1. ABSTRACT

How has media attention to science changed over the past decade? Does media attention bring scientific attention too?

To answer these questions, we collected media attention statistics from Altmetric.com for over 40,000 papers published in PNAS journal in the last 13 years. Our analysis reveals that (i) Media attention to science has slowly, but steadily increased over time, (ii) Media attention doesn’t necessarily translate into scientific attention (citations) and, (iii) It is non-trivial to predict the amount of media attention a paper gets from attributes such as authors, author affiliations, abstract, and title.

2. RELATED WORK

Darling et al. [2] studied the role of social media in the life cycle of a scientific publication and conclude that Twitter has been playing an ever increasing role in the dissemination and publicity of scientific literature. Haustein et al. and others [3, 1] propose and validate the use of new metrics (alt-metrics) that can capture the real world impact of science, including social media buzz created and number of mainstream media mentions gathered by a scientific article. Many popular journals and publishers have already embedded altmetric scores on their online webpages. Closest to our study is the work by Zoller et al [4], who compare citations and altmetric scores from a social bookmarking system. Their analysis on more than 250,000 publications shows that citations and altmetrics are weakly correlated. To the best of our knowledge, our study is the first to look at media related metrics for scientific studies and perform analysis on the relationship between popularity in main stream media and popularity in the scientific community.

3. DATA

We collected data from the Proceedings of the National Academy of Sciences (PNAS journal) for 13 years (2004–2016). This gave us 44,109 papers, which cover a wide range of disciplines, including mathematics, physics, chemistry, biology, etc. Each paper has an Altmetric score1 (a score indicating the media attention a paper gets), its coverage in the popular news media, social media (Twitter, Facebook, Reddit, etc) and scientific popularity stats like number of downloads, number of readers on Mendeley, etc. For each paper, we also obtained the number of citations from the Web of Science database. Figure 1 shows an example for one paper.2 This paper has an Altmetric score of 138, and mentioned by 11 news media outlets. Since social media attention is a relatively new phenomenon (starting around 2008), we only consider mentions in main stream news as a sign of media attention to a paper. Our datasets are available for download on the project webpage.3

4. FINDINGS

Media attention over the years PNAS publishes a new version of the journal every week. Each paper is associated with a week in which it was published. For each paper published in a week, we extract the number of mentions in news media and compute the average number of news media mentions over all papers in each week. We plot this data over a period of 654 weeks (2004–2016). Figure 2 shows the average number of media mentions for all papers (blue) and for a specific field (Neuroscience, the most frequent field in our data, consisting of 3,500 papers). Both have a slight positive slope for a linear regression line fit (Slope 0.0071 for all papers and 0.0073 for Neuroscience papers). We tested the value of the non-zero slope for statistical significance using a t-test (p < 0.0001). Similar trends were found for the other fields in our data, indicating a slow but steady increase in media attention to science over the last decade.

1http://bit.ly/2iqm7k3
2e.g. see http://bit.ly/2iHxXEc
3https://users.ics.aalto.fi/kiran/mediaAttentionScience/
dom. i.e., it is not Lasso, Linear Regression, Ridge, and Decision Trees, but experimented various standard regression models including thor’s affiliation(s), and (v) research field of the paper. We words in the paper abstract, (iii) author’s name(s), (iv) au-
gets (target variable: number of news mentions) based on features extracted from: (i) words in the paper title, (ii) words in the paper abstract, (iii) author’s name(s), (iv) au-
university (abstract) 0.08
ecient (field) 0.078
decades (abstract) 0.077
climatic (abstract) 0.074
mortality (abstract) 0.074
century (abstract) 0.074

Figure 3: Heatmap of correlations. altmetrics: Alt-
metric score, news: Number of mentions in news me-
dia, {abs, full}_dwnld: Number of downloads of the
abstract and full version of the paper.

Can we predict media attention? We built regression
models to predict the amount of media attention a paper
gets (target variable: number of news mentions) based on
features extracted from: (i) words in the paper title, (ii)
words in the paper abstract, (iii) author’s name(s), (iv) au-
the social tagging system bibsonomy.

Mediation analysis Next, we analyzed the features that are
most positively correlated with media attention. For each
feature, we computed its Pearson correlation coefficient with
the media attention a paper containing this feature gets.
From Table 1, we can get a sense of what factors could po-
tentially lead to news media attention. For example, pa-
pers with keywords ‘health’, ‘mortality’, ‘climate’, etc in the
abstract get more attention and fields like Psychology and
Sustainability, studying mental health, and climate change,
get more media attention.

Acknowledgements. This work has been supported by
the Academy of Finland project “Nestor” (286211) and the
EC H2020 RIA project “SoBigData” (654024).

5. REFERENCES
[1] L. Bornmann. Do altmetrics point to the broader
impact of research? an overview of benefits and
disadvantages of altmetrics. Journal of Informetrics,
scientific publication. Ideas in Ecology and Evolution,
6(1), 2013.
sources in the bibliometric community. Scientometrics,
the social tagging system bibsonomy. Journal of

Table 1: Top 10 correlated features with media at-
tention.

<table>
<thead>
<tr>
<th>Feature Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>psychological_and_cognitive_sciences (field)</td>
</tr>
<tr>
<td>health (abstract)</td>
</tr>
<tr>
<td>people (abstract)</td>
</tr>
<tr>
<td>united (abstract)</td>
</tr>
<tr>
<td>earth_atmospheric_and_planetary_sciences (field)</td>
</tr>
<tr>
<td>sustainability_sciences (field)</td>
</tr>
<tr>
<td>centuries (abstract)</td>
</tr>
<tr>
<td>decades (abstract)</td>
</tr>
<tr>
<td>climate (abstract)</td>
</tr>
<tr>
<td>mortality (abstract)</td>
</tr>
<tr>
<td>century (abstract)</td>
</tr>
</tbody>
</table>

Figure 2: Average number of news media mentions over time. The regression fit lines overlap.