

MathSys Research Study Group project: Exploratory analysis of research data and outputs

External partner: Researchfish – Dr. Vera Hazelwood

Academic convenor: Prof. Colm Connaughton

EPSRC are currently investing £161 M in mathematical sciences research in the UK. It is reasonable to ask why this money is better invested in mathematical research rather than in frontline public services? Providing quantitative evidence of the positive impacts of mathematical sciences research on society and on the economy is essential to answering this question. Researchfish possesses a unique data set [3] recording outputs and impacts from EPSRC-funded research grants which can help with this. The aim of this project is to perform some exploratory analyses of the Researchfish data. We will address questions such as the following:

1 - Quantifying timescales to impact

The London Mathematical Society in its submission of evidence [1] to the 2010 EPSRC-sponsored International Review of Mathematical Sciences stated that an “often overlooked feature of fundamental research is the length of time before pay-off: in the mathematical sciences the time gap can often be extremely long”. The implication is that mathematical research is somehow special in terms of the time society should expect to wait before seeing the benefits. An illustrative figure of 10 – 50 years suggested. On the other hand, a quantitative analysis of the impact case studies submitted to the 2014 REF [2] performed by Kings College London’s Public Policy Institute found that the time lag between the publication of the underpinning research and the subsequent impact in the Mathematical Sciences Unit of Assessment was broadly comparable to the time lag observed for other scientific disciplines like Physics, Chemistry and Medicine. The authors of [2] however point out that these results may have been skewed by the way in which the REF assessment process was designed. We will use the Researchfish data, which is not biased by the REF assessment criteria, to try to study this conundrum in a statistically robust way. Do we really need to continue to make the “long term” case for mathematical research or can we make a case with more immediate value? Ultimately the aim is to determine how best to demonstrate the value of mathematical sciences research to policy-makers and to the public.

2 – Visualisation of complex data sets

Researchfish is also interested in developing meaningful visualisations of their dataset, so qualitative comparisons are available “at a glance”. An example would be developing radar charts to represent Researchfish data for different levels and types of aggregation

(researcher, university, funder, country, area of research etc). It is possible that this approach would help to answer the question above¹.

3 – Predictive analytics

We are also interested to explore the predictive power of our dataset, for example, what are the key characteristics of a research award that determine its success in certain areas of impact (e.g. generation of new knowledge vs knowledge exchange or policy influences). This would enable our data to be used for strategic decisions by funders when they wish to design a research programme to achieve certain impacts. For example, a common question by funders is whether it is more efficient to fund few large grants or many smaller grants to achieve specific outcomes.

Technical skills required:

Basic descriptive and inferential statistics, dimensionality reduction, clustering and feature detection, data visualisation, random forests.

References:

- [1] EPSRC, *International Review of Mathematical Sciences PART III: Information for panel – responses to stakeholder consultation*. (2010)
- [2] King’s College London and Digital Science, *The nature, scale and beneficiaries of research impact: an initial analysis of the Research Excellence Framework (REF) 2014 impact case studies*. (2015)
- [3] Researchfish Common Question Set, 25.11.18 (attached)

¹ Due to the nature of the data, for some questions, it may be necessary to use (from a code library or developed by students) a “key word” search algorithms to identify awards meeting specific criteria.