Expertise as an aspect of author contributions

Shubhanshu Mishra Brent D. Fegley Jana Diesner Vetle I. Torvik
School of Information Sciences
University of Illinois at Urbana-Champaign
Champaign, IL 61820, USA
smishra8@illinois.edu, bfegley@aptima.com, jdiesner@illinois.edu, vtorvik@illinois.edu

Authors contribute a wide variety of intellectual efforts to a research paper, ranging from initial conceptualization to final analysis and reporting, and many journals today publish the allocated responsibilites and credits with the paper. An overarching yet unreported aspect of these responsibilities is relevant expertise, that is, past experience and knowledge about the phenomenon under study and the context/techniques used to study it. Here, we study author contributions from the perspective of relevant and complementary expertise based on past authorships "conceptual coverage" of the paper at hand. Using concepts from the the MeSH hierarchy assigned to 10.2 million papers in MEDLINE published during 1980-2009, we find that authors collectively cover the great majority of concepts, typically with one dominant author (most often in last postion but frequently 2nd-to-last) and each additional author contributing complementary expertise. For example, 2-author papers fail to cover about 20% of the concepts (i.e, are new to the authors) while 5-author papers fail to cover about 10%, on average. The relative expertise contributions on multi-author papers vary systematically by career stage and author-position, and has changed over time. We also provide an online tool that provides a temporal profile of expertise contributions for any author in the Author-ity 2009 dataset: http://abel.lis.illinois.edu/legolas

Introduction

Authorship credit allocation is a widely studied issue in the field of bibliometric analysis (Merton, 1968; Rennie, Yank, & Emanuel, 1997; Shen & Barabási, 2014; Yank & Rennie, 1999; H. Zuckerman, 1987; H. A. Zuckerman, 1968). The International Committee of Medical Journal Editors (ICMJE) has a specific set of guidelines regarding authorship and credit allocation on papers; indicating the importance of this process (International Committee of Medical Journal Editors, 2018). Bibliometric scholars usually employ methods like self-reported author contributions on journals (Bates, Anic, Marusic, & Marusic, 2004; Rennie et al., 1997; Yank & Rennie, 1999), citation patterns of authors (Shen & Barabási, 2014), or some other heuristic methods (Clement, 2014) for estimating or defining credit per author on a paper. There is wide consensus in the academic community that authorship implies responsibility and significant contribution (International Committee of Medical Journal Editors, 2018; Rennie & Flanagin, 1994). Even though many journals have recently started to require authors to report their contribution level on the article, the resulting data are usually sparse and subject to an author's interpretation of what is meant by "contribution". Further issues with the reliability of this kind of self-reported data (Bates et al., 2004; Rennie & Flanagin, 1994; Yank & Rennie, 1999) are in transparency over the type of contribution, e.g. intellectual, technical or editorial.Assessing author contribution allows for the analysis of mechanisms that drive scholarly collaborations, such as incentive schema (Leimu & Koricheva, 2005). Work in this area has shown that getting full instead of partial credit for a publications may be perceived of greater value to scholars (Katz & Martin, 1997). Other scholars have considered author contribution as a measure of collaboration strength (Newman, 2001, 2004).

Our work is focused on quantifying conceptual expertise based on overlapping concepts of a paper and its author's prior papers. This approach helps quantify complementary expertise added by each author on a given paper. It is related to previous studies on identifying topics in scientific articles and their association with authors (Blei, 2012; Rosen-Zvi, Griffiths, Steyvers, & Smyth, 2004; Steyvers, Smyth, Rosen-Zvi, & Griffiths, 2004).

Methods

Data: The Author-ity 2009 (Torvik & Smalheiser, 2009, 2018) dataset consists of disambiguated author names for MEDLINE articles published through mid 2009. Each article in MEDLINE has been tagged with at-least one Medical Subject Heading (MeSH) term. Between 1980-2009 the Authority MEDLINE subset consists of 10.2M papers authored by 7M authors of which around 90% authors have their first paper after 1980.

Computing author expertise on a paper: For each author their temporal author concept profile is constructed. This profile facilitates computing the number of prior papers of an author on a given concept in every year of their scholarly career. The expertise of an author on a concept is based on the number of prior articles by that author on that concept (we denote this by x). The expertise scores are scaled as $y = log_{10}(x + 1)$, to capture expertise difference based on order of number of prior papers. The expertise of the author for a concept on a given paper is then normalized by the max expertise on that concept on the paper. This is called the weighted expertise of that author. The conceptual expertise by an author on an article is defined as the iterative cumulative weighted measure of their expertise on each topic on an article. The resulting ranking represents the order in which each person has contributed their expertise to an article. The resulting value of conceptual expertise for identifies each author's complimentary expertise on an article after removing all expertise contributed by all the other authors who have a higher value of conceptual expertise. A demonstration of the calculations at: `http://abel.ischool.illinois.edu/legolas/coverage?pmid=15922829`.

Identifying author independence: As an author's career progresses, they change roles, venture into new areas of research, and publish with different collaborators. All of these factors affect the expertise of an author in their future articles. Collaborating with authors from different domains may initially lead to lower expertise for an author on these publications, but also gives them the opportunity to broaden their expertise profile. This evolution eventually leads to higher complimentary or top expertise contributions by the author, in their future papers. On the other hand, sustained collaborations with partners who have higher expertise on an author's areas will keep the author's contribution to a paper continuously low. This can be considered an example of "living under the shadow" of the senior author. The temporal change in conceptual expertise of an author over their career can be operationalized to study the when an author becomes an independent contributor. An author can be characterized by two types of profiles: a) **maximum expertise profile:** the proportion of their articles in which they have the maximum expertise, i.e. they have a conceptual expertise > 0 and are a top expert on that paper, b) **significant expertise profile:** the proportion of their articles in which they have the significant expertise, i.e. they have conceptual expertise > 0.

The career profile of an author is identified via a polynomial logistic model that predicts the ratio of papers per the two above mentioned type of expertise over the author's professional age in years (years since first publication).

Results

Figure 1 shows the collective contribution of all authors to the conceptual coverage, on average. The great majority of concepts are typically covered and the coverage has been going up over time. Also, each additional author adds complementary expertise by several percentage points. This might help explain the widely known phenomenon that the number of authors on papers has been steadily increasing over time.

Figure 2 shows the distribution of the position of the dominant author (the author with the highest coverage). It is not surprising that the last author is most often the one who contributes the most expertise. However, the last author is the dominant in less than 50% of the papers with four or more authors. In other words, it cannot be taken for granted that the last author contributes the most expertise. The second to last author is dominant in more than 20% of the papers, regardless of the number of authors. All the other authors are equally likely to be dominant, today. It is also clear that the role of the first author has dramatically reduced over the short time period studied here.

Figure 3 shows the temporal profile of an author's contributing expertise. The author published their first paper in 1978 and had 77 papers by mid-2009 in MEDLINE. After 8 years, the author is dominant in the majority of their articles, and after 11 years the author contributes significant expertise (dominantly or in complement to another dominant author) in more than 90% of their articles. These trends reflect a sense of independence. Similar profiles of other authors in Author-ity dataset can also be viewed at: `http://abel.ischool.illinois.edu/legolas/profile`

Acknowledgments

Research reported in this publication was supported in part by the National Institute on Aging of the NIH (Award Number P01AG039347) and the Directorate for Education & Human Resources of the NSF (Award Number 1348742). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the NSF.

References

Bates, T., Anic, A., Marusic, M., & Marusic, A. (2004). Authorship criteria and disclosure of contributions: Comparison of 3 general medical journals with different author contribution forms [Journal Article]. *JAMA*, *292*(1), 86-88. Retrieved from `http://dx.doi.org/10.1001/jama.292.1.86` doi: 10.1001/jama.292.1.86

Blei, D. M. (2012). Probabilistic topic models [Journal Article]. *Communications of the ACM*, *55*, 77. doi: 10.1145/2133806.2133826

Clement, T. P. (2014). Authorship matrix: A rational approach to quantify individual contributions and responsibilities in multi-author scientific articles [Journal Article]. *Science and Engineering Ethics*, *20*(2), 345-361. doi: 10.1007/s11948-013-9454-3

International Committee of Medical Journal Editors. (2018). *Defining the role of authors and contrib-*

utors [Web Page]. Retrieved 2018-07-31, from http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? [Journal Article]. *Research Policy*, *26*(1), 1-18. doi: 10.1016/S0048-7333(96)00917-1

Leimu, R., & Koricheva, J. (2005). Does scientific collaboration increase the impact of ecological articles? [Journal Article]. *BioScience*, *55*(5), 438-443. doi: 10.1641/0006-3568(2005)055[0438:dsciti]2.0.co;2

Merton, R. K. (1968). The matthew effect in science: The reward and communication systems of science are considered. [Journal Article]. *Science (New York, N.Y.)*, *159*, 56-63. doi: 10.1126/science.159.3810.56

Newman, M. E. J. (2001). Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality [Journal Article]. *Phys. Rev. E*, *64*, 16132. doi: 10.1103/PhysRevE.64.016132

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. [Journal Article]. *Proceedings of the National Academy of Sciences of the United States of America*, *101 Suppl*, 5200-5. doi: 10.1073/pnas.0307545100

Rennie, D., & Flanagin, A. (1994). Authorship! authorship!: Guests, ghosts, grafters, and the two-sided coin [Journal Article]. *JAMA*, *271*(6), 469-471. doi: 10.1001/jama.1994.03510300075043

Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: A proposal to make contributors accountable [Journal Article]. *JAMA*, *278*(7), 579-585. doi: 10.1001/jama.1997.03550070071041

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of the 20th conference on uncertainty in artificial intelligence* (pp. 487–494). Arlington, Virginia, United States: AUAI Press. Retrieved from http://dl.acm.org/citation.cfm?id=1036843.1036902

Shen, H.-W., & Barabási, A.-L. (2014). Collective credit allocation in science [Journal Article]. *Proc Natl Acad Sci U S A*, *111*, 12325-12330. doi: 10.1073/pnas.1401992111

Steyvers, M., Smyth, P., Rosen-Zvi, M., & Griffiths, T. (2004). Probabilistic author-topic models for information discovery. In *Proceedings of the tenth acm sigkdd international conference on knowledge discovery and data mining* (pp. 306–315). New York, NY, USA: ACM. doi: 10.1145/1014052.1014087

Torvik, V. I., & Smalheiser, N. R. (2009). Author name disambiguation in medline [Journal Article]. *ACM Trans Knowl Discov Data*, *3*(3). doi: 10.1145/1552303.1552304

Torvik, V. I., & Smalheiser, N. R. (2018). *Author-ity 2009 - pubmed author name disambiguated dataset.* University of Illinois at Urbana-Champaign. Retrieved from https://doi.org/10.13012/B2IDB-4222651_V1 doi: 10.13012/B2IDB-4222651_V1

Yank, V., & Rennie, D. (1999). Disclosure of researcher contributions: A study of original research articles in the lancet [Journal Article]. *Annals of Internal Medicine*, *130*(8), 661-670. doi: 10.7326/0003-4819-130-8-199904200-00013

Zuckerman, H. (1987). Citation analysis and the complex problem of intellectual influence [Journal Article]. *Scientometrics*, *12*, 329-338. doi: 10.1007/BF02016675

Zuckerman, H. A. (1968). Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity [Journal Article]. *American Journal of Sociology*, *74*, 276-291. doi: 10.2307/2775535

Figure 1. **Mean proportion of MeSH terms covered in Medline articles.** Articles published between 1980 and 2009, with given number of authors in the byline. Only articles with 2 or more authors are considered.

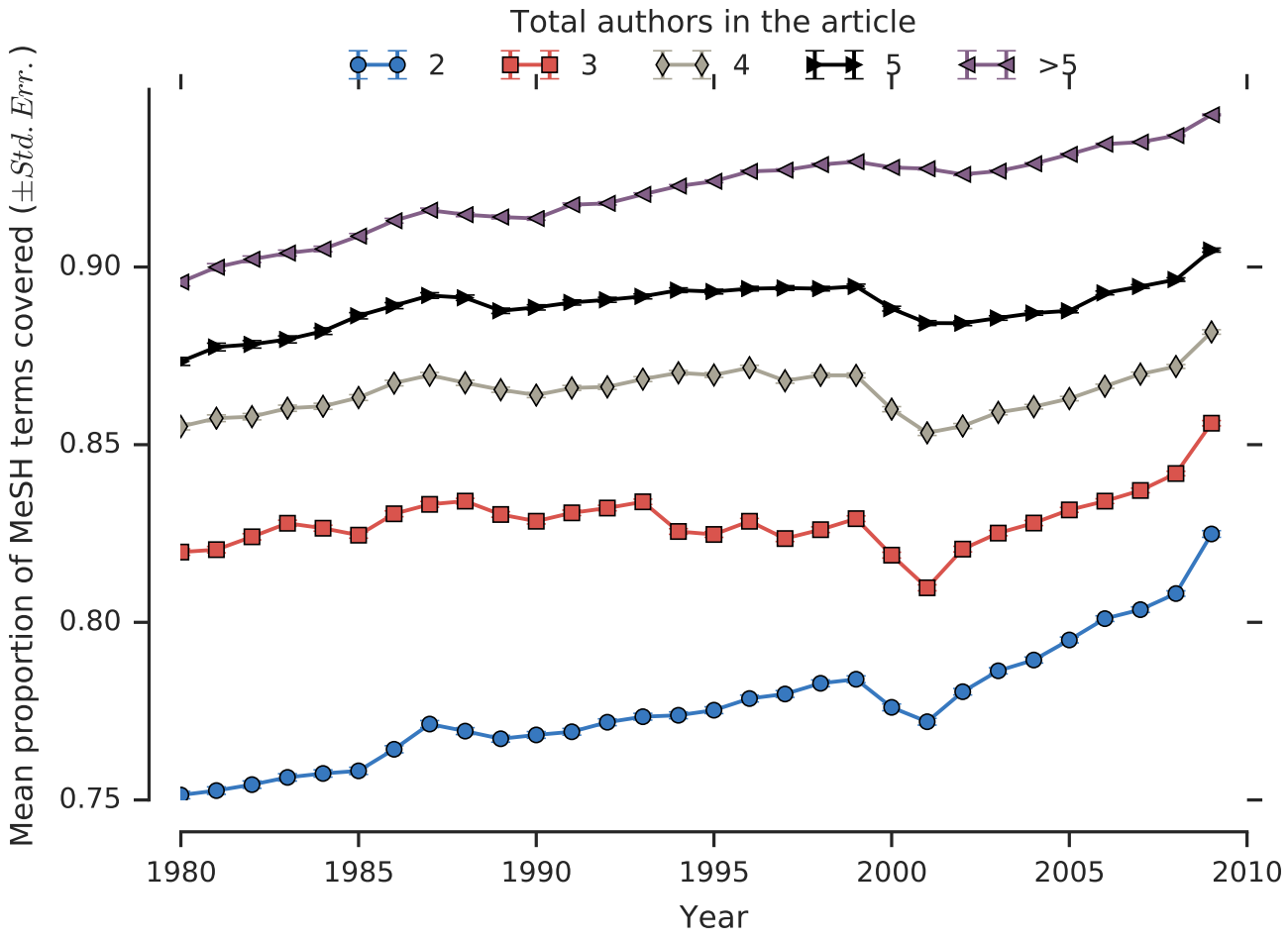

Figure 2. **Proportion of papers with maximum expertise by authors at various position in the bylines.** Papers having 2, 3, 4, 5, 6 and 7 authors.

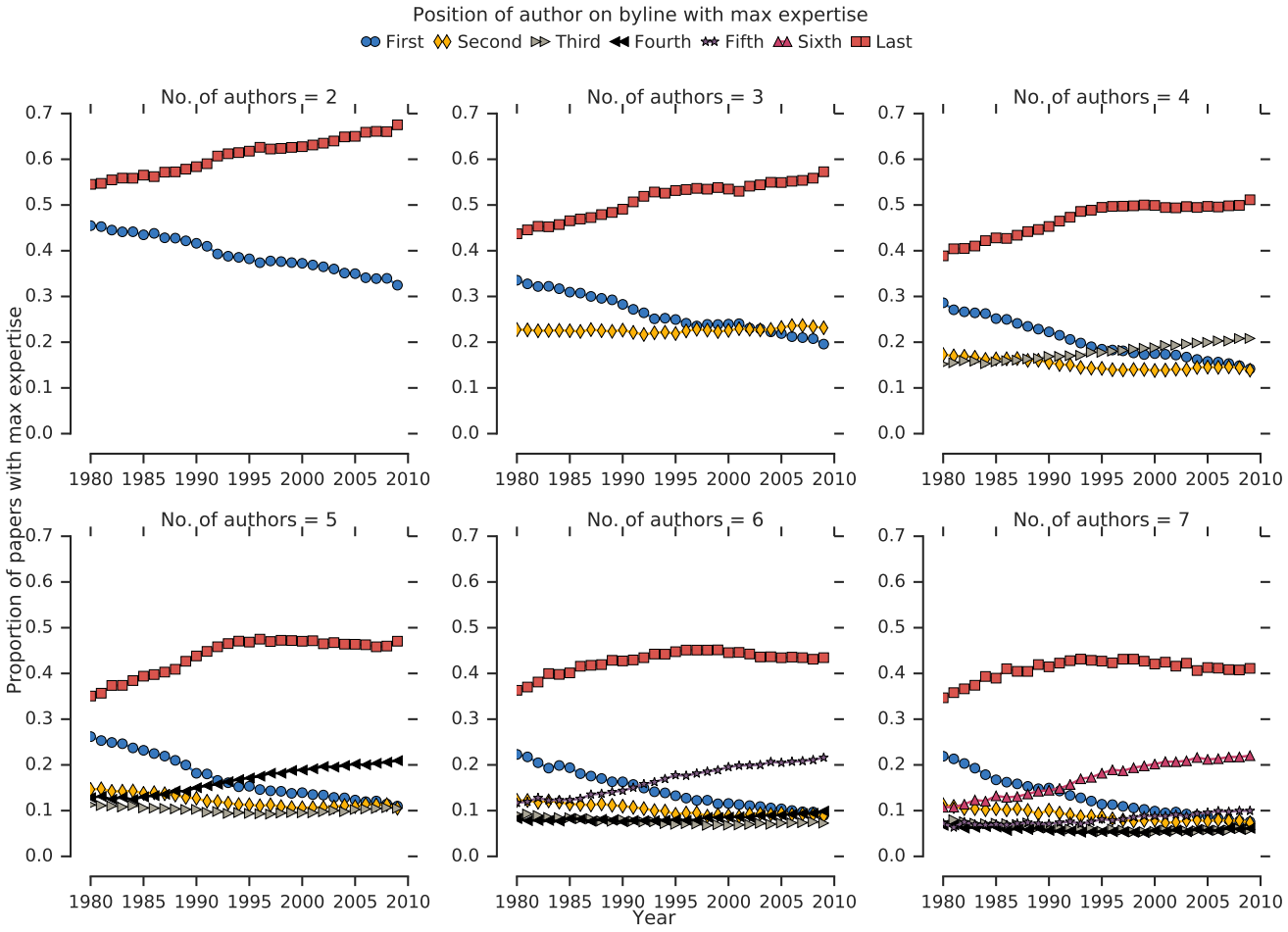

Figure 3. **Career profile of an author in PubMed using the Legolas interface.** The scatter points represent the actual proportion of articles in which the author had maximum expertise (blue) and significant expertise (black). For each type of expertise, the fit of the model profile is shown. More details at: http://abel.ischool.illinois.edu/legolas/profile?auid=207390_1