of public health, as well as basic science, clinical, and mixed basic/clinical departments within the medical school, interdisciplinary research centers that were classified administratively as departments, and the schools of Nursing and of Dental Medicine, which were not divided into departments. The distinction of basic vs. clinical vs. hybrid only relates to School of Medicine departments at Columbia University Medical Center (CUMC), where only basic science departments have Ph.D. programs. Clinical departments perform clinical services and research but cannot offer the Ph.D. Hybrid departments have Ph.D. programs and offer clinical services.

For each specific author academic rank and author department type combination we calculated numbers of articles published in high, medium, and low-impact journals. We assigned each paper into one of two categories along five separate axes: high (five or more) vs. low (four or fewer) numbers of authors; mixing of academic rank vs. single academic rank; inclusion of at least one professor vs. non-inclusion of professors; inclusion of at least one research scientist vs. non-inclusion of research scientists; and inclusion of at least one basic science author vs. non-inclusion of basic science authors. For each of these subsets we counted the articles published in high, medium, and low-impact journals.

2.3. Co-authorship network analysis

The second goal of this research was to characterize authorship networks among the most prolific departments at CUMC. To address this goal we formulated two research questions: (1) which departments have the highest and lowest ratios of within- to outside-department collaborations? and (2) which departments have the highest and lowest levels of overall structural integration?

To address these questions we employed social network analysis. Network-based approaches have been applied in mapping of knowledge domains [10], for describing basic principles that govern the structure of co-authorship networks [11], and for helping researchers understand the social aspects of scientific collaboration [12]. Within a given context, e.g. a research department, institution, or scholarly database, individuals are modeled as nodes in a network. Links are assigned between individuals with one or more types of social ties, such as participating on the same research team or service on the same grant review panel. Supplementing bibliographic data with additional author-specific information can lead to new insights [13]. However, since data about various types of social ties between researchers may not exist – or may be time consuming to collect via surveys – co-authorship is commonly used to model relations between researchers.

Although co-authorship networks incorporate just one of many possible types of social ties between researchers, they combine two key advantages. First, they are derived from the well-organized endpoint of scholarly work, the published article. Second, co-authorship data are readily available in scholarly databases. Co-authorship network models may be static, reflecting collaborative activity over a period of time (as in the current work), or dynamic, reflecting multiple periods of time.

Using author data from all publications in the MEDLINE database for the selected time frame (2002–2011), we then generated a co-authorship network for each of the 25 top departments, with authors represented as nodes, and links assigned between authors and their co-authors. We used the NetworkAnalyzer [14] tool built into the network visualization software Cytoscape [15], to count connected components and measure centralization in each network.

To address our first research question we calculated the numbers of within-department and outside-department authors within CUMC. We also calculated the total numbers of authors and publications in each department network between years 2002 and 2011.

For our second research question, there are many statistical measures of social network structure. To measure the structural integration of author-networks, we elected to measure numbers of connected components and centralization. Both are measures of overall structural integration of a network. In undirected networks, two nodes are considered connected if there is a path linking them. Within a network all nodes connected in this way form a connected component. Network centralization is a measure of overall topological structure indicating the level of structural integration in the network. It calculates the variation in the centrality scores among all the nodes in the network, while the centrality score of each node is computed according to degree of the nodes, i.e., the number of nodes connected to it. Networks that are more star-like in shape have a high centralization level close to one, while networks having a decentralized shape have a low centralization level close to zero. We used Freeman's formula [16] for calculating network centralization as follows:

$$C_D = \frac{\sum_{i=1}^{N}[C_d(n^*) - C_d(i)]}{\max \sum_{i=1}^{N}[C_d(n^*) - C_d(i)]}$$

where N is the total number of nodes, $C_d(n^*)$ is the node with the largest centrality, and $C_d(i)$ is the centrality of each node i.

3. Results

3.1. Co-authorship impact analysis

We identified a total of 187 combinations of authors of varying academic ranks (Appendix). We began by focusing on the top 20 patterns, each of which occurred 44 times or more (Fig. 1). The six most common combinations overall, in descending order of frequency, were *assistant professor and professor*; *professor*; *associate professor and professor*; *assistant professor*; *professor and research scientist*; and *two professors*. A chi-squared test for differences among journal impact categories found four of the top 20 combinations (25.0%) to be associated with statistically significant differences between the three journal impact tier categories ($p < 0.0025$ for each of these four combinations after Bonferroni correction for multiple comparisons). These four pairs were *professor and professor*, *professor and research scientist*, *assistant professor*, *and two postdocs*. Two of the four combinations were associated with publication in high-impact journals (i.e., *two professors*; *professor and research scientist*). Conversely, articles authored by an *assistant professor* with no other ranking authors were statistically significantly associated with publication in low-impact journals, as were articles authored by *two postdocs*.

The most common combination overall was *professor and assistant professor*, which occurred 519 times. Articles with this combination of author academic ranks appeared more commonly in medium- and low-impact journals than articles with other combinations, although the difference among journal tiers was not statistically significant ($p = .0048$).

To consider author academic rank patterns that occurred more rarely than the top 20, we also conducted a broader analysis of the combinations occurring 10 or more times ($n = 45$). Among these combinations, six additional patterns occurred more commonly in high-impact journals, although in some of these cases the difference in counts among journal tiers was not statistically significant after Bonferroni correction for multiple comparisons ($p < 0.00111$). Notably, five of these six patterns included at least one professor and six included at least one research scientist: *professor and research scientist* ($p < .0001$), *assistant professor and research scientist* ($p < .02$); *professor and two research scientists* ($p < .001$); *two professors and one research scientist* ($p < .02$); *two professors, assistant professor, and research scientist* ($p < .01$); and *two professors, associate professor and research scientist* ($p < .01$).

Fig. 1 shows the 20 most commonly occurring author academic rank combinations ranked by proportion of articles appearing in high-impact journals. The five combinations for which the proportion of articles appearing in high-impact journals was the highest included at least one professor. The combination for which the proportion of articles published in high-impact journals was the lowest was *two postdocs*.

Fig. 2 shows the proportions of articles published in high-, medium-, and low-impact journals for five contrasting pairs of article types. These contrasting pairs are *mixed rank* vs. *single rank*; *at least one professor* vs. *no professors*; *at least one research scientist* vs. *no research scientists*; *at least one basic science* vs. *no basic science*; and *high* vs. *low number of authors*. Differences among the journal impact factor categories were not statistically significant for *mixed rank* vs. *single rank* or for *high* vs. *low number of authors*, but were statistically significant for the other three subsets ($p < .005$ after

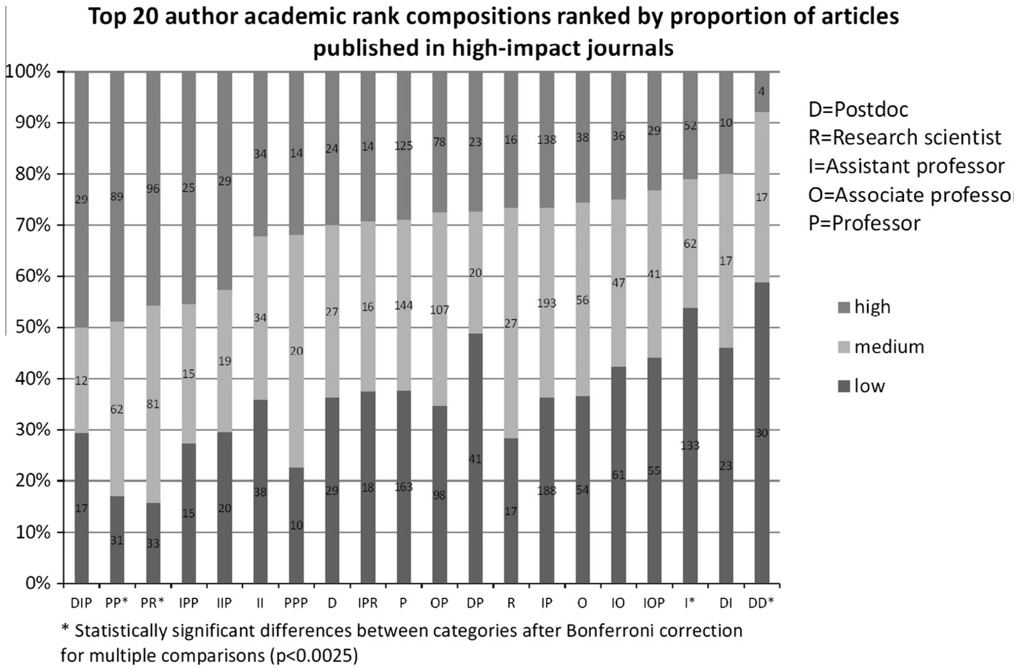

Fig. 1. Top 20 author academic rank compositions ranked by proportion of articles appearing in high-impact journals.

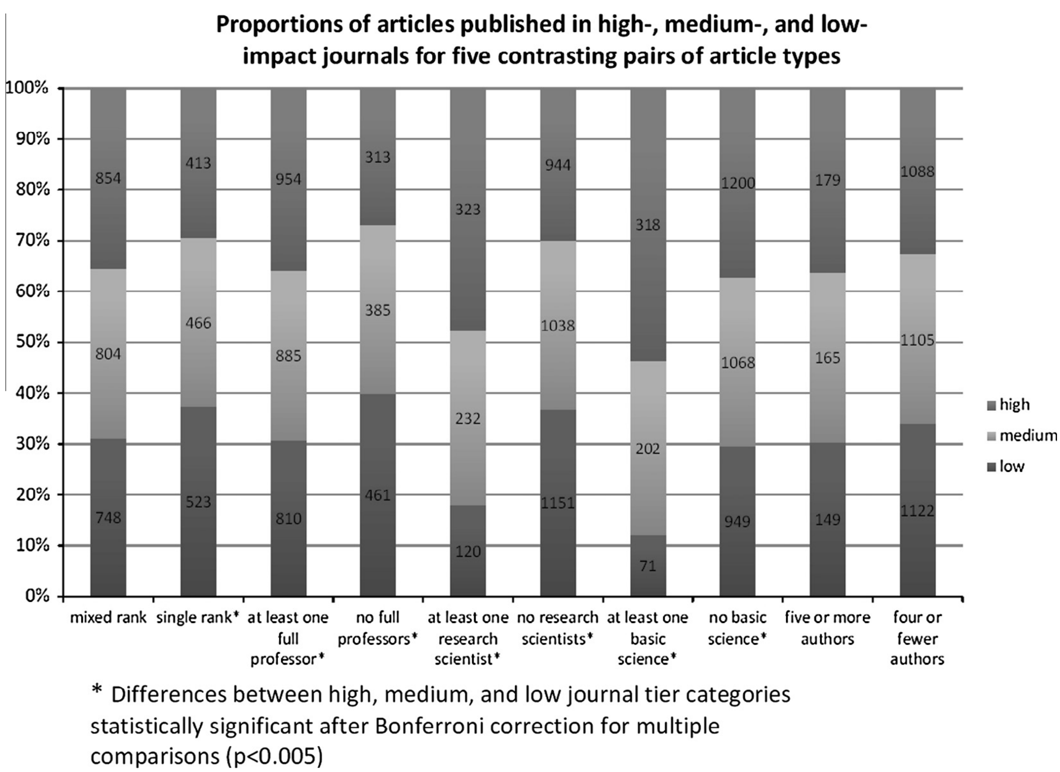

Fig. 2. Author academic rank and departmental affiliation combinations associated with publishing in high-impact journals. These include authors of mixed academic rank, participation of at least one professor, participation of at least one research scientist, participation of at least one author from a basic science department, and participation of five or more authors.

Bonferroni correction for multiple comparisons). Characteristics statistically significantly associated with publication in high-impact journals were inclusion of at least one professor, inclusion of at least one research scientist, and inclusion of at least one basic science author. The characteristic with the strongest association with publication in high-impact journals was inclusion of at least one basic science author – among the papers that included at least one author from a basic science department, 53.8% appeared in high-impact journals; conversely, only 12.0% of such papers appeared in low-impact journals.

As shown in Fig. 3(a), the average number of departments collaborating on papers by school was highest for dentistry and public health schools, and lowest for the School of Medicine. Fig. 3(b) shows that within the School of Medicine, the average number of departments collaborating on papers was highest for departments of a hybrid type that included both basic and clinical scientists, followed by departments of basic sciences, and then departments of clinical sciences.

3.2. Co-authorship network analysis

Results of the co-authorship network analysis for the top 25 prolific departments appear in Table 1 below. Although our sample included only three basic science departments, numbers of publications per author were higher, on average, than in clinical departments (average of 15.96; for clinical departments, 9.72). Levels of structural integration, as measured by centralization, varied substantially by department. Departments organized around specific medical specialties typically had the lowest levels of centralization, while interdisciplinary centers, and departments with higher numbers of authors per publication, generally had the highest levels of centralization and structural integration.

Fig. 4 provides an example of a distinct structural difference between a clinical department (i.e., Dermatology) and an interdisciplinary research institute (i.e., Taub Institute). Nodes are authors and links are assigned between authors and their co-authors. Authors affiliated with the indicated department are red; authors affiliated with other departments at the University are gray. Layout is determined by a force-directed node placement algorithm in which all nodes repel one another except linked nodes, which are drawn spatially closer to one another. As indicated by the mixing of red and gray nodes, researchers in Dermatology collaborated with others both inside and outside their department. The same was true for researchers in the Taub Institute. However, as seen in the figure, a significantly higher proportion of links in the core of the Institute's co-authorship network were with researchers in other departments. While the Taub Institute included only 33 authors overall, the department of Dermatology included 118 authors; however, as the networks included immediate co-authors, the networks for the two departments were comprised of the same number of authors (260). The Dermatology network had the lowest ratio of outside-department to within-department collaborators in our sample (1.20) whereas the Taub Institute network had the highest (6.88). The Taub Institute (full name, the Taub Institute for Research on Alzheimer's Disease and the Aging Brain) is, by design, highly interdisciplinary. The level of centralization at the Taub Institute (0.26) was also higher than that of Dermatology (0.15), indicating a more integrated structure overall.

We identified clear distinctions in authorship patterns between basic science and clinical departments (details appear in Table 1). The basic science departments, on average, also had higher ratios of outside- to within-department collaborators (average of 5.36; for clinical departments, 3.62). With regard to network structure, the basic science departments ranked lower in number of connected components (average of 2.67; for clinical departments, 9.38) and had higher levels of centralization (average of 0.26; for clinical departments, 0.19, with no measurement greater than 0.33). Given the small number of departments it was not possible to confirm these differences statistically.

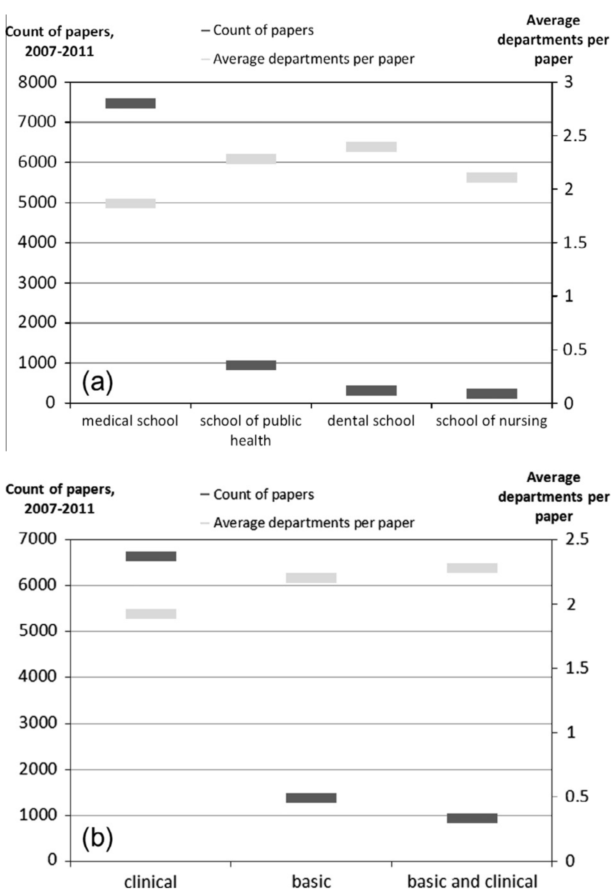

Fig. 3. (a) Counts of papers and average departments per paper, by school. (b) Counts of papers and average departments per paper by department type, in the school of Medicine.

4. Discussion

4.1. Implications

Translating the findings of basic science research for clinical and public health application is fundamental to the national CTSA program in the United States and can result in high-impact research. In this research we found inclusion of at least one basic science author to be strongly associated with publication in high-impact journals. This finding is of potential relevance to researchers in both basic science and clinical departments, as well as to university administrators and funders. For clinical researchers who are entrenched in disciplinary enclaves and have to juggle with research, clinical service, and educational demands, this finding provides yet more evidence that translational research, in addition to benefiting society [17], may also benefit one's career – teaming up with a basic science researcher may be a good strategy when pursuing publication in high-impact journals. Author combinations that included at least one research scientist were also associated with publication in high-impact journals, highlighting the importance of access to high-level dedicated research staff.

Mixing of academic rank was associated with publication in high-impact journals. However, some mixed-rank patterns involving faculty of various ranks (e.g., *professor, associate professor, and assistant professor*; and *associate professor and assistant professor*) were statistically significantly associated with publication in lower-impact journals. One possible explanation is that these author academic rank patterns are common in disciplines where authors tend to publish in lower-impact journals.

The department-level analysis of co-authorship network structure confirms significant variation across departments in levels of within- and outside-department collaboration as well as in overall structural integration. Departments organized around specific medical specialties typically had the lowest levels of centralization, while interdisciplinary centers, and departments with higher numbers of authors per publication, generally had the highest levels of

Table 1
Statistical measures of network structure for 25 departmental co-authorship networks at our university.

Department	Type	Authors in dept.	Collaborating authors in other depts.	Total number of nodes	Ratio of outside- to within-department collaborators	Publications	Publications per author	Number of connected components	Centralization
Anesthesiology	Clinical	100	249	349	2.49	842	8.42	16	0.10
Dept of Biomedical Informatics	Basic science	28	110	138	3.93	391	13.96	2	0.25
Dermatology	Clinical	118	142	260	1.20	531	4.50	12	0.15
Epidemiology	Public health	46	269	315	5.85	650	14.13	4	0.16
G H Sergievsky Ctr	Basic science	28	148	176	5.29	483	17.25	2	0.27
Medicine-Cardiology	Clinical	139	366	505	2.63	1335	9.60	8	0.19
Medicine-Contract Physicians	Clinical	35	187	222	5.34	435	12.43	13	0.09
Medicine-Endocrinology	Clinical	43	203	246	4.72	386	8.98	1	0.24
Medicine-General Medicine	Clinical	50	186	236	3.72	414	8.28	13	0.16
Medicine-Nephrology	Clinical	51	225	276	4.41	337	6.61	4	0.20
Medicine-Oncology	Clinical	37	227	264	6.14	454	12.27	3	0.33
Medicine-St. Lukes/Roosevelt	Clinical	81	170	251	2.10	471	5.81	14	0.18
Neurological Surgery	Clinical	36	150	186	4.17	317	8.81	1	0.31
Neurology-Critical Care	Clinical	21	144	165	6.86	419	19.95	4	0.32
Obstetrics & Gynecology	Clinical	83	197	280	2.37	787	9.48	12	0.11
Ophthalmology	Clinical	75	96	171	1.28	471	6.28	7	0.18
Pathology and Cell Biology	Clinical	158	644	802	4.08	1260	7.97	7	0.15
Psychiatry	Clinical	60	241	301	4.02	451	7.52	15	0.14
Psychiatry-Child Psychiatry	Clinical	65	146	211	2.25	336	5.17	12	0.18
Psychiatry-Neuroscience	Clinical	42	164	206	3.90	401	9.55	1	0.30
Psychiatry-Therapeutics	Clinical	37	145	182	3.92	351	9.49	2	0.18
Radiology	Clinical	75	317	392	4.23	456	6.08	12	0.13
School of Nursing	Clinical	76	237	313	3.12	448	5.89	25	0.19
Surgery	Clinical	146	451	597	3.09	1151	7.88	15	0.12
Taub Institute	Basic science	33	227	260	6.88	550	16.67	4	0.26
Average		67	226	292	3.92	565	9.72	8	0.19

centralization and structural integration. The fact that none of the networks had a centralization measure greater than 0.33 suggests that patterns of collaborative authorship were distributed rather than coordinated by a central group of leaders, as would be suggested by a network centralization measure greater than 0.5. Lower centralization is compatible with the paradigm of academic autonomy and the distribution of knowledge and expertise. Additional research is needed to determine whether individual departments, over time, are trending towards higher levels of structural integration. The high-level structural integration of a department indicates how departmental research activities center around a few leaders. The more integrated a department, the more centralized research activities led by a few researchers can be observed.

There are several ways in which knowledge of these differences between departments may be of relevance to discussions of mentoring and academic performance appraisal for promotion. First, they provide evidence to support commonly held assumptions about differences in authorship patterns between departments, which is important to place a researcher's scholarly portfolio within its proper disciplinary context. For example, some departments, such as dermatology and ophthalmology, had low levels of collaboration overall with researchers in other departments.

Patterns of co-authorship frame a researcher's academic track record, and publication lists are a primary area of focus for academic research performance evaluation. To place discussions of academic progression evaluation within their departmental or disciplinary context, committees on promotions and tenure must account for differences between disciplines in publication rates and practices. For example, some departments have low publication rates overall, while others have high numbers of single author papers. Likewise, authors in departments focused on a clinical specialty may prefer to collaborate within their departments, while authors in interdisciplinary research centers may be more likely to collaborate outside of their departments. Discussions of academic performance often rely on collective beliefs or anecdotal evidence about these disciplinary differences rather than on solid data. Therefore, our study demonstrates the feasibility of using co-authorship patterns to reveal the disciplinary differences to guide data-driven academic performance evaluation.

Examining the co-authorship network structure of a given department may also be helpful in identifying cases where given researchers are "at-risk", e.g. because they collaborate exclusively with one specific senior faculty member. Mentors can also use these findings to counsel junior faculty on which departments may have faculty more willing to collaborate externally, or to participate in cross-departmental research.

4.2. Limitations

This study is subject to several limitations. First, like all other association studies, we did not determine the causal relationship between authorship patterns and impact factors of the journals in which the publication appeared. The data analysis did not take into account some confounding factors, such as the number of authors per paper, the number of authors within a department, and the discipline of the leading or senior authors. For example, many basic science journals have high impact factors. If the leading or senior author is from the basic sciences, they are more likely to submit to higher impact journals, which may result in the misinterpretation that "papers with at least one author from a basic science department are significantly more likely to appear in high-impact journals". Moreover, we do not know if publishing in high impact factor journals causes a person to become a professor or if professors are more likely to publish in high impact factor journals. There is also a possibility that inclusion of a professor in the author list is associated with publication in higher impact factor journals because of editorial bias toward publishing known scholars. Most clinicians are only part-time researchers and often

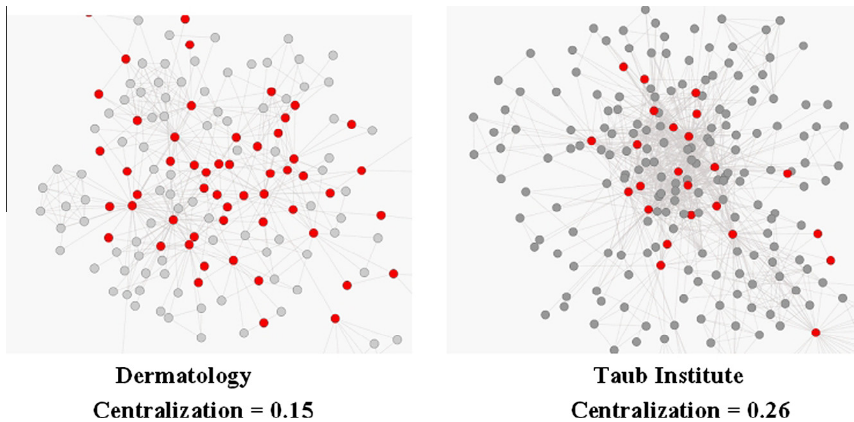

Fig. 4. Largest connected components of two departmental co-authorship networks with different structures: left – Department of Dermatology (centralization = 0.15); and right – Taub Institute (centralization = 0.26).

have little formal training in research methods; most basic scientists are full-time researchers with extensive research training. Questions such as "is the inclusion of a research scientist as a co-author associated with publication in a higher impact factor journal because those individuals have more focused research and writing time, because of their training, or because of some other factor?" warrant further investigation in future studies.

Second, the researcher profiling system used in this study, CUSP, does not retain data on authors who have left the university. Historical data may include publications by authors who are not in CUSP. The inclusion of only current employees may introduce various forms of bias into the analysis, especially if only the current rank was considered. In an effort to isolate this problem we limited our analysis of authorship patterns to a recent five-year period (2007–2011) since the standard promotion time window in our institution ranges from 7 to 11 years.

Third, we did not include students in our co-authorship pattern analysis because CUSP does not have information for students unless they were research assistants in our payroll system. The lack of information about students may change the statistics associated with co-authorship patterns. However, our goal of this paper is to present original analytical methods for combining MEDLINE and scientific profile data to identify author characteristics associated with high-impact journal publications. The data of Columbia University Medical Center were used for demonstration purposes. Others who are interested in applying this method to their institutional scientific profile systems may have the ability to address students or trainees.

Fourth, the author name disambiguation system used in this research, ReCiter, though optimized to maximize accuracy, may have introduced inaccuracies into the results. For example, in some cases an author's publications may have been incorrectly attributed to the wrong person. However, such inaccuracies would have caused false negatives rather than false positives; therefore, it is likely that our statistical significant findings would still hold or be stronger with the identification of missed publications due to the inaccuracies caused by ReCiter.

Fifth, we simply used journal impact factor to represent publication impact, which is not necessarily the best or the most sophisticated method for publication impact calculation. We are aware of related ongoing research such as "Impact Story" (http://impactstory.org), and will consider employing other methods under development to measure publication impact in future research.

Finally, our analysis was limited to co-authors within our institution, CUMC. One of our future research plans is to collaborate with other CTSAs who also have scientific profile systems to replicate this study and include within- and outside-institution co-authors for publication impact analysis. We expect such a large-scale study will require more complex collaborative data collection and cleaning. We hope that the methodologies reported in this study can inspire other CTSAs to join the collaboration.

5. Conclusions

In this demonstration study we analyzed the patterns of collaboration at one academic medical center. We found specific authorship patterns to be statistically significantly associated with publication in high-impact journals. Inclusion of professors, research scientists, and basic science researchers as authors, in particular, were all strongly associated with publication in high-impact journals. Mixing of academic rank, overall, was also associated with publication in high-impact journals; however, some specific author academic rank combinations involving professors of mixed rank were associated with publication in low-impact journals. We also found marked differences between departments in tendency for authors to publish within vs. outside departments, and in overall co-authorship network cohesion.

The results of this research provide original quantitative evidence that might be informative not only for supporting discussions of academic performance appraisal, but also for mentoring junior faculty. Further research is needed to determine whether similar patterns results would occur at other institutions, and to explore patterns of co-author combinations using other author-specific variables. Similar methods might also be used to identify specific departments, which do or do not have a history of collaboration, and to use the resulting data to inform planning of symposia or similar events that may make interdisciplinary initiatives more likely to occur.

Funding

This publication was supported by the National Center for Advancing Translational Sciences through Grant No. UL1 TR000040. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

D.D. assembled input data. M.B., S.B., J.M., S.J, C.W. designed the research methods and experiments. M.B. carried out the analyses and wrote the initial draft of the paper. C.W. made substantial edits to the paper. All authors made substantial contributions to refining the paper and approved the final draft.

Competing interests

The authors declare no competing interests.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2014.07.015.

References

[1] Mabry PL, Olster DH, Morgan GD, Abrams DB. Interdisciplinarity and systems science to improve population health: a view from the NIH Office of Behavioral and Social Sciences Research. Am J Prev Med 2008;35(2 Suppl.). S211–24.
[2] Stokols D, Misra S, Moser RP, Hall KL, Taylor BK. The ecology of team science: understanding contextual influences on transdisciplinary collaboration. Am J Prev Med 2008;35(2 Suppl.):S96–S115.
[3] Kulage KM, Larson EL, Begg MD. Sharing facilities and administrative cost recovery to facilitate interdisciplinary research. Acad Med 2011;86(3):394–401.
[4] Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science 2007;316(5827):1036–9.
[5] Hall KL, Feng AX, Moser RP, Stokols D, Taylor BK. Moving the science of team science forward: collaboration and creativity. Am J Prev Med 2008;35(2 Supplement):S243–9.
[6] Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S Am 2005;102(46):16569.
[7] Lee YJ, Boland MR, Bakken S, Weng C. Ontology alignment for linking a homegrown research networking system to VIVO: findings and implications for standards development. AMIA Jt Summits Transl Sci Proc 2013:99.
[8] Johnson SB, Bales ME, Dine D, Bakken S, Albert PA, Weng C. Automatic generation of investigator bibliographies for institutional research networking systems. J Biomed Inf 2014. 30(March). pii: S1532-0464(14)00080-X.
[9] ISI Journal Citation Reports. Philadelphia, PA: Institute for Scientific Information.
[10] Shiffrin RM, Börner K. Mapping knowledge domains. Proc National Acad Sci U S Am 2004 Apr;101(Suppl. 1):5183–5.
[11] Newman MEJ. Coauthorship networks and patterns of scientific collaboration. Proc National Acad Sci U S Am 2004 Apr;101(Suppl. 1):5200–5.
[12] Bales ME, Kaufman DR, Johnson SB. Evaluation of a prototype search and visualization system for exploring scientific communities. In: Proceedings of the american medical informatics association annual symposium, San Francisco, CA; 2009, p. 24–8.
[13] Schleyer T, Spallek H, Butler BS, Subramanian S, Weiss D, Poythress ML, et al. Requirements for expertise location systems in biomedical science and the Semantic Web. In: Proceedings of "Personal Identification and Collaborations: Knowledge Mediation and Extraction (PICKME)", workshop at the international semantic web conference; 2008.
[14] Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M. Computing topological parameters of biological networks. Bioinformatics 2008;24(2):282–4.
[15] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–504.
[16] Freeman Linton. A set of measures of centrality based upon betweenness. Sociometry 1977;40:35–41.
[17] Woolf SH. The meaning of translational research and why it matters. J Am Med Inf Assoc 2008;299(2):211–3.