Et Al(ex): Examining the impact of Alexs on the field of computer science

Alex {Bellon, Liu, Redding, Snoeren, Yen}
UC San Diego

***Abstract*—Apropos of nothing and without any sort of personal bias, we collectively investigate the impact that Alexs have had on the field of computer science. We propose the *Alex Factor* (A_{fac}), *Alex Concentration* (A_{conc}), and *Alex Index* (A_{ind}), measurements of how Alex-y a publication is. Additionally, this work aims to break the current record *Alex Index* in a computer science conference/workshop publication[1], with an author list composed of only 5 Alexs.**

I. Introduction

Science (a couple of the authors) has long (since about 2 months ago) wondered just how much impact Alexs have had on the field of computer science. Given the popularity of the name and our own experiences in a department with at least 10 Alexs, we scientifically hypothesized (guessed with no supporting evidence) that the resulting contributions must be nontrivial. There has been previous work investigating characteristics of author names, but they have mainly focused on similar surnames[2;3;4] or having exactly the same first and surnames[5;6], and were mostly focused on economics for some reason. Additionally, none of these works have focused exclusively on the name Alex (a large oversight, in our opinion).

In addition to being a popular name both in general and within computer science publications (see Section III), the name "Alex" (and its common root name, "Alexander") benefits from having dozens of derivative, diminutive and shortened forms. This bolsters the number of authors that we can claim to be "Alex"s.

We therefore provide a first analysis of computer science authorship by Alexs, in addition to comparisons against authorship of other names. We specifically propose the **Alex Factor** (A_{fac}), **Alex Concentration** (A_{conc}) and **Alex Index** (A_{ind}) as ways to compare the Alex-ness of different publications. Finally, this paper sets the record for the highest A_{ind} (5.0) in a CS conference/workshop paper.

II. Methodology

We use the DBLP[1] dataset to conduct our investigation, using only entries published in conferences and workshops (excluding journals, book chapters, reference works, etc.). We perform analysis on names as they are entered in DBLP, which means that many of our results are underestimations, as a non-trivial amount of publication entries use only the first letter of first names to identify authors rather than their entire first name. Overall, we draw from a dataset of 3402082 publications, and we restrict our analyses to works published between 1950 and 2023.

[1]There was a journal publication with 9 Alexs :/

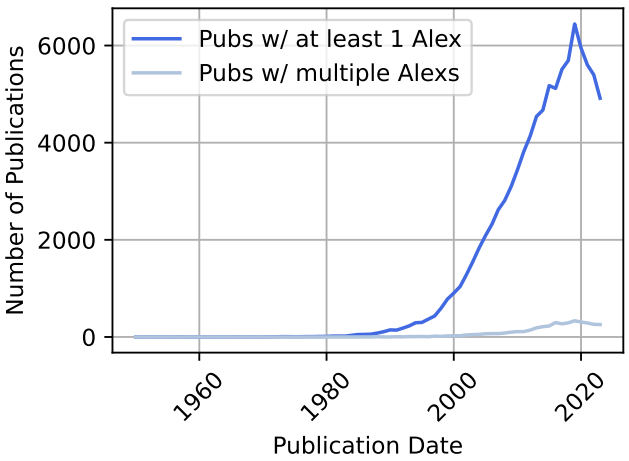


Fig. 1: Number of publications with one or more Alex authors over time

A. Alex Metrics

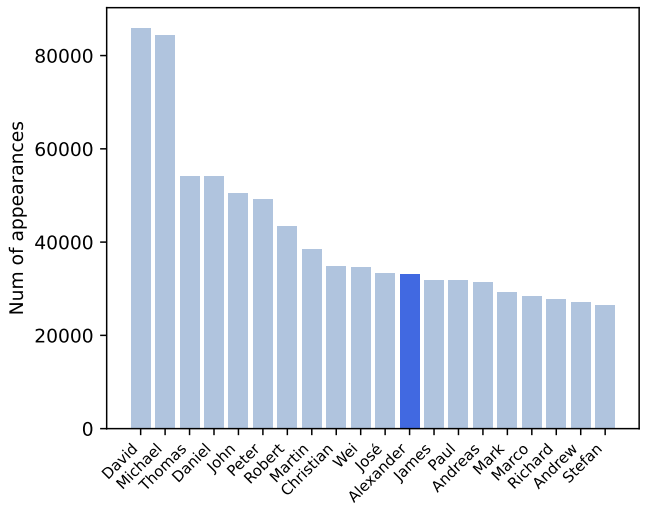
In addition to performing general analyses of author first names in computer science publications, we also propose and calculate 3 metrics: *Alex Factor* (A_{fac}), *Alex Concentration* (A_{conc}), and *Alex Index* (A_{ind}). For all three factors, we include the following names to all count as "Alex" authors, notated by Alex$'$:

- Alex
- Alec
- Aleks
- Alexei
- Alexey
- Alexia
- Alexis
- Alessio
- Alessia
- Alejandro
- Alejandra
- Aleksander
- Aleksandar
- Aleksandr
- Alessandro
- Alessandra
- Alexander
- Alexandre
- Alexandru
- Alexandra
- Alexandria

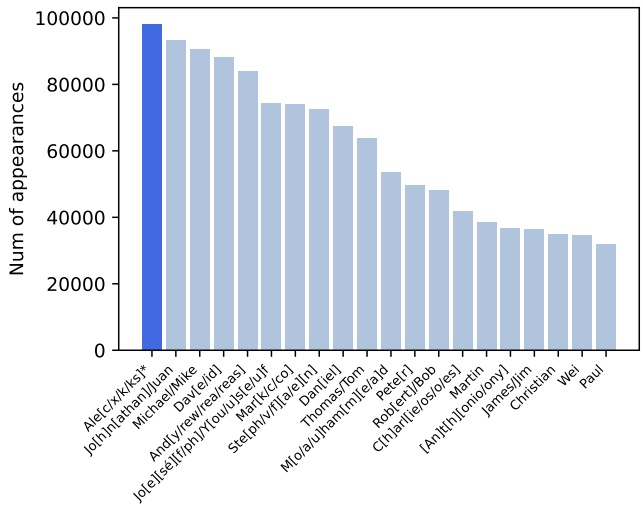
The **Alex Factor** of a paper is calculated via the following formula

$$A_{\text{fac}} = \sum_{i=1}^{n} \text{Alex}'_i$$

where n is the number of authors in the author list. The **Alex Concentration** of a paper is calculated via the following formula



(a) Non-grouped



(b) Grouped

Fig. 2: Distribution of first names in computer science papers, with and without being grouped by names from the same root

$$A_{\text{conc}} = \frac{A_{\text{fac}}}{n}$$

Finally, the **Alex Index** of a paper can be calculated with

$$A_{\text{ind}} = A_{\text{fac}} \cdot A_{\text{conc}}$$

We use these metrics to compare all the publications in our dataset to find the most "Alex"-y publications. There is still future work needed to determine whether similar metrics could be calculated for names other than Alex. Given thousands of first names exist in the world, we believe that this will be a monumental task best left to others not scrambling to submit to SIGBOVIK before the deadline.

III. Results

We calculate some general statistics about computer science paper authorship, then focus specifically on Alexs, including comparing papers by our three metrics: A_{ind}, A_{fac}, and A_{conc}.

A. General author name patterns

We plot the distributions of occurrences of different first names in Figure 2. We plot both by raw first names as they appear in DBLP, and after grouping first names by the same root name. We find that for raw names (Figure 2a), computer science authorship is dominated by Davids and Michaels, with those two names being over 20,000 publications ahead of the next most common name. Unfortunately, the highest ranked "Alex" name, "Alexander" only places 12th by number of appearances.

Accordingly, we perform a secondary analysis in order to give ourselves a more impressive ranking. In Figure 2b, we combine names based on shared root name, in order to get a more general sense of authorship name patterns. The John-derived names jump up to first place from fifth in the raw ranking, knocking down Michael- and David-derived names to third and fourth respectively. Fortunately, the Alexs have a much better showing after including all Alex variants, taking first place with almost 100,000 appearances. Additionally overall, the number of publications does not drop off as steeply after the top names.

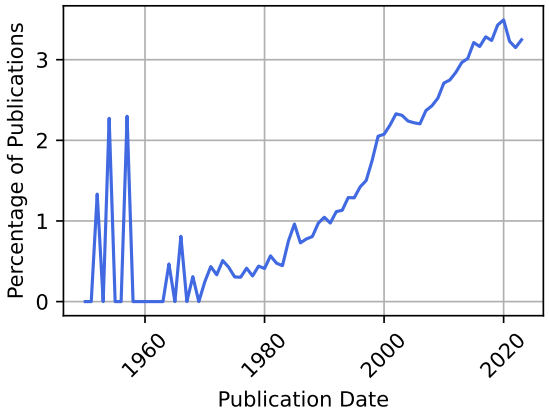


Fig. 3: Percentage of Alex publications over time

Another interesting observation is that based on the difference in rankings between Figure 2a and Figure 2b, it is more common to see longer versions of names (e.g. "Alexander" and "Michael") as opposed to shorter nicknames (e.g. "Alex" and "Mike"). This is likely due to the fact that longer names seem more fancy and official[2].

B. Alex patterns

Within the space of Alex-related names, we find multiple interesting patterns. We first investigate one of our most pressing questions: how much computer science research is produced by Alexs? In Figure 3, we plot the percentage of

[2]Source: idk, seems true

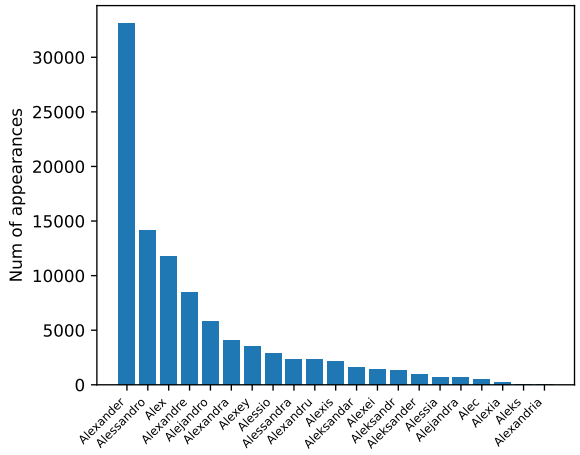


Fig. 4: Distribution of Alex variants

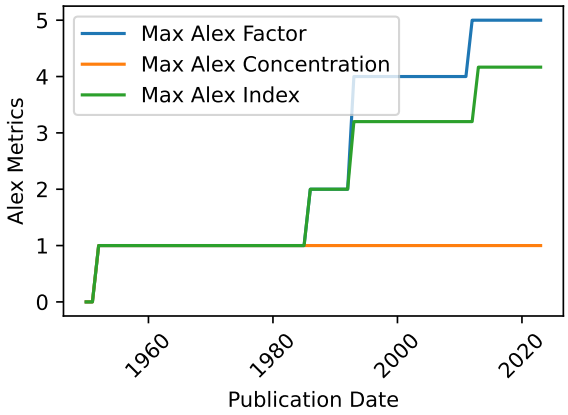


Fig. 5: Maximum Alex Factor (A_{fac}) over time

publications each year that were authored by at least one Alex. We find that aside from some large spikes before 1980 (when the overall amount of computer science papers was already low compared to today), the percentage of Alex-based publications has been steadily growing since the 1970s. Unfortunately, it seems that the Alex Apex was reached in 2020, and has been declining ever since. We hope to make a contribution to reversing this pattern with the publication of this work.

We next investigate which variants of "Alex" appear most often (Figure 4). As expected from our general analysis, "Alexander" is the most popular, followed distantly by "Alessandro" and "Alex". We see that "x"-based "Alexander" variants (Alexander, Alexandre, Alexandra) are more popular than "k"-based variants (Aleksandar, Aleksandr, Aleksander), and there are almost double the number of "Alexey"s as "Alexei"s. These are of no importance to research output or contribution, we just thought the figure looked cool :).

C. Alex metrics

We finally calculate our three metrics, A_{fac}, A_{conc}, and A_{ind} for computer science publications as a whole. The changes in maximum values of each of these metrics over time can be found in Figure 5.

The maximum A_{fac} quickly rose to 1 early in computer science research, in the 1950s. It stayed stagnant for over 3 decades (which honestly we find pretty surprising), with the first computer science conference paper featuring 2 Alexs being published in the late 1980s. From there, competition quickly increased, with the maximum A_{fac} soaring up to 4 before the millennium. Finally, in the 2010s, we reached the current record of 5, which we tie with this work. We do note that we are the first work to have 5 authors with the same variant of "Alex", as the other two papers included "Alexander"s, "Alexey"s, etc. .

For A_{conc}, the highest possible record for maximum Alex Concentration was set early on in 1952, with a single-author Alex publication. There is still future work to be done to determine if it is possible to have more Alexs than authors in order to achieve an $A_{\text{conc}} > 1.0$.

We finally come to A_{ind}, which is definitely not a measurement carefully crafted such that this work would win. A_{ind} rose in line with A_{fac} until the 1993, where it stagnated for a few decades at 3.2. It continued to rise with A_{fac} in 2013, where it still stands at $4.16\overline{6}$. It is with this work, over a decade later, that we now push the record A_{ind} to 5.0.

IV. Conclusion

Overall, we find that while Alexs may not have the most number of publications (until grouped with all Alex-variants), maybe the real number of publications are the friends we made along the way.

References

[1] Dagstuhl, S. Dblp: computer science bibliography, 2023.

[2] Einav, L., and Yariv, L. What's in a surname? the effects of surname initials on academic success. *Journal of Economic Perspectives 20*, 1 (2006), 175–188.

[3] Goodman, A. C., Goodman, J., Goodman, L., and Goodman, S. A few goodmen: Surname-sharing economist coauthors. *Economic Inquiry 53*, 2 (2015), 1392–1395.

[4] Gray, P. G. Initial letters of surnames. *Journal of the Royal Statistical Society Series C: Applied Statistics 7*, 1 (1958), 58–59.

[5] Otto, P., and Otto, P. Impact of academic authorship characteristics on article citations. *Revstat Statistical Journal 20*, 4 (2022), 427–447.

[6] Wirth, V., and Wirth, V. Author-unification: Name-, institution-, and career-sharing co-authors. *arXiv preprint arXiv:2303.17328* (2023).