Some practical difficulties in creating field-normalized altmetrics

“There is only one thing in life worse than being talked about, and that is not being talked about.” Oscar Wilde

Michael Taylor

Introduction

- Researching subject-based variations in different forms of altmetrics for PhD
- One practical outcome is to underpin approaches to creating subject-normalized altmetrics
- This is particularly important as altmetrics are being increasingly examined for research evaluation: users are expressing an interest in analysis
- Furthermore, two distinct use-cases:
  - to read “off the page” when an article is “highly shared”;
  - to perform more nuanced evaluation work, typically with a portfolio
- Citation has a long history of subject normalized metrics, eg, FWCI, NCI, and ‘top-slicing’ – ie, analysing the top 10% for citation
- It is tempting to simply follow these approaches, but these metrics make assumptions about the data that don’t necessarily follow

Data set

- 9610 articles drawn from the NIH's publication database
- Classified into 13 subject areas on an article-by-article basis using machine learning (not journal classification)
- Subject areas > 200 articles are analysed:
  - Biochemistry and Cell Biology; Cardiorespiratory Medicine and Haematology; Clinical Sciences; Genetics; Immunology; Medical Microbiology; Neurosciences; Oncology and Carcinogenesis; Other Physical Sciences; Paediatrics and Reproductive Medicine; Physical Chemistry; Psychology; Public Health and Health Services

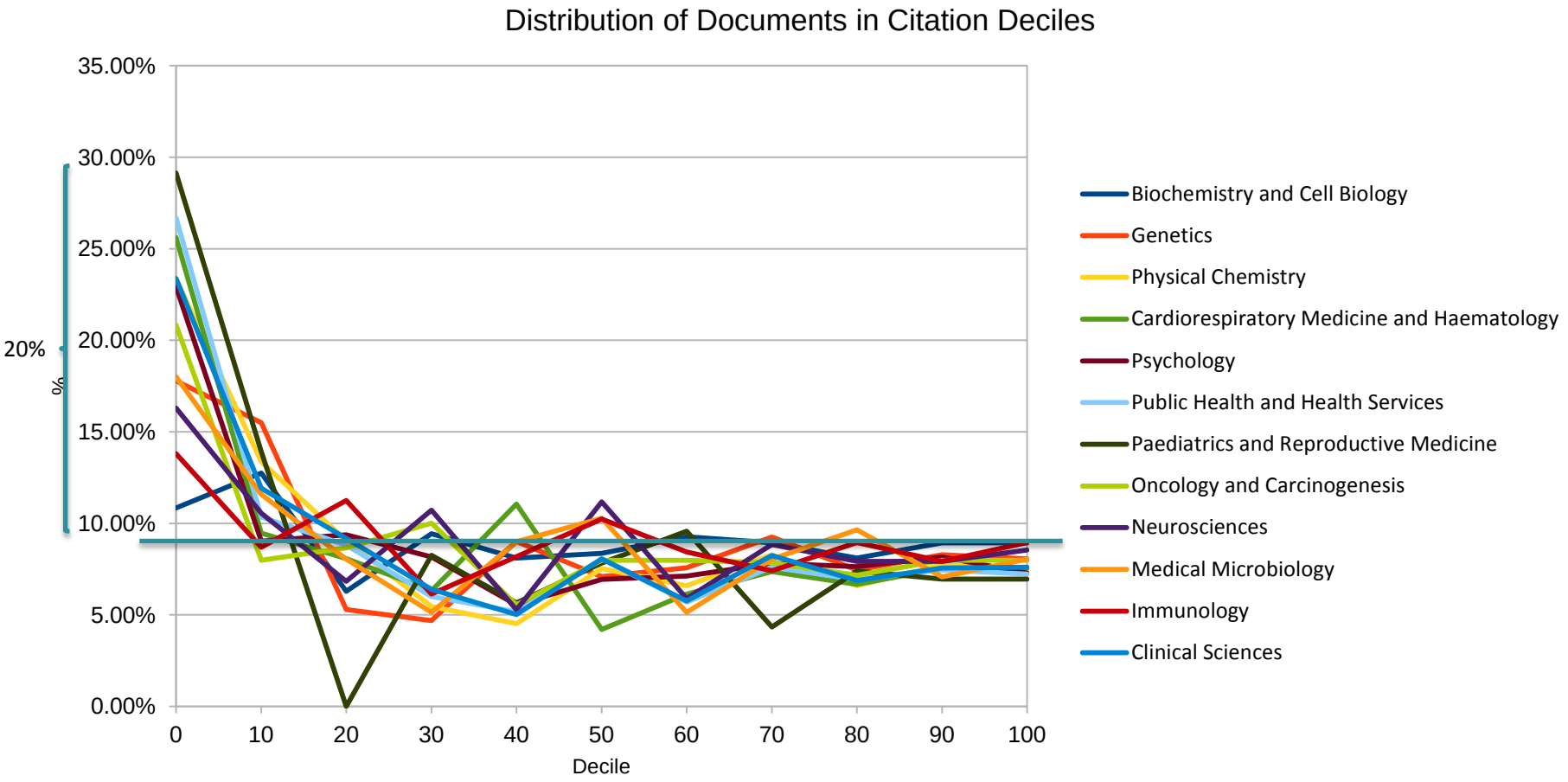
- Citations from NIH's icite tool
- Mendeley data direct from Mendeley's API
- Altmetric data from Altmetric.com's API

- Costa et al "Exploring paths for the normalization of altmetrics: applying the Characteristic Scores and Scales" http://altmetrics.org/wp-content/uploads/2016/09/altmetrics16_paper_6.pdf
- Taylor "Engineers Don't Blog" – poster at 2AM - https://figshare.com/articles/Engineers_Don_t_Blog_and_Other_Stories_why_Scopus_uses_subject_area_benchmarking_/1568135

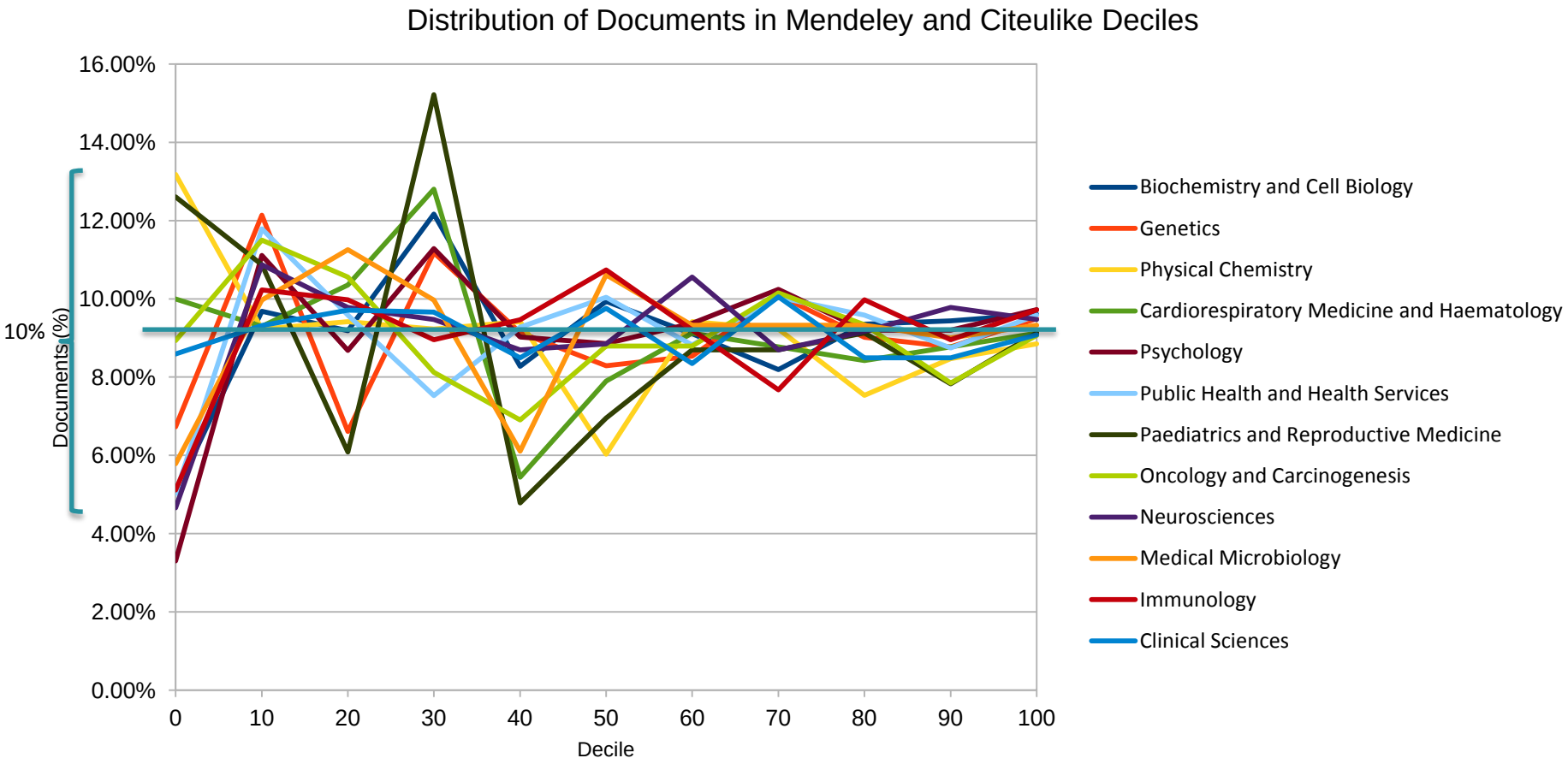
Methodology

- Four data points considered:
  - Citations
  - Mendeley readers & Citeulike saves
  - Altmetric Attention Score
  - Social Network activity (Twitter, Facebook, Reddit, Google+)
- Articles with zero values in each four data are placed in a distinct category
- Articles with non-zero values are placed into deciles (NB collisions are an issue, eg, Social Network count of 1 straddles more than one decile
- Examining the number of documents in each classification / decile, and the mean values within that decile
  - How does ‘zero’ compare to a decile?
  - Are deciles reasonably evenly distributed?

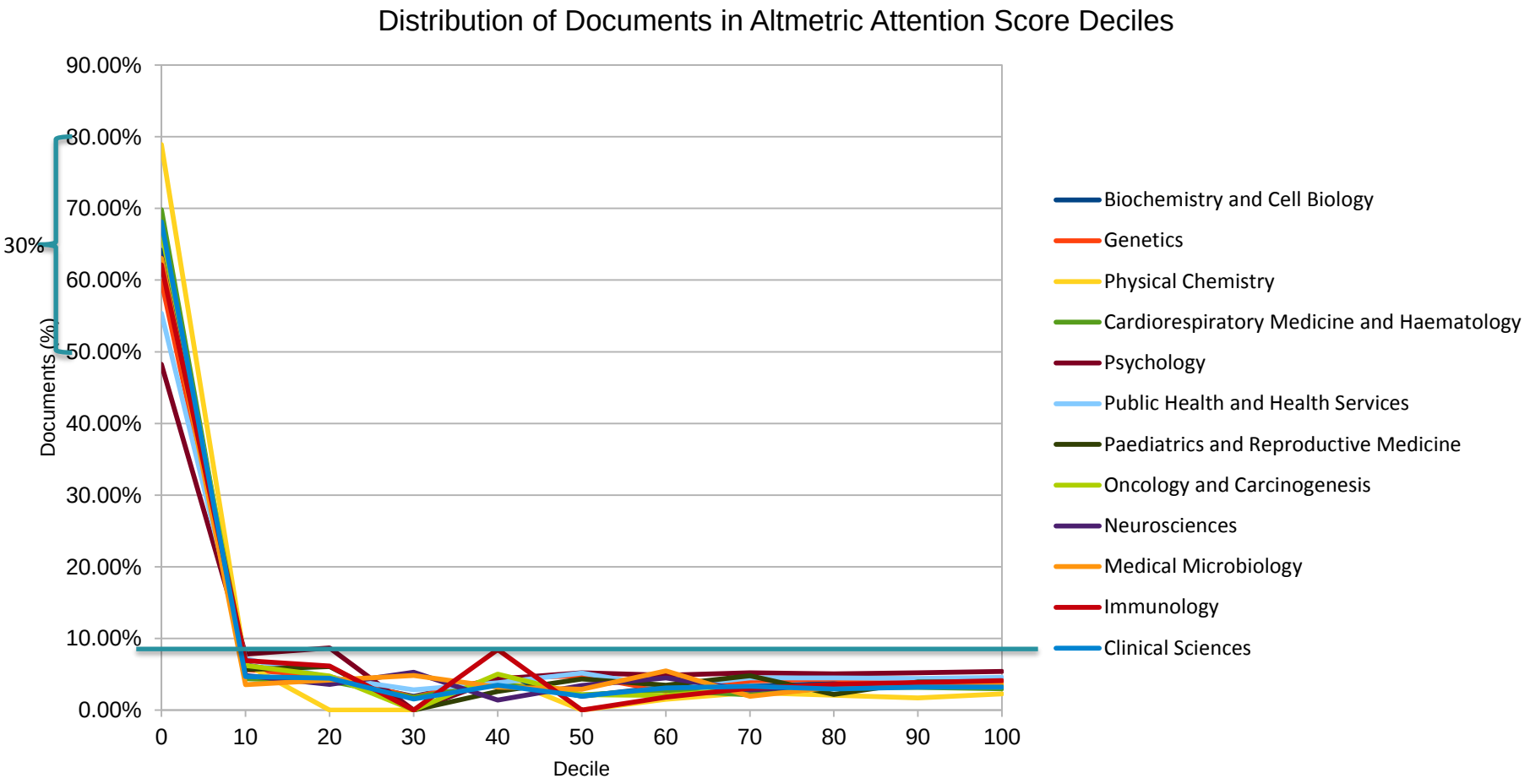
articles in deciles - citation



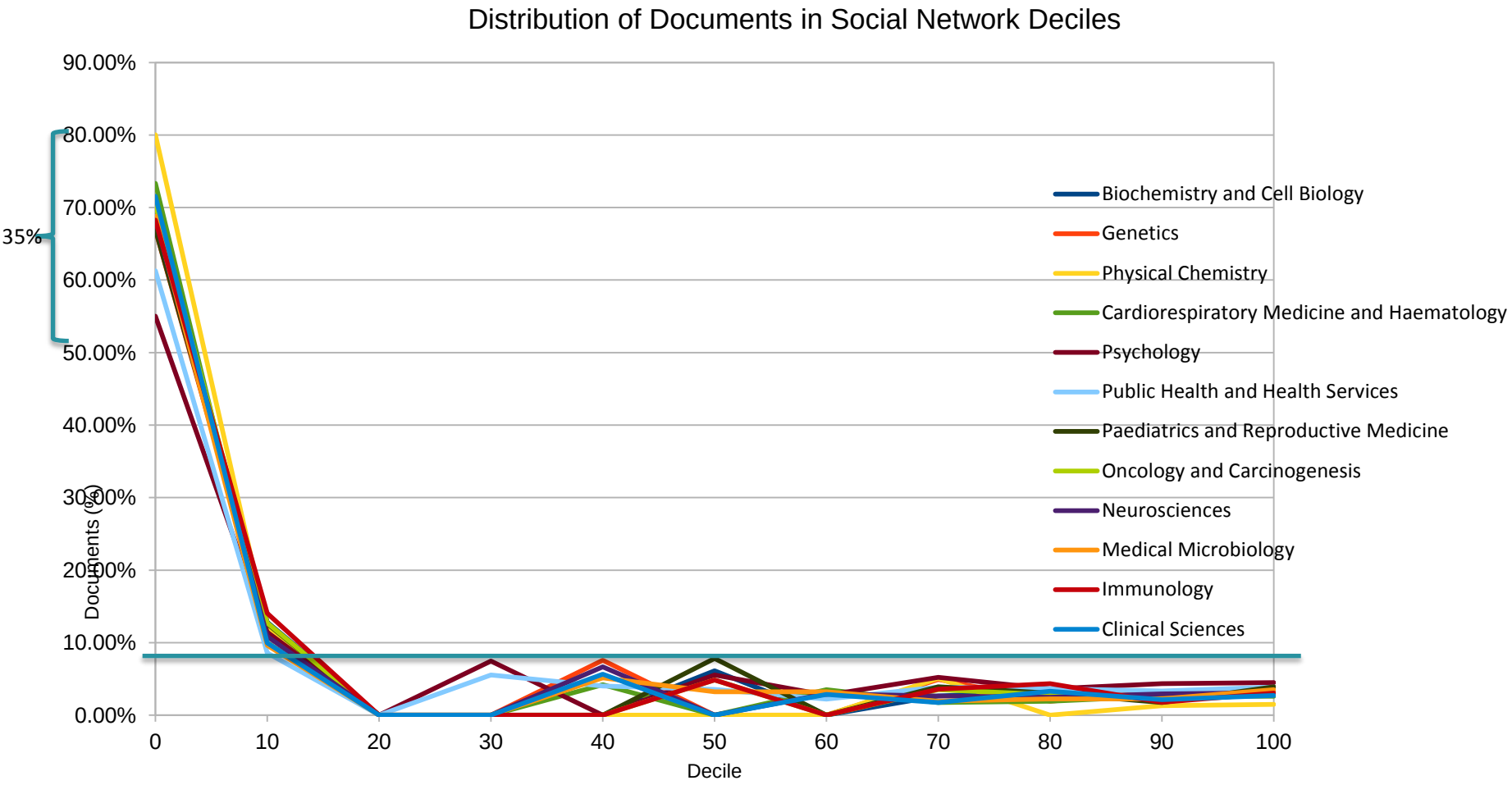
articles in deciles – Mendeley & Citeulike



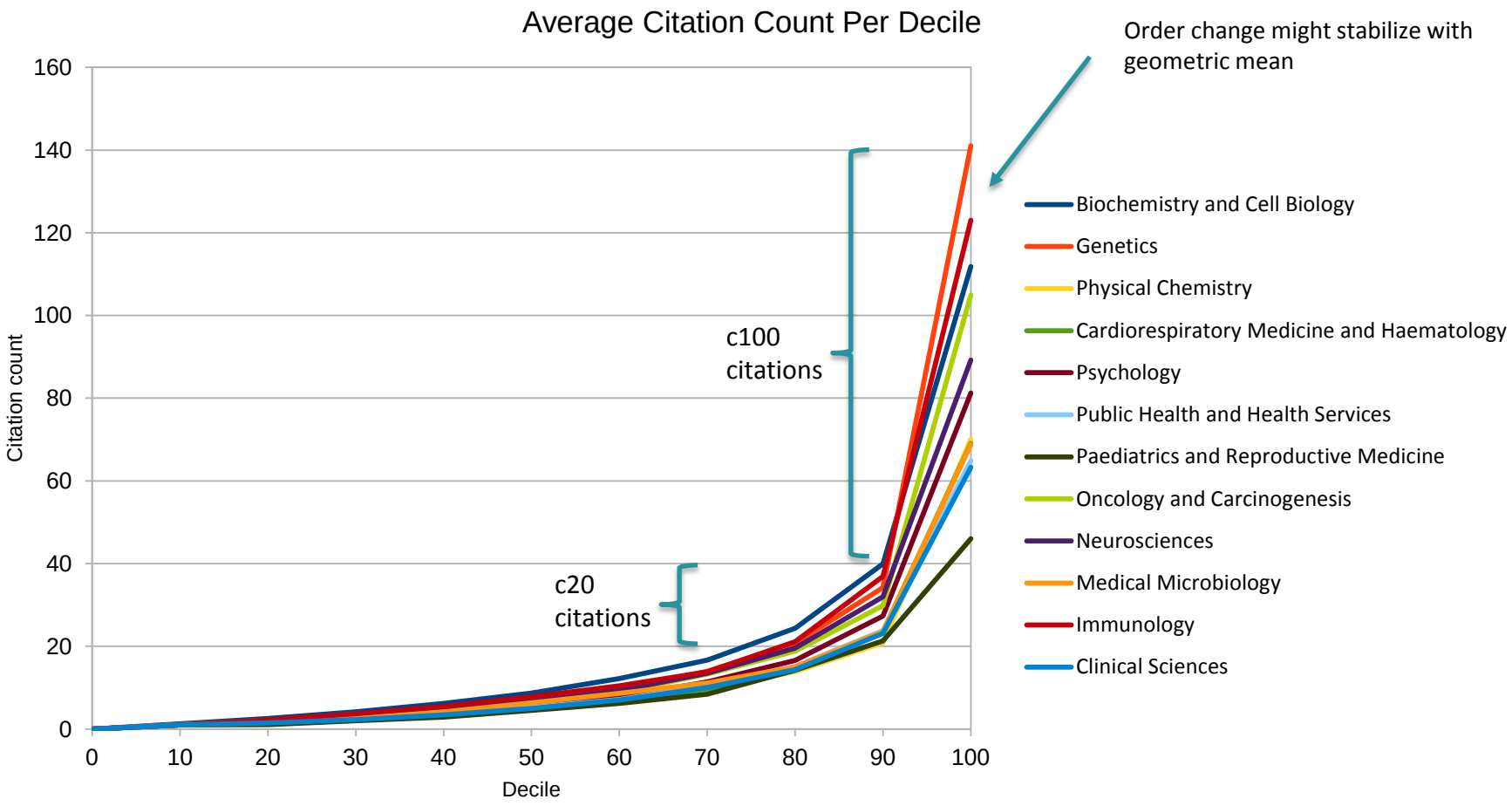
articles in deciles – Altmetric Attention Score



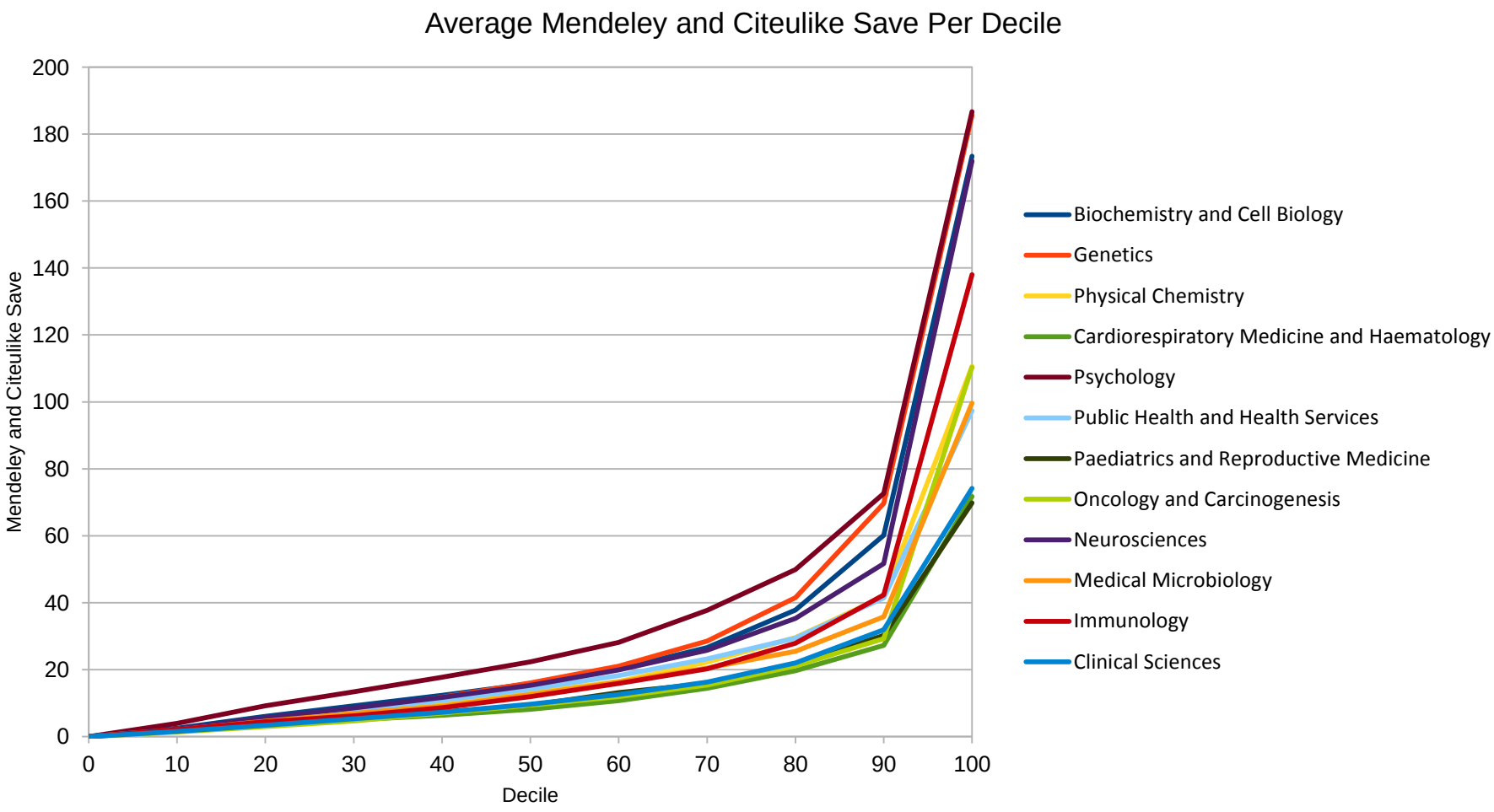
articles in deciles – Social Network Activity



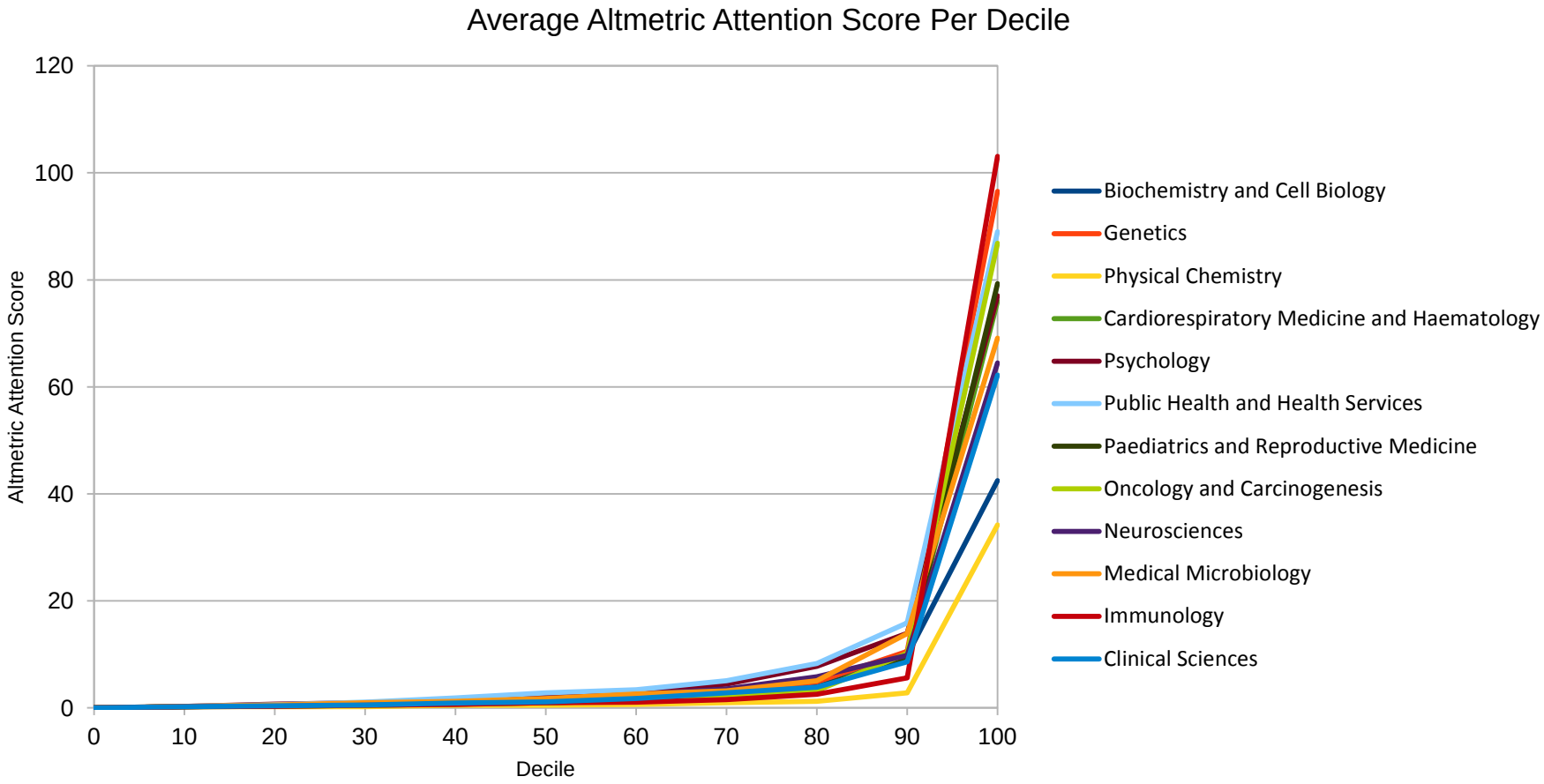
Mean values in deciles - citation



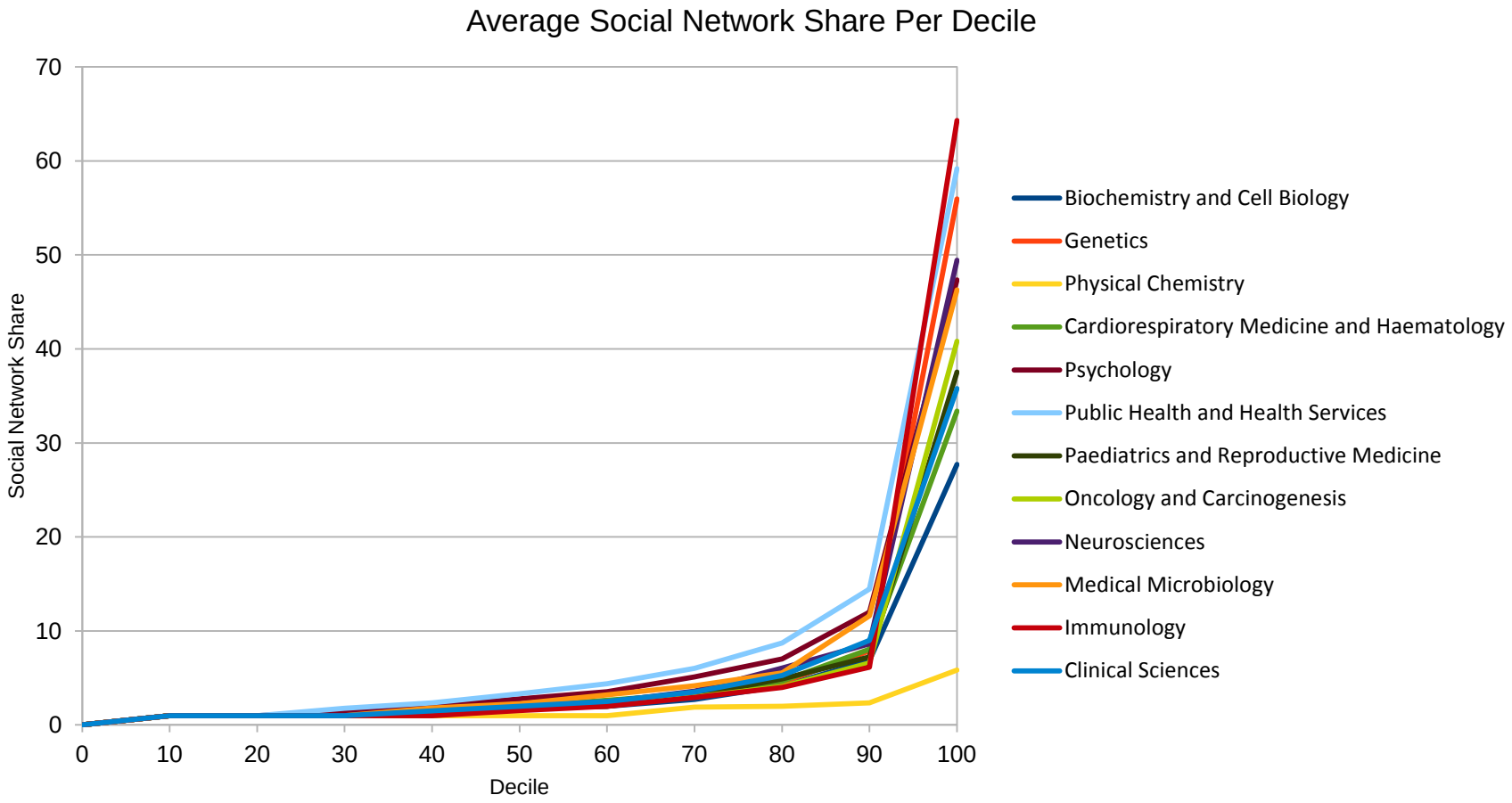
Mean values in deciles – Mendeley & Citeulike



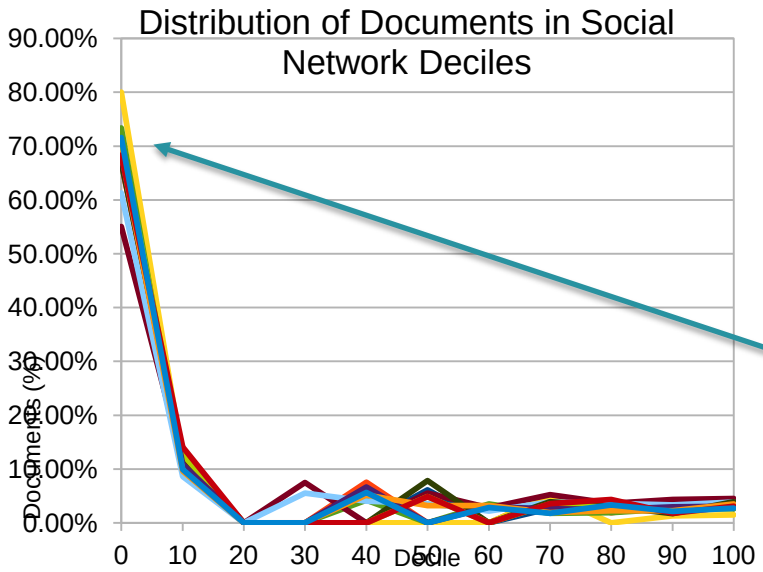
Mean values in deciles – Altmetric Attention Score



Mean values in deciles – Social Network



Observations using Social Network data



Only limited similiarity between high log growth and low zero %

Is there equivalence between a mean score of 5, and a mean score of 65?

Upper-range trend approximates to log growth ($y = x^{m_2}$)

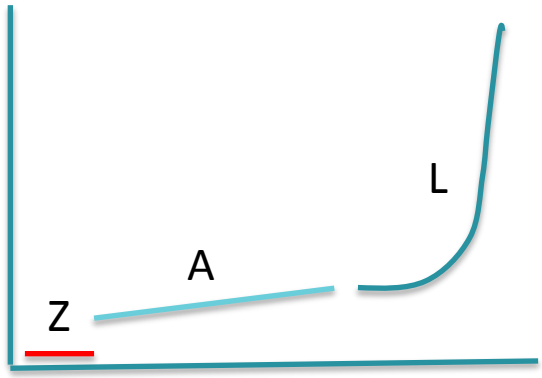
Mid-range trend approximates to arithmetic growth ($y = m_1 x$)

To some extent, these observations can be made to all four data source, and for all disciplines

The order of the disciplines in the upper decile is different to mid-range deciles

Difference between max and min in upper decile is multiple of mid-range

Research questions



- Does each data source and each discipline exhibit three Zero, Arithmetic, Log (ZAL) features?
- Can we identify the range where log growth exceeds arithmetic?
  - (Look at linear and log divisions, regression analysis, confidence)
- Are growth states independent? Does an article with log growth require an arithmetic stage? Is it an error to see arithmetic and log growth as being the same type of activity?
- Can we reliably quantify m_1 and m_2 for the different disciplines? (And probability P for zero-values)
- Should platforms allow for analysis using ZAL values?
- For different data sources, the balance between ZAL importance will shift: eg, Z is less important for Mendeley, L will rarely be an issue for policy documents. How can we support analysts to use the most relevant metric?
- Does it matter if very few chemistry papers are ‘highly shared’, compared to psychology, compared to a baseline of 10%