An Experimental Study of the Maximal Number of Authors for a Research Paper*

Antoine Amarilli Arthur Milchior Pablo Rauzy Florian Praden
Ludovic Patey David Naccache Nina Soleymani Antoine Rauzy Assia Tria
Marc Jeanmougin Martin Retaux Matthijs Hogervorst Alice Boilève
Yannick Bonthonneau Jordi Brahamcha Çağri Sert Max Fathi
David Gontier Laetitia Citroen Floriane Dardard Enrique Pujals
Marc Bagnol Mathieu Delorme Silvain Rideau Guillaume Seguin
Fabrice Ben Hamouda Bastien Mallein Stéphane Caron Michael Mathieu
Christophe Rippert Hang Zhou Yoann Bourse Alejandro Kocsard
Aurélien Puiseux Sévan Kazandjian Jacques-Henri Jourdan Claude Perdigou
Elie Gouzien Thibaut Verron Tahina Ramananandro Guillaume Cebron
Valentin Métillon Tristan Tissot Florian Blanc Antoine Fermé Simon Lepot

Florence Clerc Nicolas Daviaud Édouard Thiel Nicolas Normand
Cyril Lavedrine Jean-Paul Delahaye Baptiste Haddou Florent Lachaussée
Mehdi Bouaziz Léa Lejeune Mathis Funk Hervé Vennin-Rendos
Remy Leone Antonin Prunet Marie Mossé Maxime Gheysens
Sarah Dichy-Malherme Noé de Rancourt Pierre Cagne Ismaël Bouya

Denise Maurice Christophe Travert Adeline Pierrot Émilia Robin-Hivert
Louis Jachiet Mikaël Rabie Oscar Blumberg Dominika Lesnicki
Raphaël Rose-Andrieux Vincent Brault Guilhem Gamard Chia-Tche Chang
Tony Jin Olivier Blazy Nicolas M. Thiéry Albane Saintenoy
Bastien Gorissen Arthur Forey Gregor Kolbe Stefan Biedzinski Jack Amiel
Peter Bloomfield Pierre Gaillard Jérome Casse Emilien Joly
Noémie Elgrishi Lingjiong Zhu Charles Ollion
Vianney le Clément de Saint-Marcq Jean-Baptiste Mairy Jean Ollion
Nicolas Boumal Romain Hollanders Wonchan Lee Hakjoo Oh Sham Tlili
Jean-Francois Rupprecht Virgile Andreani Pierre Wolinski Mathieu Renauld
Baudoin Collard Patricia Mirabile Xavier Koegler Blaise Li Mehdi Tibouchi

Émeric Tourniaire Marc Mezzarobba Alin Bostan François Garillot
Nicolas Rolin Olivier Duchenne Florent Hivert Alexandre Benoit
Vivien Ripoll Marie Denizet Bernardo David Abel Lacabanne
Florent Chuffart Nicolas Brisebarre Florent de Dinechin Paolo Ienne
Thomas Gallouet Amaury Pouly Pierre-Marie Pédrot Marthe Bonamy
Tania Louis Fathi Ben Aribi Baptiste Libé–Philippot Ilia Smilga
Vincent Chatalic Céline Ménard Lucie Ducloué Yoann Grimaud
Quentin Delacour Thomas Scaffifi Martin Bodin Anne Fichen
Sophie Mitjavile Paul Mercat Claude Loverdo Thomas Haettel
Alexander Hurst Bruno Le Floch Andrei Smilga Mourad Gouicem
Marco Lazzari Antoine Chatalic Robin Stephenson Ravi Bajaj
Martin Potier Julian Chaidez Reuben Saunders thomas fayolle Jesse Selover
Nick Benson

Abstract

The importance of research papers as the main medium for scientific communication makes it useful to find bounds on important parameters such as the number of authors. Finding the minimal value of this parameter is easy, but finding the maximal number is harder. Theoretical approaches have already yielded upper bounds for this number; in this paper, we provide an experimental lower bound on the maximal number of authors for a research paper.

1 Introduction

Authors are an important aspect of most research papers to date; the structure and layout of major scientific journals such as *Science* or *Nature* show that almost every paper currently published includes an author list. The order of names in the author lists have already been studied by [CSS05]; the object of our study is the number of authors in author lists.

The length of the author list varies from paper to paper. A quick glance at several papers of *arXiv* [Gin91] suggests that most papers fall within the $[1, 100]$ range. However, it is conceivable that exceptional papers such as [$\mathrm{V}^{+}01$] or [$\mathrm{A}^{+}10$] could fall out of this range. We therefore turn to the question of finding an upper and a lower bound for the maximal number of authors of a research paper, which we will denote by ♏.

Finding a lower bound is easy: since ♏ is a nonnegative integer, it is necessarily ≥ 0. This is caused by the fact that papers do not mention the (sometimes very large) list of people who tried to prevent the paper writing process to complete. Moreover, this bound is tight, because it is achieved by [Ano95].

The question of finding an upper bound has received less attention to date. The seminal paper about this question is [Lip10] which suggests several possible bounds. However, these bounds are not proven to be tight. Tight lower bounds on ♏ can be obtained by looking at existing research papers and finding examples with a long list of co-authors; however, it appears that building an specific paper with a large list of co-authors is a more productive approach to reach high values of the author list length and thus provide tight lower bounds for ♏.

Of course, we aren't interested in trivial cases in which we make up fictious names for co-authors, or use the name of existing people without ensuring that they are willing to co-sign. The author list of our paper should only contain names associated to exactly one (living or dead) human being who effectively accepted to co-sign the paper.

2 Experimental protocol

Remember that our goal is to build a paper $\wp$ with a list of co-authors which is as long as possible. The process used in the construction of $\wp$ is as follows: we navigate in the real world (section 2.1) to find co-authorship candidates (CACs); we use social interaction (section 2.2) to convince the CACs to co-sign $\wp$; having repeated this process a large number of times, we compile (section 2.3) the list of co-authors and produce the final version of $\wp$ which, by construction, has a large number of co-authors.

2.1 CACs localization process

The main algorithm for locating CACs is a breadth-first traversal of the social graph using the main authors as origins, restricting ourselves to individuals who are suitable CACs with regards to professional orientation, familiarity with paper publication processes and sense of humor. The use of a breadth-first research allows us to ask co-authors to collect signatures from their friends; moreover, it lends itself to a straightforward parallelization scheme using distributed algorithm theory.

As a first heuristic, we can expect high concentrations of CACs to be located around major universities and research institutes. It is therefore worthwile to induce modifications of the social graph through a physical relocation process of CAC detectors in the geographical euclidean space. The most effective method to achieve this with regards to the relocation amplitude (ie. the movement distance in the euclidean space, or "real world") is the procedure of *international professional travel* (IPT), which can be implemented in several ways: university exchanges, internships, and conferences.

*This work has been partly supported and sponsored by the Club Inutile of the École normale supérieure, Paris.

2.2 Social handling of CACs

When a CAC is encountered, a social exchange channel is initiated through the use of the conventional message start marker "Hi!", as observed in the population of English-speaking human beings. Naturally, the process is adapted for non-English-speaking CACs through translation to a human language which the CAC is able to comprehend. Upon receiving a suitable reply from the CAC, the request is formulated using suitable register of language and verbal modality ("Would you like...") followed by a concise description of the requested task ("co-sign a paper"). Errors encountered before this stage are usually fatal (see transcript 1).

The high unexpectedness level of such a request usually triggers a questionning round from the CAC in order to rule out transmission errors and to obtain a understandable description of the experimentator's motives. As an answer, we provide the CAC with a description of our undertaking, including, if necessary, a draft preview of $\wp$. Possible failure causes at this point are an unsufficient familiarity with the paper publication process to understand the request (see transcript 2) or a refusal prompted by an insufficient sense of humor or faulty degree-detection algorithm from the CAC (see transcript 3). This reflects an error in the process used at the previous step to weed the social graph; it also reveals that not all human beings can sign a research paper, which means that the value of ♏ is notably inferior to the total existing human population count and that the bounds proven in [Lip10] can be drastically improved (ie. they aren't tight at all).

When no errors occur, the CAC is willing to co-sign and can be added as a co-author of $\wp$, thus incrementing the value of our lower bound for ♏.

2.3 Final compilation

TODO: Keep this in sync with the rest.

The author list of the final version of $\wp$ is built by arranging in alphabetical order the names of the CACs who finally accepted to co-sign the paper.

We also added the names "Antoine Amarilli", "Arthur Milchior" and "Pablo Rauzy" at the beginning of the author list. These three authors can be added "for free": because of their involvment in carrying out the experiment and compiling $\wp$, they can be considered as co-authors and add 3 additional names to the author list of $\wp$.

Several practical difficulties arise when we try to create the final version of $\wp$: because we want $\wp$ to qualify as a valid research paper, we must provide a few items, such as content. To pad $\wp$ to a reasonable length, we ...

TODO: Add a description of the whole paper once completed.

3 Results

TODO: Keep this in sync with the rest: final number of co-authors achieved, and publishing process.

The final version of $\wp$ has been published as [AMR$^+$11].

4 Possible extensions

The CAC localization procedure described here isn't limited to the number of co-authors we achieved, but scales perfectly to larger numbers. It can be conjectured that running this process with more time or more funding could result in a larger lower bound for ♏.

The localization procedure also depends on several factors which could be optimized, such as the researcher's social skills and general ability to interact with other human beings in a suitable manner. This point is still the subject of active research.

Another possibility for future efforts would be to enforce tighter constraints on the notion of "paper" or of "co-author". For instance, we could require co-authors to actually contribute something to the final paper in addition to their names, or, at least, to have some involvment with the paper rather than just signing it. (However, it should be remembered that such a practice is quite commonplace when doing actual research.)

Notice also that [AMR$^+$11] has an important effect on the graph ♄ induced by the co-signature relation over scientific papers and on the particular example of Erdős numbers which has been particularly well studied by recent research efforts. More specifically, it can be observed that the Erdős number of all authors of [AMR$^+$11] is now...

TODO: Keep this in sync with the Erdős number.

[AMR+11] therefore serves as a kind of “hub” in ち: it can be conjectured that research efforts such as ours significantly improve the connectivity of ち.

5 Conclusion

TODO: Write conclusion last.

References

[A+10] G Aad et al. Charged-particle multiplicities in *pp* interactions at $\sqrt{s} = 900$ gev measured with the atlas detector at the lhc. oai:cds.cern.ch:1249427. *Phys. Lett. B*, 688(arXiv:1003.3124. CERN-PH-EP-2010-004. 1):21–42. 40 p, Apr 2010. Comments: 40 pages (pages 24-40 appendix with ATLAS collaboration author list), 4 figures.

[AMR+11] A. Amarilli, A. Milchior, P. Rauzy, et al. An experimental study of the maximal number of authors for a research paper. *TODO: journal.*, 2011.

[Ano95] Anonymous. Regulatory status of methyl bromide and priority review of methyl bromide alternatives. *Pesticide Regulation (PT)*, Notice 95(4):1–4, 1995.

[CSS05] J. G. Cham, M. A. Slackenerny, and B. S. Smith. Author list: A context-changing model of infering nano-informatic structures. *Intl. Journal of Temporal Deflective Behavior*, 2(562):138, 2005.

[Gin91] Paul Ginsparg. arxiv, 1991. http://arxiv.org/.

[Lip10] D. Lipton. Limits on the number of co-authors? *Gödel’s Lost Letter and P=NP*, 2010. `http://rjlipton.wordpress.com/2010/04/28/limits-on-the-number-of-co-authors/`.

[V+01] J. C. Venter et al. The sequence of the human genome. *Science*, 291(5507):1304–1351, 2001.

Experimentator: Hi!
CAC: Go away!

Transcript 1: Unsuccessful social interaction with a CAC: behavioral incompatibility between experimentator and CAC.

Experimentator: Hi!
CAC: Hi!
Experimentator: Would you like to co-sign a paper?
CAC: A *what*?
Experimentator: A paper.
CAC: What's that? Like, blank paper?
Experimentator: You know, a, uh... Oh, never mind.

Transcript 2: Unsuccessful social interaction with a CAC: the CAC was unfamiliar with the paper publication process.

Experimentator: Hi!
CAC: Hi!
Experimentator: Would you like to co-sign a paper?
CAC: What do you mean?
Experimentator: We are writing a paper about the maximal possible number of co-authors for a research paper. Would you like to sign?
CAC: Seriously? That's *so* stupid!
Experimentator: Do you want to sign?
CAC: No way!

Transcript 3: Unsuccessful social interaction with a CAC: the CAC does not implement a sense of humor suitable for this request.

Experimentator: Hi!
CAC: Hi!
Experimentator: Would you like to co-sign a paper?
CAC: What do you mean?
Experimentator: We are writing a paper about the maximal possible number of co-authors for a research paper. Would you like to sign?
CAC: Seriously?
Experimentator: Yep. So, do you want to sign?
CAC: Yeah! Sure!
Experimentator: OK. Thanks!

Transcript 4: Successful social interaction with a CAC.