Citation Distance Matters: Towards a New Metric for Evaluating Journal Quality

Kate Barnes[1], Levente Török[2], and Roland Molontay[13]

[1] Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary
[2] Snap Inc. Budapest, Hungary.
[3] HUN-REN–BME Stochastics Research Group, H-1111, Budapest, Hungary
k_barnes@coloradocollege.edu, toroklev@gmail.com, molontay@math.bme.hu

1 Introduction

The primary metric for evaluating journals, the Journal Impact Factor (JIF), simply reflects the yearly mean number of citations received by articles [5]. Pure citation-counting metrics fail to distinguish the quality of the citations themselves. Some citations, such as most self-citations, are trivial while others, such as citations by high-impact works, are crucial. Furthermore, pure citation-counting metrics are easy targets for *citation gaming*, the coordination of groups of journals or authors who cite each other disproportionately to artificially inflate their scores.

Citations coming from far away in the academic network may indicate quality. The further a journal's reach, the larger its impact on science in general. Conversely, low average citation distances may identify low-quality or misbehaving journals. Here we compare five metrics calculated in the journal citation networks themselves. We find that journals suspected of citation gaming have lower average citation distances, while highly ranked journals have higher average citation distances. Journal impact metrics, such as the JIF, could be improved by weighting citations by their length in citation networks.

2 Data and Methods

2.1 Journal Citation Networks

We collected data from Microsoft Academic Graph (MAG), a longitudinal bibliographic databases of academic articles [1]. The data includes information about papers published between 2013 and 2019 and all papers cited by them. Academic citation networks vary with time [6], so we built 7 directed journal citation networks, one for each year of data, in which vertices are journals. A directed edge was added between two journals, $\{j_{citing}, j_{cited}\}$, if there is at least one paper from j_{citing} that cites j_{cited}. Edges are weighted as the number of citations from j_{citing} to j_{cited} over the total number of citations made by j_{citing}. Note that journal networks for a given year were built from the set of papers published in that year as well as all papers that they cite, therefore the networks span prior years. In total, we analysed 48,821 journals. Subject categories were assigned to the MAG data using the Web of Science, as shown in figure 1.

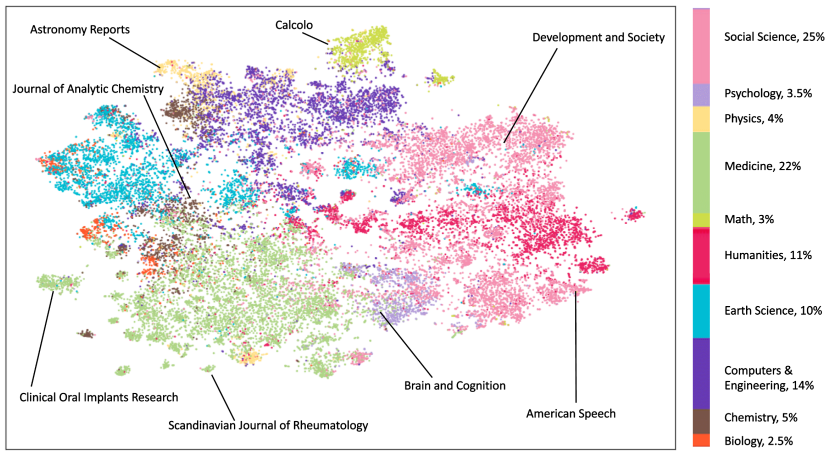

Fig. 1: Two dimensional t-SNE embedding of the MAG 2013 journals network, with located example journal titles. Distribution of disciplines in color legend.

2.2 Calculating Journal Citation Network Metrics

We explored 5 metrics assessing a journal's average distance from those that cite it. Two metrics use node embeddings, node2vec for d_1 and the t-distributed stochastic neighbor embedding (t-SNE) for d_2, and have the advantage of accounting for the overall network structure. The other metrics consider journals' immediate citations: the intersection over union (IoU) of journal reference lists (d_3), the similarity of journals' out-going and incoming citations (d_4) and the subject entropy of journals' incoming citations (d_5). All metrics were min-max normalized by discipline to account for variability in typical citation distances between disciplines. The result is 5 metrics describing each journal, all taking values between 0 and 1.

The **node2vec embedding** method is a way to encode nodes in a network as vectors. It uses random walks to describe each node in terms of its surrounding community. We initialized 200 random walks of length 30 starting from each node to embed the journals as 64-dimensional vectors. The node2vec embeddings for each journal were compared using the algorithm's distance metric [2].

We used **t-SNE** to reduce the node2vec embeddings to 3-dimensional vectors. Calculating distances in lower dimensional space may help with noise reduction. d_2 is defined as the Euclidean distance between points in the t-SNE space. This measure may capture large, perhaps interdisciplinary, distances but t-SNE is known to misrepresent close distances [4].

The **IoU** measure, d_3, highlights that two journals building on the same set of information are likely to be similar. This measure is calculated with the outgoing citations made by journals. In equation 1, R_1 is the set of journal ids cited by journal 1, and R_2 is the set of journal ids cited by journal 2. C_{out1} is the vector of edge weights for journal 1's outgoing citations, and C_{out2} the vector of edge weights for journal 2's outgoing citations. Both C vectors are ordered by journal id in order to compute the dot product.

$$d_3 = 1 - \frac{C_{out1} \cdot C_{out2}}{|R_1 \cup R_2|} \quad (1)$$

Distance measures d_1, d_2 and d_3 are calculated for every edge between a journal and those that cite it. The final metric for the journal is the average of these calculations.

d_4 compares journals' outgoing and incoming citations, the **citing and cited** journals. In citation stacking, anomalous pairs or groups of journals cite each other disproportionately. d_4 measures the amount of a journal's incoming citations from journals that it also cites frequently. In equation 2, C_{out} is the vector of edge weights for the journal's outgoing citations, and C_{in} is the vector of edge weights for the same journal's incoming citations. Both vectors are ordered by journal id. Then, d_4 is defined using the dot product.

$$d_4 = 1 - C_{out} \cdot C_{in} \quad (2)$$

d_5 captures the **subject entropy** of the journals' incoming citations. We expect high entropy to correspond with high citation distances due to more interdisciplinary citations. d_5 is defined as the Shannon's entropy of the distribution of disciplines that cite a journal. In equation 3, p_i, is the proportion of the given journal's references coming from discipline i, and this is calculated for all n disciplines.

$$d_5 = -\sum_{i=1}^{n} p_i \log(p_i) \quad (3)$$

2.3 Validation Data

Average citation distances were compared in terms of four validation data sets. Two data sets identify low quality journals: journals suspended by the Journal Citation Reports (JCR) due to fraudulent behaviour, and journals suspected of citation gaming by the CItation Donors and REcipients (CIDRE) algorithm [3]. The other two data sets identify high quality journals: those ranked in the upper quartile of all journals by Scimago (SJR), and the top ranked journals by the Norwegian Register of Scientific Journals (NRJ), compiled by a panel of experts. In the JCR and CIDRE data sets, low quality journals are labeled 1 and normal journals are labeled 0. In the SJR data, journals in the upper quartile ($Q1$) are labeled 1 and the rest ($Q2$, $Q3$ and $Q4$) are labeled 0. In the NRJ data the highest ranking journals (2) are labeled 1 and the other rankings (1 and 0) are labeled 0. All of the validations sets are imbalanced, with 5-25% of the data in class 1.

3 Results and Discussion

Table 1 shows t-tests and ROC-AUC results comparing high and low quality journals. The ROC-AUC is a performance metric for quantifying the ability of a feature to distinguish between binary classes. The direction of the differences between group means are as expected. High quality journals score higher, indicating they receive citations from father away in the journal citation networks. Low quality journals score lower overall, suggesting local rather than broad impact in the journal citation network.

	SJR			NRJ			CIDRE			JCR		
Metrics	0	1	AUC	0	1	AUC	0	1	AUC	0	1	AUC
node2vec, d_1	0.432	0.474	0.568	0.450	0.464	0.521	0.452	0.325	0.552	0.452	0.441	0.524
t-SNE, d_2	0.265	0.299	0.584	0.278	0.302	0.557	0.281	0.267	0.528	0.281	0.264	0.541
IoU, d_3	0.992	0.998	0.589	0.991	0.999	0.557	0.995	0.993	0.512	0.995	0.986	0.598
citing-cited, d_4	0.985	0.998	0.536	0.989	0.996	0.484	0.991	0.990	0.68	0.991	0.962	0.704
subj entropy, d_5	0.508	0.562	0.575	0.529	0.565	0.553	0.534	0.471	0.603	0.534	0.529	0.511

Table 1: ROC-AUC and differences in group means on 4 validation sets. Shaded cells indicate significant results, $p \leq 0.05$. SJR and NRJ top-ranked journals are labeled 1. Journals suspected of misbehaving by CIDRE and suspended by the JCR are labeled 1.

The disciplinary entropy measure, d_5, performed well. These results indicate that high quality journals have an impact outside of their own field. Furthermore, disciplinary entropy correlates with the graph distance measures (Spearman's coefficients d_1: $\rho = 0.47$, d_2: $\rho = 0.48$, d_3: $\rho = 0.36$), showing that longer distance citations are indicative of interdisciplinarity. Mean citation distances also differ between disciplines (ANOVA $p < 0.001$), but we normalized by discipline in order to account for this.

We are expanding this analysis to consider other ways of quantifying citation distance, as well as temporal analyses of how citation distances changed from 2013-2019. Finally, we will weight journals' citation counts with the best performing distance measures to evaluate the incremental predictive power of citation proximity over simple citation counts.

Summary. Highly ranked journals have higher average citation distances, while journals suspected of malpractice have lower average distances. These 5 metrics show promising results that citation distance is a proxy measure for journals' scientific quality.

References

1. Färber, M., Ao, L.: The microsoft academic knowledge graph enhanced: Author name disambiguation, publication classification, and embeddings. Quantitative Science Studies 3(1), 51–98 (2022)
2. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 855–864 (2016)
3. Kojaku, S., Livan, G., Masuda, N.: Detecting anomalous citation groups in journal networks. Scientific Reports 11(1), 14524 (2021)
4. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research 9(11) (2008)
5. McKiernan, E.C., Schimanski, L.A., Muñoz Nieves, C., Matthias, L., Niles, M.T., Alperin, J.P.: Use of the journal impact factor in academic review, promotion, and tenure evaluations. Elife 8, e47338 (2019)
6. Varga, A.: Shorter distances between papers over time are due to more cross-field references and increased citation rate to higher-impact papers. Proceedings of the National Academy of Sciences 116(44), 22094–22099 (2019)