



OPEN

Impact of medical technologies may be predicted using constructed graph bibliometrics

Lawrence Jiang¹✉, Ashir Raza¹, Abdel-Badiah El Ariss², David Chen³, Nicole Danaher-Garcia⁴, Jarone Lee^{2,5} & Shuhan He^{2,4,5,6}

Scientific research is driven by allocation of funding to different research projects based in part on the predicted scientific impact of the work. Data-driven algorithms can inform decision-making of scarce funding resources by identifying likely high-impact studies using bibliometrics. Compared to standardized citation-based metrics alone, we utilize a machine learning pipeline that analyzes high-dimensional relationships among a range of bibliometric features to improve the accuracy of predicting high-impact research. Random forest classification models were trained using 28 bibliometric features calculated from a dataset of 1,485,958 publications in medicine to retrospectively predict whether a publication would become high-impact. For each random forest model, the balanced accuracy score was above 0.95 and the area under the receiver operating characteristic curve was above 0.99. The high performance of high impact research prediction using our proposed models show that machine learning technologies are promising algorithms that can support funding decision-making for medical research.

Scientific research is a driving force of innovation designed to expand the frontiers of human knowledge and improve economic and social progress. However, research policy and exploration of promising research directions are shaped in part by the decisions of funding bodies, such as governments and universities, as well as for-profit and nonprofit private entities, to fund these research studies^{1,2}. Determining which proposed research projects are funded based on impact remains a dynamic process that involves a combination of peer review and quantitative research metrics^{2,3}. The funding decision process requires transparency in the way public research funds are allocated based on peer review and metrics. Transparency is needed to ensure reproducibility through increased use of publicly available data for responsible decision-making^{4,5}.

Standardized citation metrics of research articles in a scientific field and scientist profiles may be used in part to inform the decision-making to fund new scientific projects^{6–8}. Using the number of citations alone as the sole quality indicator of research is limited due to its narrow scope that only measures the uptake of the work by other researchers^{9,10}. Moreover, citations are lagging indicators of research impact that vary widely by journal and scientific field¹¹. The limitations of relying solely on standardized citation metrics in the funding decision process can have a significant impact on the prospective development of science. Funding bodies may overlook potentially impactful research projects that are not immediately recognized by the scientific community. Additionally, researchers may prioritize producing work that is more likely to be cited, rather than pursuing research that is more innovative or impactful, which can slow scientific progress¹².

Recent research to extract signals from network science-enabled knowledge graphs has been used to quantify domain knowledge in health and life sciences¹³, materials science¹⁴, and drug discovery¹⁵. Recent work by Weis and Jacobson¹⁶ demonstrated the promising performance of leveraging knowledge graph dynamics and machine learning algorithms to a biomedical-focused dataset to identify innovative research of likely future importance. Additionally, machine learning algorithms have been used to identify seminal research of likely future significance in drug discovery. To evaluate the performance and reproducibility of using machine learning-based classification of high-impact research studies, we designed supervised random forest models trained on graph bibliometrics to predict high-impact research studies.

¹Duke University, Durham, USA. ²Department of Emergency Medicine, Massachusetts General Hospital, Boston, USA. ³Temerty Faculty of Medicine, University of Toronto, Toronto, Canada. ⁴School of Healthcare Leadership, Institute of Health Professions, Boston, USA. ⁵Trauma, Emergency Surgery, Surgical Critical Care, Massachusetts General Hospital, Boston, USA. ⁶Lab of Computer Science, Massachusetts General Hospital, Boston, USA. ✉email: lawrence.jiang@duke.edu

Methods

We collected metadata on 1,485,958 publications with non-null author, title, and ISSN data from 40 high-impact medical journals, listed in Table 1, between 1980 and 2020¹⁷. Data was collected from the APIs of Lens Lab and Elsevier. Since the data was collected in 2021, 2020 was the last year with complete data at the time. Articles without sufficient data on the date they were published were removed. A schematic on the search process of publications included in this study is shown in Fig. 1.

We used Neo4j database and Weis and Jacobson methodology to create nodes and edges for each publication, author, and institution¹⁶. From the processed database, we generated the same 28 bibliometrics shown in Table 1 of the Weis and Jacobson methodology for each article in our training dataset. The bibliometrics are also available as Table 2 of this article. These constructed graph bibliometrics serve as input features for the proposed

The New England Journal of Medicine	The Lancet
JAMA—the Journal of the American Medical Association	The BMJ—the British Medical Journal
Annals of Internal Medicine	JAMA Internal Medicine
PLOS Medicine	Lancet Oncology
World Psychiatry	Lancet Neurology
Journal of Clinical Oncology	European Heart Journal
JACC—Journal of the American College of Cardiology	Lancet Infectious Diseases
Lancet Diabetes and Endocrinology	Circulation
Lancet Respiratory Medicine	Gastroenterology
Gut	JAMA Oncology
European Urology	JAMA Psychiatry
American Journal of Psychiatry	Circulation Research
Hepatology	American Journal of Respiratory and Critical Care Medicine
Blood	Journal of Allergy and Clinical Immunology
Annals of the Rheumatic Diseases	JNCI—Journal of the National Cancer Institute
Journal of Hepatology	Intensive Care Medicine
Diabetes Care	Annals of Oncology
Leukaemia	Lancet Psychiatry
European Respiratory Journal	Brain
JAMA Pediatrics	JAMA Neurology

Table 1. Journals in dataset.

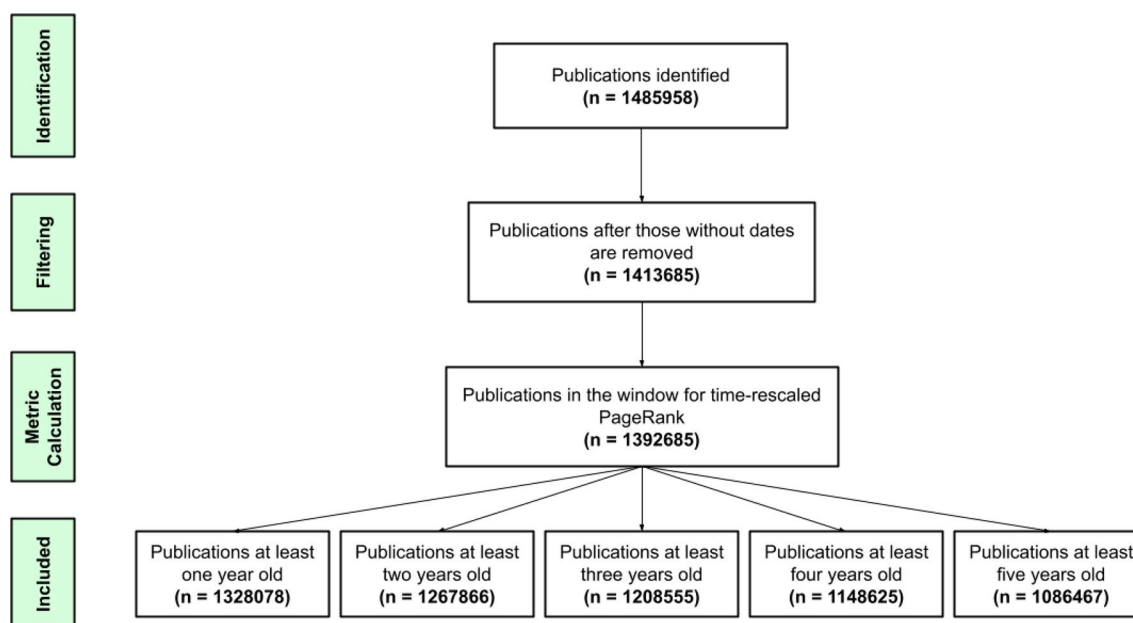


Figure 1. Eligibility of publications. On the date the data were pulled, there were 1,485,958 publications in the 40 journals between 1980 and 2020. Of those, we did not have sufficient information on the date of publication for 72,273 articles. In calculating time-rescaled PageRank, 21,000 randomly selected publications were placed outside the sliding window.

Variable	Description
Citations per paper	Mean number of citations per paper for papers the author has published
Δ Citations per paper	Change in the mean number of citations per paper for the author over the preceding 2 years
Citations per year	Average number of citations per year for papers the author has published
Maximum citations	Maximum number of citations a paper has received out of all the papers the author has currently published
Rank citations per year	Rank of the author among all other authors in terms of mean citations per year
Total citations	Number of citations author has received
Δ Total citations	Change in the total number of citations for the author over the preceding 2 years
Total papers	Total number of papers published by the author
Δ Total papers	Change in the total number of papers over the preceding 2 years
Citations	Citations collected in the current year
Adopters	Number of unique citing authors in the current year
Author age	Number of years since the year of publication of the author's first paper
h-index	Author's h-index
Δ h-index	Change in the author's h-index over the past 2 years
Recent coauthors	Number of coauthors the author has had in the current and immediately preceding year
Δ Mean journal citations per paper	Two-year change in the mean number of citations per paper of the journals the author has published in
Mean journal citations per paper	Mean number of citations per paper for the journals the author has published in
Δ Mean journal h-index	Two-year change in the mean h-index for the journals the author has published in
Mean journal h-index	Mean h-index for the journals the author has published in
Mean journal maximum citations	Mean of the maximum number of citations any paper published in a journal has received for each journal the author has published in
Mean journal rank citations per paper	Rank of journal in which the author has published, as determined by the mean number of citations per paper
Mean Δ journal total papers	Change in the mean of the total number of papers published in journals the author has published in
Total journals	Total number of journals published in by the author
Mean journal total papers	Mean of the total number of papers published in journals the author has published in
Learned network embedding	Unsupervised embedding of local network structure calculated via application of the node2vec algorithm on the citation graph
Time-rescaled node centrality	Time-balanced network centrality calculated using the full citation network
Unweighted PageRank	PageRank score of author, calculated on the unweighted coauthorship network
Weighted PageRank	PageRank score of author, calculated on the weighted coauthorship network

Table 2. Input features. Adapted from Ref.¹⁶.

machine learning based classifier of high-impact publications. A diagram of the general workflow of the data pre-processing pipeline is shown in Fig. 2.

We calculated the Node2vec, a vector representation for nodes based on random walks, for each author node using Neo4j's built-in function. We labeled publications in the top 5% of one feature, time-rescaled PageRank, as high-impact. As a node centrality metric that accounts for temporal bias, time-rescaled PageRank allows for distinction of high-impact articles controlling for their publication dates²⁸. To calculate the time-rescaled PageRank metric, we used a sliding average and standard deviation in a sliding window to rescale each publication temporally. In the setup of the sliding window, we placed 21,000 randomly selected publications outside the window.

Using the Python package scikit-learn 1.1.2, we trained random forest supervised classification models with different amounts of post-publication data. We applied imbalanced-learn's synthetic minority oversampling (SMOTE) function to the data using default parameters to account for the imbalanced nature of high and low-impact publications. In total, we trained ten different supervised random forest models. Six models were trained to predict high-impact publications five years after publication and four models were trained to predict high-impact publications using post-publication data from the year of publication. We used a hyperparameter grid search with a tenfold cross-validation procedure to identify the optimal hyperparameter values based on guidance from Van Rijn and Hutter for max features, min samples leaf, and split criterion³⁶. After applying SMOTE, we split the data into training and testing sets with a test size proportion of 0.25. Then, we evaluated models on the test set with balanced accuracy (BA) and area under the receiver operating characteristic curve (ROC AUC). We additionally evaluated variable importance with a mean decrease in impurity test on the 0-year data, 5-year label model.

Results

We evaluated random forest models on the test split with BA and ROC AUC scores. Overall, the ROC AUC and BA scores for each model was above 0.99 and 0.95 respectively (Table 3).

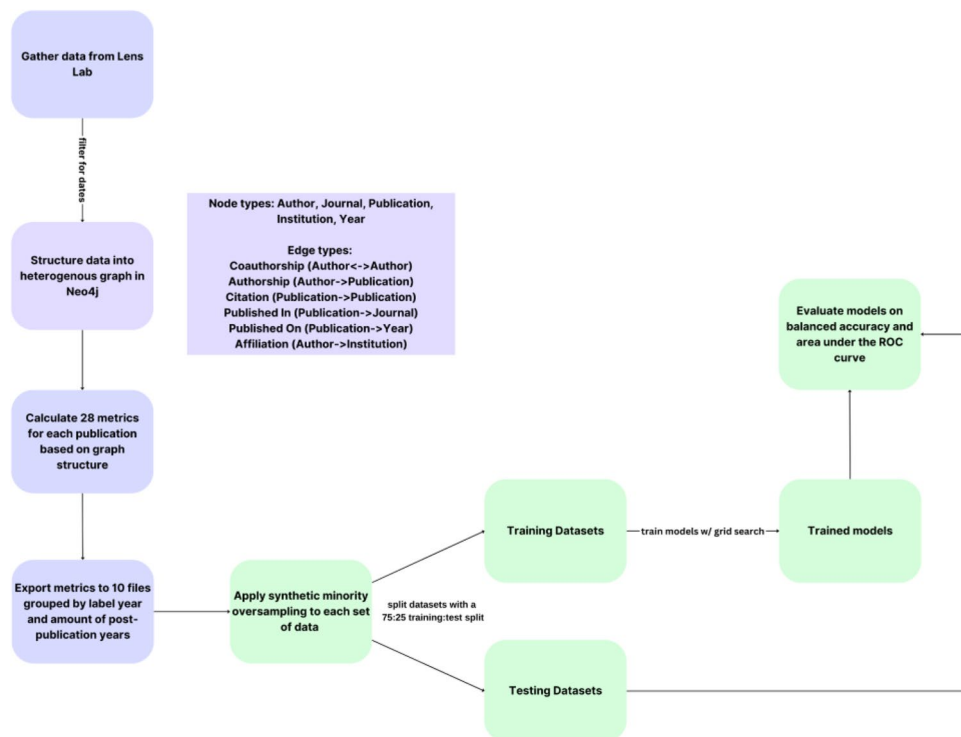


Figure 2. Workflow schematic.

Post-publication years	Label year	Total	True positives	True negatives	False positives	False negatives	Balanced accuracy	Receiver operator characteristic
0	5	516,072	252,452	240,696	17,149	5775	0.955564	0.991447
1	5	516,072	254,183	245,824	12,021	4044	0.968859	0.995998
2	5	516,072	255,667	249,337	8508	2560	0.978545	0.997989
3	5	516,072	256,540	252,953	4892	1687	0.987247	0.99173
4	5	516,072	257,257	255,434	2411	970	0.993447	0.999731
5	5	516,072	258,227	257,840	5	0	0.99999	1.0
1	1	617,046	302,923	297,771	10,808	5544	0.973501	0.997095
2	2	594,279	293,639	291,028	6308	3304	0.983829	0.998922
3	3	569,569	282,885	281,004	3921	1759	0.990029	0.999659
4	4	543,605	270,620	270,425	1580	980	0.995292	0.999899

Table 3. Outcomes.

Figure 3 showed that the models generally had higher BA and ROC AUC scores relative to the number of post-publication years. We note that there was a drop in ROC AUC in the 3-year post-publication/5-year label model, suggesting lower performance of the model at the three-year mark.

Figure 4 showed that the proportions of false classifications for each model were low. For models with a label five years post publication, the lowest proportion of false positives and false negatives were 0.0332 and 0.012, respectively, for the zero-year model. However, as more data were made available to models, the proportion of false classifications decreased.

The results of our mean decrease in impurity test are detailed in Table 4. The most important variable was rescaled PageRank. This is to be expected as it is the variable used to decide the label. However, its importance is still relatively small at 0.166. The importance of the next four variables sums up to a similar value: 0.158. Notably, these four variables—citations per year, author h-index, maximum citations, and total papers published, are not network variables, but are instead common metrics used to discern the potential impact of a publication. Weighted PageRank and unweighted PageRank, two other network variables, are less important, but still significant, at 0.012 and 0.011 respectively.

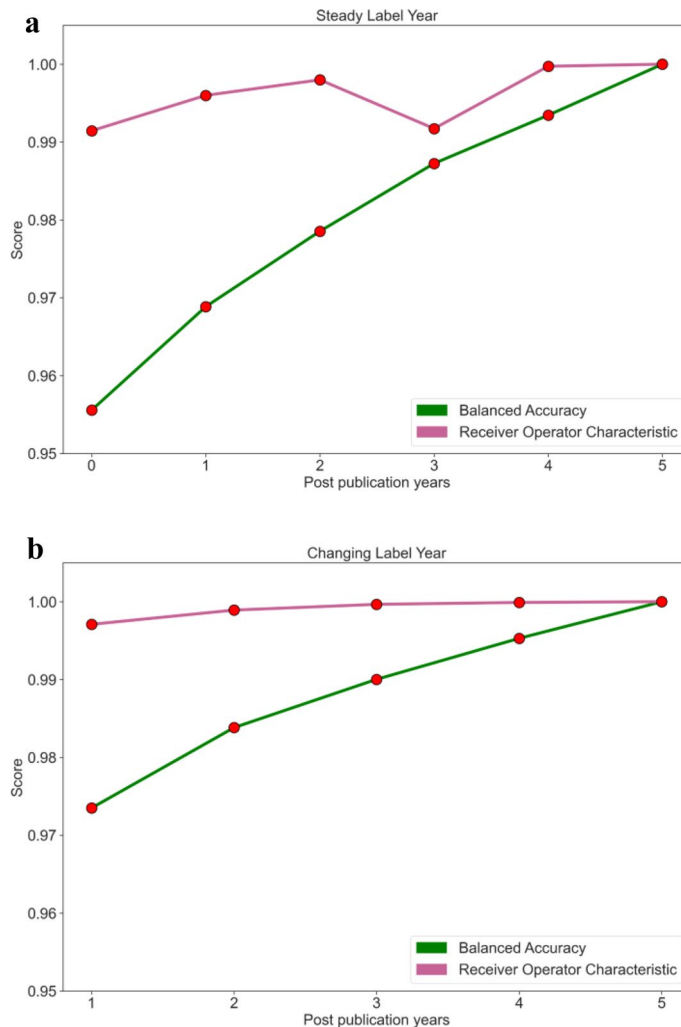


Figure 3. Performance metrics by models. Graphs of BA and ROC AUC for trained models. In (a), the data have a label 5 years after publication. In (b), the data have a label the same year as the amount of post-publication data made available to the model. (a) Corresponds to the models in rows 1–6 of Table 3. (b) Corresponds to the models in rows 6–10 of Table 3.

Discussion

The pursuit of scientific research is intricately tied to the progress of human society, and it is shaped by decisions made by various funding bodies, such as governmental organizations, universities, and both for-profit and non-profit private entities, who provide financial support for these research studies^{1,2}. The National Institutes of Health (NIH), for instance, allocated a budget of 33.34 billion dollars towards scientific research in 2022¹⁸, highlighting the significant investment made in this area. To ensure the effective utilization of public funds, it is crucial to allocate resources in a way that maximizes favorable outcomes. An analysis done by Fang and Casadevall¹⁹, observed that biomedical research fund allocation in the United States is inadequate for prioritizing which applications to fund. The present evaluation methodology for impact implicitly relies on quantitative metrics, including the number of publications, citations, and the impact factor of the journals in which researchers publish, as well as related measures such as the h-index. For example, prior work has observed a positive correlation between bibliometrics and NIH funding^{20,21}. Furthermore, another study found that NIH-funded researchers had higher h-indices and citation rates than non-funded researchers, and that these differences were more pronounced in certain fields, such as immunology and neuroscience²². These articles shed light on some of the potential biases associated with the current bibliometric measures.

There is also growing concern about the potential misuse and abuse of bibliometric measures. In some instances, researchers may engage in self-citation, which artificially inflates their citation counts, or they may publish multiple papers on the same topic to boost their h-index. Thus, relying solely on conventional citation metrics may introduce biases that impede our progress in identifying and pursuing impactful research areas^{7,23,24}. In fact, the use of time-rescaled measures of node centrality is an important consideration in knowledge graph analysis for objective decision-making. Its use has already shown promising results in various fields.

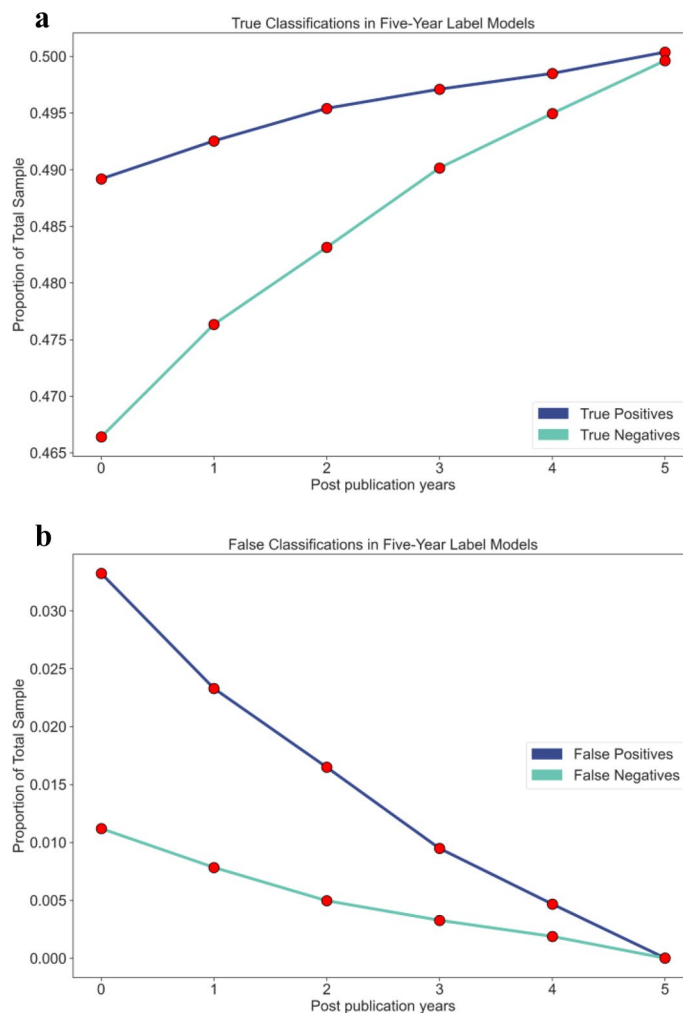


Figure 4. Classification proportions by models. Graphs of classification proportions for trained models with a label five years after publication. Both figures correspond to the models in rows 1–6 of Table 3.

Variable	Importance value
Rescaled PageRank	0.166
Citations per year	0.049
Author h-index	0.043
Maximum citations	0.042
Total papers	0.024
Citations	0.024
Citations per paper	0.021
Δ Total citations	0.018

Table 4. Variable importance. The eight most important variables on the 0-year data, 5-year label model, in descending order.

According to a study by Kumar et al., the centrality of nodes in a social network can vary over time, making it necessary to use time-rescaled measures of centrality to accurately assess the importance of nodes²⁵. A study by Jiang et al. used time-rescaled measures of centrality to analyze the evolution of the Chinese stock market network, identifying key nodes and potential sources of systemic risk²⁶. In the context of scientific research, a study by Li et al. demonstrated the effectiveness of using time-rescaled measures of centrality to track the evolution of a knowledge graph in the field of neuroscience, identifying the emergence of new research topics and potential areas for collaboration²⁷.

As such we demonstrate the impact of the network framework in predicting high-impact publications. Compared to the models used in Weis and Jacobson, our models showed more favorable performance results when predicting high-impact clinical medicine studies compared to biotechnology studies¹⁶. In our one- and two-year models for a label year five years post publication, the balanced accuracies are approximately 0.969 and 0.979. On the dataset used in Weis and Jacobson, these numbers were 0.77 and 0.87 respectively.

We observed that the ROC AUC score of each model was greater than 0.99, which indicates that the models have a high level of accuracy in distinguishing between high-impact and low-impact publications (Table 3). The BA scores, which were greater than 0.95, indicate that the models performed well in identifying high- and low-impact cases. Figures 3 and 4 demonstrated that the ROC AUC and BA scores increased with an increase in the amount of data used in the models, except for the three-year post-publication model that predicted impact five years after publication. This is expected since the larger the dataset, the more information was available for the model to learn from; however, the exception observed can be explained by several possible factors including data availability, model design, sample bias, random variation, or noise.

The need for implementation of a time-rescaled measure of node centrality arises from the fact that networks are dynamic and constantly evolving. Using a static measure of node centrality that does not take into account the temporal aspect of the network may not accurately reflect the node's current importance or influence. For instance, a node that was highly central in the network in the past may have lost its importance due to changes in the network, and a node that was less central in the past may have become more important due to new connections or changes in behavior. By using a time-rescaled measure of node centrality, we can better capture the temporal dynamics of the network and get a more accurate representation of each node's current importance or influence^{28,29}.

In addition, by representing papers as nodes and their citations as edges, the use of node2vec can learn embeddings that capture the relationships between papers based on their citation patterns. These embeddings can be used to identify clusters of related papers or to detect influential papers that have had a significant impact on the field. By using node2vec to learn embeddings that capture structural and contextual features of nodes in the network, and combining this with time-rescaled measures that reflect the frequency and recency of node involvement in network activity, it is possible to obtain a more nuanced and dynamic understanding of node importance. This approach can help to address some of the limitations of each method when used in isolation, such as the limited interpretability of node2vec embeddings and the sensitivity of time-rescaled measures to hyperparameters and time window size³⁰.

The nature of data is often nonlinear, and the implementation of machine learning has led to the discovery of more meaningful outcomes than previously achievable. Machine learning-assisted prediction models have become increasingly prevalent in various scientific fields, particularly in this area of study, and are expected to have a significant impact on decision-making processes. The framework used in this study serves to confirm the reproducibility of the model presented by Weis and Jacobson.

Nonetheless, some limitations arise when using this framework, such as the requirement for historical data, which may not always be available or accessible. In addition, constructing a comprehensive and accurate knowledge graph can be difficult or impossible in some fields, which may limit the applicability of the approach³¹. Another limitation is that the approach may not capture all factors that contribute to impactful research, such as novelty, content, or relevance to current research trends. The knowledge graph structure alone may not always be sufficient to predict the impact of a given paper or research area³². Moreover the approach is designed to identify impactful research at the level of individual papers, authors, or research subjects, and may not be directly applicable to other levels of analysis such as journals, conferences, or research communities. Its effectiveness in other contexts may therefore be limited³³. The model assumes that the knowledge graph is a self-contained representation of the research domain, but impactful research may be influenced by external factors such as funding or policy decisions, which may not be reflected in the knowledge graph³⁴. The model may be less effective in domains with complex knowledge graphs or high levels of noise, as it may be difficult to distinguish between significant and insignificant changes in the graph³⁵. Future research endeavors should prioritize the prospective validation of our results over time and emphasize the application of our findings specifically to individual research communities or domains.

In the context of defining high-impact papers using a network feature, it is noteworthy to mention that these network features are integral components of the models. This might suggest a potential redundancy, where network features could be perceived as predicting themselves. Nonetheless, forecasting the future state of a network feature, such as the time-rescaled PageRank, is not merely contingent on its current state. Analogously, predicting a child's future height requires considerations beyond their present height, encompassing factors like nutritional intake and environmental exposures. Thus, the models are enhanced by the incorporation of both network and non-network features. As indicated by Weis and Jacobson, the precise utilization of each variable by the models remains elusive, but insights can be gleaned from the mean decrease in impurity test. It is evident that while time-rescaled PageRank holds significant importance, the subsequent four salient factors are non-network features, cumulatively equating to a comparable importance as the time-rescaled PageRank. This underscores the indispensable nature of both feature types, suggesting that their combined influence fosters a more refined model than either feature type in isolation.

While the efficacy of time-rescaled measures of node centrality has been established, the dynamic intricacies of knowledge graphs necessitate further refinement. The methodologies employed in model construction may not always be optimal. For example, the aggregation of features rooted in author metadata was achieved by averaging said metadata for each publication. A more nuanced aggregation approach, taking into account the author's position in the list, might offer a richer data capture. Moreover, although defining high-impact papers using a feature type present in the model is defensible, a redefined criterion might be more fitting.

Beyond prediction of high-impact studies from bibliometrics, future research is needed to explore the potential to predict high-impact grants that can fund one or more studies. Prospective testing of our study's methodology applied to research grants can evaluate the extent to which predictive models trained on graph bibliometrics can inform decision-making of funding allocation for multiple studies.

The ramifications of employing time-rescaled measures of node centrality in decision-making warrant thorough assessment. Subsequent studies should delve into the influence of these measures on the precision and efficacy of decision-making, especially in the realm of funding opportunities, and proffer guidelines for their judicious application. Additionally, there is a compelling need for pioneering methods to augment the accuracy of time-rescaled centrality computations. Comparative analyses between these innovative techniques and established ones will be instrumental in discerning their efficacy and relevance.

Data availability

The dataset was generated by gathering publication data from Lens Lab, available at <https://www.lens.org/>. Additional data on publication dates were gathered from Elsevier, available at <https://dev.elsevier.com/>.

Code availability

Open-source Jupyter notebooks for model training and evaluation, as well as graph construction and metric extraction, can be found at <https://github.com/ljdj20/med-predictions>.

Received: 24 June 2023; Accepted: 16 January 2024

Published online: 29 January 2024

References

- Bloch, C. *et al.* Developing a methodology to assess the impact of research grant funding: A mixed methods approach. *Eval. Program Plann.* **1**(43), 105–117 (2014).
- Hicks, D. Performance-based university research funding systems. *Res. Policy* **41**(2), 251–261 (2012).
- Bloch, C. & Sørensen, M. P. The size of research funding: Trends and implications. *Sci. Public Policy.* **42**(1), 30–43 (2015).
- Moed, H. F. The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Sci. Public Policy.* **34**(8), 575–583 (2007).
- Chen, D. Open data: Implications on privacy in healthcare research. *Blockchain Healthc. Today.* (2020).
- Ioannidis, J. P., Hozo, I., & Djulbegovic, B. Federal funding and citation metrics of US biomedical researchers, 1996 to 2022. *JAMA Netw. Open.* **5**(12), e2245590 (2022).
- Boyack, K. W. & Jordan, P. Metrics associated with NIH funding: A high-level view. *J. Am. Med. Inf. Assoc.* **18**(4), 423–431 (2011).
- Ioannidis, J. P., Baas, J., Klavans, R. & Boyack, K. W. A standardized citation metrics author database annotated for scientific fields. *PLoS Biol.* **17**(8), e3000384 (2019).
- Nicolaisen, J. Citation analysis. *Annu. Rev. Inf. Sci. Technol.* **41**(1), 609–641 (2007).
- Seglen, P. O. Why the impact factor of journals should not be used for evaluating research. *BMJ.* **314**(7079), 497 (1997).
- Higham, K. W., Governale, M., Jaffe, A. B. & Zülicke, U. Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields. *J. Informetr.* **11**(4), 1190–1200 (2017).
- Bendisoli, S., & Garfinkel M. Dealing with the limits of peer review with innovative approaches to allocating research funding. EMBO Science Policy Programme (2021).
- Ernst, P., Meng, C., Siu, A., & Weikum, G. Knowlife: A knowledge graph for health and life sciences. In *2014 IEEE 30th International Conference on Data Engineering 2014 Mar 31* (pp. 1254–1257). IEEE.
- Mrdjenovich, D. *et al.* Propnet: A knowledge graph for materials science. *Matter.* **2**(2), 464–480 (2020).
- Zeng, X., Tu, X., Liu, Y., Fu, X. & Su, Y. Toward better drug discovery with knowledge graphs. *Curr. Opin. Struct. Biol.* **1**(72), 114–126 (2022).
- Weis, J. W. & Jacobson, J. M. Learning on knowledge graph dynamics provides an early warning of impactful research. *Nat. Biotechnol.* **39**(11), 1300–1307. <https://doi.org/10.1038/s41587-021-00907-6> (2021).
- Catalá-López, F. *et al.* Global mapping of randomized trials related articles published in high-impact-factor medical journals: A cross-sectional analysis. *Trials* **21**(1), 34. <https://doi.org/10.1186/s13063-019-3944-9> (2020).
- National Institutes of Health. FY 2022 By the Numbers: Extramural Grant Investments in Research—NIH Extramural Nexus (2022).
- Fang, F. C., & Casadevall, A. Research funding: The case for a modified lottery. *mBio* **7**(2), e00422-16 (2016).
- Le, N. H., Sun, H., Hui, C., Kim, J. & Kim, M. A correlation between National Institutes of Health funding and bibliometrics in neurosurgery. *World Neurosurg.* **156**, e1005–e1013. <https://doi.org/10.1016/j.wneu.2021.12.015> (2022).
- Green, J. P. & Bowman, M. A. Associations between academic performance of division chiefs and receipt of National Institutes of Health funding. *JAMA* **287**(6), 895–898 (2002).
- Ioannidis, J. P., Boyack, K. W. & Klavans, R. Estimates of the continuously publishing core in the scientific workforce. *PLoS One* **9**(7), e101698 (2014).
- Ioannidis, J. P., Hozo, I. & Djulbegovic, B. Federal funding and citation metrics of US biomedical researchers, 1996 to 2022. *JAMA Netw. Open* **5**(12), e2245590 (2022).
- Ioannidis, J. P., Baas, J., Klavans, R. & Boyack, K. W. A standardized citation metrics author database annotated for scientific field. *PLoS Biol.* **17**(8), e3000384. <https://doi.org/10.1371/journal.pbio.3000384> (2019).
- Kumar, R., Novak, J., Raghavan, P., & Tomkins, A. Structure and evolution of online social networks. In *Link mining: Models, algorithms, and applications* (pp. 337–357) (Springer, 2016).
- Jiang, Z. Q., Zhou, W. X. & Sornette, D. Time-rescaled measures of node centrality to identify systemic risk and critical nodes in financial networks. *J. Econ. Dyn. Control* **111**, 103763. <https://doi.org/10.1016/j.jedc.2019.103763> (2020).
- Li, Y., Wang, H., Xu, H., Cui, Z. & Tang, Y. Knowledge graph analysis on the evolution of neuroscience research. *Front. Neurosci.* **15**, 686025. <https://doi.org/10.3389/fnins.2021.686025> (2021).
- Mariani, M. S., Medo, M. & Zhang, Y.-C. Identification of milestone papers through time-balanced network centrality. *J. Informetr.* **10**, 1207–1223. <https://doi.org/10.1016/j.joi.2016.09.005> (2016).
- Xu, S., Mariani, M. S., Lü, L. & Medo, M. Unbiased evaluation of ranking stent performance in science and technology citation. *Informetrics* **14**, 101005 (2020).
- Grover, A., & Leskovec, J. Node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge Discovery & Data Mining* (pp. 855–864). ACM. <https://doi.org/10.1145/2939672.2939754> (2016).

31. Kucuktunc, O., Cambazoglu, B. B. & Aykanat, C. A survey on scholarly data: From big data perspective. *J. Informetr.* **11**(3), 869–889. <https://doi.org/10.1016/j.joi.2017.06.005> (2017).
32. Leydesdorff, L. & Rafols, I. Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *J. Informetr.* **5**(1), 87–100. <https://doi.org/10.1016/j.joi.2010.09.002> (2011).
33. Mingers, J. & Leydesdorff, L. A review of theory and practice in scientometrics. *Eur. J. Oper. Res.* **246**(1), 1–19. <https://doi.org/10.1016/j.ejor.2015.04.002> (2015).
34. Thelwall, M. Are citation classics more robust than the average paper? A test using graph theory. *J. Informetr.* **11**(1), 1–12. <https://doi.org/10.1016/j.joi.2016.11.005> (2017).
35. Lerner, J. & Stern, S. The empirical impact of a decade of venture capital. *Rev. Financ. Stud.* **30**(9), 2993–3036. <https://doi.org/10.1093/rfs/hhx031> (2017).
36. Van Rijn, J. N., & Hutter, F. Hyperparameter importance across datasets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2367–2376. ACM. <https://doi.org/10.1145/3219819.3220058> (2018).

Author contributions

S.H. proposed and advised for the project. L.J. wrote the methods and results section of the manuscript, conducted all parts of the project, including data gathering, manipulation, and analysis, and is the corresponding author. L.J. and A.R. designed and generated the machine learning models and prepared all figures and tables. A.-B.E.A. wrote the abstract and discussion of the manuscript. D.C. wrote the introduction. D.C., N.D.G., and J.L. revised the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024