Presentation and Success in Computer Science Research

William Howard-Snyder
Department of Computer Science
University of Washington
Seattle, WA 98102
howarwil@cs.washington.edu

Eden Chmielewski
Department of Computer Science
University of Washington
Seattle, WA 98102
edenski@cs.washington.edu

Anjali Agarwal
Department of Computer Science
University of Washington
Seattle, WA 98102
anjalia@cs.washington.edu

Matthew Jack
Department of Computer Science
University of Washington
Seattle, WA 98102
matvj250@cs.washington.edu

Qirui Wang
Department of Computer Science
University of Washington
Seattle, WA 98102
qw43@cs.washington.edu

Abstract

We aim to better understand the relationship between the presentation of computer science (CS) research papers and their success. This is an important problem to study because the way in which research is presented is largely within an author's control, and making explicit the factors which contribute to a paper's success can help researchers, especially young ones, communicate their results more effectively to the scientific community. Analyzing this problem, however, involves many challenges, such as storing and analyzing data on millions of research papers, operationalizing paper success, and disentangling the effect of presentational factors from non-presentational factors. Previous work has analyzed the relationship between presentational factors and citation count, but they focus on linear relationships and Library Information Science papers, whereas we consider more complex relationships and focus on CS papers. For many of the presentational factors that we considered, there was an optimal range where the papers that were within that range had significantly higher citations than those outside of it.

1 Introduction and Motivation

Writing research papers is one of the primary ways that researchers communicate their ideas to the rest of the scientific community. Papers offer a compact, familiar format to digest results, learn new methods, and review related work. Because of their important role in scientific communication, one's research papers are often a deciding factor in whether one is accepted for an academic position, such as becoming a professor. Despite their importance for career opportunities, there are many *external factors* outside of a researcher's control that affect the number of papers they publish and the influence of their work. For example, the size of their research field, the reputation of their affiliated

Preprint. Under review.

research institutes, the popularity of their research topic, and whether Reviewer #3 had their morning coffee before reading their paper.

In this work, we examine *presentational factors*, which are strictly within a paper writer's control and have to do with how the paper is presented. These include writing style and composition. We aim to understand how these factors vary with paper success. Our precise research question is: *What is the relationship between presentational factors of CS paper abstracts and their success?* We tackle this question with a data-driven approach by analyzing over 5 million CS research paper abstracts from Semantic Scholar (Section 2). We use these results to derive writing tips for researchers so they can maximize the success of their work.

One of the main challenges in answering this question is disentangling the effects of the presentational factors from non-presentational factors. Such non-presentational factors include external factors, such as popularity of the co-authors or research institution, and non-presentational internal factors like research area, experimental design, and significance of results. With respect to presentational factors, the non-presentational factors appear as variance in the data, complicating the relationship between success and presentation. Moreover, researchers will often spend weeks crafting the perfect prose for their papers and go through many drafts to maximize the presentational quality of their work. With relatively small differences in quality of writing between papers, non-presentational factors like significance of results will likely have a larger effect on a paper's success than presentational factors. Nevertheless, it is important to characterize and make explicit this relationship so that researchers, especially those with less paper-writing experience, can be aware of the implicit presentational criteria that will affect their future career opportunities.

Previous works have investigated how presentational factors influence the success of academic papers [6, 5]. However, these studies analyze a much smaller number of research papers (~200 papers) and focus on a different field. Other work has made an effort to quantify the success of academic research papers from their citation graph. For example, Chen et. al (2007) used Google PageRank to measure influence [1] and Giatsidis et al. (2019) used graph density-related metrics [2]. These works try to account for the fact that papers with low citation counts are not necessarily less successful than those with high citation counts. However, it is not clear what assumptions are made about the data with these metrics, nor how they will affect the results. For this reason, we opt for a measure of paper success that is derived from raw citation count.

We identify two main classes of presentational factors. Namely, *readabilty-based factors* (Section 3.2), which account for the complexity of the syntax and vocabulary used in a paper's abstract, and *content-based factors* (3.3), which account for the inclusion of specific types of content, specifically positive and argumentative language. We find that there is a parabolic relationship between each of the presentational factors that we considered and paper success (Section 4). From these results, we derive principled writing tips and discuss what factors might cause these parabolic relationships (Section 5). Finally, we consider the negative ethical implications of our work and present ways to mitigate our negative impact (Section 7).

2 Dataset Description

This work utilizes the Semantic Scholar Academic Graph dataset, which contains information on over 214 million academic research papers, including titles, abstracts, year published, and fields of study. Our objective was to examine CS papers, which prompted us filter by papers that were tagged with the "Computer Science" field of study attribute. The resulting dataset contained 15.5 million papers.

To ensure the quality and completeness of our analysis, we performed several additional data filtration steps. First, we excluded papers that did not have open-access abstracts, as abstract data was crucial to analyzing our presentational factors. Additionally, to evaluate paper success, we employed a 5-year citation count metric (refer to Section 3.1). Using this metric required us to consider papers published prior to 2018, prompting another filtration step.

A final critical step involved pruning documents that were disconnected from the academic graph, i.e., those neither citing any CS papers nor being cited by any CS papers. Upon manual inspection we discovered that these documents generally fell into two categories: either they were mistakenly labelled as CS documents despite belonging to non-CS disciplines, or they were non-paper entities such as patent documents or books. After this pruning process, our refined dataset was comprised of 5.7 million CS papers, accounting for 37% percent of our original CS paper dataset.

In addition to the Semantic Scholar dataset, we acquired a set of influential papers to validate any findings we presented about influence. To compile this dataset, we scraped information from two online sources: the Best Paper Awards in Computer Science collection compiled by Professor Jeff Huang, and a curated list of important publications in Computer Science from Wikipedia (Section 9).

3 Analytical Approach

At a high-level, our analysis involved operationalizing success and several presentational factors, computing these for each of the 5.7 million abstracts, and performing statistical tests to evaluate our hypothesized relationships between paper success and each factor. In the following sections, we describe in detail how we operationalized these constructs and identify the challenges that each presented.

3.1 Operationalization of "Success"

We operationalized the success of a CS paper as the number of citations it accrued from other CS papers within five years of publication, which we refer to as the *5-year citation count*. Previous methods predominantly relied on raw citation counts as a measure of success [6]. However, the use of 5-year citation count addressed the potential bias against more recent publications, ensuring that they were not disadvantaged by having fewer years to accumulate citations compared to older papers.

While any n-year cutoff could have addressed this bias (e.g., 3-year or 10-year), our choice of $n = 5$ aims to strike a balance between accurately representing citations per paper and analyzing a large paper dataset. A higher n captures more citations for a particular paper, but also necessitates limiting our analysis to papers predating $2023 - n$. Setting $n = 5$ enabled us to include all papers published prior to 2018 in our analysis, representing 75.5% percent of the original dataset. Moreover, this choice aligns with the tendency for a paper's growth in citation count to decelerate after its initial five-years span (Figure 1(a)). The claim that 5-year citation count effectively captures a paper's citations is further substantiated by a Pearson correlation coefficient of 0.74 between 5-year citation count and raw citation count.

To evaluate the validity of this success metric, we compared its performance on a curated list of influential CS papers from Wikipedia and the broader dataset of CS papers. Our investigation revealed that papers in the Wikipedia dataset had a median 5-year citation count of 45, whereas the broader dataset had a median count of 0 (Figure 1(b)). This indicates that the chosen metric effectively distinguishes successful papers.

3.2 Readability

We operationalized readability as the Gunning FOG Index (FOGI), which is a reading comprehension formula used in linguistics and education research [3]. The formula estimates the number of years of formal education required to comprehend a particular piece of text as follows:

$$\text{FOGI} = 0.4 \left(\left(\frac{\text{\# words}}{\text{\# sentences}} \right) + 100 \cdot \left(\frac{\text{\# complex words}}{\text{\# sentences}} \right) \right) \quad (1)$$

where a word is considered complex if it has more than two syllables. As the FOGI of a piece of text increases, that text will have longer sentences with more complex words. This generally corresponds to less readable text, so FOGI and readability are inversely related. Of course, FOGI is not a perfect measure of readability. For instance, the simple sentence "These bananas are interesting." has a FOGI of $0.4\,(4/1 + 100 \cdot 2/1) = 81.6$. To evaluate the construct validity of FOGI as a measure of readability, we measure FOGI over time to see if it captures the pre-established trend that readability of academic abstracts is decreasing over time [4].

We make the following hypothesis about the relationship between readability and success:

Hypothesis 1: A paper's success *increases* with readability (i.e., decreases with FOGI).

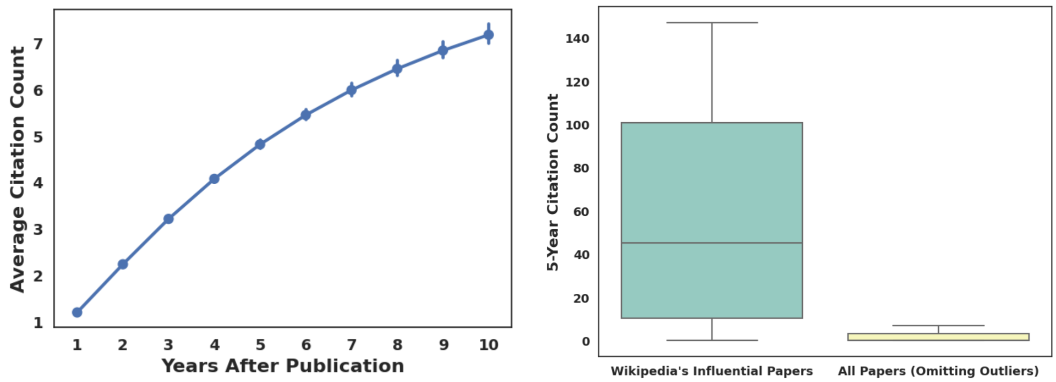

Figure 1: (a) Pointplot of the number of years after publication versus average citation count. Bars at each point are generated by the bootstrapped 95% confidence interval. (b) Boxplots of 5-year citation count of papers in Wikipedia's influential paper dataset, against the broader 5.7 million paper dataset. (Outliers are excluded from the boxplots. Outliers are calculated as datapoints that fall above the third quartile plus 1.5 times the inner quartile range.)

We make this hypothesis on the basis that papers with high readability (or low FOGI) are more accessible to a wider audience. The more accessible a paper is, the more likely it is to be read, and the more people that read a paper, the more times that paper gets cited.

3.3 Content-Based Presentational Factors

The two content-based presentational factors we investigate in this paper are positive and argumentative language. For both presentational factors, we used a combination of ChatGPT query results and manual inspection of a random subset of abstracts to create a set of words or short phrases that reflect the desired sentiment. Both sets are included in the appendix (Section 10).

For positive language, we gave ChatGPT the following prompt: "Expand the current list of positive words that I have below and write your output in python list format (speak like an academic): ['novel', 'contribution']". We then refined the resulting list by specifically asking ChatGPT for distinct words and critiquing its output, for example, "Words like 'serene' and 'ecstatic' would not be used in a research paper. Try again". As of December 2nd, 2023, this query generates a list of positive words such as "groundbreaking" and "valuable", with some irrelevant words like "joyful" that were manually removed.

For argumentative language, we gave ChatGPT the following prompt: "List argumentative n-grams found in academic literature for 1 <= n <= 5." We then followed this prompt with another: "Make it more aggressive." As of November 29th, 2023, this query generates a list of argumentative language such as "lacks empirical evidence". We include this phrase and others in our argumentative language set.

Henceforth, the positive and argumentative language counts refer to the number of instances of the positive and argumentative language in our positive and argumentative language sets in a paper's abstract, respectively. To assess the construct validity of our positive and argumentative language counts, we conducted a manual inspection of a random sample of approximately 100 papers. During this review, we compared the number of instances we identified as positive and argumentative language to the respective positive and argumentative language counts in our dataset. We found a high degree of similarity between these instances and our counts, signifying the effectiveness of our sets in capturing the intended positive and argumentative language present within the abstracts of CS research papers. However, this review did prompt us to further refine our sets. For instance, following this evaluation, we made adjustments such as including the term "speedup" and excluding the term "clear" from the positive language set.

Positive language was operationalized as the percentage of words in an abstract that were in the positive language set, hereby called positive language percentage. Calculating this began with finding

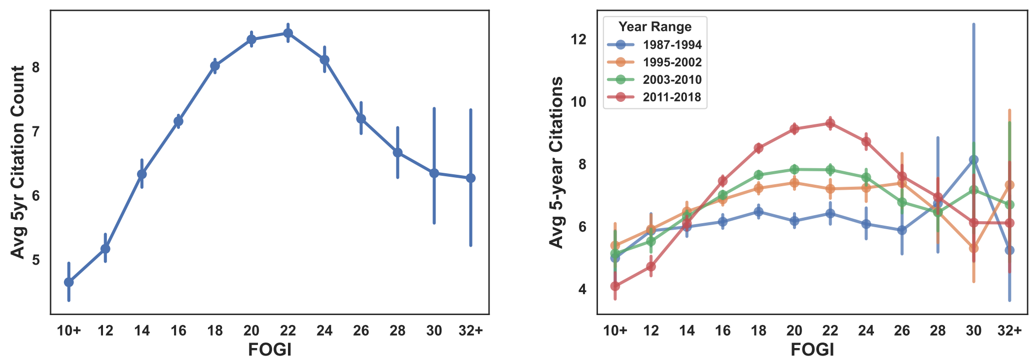

Figure 2: (a) Pointplot of 5-year citation count versus FOGI. Each point x is the average 5-year citation count for all abstracts with a FOGI in $[x, x+2)$. The point at 10+ includes all abstracts with FOGI less than 10 and 32+ includes all abstracts with FOGI greater than 32. Bars at each point are generated by bootstrapping 95% confidence intervals. (b) Same pointplot described previously, but computed on subgroups of the data partitioned by year.

the number of instances of each positive word in every abstract. This was done using Python's `string.count()` method, so variants of these positive words were included as well, e.g. "efficently" was counted along with "efficient". Words that contained our positive words, such as "coefficient" containing "efficient", were not counted. After these counts were calculated for every abstract, each count was divided by their respective abstract length and rounded to the nearest hundredth, which was necessary for result interpretability. Finally, these values were multiplied by 100 to represent percentages.

We validated these operationalizations by interpreting how they changed in relation to abstract length (e.g. positive language percentage rather than positive word count) and in relation to the year that a paper was published.

We make the following hypotheses about the relationships between success and our content-based presentational factors:

Hypothesis 2: A paper's success *increases* with the positive language percentage of its abstract.

We made this hypothesis because we believed a positive sounding abstract indicates that a researcher has produced novel and exciting results, which others in the field will want to read.

Hypothesis 3: A paper's success is *parabolically correlated* with the argumentative language count of its abstract.

We hypothesize this relationship because we believe that research papers that are presented as challenging existing ideas will gain attention and garner broader interest. However, excessive use of argumentative language might diminish a paper's credibility.

4 Results

4.1 Readability

We find that **Hypothesis 1** is incorrect. Citation count does not increase linearly with readability. Instead, we find that the relationship between 5-year citation count and FOGI is parabolic (Figure 2(a)). The papers that have a FOGI between 20 and 22 receive at least 2 more citations within five years of publication on average than papers with FOGI greater than 30 or less than 14.

We also observe that this parabolic relationship has gotten more pronounced over time. Figure 2(b) re-plots the previous point plot, but with the data partitioned into four 8-year chunks starting at 1987 and ending at 2018. The peak average 5-year citation count for papers published from 2011-2018 (red) is almost three citations higher than the average citation count for papers with a FOGI less

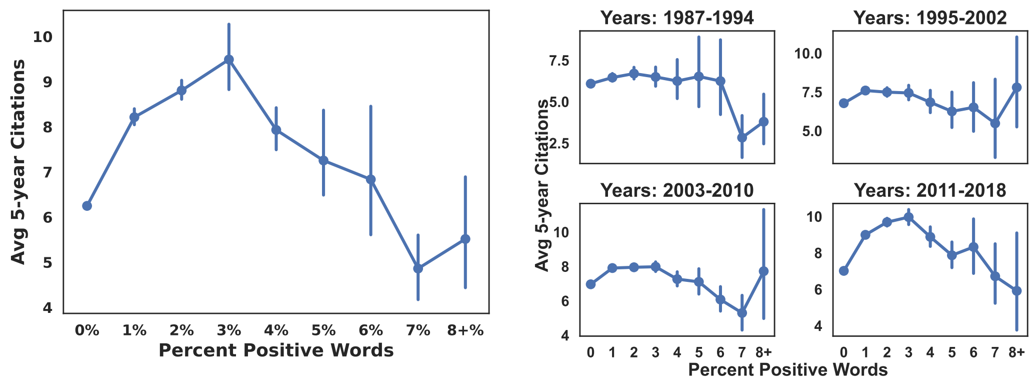

Figure 3: (a) Pointplot of average 5-year citation count estimate vs. positive word percentage. Bars at each point are generated by the bootstrap 95% confidence interval. (b) Same pointplot described previously, but computed on subgroups of the data partitioned by year.

than 14 or more than 30; whereas this difference decreases to about 1 citation for papers published between 2003-2010 and 1995-2002 (green and orange); and for papers published between 1987-1994 there is no significant difference in average citation count for FOGI from 12 to $32+$ (blue).

4.2 Positive Language

We randomly sampled 2 million CS papers that we had the abstracts of, and performed the earlier-mentioned calculations of positive language percentage. No further grouping was performed except for abstracts with positive word percentages of 8% and beyond being grouped into the "8+%" category, as the number of abstracts with each percentage dropped precipitously from there.

We find that **Hypothesis 2** is incorrect; the relationship between paper success and positive language percentage is actually parabolic (Figure 3(a)). Papers with abstracts with 3% positive language have an average of 3.24 more citations within 5 years of publication than those with no positive language, and an average of 3.98 more citations on average within 5 years of publication than those with 8+% positive language.

When stratified by year published (using the same 8-year buckets from Section 4.1) to analyze the validity of this presentational factor, we find that 3% positive language can only be distinguished as the most optimal value in the 2011-2018 year bucket (Figure 3(b)). However, a positive language percentage of 1-3$\%$ consistently appears more optimal than 0% or 8+$\%$ for all year buckets, maintaining the overall parabolic relationship found previously.

4.3 Argumentative Language

We randomly sampled 2 million CS papers we had abstracts for, and grouped them into three buckets according to their argumentative language counts. These groups are defined as:

- No argumentative language (0 instances)
- Some argumentative language (1-4 instances)
- Lots of argumentative language (5+ instances)

Approximately 90% of our abstracts fall into the first group. We find that **Hypothesis 3** is correct, success is parabolically correlated with argumentative language. In particular, the mean citation counts shown in Figure 4(b) are $\mu_{\text{none}} \approx 7.001$, $\mu_{\text{some}} \approx 9.987$, and $\mu_{\text{lots}} \approx 5.338$, indicating that the papers with 1-4 instances of argumentative language ("some") accrue an average of 3 citations more than papers with no argumentative language, and ≈ 4.7 citations more than papers with 5+ instances of argumentative language ("lots"), in the span of 5 years following their publication. The confidence intervals for each estimate are derived using the bootstrap 95% confidence interval. We include a

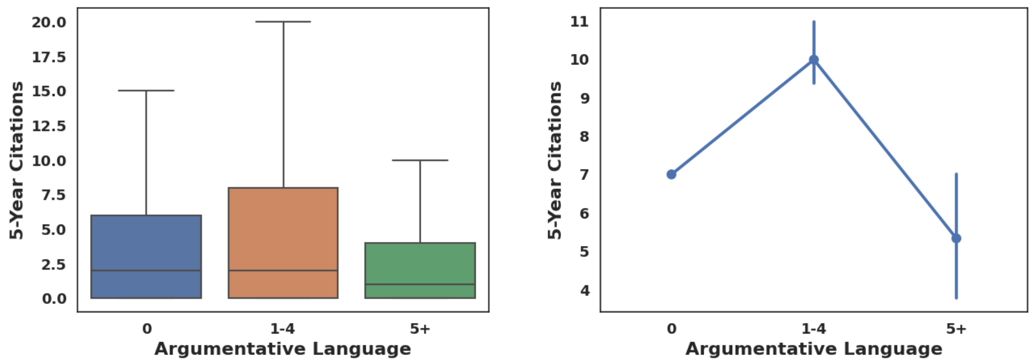

Figure 4: (a) Boxplot of 5-year citation count vs. argumentative language count partitioned by amount (none, some, lots). (b) Pointplot of 5-year citation count vs. argumentative language count. Uses the same partitioning scheme as the boxplot. Bars at each point are generated by the bootstrap 95% confidence interval.

boxplot of 5-year citation count vs. argumentative language count to demonstrate the distribution of the data.

We also find that our results are robust to the length of the abstract. When normalizing our argumentative language counts to find the argumentative language count per average abstract length, the parabolic trend is in fact even more pronounced.

5 Discussion

5.1 Readability

The optimal FOGI score of 20-22 roughly corresponds to a PhD-student's level of reading comprehension (e.g., 13-17 is college-level, 18-23 is PhD, 24-25 is post doc, and 26+ is professor). This suggests that when writing an abstract, one should aim for a PhD-student level of reading comprehension. This involves using moderately complex words (with 1.5-2 syllables on average) and moderately long sentences (with 22-28 words on average).

We were initially surprised by the parabolic relationship between readabilty and success (Figure 2(a)). We speculate that this is the result of competing causal factors that affect citation count. The first is *accessibility*, which decreases with FOGI. The more years of education that are required to comprehend a particular piece of text, the less accessible it is. The second is *expressibility*. More complex language is more capable of expressing more complicated concepts. So, in general expressibility increases with FOGI. For FOGI < 20, the language starts to get significantly less expressive, and for FOGI > 22, the language makes the papers significantly less inaccessible.

Our last result (Figure 2(b)) was that the parabolic relationship has become more pronounced over time. We have two hypotheses explaining this change in effect. Out first hypothesis is that PhD students are making up a larger proportion of the field, so writing to their level of reading comprehension is more important now than in the past. The second hypothesis is that the field of CS has become more over-saturated, leading to a higher bar in what gets cited.

5.2 Positive Language

According to the data, 3% is the optimal percentage of positive language to include in a CS paper's abstract. The average length of an abstract from our dataset is ~145 words, so on average, an optimal abstract would contain $0.03 \times 145 \approx 4$ positive words.

We were surprised by this result. The fact that including some positive language is correlated with higher 5-year citation count makes sense. A neutral-sounding abstract may cause readers to question if the paper describes anything new or exciting, in other words, anything worth reading. From our findings, it seems that an abstract can also have too much positive language. However, this makes sense, as an abstract that comes off too positive could be interpreted as artificially or exaggeratedly optimistic, which could call the paper's validity into question. In practice, our findings indicate a moderate amount of positive language is best for a paper's abstract. For example, a researcher may want to clearly state that an idea of theirs is novel, an experiment was successful, a finding was beneficial, and so on, without overemphasizing these statements. This could entice potential readers, convincing them that they will gain something from reading the paper.

In addition, our year stratification for positive language percentage shows that the average number of 5-year citations for all papers increases with every year bucket, especially in 2011-2018 (Figure 3(b)). This could be reflective of computer science as a field expanding over time. As such, this does not appear to be a major confounder with our findings.

A shortcoming of this operationalization is the disparity between the number of papers in each category. For example, in the random sample of abstracts used in Figure 3, 1 million abstracts have an positive word percentage of 0%, while only 1100 have a positive word percentage of 8+% or more. This lowers our confidence in the data groupings representing higher percentages. In addition, some positive words found through manual abstract review had to be excluded from our positive word list, as we found they were used in other scenarios that did not fit under our idea of positivity, for example, talking about how a subfield lacked a "good" solution to something. Combined with the fact our positive word set only contains 106 words, our positive word set could be more exhaustive.

5.3 Argumentative Language

The optimal argumentative language count is "some" or 1-4 instances of argumentative language. This suggests that when writing an abstract, one should include a small amount of argumentative language. However, "lots" or 5+ instances is highly correlated with a significantly lower mean 5-year citation count, indicating that it is possible to include too much.

We were not surprised by this result. We speculate that the parabolic trend we observe is because when abstracts include 5+ instances of argumentative language, the paper comes off as too aggressive and is unlikely to be cited. Further, when abstracts include no argumentative language, it is unclear how the paper differs from earlier work, and as such, why the paper is important. While reading abstracts, we commonly saw argumentative language being used to explain the relevance of a paper in the context of previous work. To illustrate with an example, an author might include the phrase "Previous work fails to address...". This makes it clear to the reader that the paper they're reading is a novel work, and precisely defines what the author's contribution will be.

A shortcoming of this operationalization is that papers with 5+ instances of argumentative language were relatively uncommon, so we have less confidence in our mean estimate for these papers. In our random sampling of 2 million CS papers with abstracts, we found ~1.8 million had 0 instances, ~140 thousand had 1-4 instances, and only ~10 thousand had 5+ instances. It's also possible there exist argumentative words and short phrases that we did not include in our argumentative language set. If we had more time and resources, it would be interesting to use NLP and data mining to try and build a better argumentative language set.

6 Related Work

Several studies have investigated the factors influencing the success of academic papers. Tahamtan et al. (2016) conducted an extensive review of 198 papers, exploring the factors affecting citation counts across many academic disciplines [6]. The factors identified in this work were classified categorically as "paper-related," "author-related," or "journal-related". Within paper-related aspects, Tahamtan et al. explored constructs such as the novelty of findings, methodology, and study design. This work motivated our focus towards on *presentational* factors, such as readability and language choices, as these areas seemed less explored within existing literature.

Building on the work of Tahamtan et al., Soheili et al. (2022) conducted a comprehensive mixed-method study of factors influencing citation count [5]. Their research was restricted to the field of Library and Information Science. The findings of Soheili et al. included a statistically significant positive correlation between the quantity of keywords in paper titles and paper citation counts. Additionally, the researchers uncovered a positive correlation between including standardized contact information in papers and citation counts. Inspired by the work of Soheili et al., we integrated word count analysis into our study. However, we advanced upon their research by analyzing nonlinear relationships instead of only linear relationships through correlation tests.

Chen et al. (2007) discussed metrics beyond raw citation for assessing paper success, employing Google's PageRank algorithm to measure the influence of papers in the Physical Review family of journals [1]. Similarly, Giatsidis et al. (2019) explored graph density-related metrics, notably defining "core influence" in their study [2]. These works prompted us to evaluate the limitations of raw citation count, driving our investigation towards 5-year citation count as a more nuanced measure of paper success.

7 Ethical Considerations

7.1 Data Access and Use

The citation data utilized in this study was obtained through Semantic Scholar's Academic Graph API. Semantic Scholar granted us consent to use this API specifically for academic research purposes. In addition, abstract texts analyzed in this study were open-access and are publicly available. That is, no abstract data behind paywalls or access restrictions was used.

7.2 Implications of Findings

The findings of this research suggest that the manner in which an academic research paper is *presented* has both a statistically significant and practically significant impact on the paper's success. While these results align with certain expectations about academic research, they do raise concerns about potential disparities in academia. Specifically, these results suggest that researchers whose writing style deviates from the above recommendations might be at a disadvantage, despite strong scholarly contributions. For instance, a researcher who learned English as a second language may encounter greater hurdles in garnering recognition for their work compared to their peers.

To mitigate this concern, it is important that guidance on paper-writing conventions is made explicit and widely available. This research serves as one such effort. In addition, we recommend that conferences release clear style guidelines, highlighting any “hidden” expectations that successful researchers adhere to. This transparency can help level the playing field within academic discourse.

8 Conclusion

In this comprehensive analysis of over 5 million computer science research paper abstracts, we analyzed papers' presentational factors and their impact on paper success. Our exploration of readability, positive language, and argumentative language revealed several insights, challenging some of our hypotheses while validating others.

In particular, our findings uncovered a parabolic relationship between each presentational factor and paper success. With regards to readability, papers with moderately complex abstract text, exhibiting a FOGI between 20 and 22, emerged as the most successful. The presence of positive and argumentative language in abstracts followed a similar trend. Abstracts with approximately 3% of their language as positive and 1-4 instances of argumentative language garnered greater success.

It is important to acknowledge the limitations of this work. Presentational factors, while impactful, are not the only factors that influence the success of academic research papers. The novelty/significance of results, popularity of authors, and reputation of institution play a substantial role as well. Disentangling the effects of presentational and non-presentational factors is a challenge, and this work by no means argues causation.

Nevertheless, our findings offer valuable insights for researchers aiming to maximize the impact of their work. Researchers may use the optimal thresholds discussed above as guidelines, tailoring their writing styles to strike a balance in readability, positivity, and argumentativity.

9 Data availability

We aquired paper data through the Semantic Scholar API: https://www.semanticscholar.org/product/api. We acquired Best Paper Awards in Computer Science Collection by scraping the awards listed here: https://jeffhuang.com/best_paper_awards/. We acquired Wikipedia's List of Important Publications in Computer Science from here: https://en.wikipedia.org/wiki/List_of_important_publications_in_computer_science

References

[1] P. Chen, H. Xie, S. Maslov, and S. Redner. Finding scientific gems with google's pagerank algorithm. *Journal of Informetrics*, 2007.

[2] Christos Giatsidis, Giannis Nikolentzos, Chenhui Zhang, Jie Tang, and Michalis Vazirgiannis. Rooted citation graphs density metrics for research papers influence evaluation. *Journal of Informetrics*, 2019.

[3] Robbie Gunning. The technique of clear writing. 1968.

[4] Pontus Plaven-Sigray, Granville James Matheson, Bjorn Christian Schiffler, and William Hedley Thompson. The readability of scientific texts is decreasing over time. *eLife*, 6, 2017.

[5] Faramarz Soheili, Ali Akbar Khasseh, Heidar Mokhtari, and Mozhgan Sadeghi. Factors affecting the number of citations: A mixed method study. *Journal of Scientometric Research*, 2022.

[6] Iman Tahamtan, Askar Safipour Afshar, and Khadijeh Ahamdzadeh. Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107, 2016.

10 Supplementary Material

10.1 Positive Language Set

["novel", "innovative", "groundbreaking", "pioneering", "insightful", "ingenious", "remarkable", "significant", "meaningful", "valuable", "exemplary", "outstanding", "exceptional", "promising", "advantageous", "beneficial", "profound", "impactful", "constructive", "productive", "commendable", "praiseworthy", "laudable", "meritorious", "stellar", "splendid", "excellent", "superb", "optimal", "superlative", "first-rate", "top-notch", "positive", "optimistic", "optimizes", "encouraging", "inspiring", "motivating", "hopeful", "successful", "accomplished", "achievable", "feasible", "realistic", "accessible", "manageable", "capable", "competent", "proficient", "effective", "efficient", "admirable", "notable", "respectable", "honorable", "prestigious", "distinguished", "esteemed", "venerated", "respected", "acknowledged", "recognized", "celebrated", "satisfying", "fulfilling", "rewarding", "credible", "robust", "valid", "rigorous", "comprehensive", "thorough", "well-designed", "methodical", "meticulous", "precise", "concise", "persuasive", "convincing", "compelling", "grounded", "solid", "substantial", "meaningful", "valuable", "noteworthy", "strong", "outperform", "improve", "verify", "verified", "achieve", "speedup"]

10.2 Argumentative Language Set

["disagree", "lacks empirical evidence", "incorrectly assumes", "lack evidence", "incorrect", "do not agree", "lack of evidence","lacks evidence", "flawed", "misguided", "does not address", "fails to adequately address", "fails to address", "lacking", "erroneous assumption", "erroneous assumptions", "lacks credibility", "disproves", "irrelevant", "incorrect assumptions", "questionable", "methodological weakness", "lack of understanding", "disprove", "disproves", "falsely claimed", "not supported by", "weakness", "previous approaches", "existing works", "existing work", "existing research", "previous work", "most strategies", "existing approaches", "previous methods", "insufficient", "little proof", "biased", "incomplete", "groundless", "most research", "prior research", "disregards", "prior work", "fails to", "existing methods", "poorly understood", "lack of", "instead of", "we instead", "instead we", "incorrectly", "existing work"]