Abstract—**Citation-based research performance reporting is contentious. The methods used to categorize research and researchers are misleading and somewhat arbitrary. This paper compares cohorts of social science categorized citation data and ultimately shows that assumptions of comparability are spurious. A subject area comparison using research field distributions and networks between a 'reference author', bibliographically coupled data, keyword-obtained data, social science data and highly cited social science author data shows very dissimilar field foci with one dataset very much being medically focused. This leads to the question whether subject area classifications should continue to be used as the basis for the plethora of rankings and lists that use such groupings. It is suggested that bibliographic coupling and dynamic topic classifiers would better inform citation data comparisons.**

Keywords—bibliographic coupling, research performance, social networks, citations, social science

I. INTRODUCTION

Research performance assessment goes hand-in-hand with citation counting but this is not without controversy [1-3]. This practice of counting citations is used explicitly to infer such things as research quality, impact, influence and, to the dismay of many, informs many author and university ranking systems [4, 5]. In short, citation counting is a blunt instrument used to describe a myriad of research assessment indicators. This is not likely to change anytime soon because document citation counts are easily accessible via citation databases such as Scopus and Web of Science (WoS).

Using citation counting to compare the quality, influence or value of two or more documents, authors or, institutions is problematic. This is evidenced by the myriad of factors or weightings (i.e. adjustments) that have been applied to raw citation counts with the hope of making the measures fairer or more meaningful [6-9]. Chief among these adjustments are time and field-based factors. The objective of adjusting citations for time has been described by [7] and is rightfully attempting to ensure that the influence of time can be accounted for – older documents have had more time to accrue citations. According to the Snowball metrics recipe book field-weighted citation impact takes into account the differences in research behaviour across disciplines [10]. The recipe book reiterates the accepted notion that "*researchers working in fields such as medicine and biochemistry typically produce more output, with more co-authors and longer reference lists, than researchers working in fields such as mathematics and education; this is a reflection of research culture, and not performance*". These behavioural traits are important when comparing between disciplines.

While time and field adjustments are understandable, it is the latter that are thought-provoking because it gives the end-user of citation data the impression that all outputs in a chosen field can or should be compared in the same way because they have been *weighted*. This is not the case because citation rates are multi-factorial and result from more than just the broad research field you happen to be pigeon-holed into [11]. Dynamics such as niche or emerging research areas [12], the country of the transcript [13], co-authorship networks [14, 15] all play a role even if classified within the same designated field.

The two most widely referred citation databases in the scientific literature are Scopus and WoS. These two citation databases refer to over 300 and 250 subject areas respectively. However much of the citation related research performance data is reported at broad subject category levels rather than more granular sub-category levels. This is the case for well-publicised reports such as the world's most "Highly Cited Researchers" list (Clarivate) which uses the WoS database and international ranking systems such as "QS World University Ranking" (QSWorld) and "The Times High Education World University Rankings" (Times) which rely on Scopus data for rankings.

A consequence of using broad and relatively fixed subject fields is that you end up comparing apples and oranges even if citations are *field-weighted*. This paper will demonstrate that the scientific community need to approach citation-derived lists, tables and rankings with a high degree of caution. Caution is vitally important as various bodies unwittingly accept these apples and oranges comparisons without question because they are presented as a nice package.

The approach taken in this study uses data from the well publicized 2017 Highly Cited Researcher (HCR) compilation by Clarivate. The study will compare a sample of HCRs classified in the field "Social Sciences, general" against other social scientists by gathering associated subject area data. The aim of the study and the reason this study is important is to highlight how easily it is to be misled by publish lists and to echo a proven method of minimising apples and oranges comparisons. The study contributes to ongoing efforts to produce meaningful comparison data by continuing to question the validity of current ranking lists and tables.

A classic example of misleading information, the 2017 HCR lists amongst its highly-cited researchers in the field "Social Sciences, general" Professor Adrian Bauman. Professor Bauman is the Sesquicentenary Professor of Public Health within the Sydney School of Public Health at the University of Sydney, Australia. The university's website lists Professor Bauman as being involved in a "*diverse research program on chronic disease prevention and methods research*" with themes and keywords including: Obesity, Diabetes and Cardiovascular Disease; Healthy Ageing; Lifespan; Physical activity; Public health; Epidemiology; Preventive health services; Research translation.

WoS data on Professor Bauman shows that as at July 2018 46% of his indexed outputs were classified in the WoS category of "Public, Environmental & Occupational Health". This WoS category happens to be part of the Social Science Citation Index and therefore partly mapped to the subject area of "Social Sciences, general" in the list of Essential Science Indicators (or ESI). Interestingly, Scopus (the other main citation provider) has a very similar field "Public Health, Environmental and Occupational Health" which is mapped to "Medicine" in their 'All Science Journal Codes' (ASJC) schema. ASJCs are used throughout this study.

II. METHOD

The WoS categorisation results in the 2017 "Social Sciences, general" HCR list being seemingly dominated by medicine-based researchers. To investigate this possibility, 29 of the 2017 "Social Sciences, general" HCR list out of the 188 given were arbitrarily chosen. Using Scopus as the citation provider, their research output categorisations were analysed; the 29 selected included Professor Bauman from the lead author's home institution. The resultant dataset was called the 'HCR dataset'.

Bibliographic coupling (BC) is a technique that identifies researchers who have cited the same research, thus inferring similar research topics. This technique is not new and was introduced by M. M. Kessler of MIT in 1963 [16]. Using Scopus as the citation provider, a prominent social science researcher from the lead author's home institution was chosen as the reference author (RA). All 2012 to 2017 references for RA's 2013 to 2018 research outputs were established. All other 2013 to 2018 research outputs that cited the same 2012 to 2017 references were found and created the 'BC dataset'.

Another dataset involving the same time period was created via Scopus using the top 6 author keywords found in RA's outputs; this was termed the 'KW dataset'. The 6 author keywords used were "public transport", "flexible transport service", "neighbourhood" (American and English spellings), "bus rapid tran*" (and "BRT"), "light rail tran* (and "LRT"), "geographically weighted regression" (and "GWR"). The "*" wildcard was used to capture a variety of 'tran' words e.g. 'tran*sit*', 'tran*sport*', 'tran*sportation*' etc…

Finally, 3 arbitrarily chosen but prolific 'Social Science' researchers listed in the Scopus database had their 2013 to 2018 research outputs obtained; this formed the final comparison dataset 'SocSci dataset'.

In summary, the data gathering resulted in FIVE comparison datasets sets of varying sizes: 1. RA dataset (68 outputs), 2. BC dataset (3,572 outputs), 3. KW dataset (7,548 outputs), 4. SocSci dataset (296 outputs) and 5. HCR dataset (2,320 outputs).

These datasets were investigated for their ASJC distributions and networks. The ASJC networks were created using Gephi network analysis software and examined network centralities and modularity. Output ASJC information is found in the Scopus data and thus became the ASJC network nodes and links per paper per ASJC.

III. RESULTS

The first result of note showed that 90% of the arbitrarily chosen HCRs had the majority of their outputs within the WoS category 'Public, Environmental & Occupational Health'. This is problematic because it suggests the 2017 HCR "Social Sciences, general" list is a cluster of medical practitioners rather than social scientists. Social Science is the study of human society and social relationships and not specifically a medical field but nor does it explicitly exclude medicine. Nevertheless the purpose of this study is not to critique definitions or subject categories per se, but it is to highlight that in order to compare the citation counts of researchers and institutions it is important to compare like with like and that *field-weightings* are likely to be inadequate to be useful and have meaning.

When comparing the ASJC distributions of all 5 datasets the results showed that there was little relationship between the ASJC distributions of the HCR dataset versus the other 4 datasets when using the RA top 10 ASJCs as the reference (Fig. 1 and Table 1). The RA dataset had the closest relationship with the BC Dataset ($r = 0.84$, $p < 0.05$), KW dataset ($r = 0.71$, $p < 0.05$) and the SocSci Dataset ($r = 0.6$, $p < 0.05$).

When investigating the ASJC networks for each dataset the leading nodes (i.e. ASJCs) were identified by ranking network centralities. Significant ASJC clusters were identified using modularity class. Table 2 lists the overall numbers for each dataset. Node count tells the reader how many ASJCs are represented in each dataset with the KW dataset being the largest and most diverse.

The results of the node rankings for each dataset is given in Table 3 and is presented at the *major* field level for each ASJC in the top 10 of each dataset. The data is presented in this way because this is the field/subject level most commonly used in published lists and ranking systems, as mentioned in the introduction. Table 3 highlights the prominence of the social science and medicine related fields for each dataset. The results show that 7 of the top 10 fields for the HCR dataset are within Scopus classified medical fields whereas 6 of the top 10 for the SocSci dataset are within fields classified by Scopus as social science. Our reference author (RA dataset) has no medical fields in the top 10 but has 4 social science fields (Table 3). Also, for each dataset, when comparing the *unique* major field occurrences in Table 3, 4 out 6 RA major fields are found in the BC and SocSci datasets, 2 in KW dataset and 1 in HCR dataset. This suggests a better (superficial) match between RA, BC and SocSci datasets.

IV. DISCUSSION

Citation data is seemingly ubiquitous and, as mentioned in the introduction, it is used to inform a multitude of research performance criteria. The results comparing ASJC information across datasets that should be related has highlighted that their 'major' fields of research are decidedly different, let alone their 'minor' fields

In the case of the 2017 HCR list for 'Social Science, general' the data in this study shows that most of the ASJCs lie within Scopus classified medical fields. Given it is accepted that medical related research gains a higher rate of citation compared to fields within social science, is it any wonder than that the Social Science HCR list is dominated by medical researchers. A prolific social science researcher such as RA in this study will find it challenging to appear on a social science HCR list by Clarivate.

It is acknowledged that a quirk of the WoS categories is the field 'Public, Environmental & Occupational Health' classified as a 'Social Science' but this situation also highlights the somewhat arbitrary and differing nature of citation-based lists and rankings. Furthermore, if an output is even partly related to a social science then this should be evident in the network data showing links between 'medicine' fields and 'social science' fields. It is important to understand that outputs are assigned subject fields based on the outlet (journal etc.) it is published in; these outlets can have one or several fields associated with them – which is the one purpose of *field-weightings*. This multi-ASJC association allows research field links to be generated in network analysis. The HCR dataset network is shown in Fig. 2 and shows the dominance of the medical fields (shown red) compared to social science fields (blue). In the figure, node sizes are proportional to the number of connections and shows how well-coupled medical fields are within this dataset. Compare the HCR network with the Scopus SocSci network (Fig. 3) and its dominance of the 'blue' social science nodes. If the HCR data is theoretically representing social science, the data presented here suggests it falls short. In the same way, when an institution is said to be the highest ranked in a particular field the question needs to be asked – what are they actually being compared and ranked in?

The RA in this study has 86% of their research outputs within the ASJC field of 'Transportation'. Whilst transportation is a sub-field of the social sciences our RA will be hard pressed to fly above the social science pack when some of the pack researches in much more popular medical fields. The consequence of this is that individual researchers and institutions alike should be assessed, not on some arbitrary broad field classification but, on the type of research they are actually performing. If this means that named and/or classified fields are dynamic and change year on year, then so be it. The systems exist and the literature tells us that nuanced citation analysis is possible [17].

V. CONCLUSION

The time for institutions to disregard broad and misleading citation based lists and rankings is near. In this study the folly related to data in one broad based field, 'social science', has been demonstrated. If understanding individual or institutional performance is an aim then the correlation results of the RA and BC datasets suggests that bibliographic coupling provides a much more accurate picture of an individual's standing rather than subject area derived analysis. It is likely that this would also be true of institutional level comparisons. This study outlines an uncomplicated process in order to approach desired comparisons. The techniques in this study are not new but the importance of highlighting fallacious research performance reporting cannot be understated. Future work in this area could concentrate on establishing author publication profiles to facilitate like-with-like comparisons.

ASJC Distribution

ASJC	Class Name	RA Dataset	BC Dataset	Group KW dataset	SocSci Dataset	HCR Dataset
3313	Transportation	86.17%	33.35%	7.64%	16.26%	0.85%
3305	Geography, Planning and Development	21.28%	19.66%	12.89%	28.57%	1.82%
2205	Civil and Structural Engineering	20.21%	15.46%	3.54%	7.64%	0.12%
2001	Economics, Econometrics and Finance (misc..	18.09%	1.62%	0.68%	0.25%	
2308	Management, Monitoring, Policy and Law	10.64%	3.56%	2.87%	3.45%	0.25%
3322	Urban Studies	10.64%	7.18%	7.10%	9.85%	0.32%
3303	Development	8.51%	4.24%	2.57%	7.64%	0.32%
1803	Management Science and Operations Resea..	8.51%	6.86%	5.04%	2.22%	
2300	General Environmental Science	6.38%	7.56%	3.13%	7.88%	0.95%
3300	General Social Sciences	4.26%	2.04%	2.12%	15.27%	0.60%

Fig. 1. Dataset ASJC distributions

TABLE 1 - CORRELATION TABLE FOR THE HCR DATASET

Pearson's correlation (r)	
vs	*HCR Dataset*
RA Dataset	-0.0618
BC Dataset	0.3386
KW dataset	0.4103
SocSci Dataset	-0.1345

TABLE 2 - OVERALL NETWORK NUMBERS

Dataset	*Nodes*	*Edges*
RA	30	220
BC	230	12,024
KW	290	26,363
SocSci	89	986
HCR	212	5,840

TABLE 3 – MAJOR FIELD ASJC NODE RANKING FOR EACH DATASET TOP 10

RA	*BC*	*KW*	*SocSci*	*HCR*
Social Sciences	Social Sciences	Social Sciences	Social Sciences	Medicine
Social Sciences	Social Sciences	Computer Science	Social Sciences	Medicine
Engineering	Medicine	Computer Science	Social Sciences	Medicine
Economics,Econometrics and Finance	Engineering	Social Sciences	Social Sciences	Medicine
Social Sciences	Environmental Science	Social Sciences	Economics,Econometrics and Finance	Nursing
Decision Sciences	Decision Sciences	Computer Science	Engineering	Medicine
Business,Management and Accounting	Social Sciences	Computer Science	Economics,Econometrics and Finance	Medicine
Social Sciences	Engineering	Medicine	Social Sciences	Social Sciences

| Environmental Science | Medicine | Engineering | **Social Sciences** | Medicine |
| Engineering | Environmental Science | **Social Sciences** | Environmental Science | Health Professions |

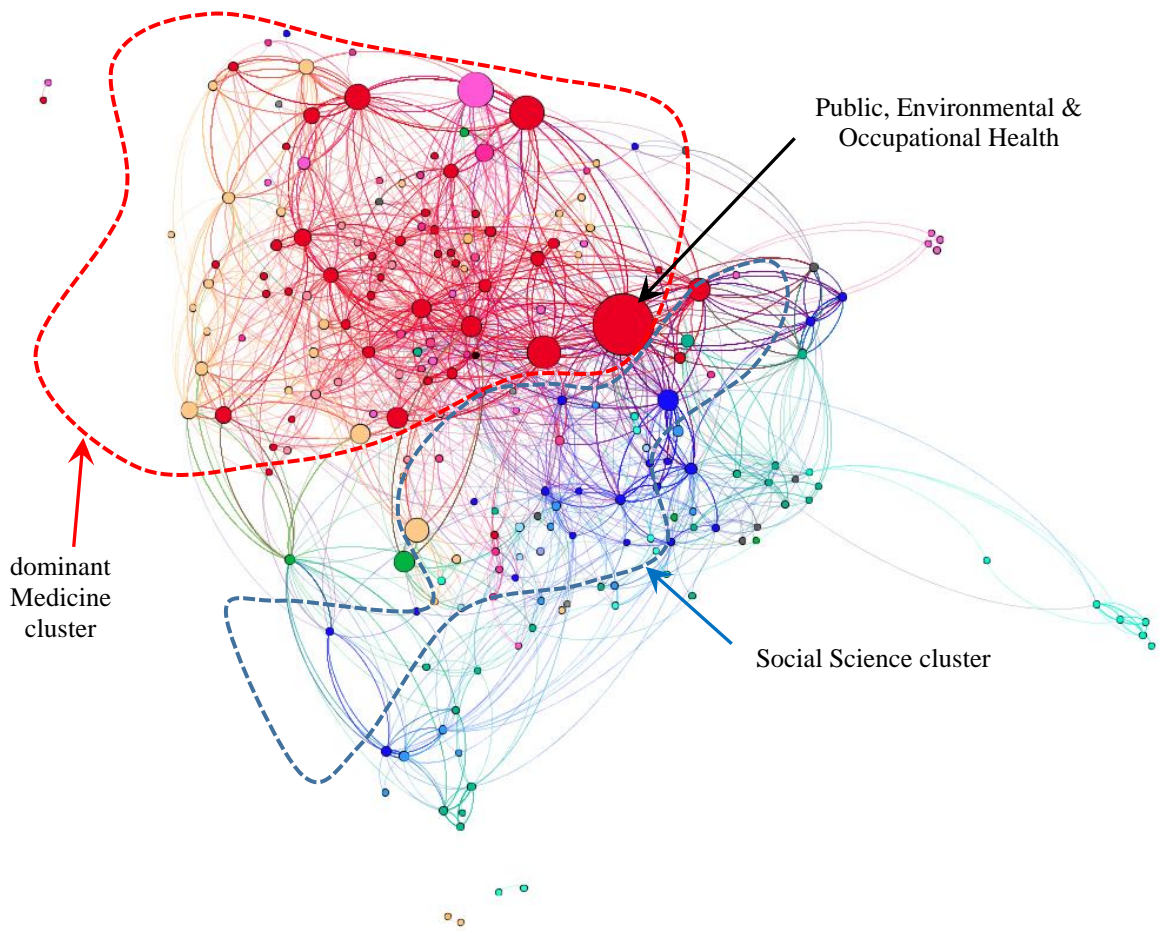
Fig. 2. HCR dataset ASJC network

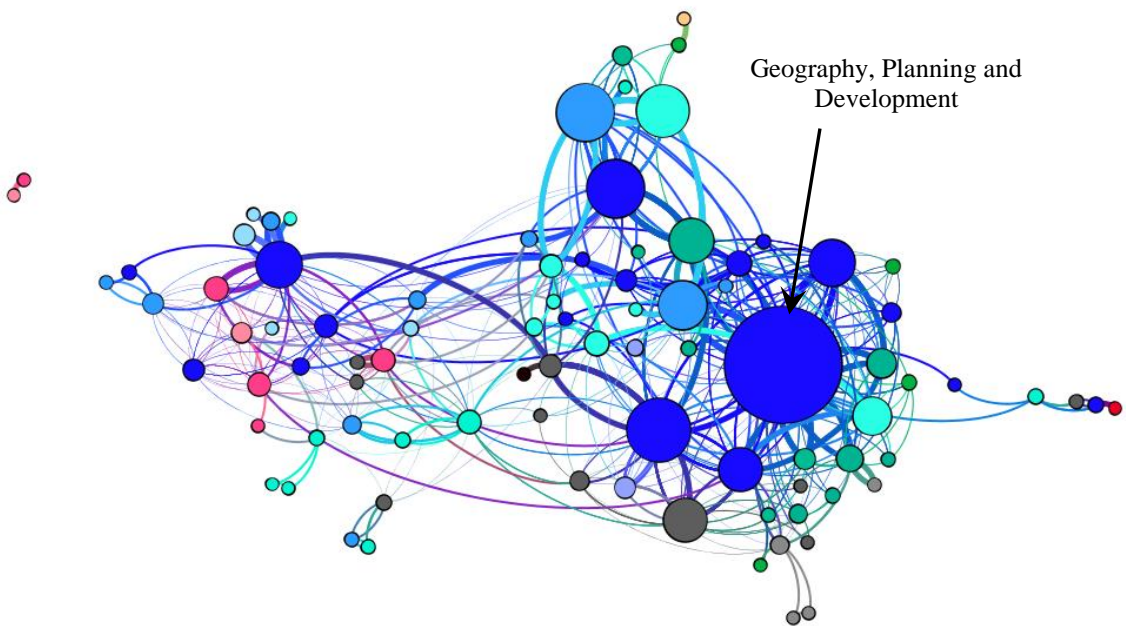Fig. 3. SocSci dataset ASJC network (blue nodes denote Social Science related)

REFERENCES

[1]   Seglen, P.O., Citation rates and journal impact factors are not suitable for evaluation of research. Acta Orthopaedica Scandinavica, 1998. 69(3): p. 224-229.

[2]   MacRoberts, M.H. and B.R. MacRoberts, Problems of citation analysis. Scientometrics, 1996. 36(3): p. 435-444.

[3]   Martin, B.R., The use of multiple indicators in the assessment of basic research. Scientometrics, 1996. 36(3): p. 343-362.

[4]   Vernon, M.M., E.A. Balas, and S. Momani, Are university rankings useful to improve research? A systematic review. PLoS One, 2018. 13(3): p. e0193762.

[5]   Linton, J.D., R. Tierney, and S.T. Walsh, Publish or Perish: How Are Research and Reputation Related? Serials Review, 2011. 37(4): p. 244-257.

[6]   Blagus, R., B.L. Leskošek, and J. Stare, Comparison of bibliometric measures for assessing relative importance of researchers. Scientometrics, 2015. 105(3): p. 1743-1762.

[7]   Jin, B., Liang, L., Rousseau, R., Egghe, L. The R- and AR-indices: Complementing the h-index. Chinese Science Bulletin, 2007. 52(6): p. 855-863.

[8]   Yan, E. and D. Ying, Weighted citation: An indicator of an article's prestige. Journal of the American Society for Information Science & Technology, 2010. 61(8): p. 1635-1643.

[9]   Egghe, L., Theory and practise of the g-index. Scientometrics, 2006. 69(1): p. 131-152.

[10] Colledge, L., Snowball Metrics Recipe Book v7. Vol. 3. 2017: Elsevier. 161.

[11] Tahamtan, I., A. Safipour Afshar, and K. Ahamdzadeh, Factors affecting number of citations: a comprehensive review of the literature. Scientometrics, 2016. 107(3): p. 1195-1225.

[12] Daim, T.U., Rueda, G., Martin, H., Gerdsri, P. Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technological Forecasting and Social Change, 2006. 73(8): p. 981-1012.

[13] Xia, X., Wang, Z., Wu, Y., Ruan, L., Wang, L. Country of authorship and collaboration affect citations of articles by south and east asian authors in agronomy journals: A case study of china, japan, and india. Serials Review, 2014. 40(2): p. 118-122.

[14] Chuan-yi, W., Chengzhe, and C. Chen, The Effect of Co-authorship Network on Research Performance of Scholars. 2017: p. 1-5.

[15] Xu, G., X. He, and A. Butler. What comes first the co-Authorship network or the citation? 2017.

[16] Kessler, M.M., Bibliographic Coupling Between Scientific Papers. American Documentation, 1963. 14(1): p. 10-25.

[17] Uddin, A., Bhoosreddy, J., Tiwari, M., Singh, V.K. A Sciento-text framework to characterize research strength of institutions at fine-grained thematic area level. Scientometrics, 2016. 106(3): p. 1135-1150.