A naive model for the assessment of paper views in academia.edu network

H.M. de Oliveira, PGOM, LD
Federal University of Pernambuco, UFPE,
Statistics Department, E-mail: hmo@de.ufpe.br
CCEN-UFPE, Recife, Brazil

Abstract—This short note presents some curve fitting to model the distribution of the number of paper views in terms of the number of uploaded papers in academia.edu. Even though it is a very prosaic study, it proposes a relation of type \( \% \text{views} = \alpha \log_{10}(\% \text{papers}) + \beta \). This allows, for example, to evaluate how much a fixed percentage, say 20\%, contributes in relation to the total number of views of the researcher’s contributions. This approach may give an indicative something similar to Pareto’s 80/20 rule, the law of the vital few. The influence of the research area is not addressed. Adjustments were carried out just for a couple of researchers who kept public their analytics of papers access. The results should probably only be applied (if possible) to scientists with large numbers of articles deposited on the network – say a few dozen papers. There is a caveat that it is not claimed that the results are general: it is only an initial investigation of limited scope, yet rather provocative.

Index terms— assesement of paper views, Pareto-like laws, scientific social networks, academia.edu.

I. DATA TAKEN INTO CONSIDERATION IN THIS EXPLORATORY INVESTIGATION

Academic social networking sites are a recent phenomena [3], [4], [6]. From an unpretentious search among quite a few profiles in academia [7], we selected the analytical access data from a few web pages where researchers chose to keep public their access statistics. This was the only selection criterion (unfortunately, the academia does not provide the top access list.) Only by way of a first provisional investigation, the access data in May 2019 concerning:

- Abdus Saboor (5,596 Total Views, top 4%)
- Gregory Chaitin (45,863 Total Views, top 2%)
- Hélio M. de Oliveira (24,539 Total Views, top 0.5%)
- Vladimir Klein (15,558 Total Views, top 4%)

were considered. This list includes an internationally renowned scientist. However, by no means this has value judgments about the quality or relevance of the selected research.

The adjusted data were obtained by accessing:
1) https://kust.academia.edu/DrAbdusSaboor
2) https://independent.academia.edu/GregoryChaitin
3) https://independent.academia.edu/VladimirC3%ADKlein
4) https://ufpe.academia.edu/hmdeoiveira

and clicking on the cup field \( \Phi \) top 4\%, the selecting papers. In the sequel, all papers are sorted in descending order by the number of times they were viewed.

Table I.1: The function \( \phi_X(x) = \% \text{viewed} \) denotes the cummulative relative frequency of the number of views in terms of a pack of of \( x \) more viewed papers (e.g. for \( x = 5 \), \( \phi(5) \) provides the percentage of total views “covered” by the 5 most viewed papers.)

<table>
<thead>
<tr>
<th>Researcher with academia</th>
<th>( \phi_X(x) = % \text{viewed} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saboor</td>
<td>0.29596 ln(x) − 0.069076</td>
</tr>
<tr>
<td>Chaitin</td>
<td>0.25959 ln(x) + 0.1294</td>
</tr>
<tr>
<td>Klein</td>
<td>0.2792 ln(x) + 0.42545</td>
</tr>
<tr>
<td>de Oliveira</td>
<td>0.16404 ln(x) + 0.10136</td>
</tr>
</tbody>
</table>

II. A SIMPLIFIED AND STRAIGHTFORWARD MODEL TO ASSESS THE PERCENTAGE OF VIEWS IN TERMS OF THE PERCENTAGE OF UPLOADED FILES

From the fitting of the distributions to the number of views (as a percentage), based on the #papers considered (the slopes from Table I.1), it is possible to propose an approximate model

\[
\frac{d \% \text{views}}{d \# \text{papers}} \approx \frac{K}{\# \text{papers}},
\]

where typically \( K \approx 0.3 \) (an order of magnitude).

In this way, it is possible to establish a typical relationship between \( \% \text{views} \) and percentage of papers taken into account as

\[
\% \text{views} = \alpha \log_{10}(\% \text{most viewed papers}) + \beta,
\]

where \( \alpha \) and \( \beta \) are constants. These relationships were derived in each case, and results are shown in Table II.2.

Table II.2: Percentage of the total number of views in terms of the percentage of uploaded papers in academia.edu. Here, \( n \) is the total number of uploaded papers in the page, and \( R^2 \) is the determination coefficient of the log fitting.

<table>
<thead>
<tr>
<th>researcher</th>
<th>%views × %papers</th>
<th>n</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saboor</td>
<td>68.147 log_{10}(% \text{papers}) − 39.630</td>
<td>32</td>
<td>0.9643</td>
</tr>
<tr>
<td>Chaitin</td>
<td>59.973 log_{10}(% \text{papers}) − 17.463</td>
<td>31</td>
<td>0.9941</td>
</tr>
<tr>
<td>Klein</td>
<td>64.288 log_{10}(% \text{papers}) − 21.743</td>
<td>10</td>
<td>0.9421</td>
</tr>
<tr>
<td>de Oliveira</td>
<td>45.971 log_{10}(% \text{papers}) + 9.962</td>
<td>273</td>
<td>0.9411</td>
</tr>
</tbody>
</table>

All curves were included in Figure 1 along with curve fitting, in order to provide a visual observation of the adherence of the models. As previously mentioned, let us evaluate how much a fixed percentage, say 20\% \( \Pi \), contributes in relation to the total number of views of the researcher’s contributions.
A typical expected behavior, say, the one that follows the Pareto law, leads to an expression of the type

$$\%\text{views} = 30 \log_{10}(\%\text{most viewed papers}) + 40, \quad (3)$$

for which 20% of most viewed papers accounts for about 80% of the total number of views. According with Table II.3, this high concentration of interest in a few papers is not strictly observed.

Table II.3: How much the most viewed 20% papers contributes to the total number of views. We have some parallel with Pareto’s 80/20 rule. The last column shows the degree of agreement with the Pareto model.

<table>
<thead>
<tr>
<th>researcher</th>
<th>amount %/%</th>
<th>degree of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saboor</td>
<td>49/20</td>
<td>61%</td>
</tr>
<tr>
<td>Chaitin</td>
<td>60/20</td>
<td>75%</td>
</tr>
<tr>
<td>Klein</td>
<td>62/20</td>
<td>78%</td>
</tr>
<tr>
<td>de Oliveira</td>
<td>70/20</td>
<td>88%</td>
</tr>
</tbody>
</table>

The higher the ratio (relative to the fixed percentage 20%), the higher the concentration of the number of views in a small number of relevant papers. This means that there are much more articles with a small number of hits. When the total number of contributions is relatively small (e.g. \( n \leq 30 \)), the usual 80/20 ratio may not be observed. Furthermore, the model fitting seems to show lesser degree of goodness of fit.

![Figure 1](image1.png)  
(a) Academia analytics for Abdus Saboor.  
(c) Academia analytics for Vladimír Klein.  
(d) Academia analytics for Hélio M. de Oliveira.

III. CONCLUDING REMARKS

This note is merely an ingenuous and unsophisticated investigation addressing the number of views of papers available in academic social networking sites [5]. The objective is to better evaluate how the availability of articles is explored in networks such as Academia.edu [7], ResearchGate, Mendeley, academic microsoft, DBLP computer science bibliography or Arnetminer [2]. Probably the number of hits should not have a substantially different distribution. It is a correspondence and not a standard scientific article. Finally, we remark that the model fit for a particular user can as well provide parameters to gain a better understanding the researcher’s profile.

Unlike other networks, academia.edu has the analytics, a tool that makes this analysis possible. Similar data are not available on the other platforms. More in-depth analyzes should be conducted, with data for many researchers and taking into account the influence of the researcher’s field.
REFERENCES


