



Partial alphabetical authorship in medical research: an exploratory analysis¹

Philippe Mongeon^{*}, Elise Smith^{*}, Bruno Joyal^{**} and Vincent Larivière^{*,***}

^{*}philippe.mongeon@umontreal.ca; elise.smith@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, succursale Centre-ville Montréal, QC, H3C 3J7 (Canada)

^{**}bruno.joyal@mail.mcgill.ca

Department of Mathematics and Statistics, McGill University, Burnside Hall, Room 1018, 805 Sherbrooke W. Montreal, QC, H3A 0B9 (Canada)

^{***}vincent.lariviere@umontreal.ca

Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, Montréal (Canada)

INTRODUCTION

Co-authorship has been rising throughout the twentieth Century (Larivière, Sugimoto, Tsou, & Gingras, 2015; Wuchty, Jones, & Uzzi, 2007), to a level that has reached, in certain disciplines, what Cronin (2001) called *hyperauthorship*. With large numbers of scholars contributing to articles comes the challenge of determining the author's order. In medical research, the first(s) and last(s) author positions are attributed to those who made the most important contributions. Typically, the first and last authors are, respectively, junior and senior scholars. Between these two poles are generally listed authors whose contributions are considered less substantial – often lab technicians (Pontille, 2004). In projects involving a large number of researchers, determining the order of authors who made marginal contributions to the research may become particularly difficult. As a result, research teams might choose to order these authors alphabetically, while maintaining a contribution-based order for the others. The trend of partial alphabetical ordering shown by Waltman (2012) seems to support this hypothesis.

However, identifying articles that show traces of such alphabetical order is far from trivial, as one has to distinguish alphabetical order occurring by chance from voluntary alphabetical order (Zuckerman, 1968). Waltman (2012) mitigated this limitation in his analysis the evolution of the frequency of alphabetical ordering, using the probability of chance alphabetical authorship as the margin of error. He also investigated the use of partial alphabetical ordering using an alphabetization score, which is the number of alphabetically ordered consecutive author pairs divided by the total number of consecutive author pairs in the authors list. We take the investigation of partial authorship a step further by identifying and analysing the frequency and the size of subsequences of alphabetically ordered authors. Our study aims to provide answers to the following research questions:

- 1) How prevalent is partial alphabetical order in the medical literature?

¹ This work was supported by the Canada Research Chairs Program and by the Social Sciences and Humanities Research Council of Canada

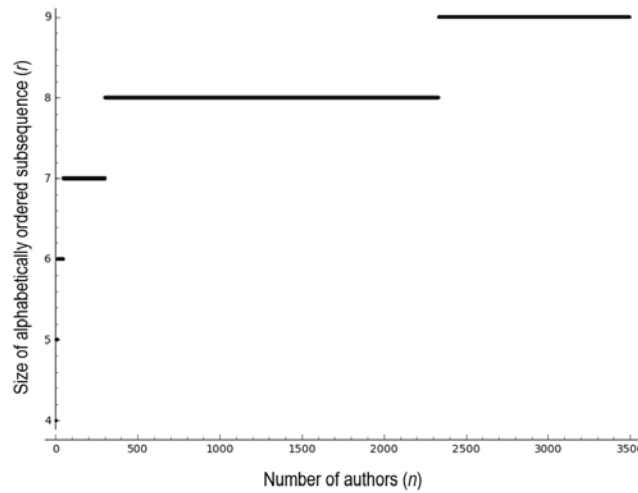
- 2) For articles where an alphabetically ordered subsequence of authors is found, what proportion of authors are listed outside (before or after) that subsequences?
- 3) How does the use of partial alphabetical order (found in questions 1 and 2) evolve over time and as a function of the number of authors per articles?

DATA AND METHODS

Articles in biomedical research and clinical medicine (NSF field classification) with between 6 and 100 authors were retrieved from Thomson Reuters' Web of Science for the 2005-2014 period ($n = 1,967,776$). Like Waltman (2012), we identified in alphabetically ordered pairs of consecutive authors, starting from the first author and moving down the list. An alphabetical subsequence of r authors is formed by consecutive alphabetically ordered pairs. However, the probability of finding an ordered subsequence of r authors by chance increases with the total number (n) of authors. Thus, to mitigate this effect, we calculated a threshold value of r for each value of n , which is the required size of the alphabetical subsequence so that the probability of chance alphabetical order is 5% or lower.

To do so, we define $b_n(r) = n! - s_n(r)$, which is the number of permutations of n authors that do *not* contain an alphabetically ordered subsequence of r authors. One can show by a combinatorial argument (See Elizalde and Noy, 2003) that the exponential generating function

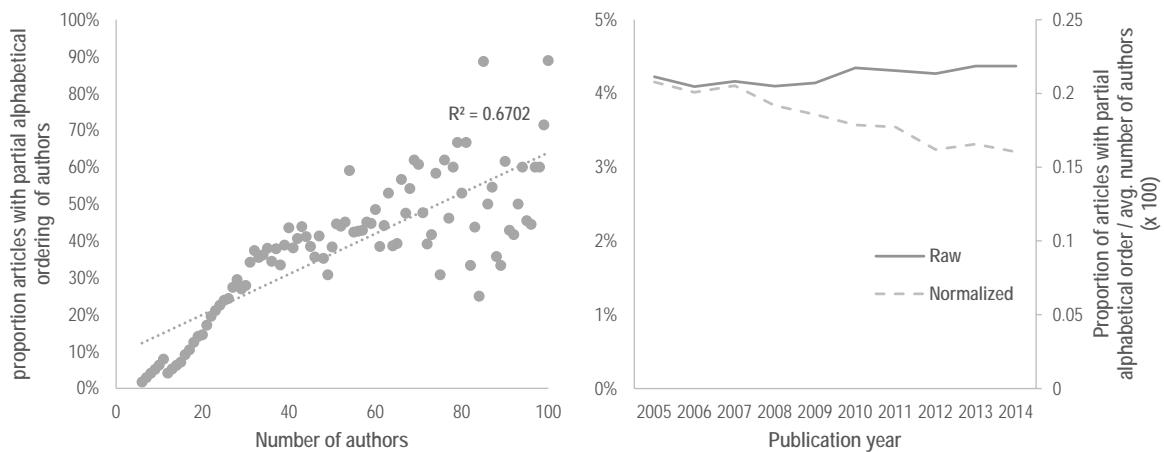
Figure 1: Minimum size r for which the probability of chance alphabetical order is 5% or less as a function of n



RESULTS AND DISCUSSION

Figure 2 (left panel) shows a strong correlation between the number of authors of an article and the frequency of partial alphabetical ordering. While this practice is quite rare for articles with a low number of authors, it becomes more frequent as the number of authors increases.

Figure 2: Prevalence of partial alphabetical ordering of authors, as a function of the number of authors (left panel) and of publication year (right panel)

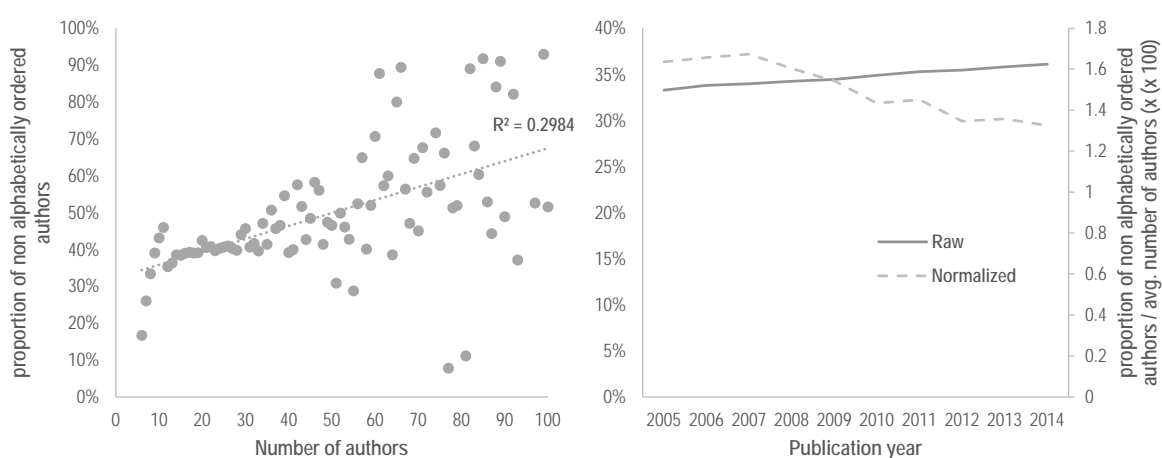


The right panel of Figure 2 shows two opposing trends, depending on the calculation method used. The raw measure indicates that the overall proportion of articles in which partial alphabetical ordering is used tends to grow over time. However, after dividing this proportion by the average number of authors per articles for each year, we see that when the number of authors is fixed, the use of partial alphabetical ordering of authors tends to decrease over time.

For articles where an alphabetically ordered subsequence of authors was found, let us divide the authors in two groups: those listed in alphabetical order—analysed above—and those who are not. While this is an oversimplification of a complex reality, let us also assume that the

authors in the first group are those who made minimal contributions to the work and that the authors in the second group made more substantive contributions. The left panel of figure 3 suggests that the growth in the number of authors is, at least in medical research, mostly due to an increase in substantial contributors. As in figure 2, the right panel of figure 3 shows two opposing trends suggesting that the observed increase in the proportion of substantial contributors over time might be caused by the increasing mean number of authors per articles over time.

Figure 3: Proportion of non-alphabetically ordered authors in articles containing a subsequence of alphabetically ordered authors, as a function of the number of authors (left panel) and of publication year (right panel)



Note: Results are based on the subsample of 80,915 articles for which a *single* alphabetically ordered subsequence was found.

CONCLUSION AND OUTLOOK

These findings provide original insights on the prevalence of partial alphabetical order in medical research and on its evolution over the last decade. We see that as the number of authors per article increases, the prevalence of partial alphabetical order also increases. A plausible explanation is that a high number of authors complexifies their ordering, so only the most important contributors are ordered according to their contribution while the others are ordered alphabetically. Or, since more credit is usually given to the first and last authors, researchers might find that ordering a high number of middle authors by contribution is simply not worth the time.

We also found that a higher proportion of authors are ordered by contribution as the number of authors increases. This suggests that the increase in the average number of authors per article in medical research might be due more to an increase in the number of authors who make substantial contributions rather than an increase of the less substantial contributors being listed as authors. However, our results suggest that when keeping the number of authors fixed, the proportion of authors who made substantial contributions seems to be slightly decreasing over time.

As in Waltman (2012), a limitation of this study is that compound names are concatenated into one string, which might cut some alphabetically ordered subsequences, resulting in a

minor underestimation of the prevalence of partial alphabetical ordering or the subsequences' size. Further developments of this work will address this limitation, and will also provide a broader picture by expanding the period and including other disciplines. This work will provide an enriched understanding of the use of partial alphabetical ordering of authors in science and might have important implications for research evaluation.

REFERENCES

- Elizalde, S., Noy, M. (2003). Consecutive patterns in permutations. *Advances in Applied Mathematics*, 30(1-2), 110-125. Doi:10.1016/S0196-8858(02)00527-4
- Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, 52(7), 558–569. doi:10.1002/asi.1097
- Larivière, V., Sugimoto, C.R., Tsou, A., Gingras, Y. (2015). Team Size Matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7): 1323–1332.
- Pontille, D. (2004). *La signature scientifique □: une sociologie pragmatique de l'attribution*. Paris: CNRS.
- Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, 6(4), 700–711. doi:10.1016/j.joi.2012.07.008
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827), 1036–1039. doi:10.1126/science.1136099
- Zuckerman, H. A. (1968). Patterns of Name Ordering Among Authors of Scientific Papers: A Study of Social Symbolism and Its Ambiguity. *American Journal of Sociology*, 74(3), 276–291. doi:10.2307/2775535