Received: 22 July 2019 | Revised: 8 July 2020 | Accepted: 18 August 2020

DOI: 10.1002/asi.24415

RESEARCH ARTICLE

JASIST WILEY

Understanding and predicting future research impact at different career stages—A social network perspective

Zhiya Zuo[1] | **Kang Zhao**[2]

[1]Department of Information Systems, City University of Hong Kong, Kowloon Tong, Hong Kong

[2]Department of Business Analytics, University of Iowa, Iowa City, IA

Correspondence
Zhiya Zuo, City University of Hong Kong, Kowloon Tong, Hong Kong SAR
Email: zhiyazuo@cityu.edu.hk

Funding information
Digital Innovation Laboratory of Department of Information Systems, City University of Hong Kong, Grant/Award Number: Active Researcher Fund

Abstract

Performance assessment is ubiquitous and crucial in people analytics. Scientific impact, particularly, plays a significant role in the academia. This paper attempts to understand researchers' career trajectories by considering the research community as a social network, where individuals build ties with each other via coauthorship. The resulting linkage facilitates information flow and affects researchers' future impact. Consequently, we systematically investigate the career trajectories of researchers with respect to research impact using the social capital theory as our theoretical foundation. Specifically, for early-stage and mid-career academics, we find that connections with prominent researchers associate with greater impact. Brokerage positions, in addition, are beneficial to a researcher's research impact in the long run. For senior researchers, however, the only social network feature that significantly affects their future impact is the reputation of their recently built ties. Finally, we build predictive models on future research impact which can be leveraged by both organizations and individuals. This paper provides empirical evidence for how social networks provide signals on researchers' career dynamics guided by social capital theory. Our findings have implications for individual researchers to strategically plan and promote their careers and for research institutions to better evaluate current as well as prospective employees.

1 | INTRODUCTION

Performance assessment is ubiquitous, ranging from sports and entertainment, to academia and industry. In the era of big data, human resources management has evolved into *people analytics* (Arellano, DiLeonardo, & Felix, 2017; Leonardi & Contractor, 2018), a data-driven analytics approach that improves the efficiency as well as the efficacy of talent acquisition as well as retention. Scientific impact, a special case of performance, plays a significant role in the evaluation of scientific research. Ever since the successful construction of bibliographic databases storing millions of citation records (Garfield, 1955), citation-based measures have become one of the norms to evaluate various types of research impact. While there are inevitable downsides such as the involvement of non-meritocracy factors, citations are still relatively reliable measures for research impact (Radicchi, Weissman, & Bollen, 2017). Being both readily available and well grounded, citation dominates research evaluation and assessment at various levels of analyses. At the level of disciplines, citation measures are one of the fundamental tools to probe into an academic discipline's evolving trajectories and future development (Zuo, Qian, & Zhao, 2019; Zuo, Zhao, & Eichmann, 2017); From the national level, such data is crucial for understanding innovations that in turn affect economic developments Ponomariov & Toivanen, 2014); Down to the individual level, decisions on hiring and placement (Way, Larremore, & Clauset, 2016; Zuo, Zhao, & Ni, 2019), promotion and tenure (Bertsimas, Brynjolfsson, Reichman, & Silberholz, 2015; Kelly & Jennions, 2006), awards (McNutt, 2014), and funding (Bornmann & Daniel, 2006; Hornbostel, Böhmer, Klingsporn, Neufeld, & von Ins, 2009) largely rely on accurate evaluations of not only how productive a researcher is, but also the impact of her research. Furthermore, despite the importance of individual researchers' academic performance accumulated by past achievement in their careers, it is equivalently, if not more, pivotal to better understand and even to predict if they can produce future research that is impactful.

In fact, empirical research has shown the predictability of researchers' future h-index (Acuna, Allesina, & Kording, 2012). Nonetheless, h-index is a nondecreasing metric that contains information throughout one's career. In addition, it takes time for a paper to attract citations, which means older papers are more likely to get more citations. In other words, h-index is often highly biased toward earlier work whereas recent publications weigh less. As a result, the outcome of a predictive model that predicts future h-index is therefore forecasting the impact of *one's past work and future work*, the former of which may even overshadow the latter. Such cumulative nature of commonly used citation-based metrics is well acknowledged, and some researchers have called for researchers' attention and caution on defining *future impact* (Penner, Pan, Petersen, Kaski, & Fortunato, 2013). Moreover, most previous work on individuals' research impact fails to distinguish researchers at different career stages even though such contextual differences can largely affect the underlying mechanism of receiving citations and generating scholarly impact. Furthermore, many extant studies lack theoretical foundations that could improve model validity and reduce spurious and inflated correlations (García-Pérez, 2013).

Therefore, this research aims at understanding the career trajectories of researchers by considering the academic community as a social network, where individual actors connect with each other to collaborate on research projects. The resulting network then facilitates the flow of information and social capital, contributing to the career development of individuals. Consequently, we systematically investigate the dynamics of individuals' research impact in both short and long-term periods via the lens of social capital theory (Nahapiet & Ghoshal, 1998). Addressing aforementioned research gaps, our contributions are threefold: First, as called by Penner et al. (2013), we conduct a finer-grained analysis of researchers' future impact. In particular, we adopt the future citation counts of one's future papers as the target of analysis. Unlike cumulative metrics such as h-index, the new target is statistically independent from one's past achievement. In practice, despite the value of past performance, future impact is often what many organizations, including research institutions, care more about when hiring or evaluating an individual. We also construct prediction models whose targets are researchers being "stars" versus "nonstars" depending on whether one's future citation is among top tiers or not. Compared to previous models for precise predictions of numeric citation values, our approach enables a more practical and intuitive solution for research assessment—after all, stakeholders often care about whether one can produce promising work that attracts more citations than peers within the same discipline, instead of wanting to know the exact number of citations. Second, our analysis adds a temporal perspective and investigates the moderating role of career ages on researchers' future impact. Leveraging time intervals between one's first and last publications as a proxy for career age, we model the temporal dynamics of researchers' future citation impact separately for different cohorts and reveal different patterns for early-stage, mid-career, and senior researchers. Third, we base our analysis of researcher-level future impact on the social capital theory (Nahapiet & Ghoshal, 1998). Compared to the majority of past studies focusing on the predictive performance, our study aims at balancing the trade-off between accurate predictions and theoretical foundations backing the proposed model. As the social capital theory has been adopted in many other contexts, building constructs based on the theory not only strengthens the validity and generalizability of our results, but also helps us better understand how our work fits in the bigger picture of research on individual performance.

2 | RELATED WORK

2.1 | Measuring scientific impact

While publication count is a straightforward metric to measure researchers' academic performance, the sheer quantity is different from the actual impact of these publications (Bornmann & Tekles, 2019; Kaur, Ferrara, Menczer, Flammini, & Radicchi, 2015). Citation count, on the other hand, is often considered the most popular and *relatively* reliable source for the quantification of academic merit and quality (Radicchi et al., 2017). Along with the growing electronic bibliometric databases such as Web of Science, Scopus, and Google Scholar, the easily accessible citation records have enabled extensive research on developing new metrics to capture publication quality. Besides well-established citation impact measures at the paper level (citation counts) and the journal level (impact factor; Garfield, 1999), there have been numerous studies proposing different measures for the author level impact. Among them, h-index (Hirsch, 2005) has attracted the most attention. Formally, a researcher has an h-index of h if h out of all her N publications has received at least h citations while the remaining N-h publications have fewer than h citations. The advantage of h-index is its ability to simultaneously capture both the quantity and quality of research output. h-index has witnessed a "hot streak" in its popularity as evidenced by the quick adoption by Nature, Science, and Web of Science (Bornmann, Mutz, & Daniel, 2008), despite critiques and proposed variants aiming to address its downsides (Dorogovtsev & Mendes, 2015; Egghe, 2006; Kosmulski, 2006).

However, the use of citation-based metrics is by no means perfect. First of all, as citations can happen due to various reasons (e.g., criticizing, disputing or acknowledging previous work; Bertin, Lariviere, & Sugimoto, 2016), the lack of citation contexts in modern bibliometric databases has made a proper citation analysis difficult. Additionally, various citation patterns such as "sleeping beauties" (Sugimot & Mostafa 2018).[1] (i.e., delayed recognition for influential papers that may be ahead of their time; Ke, Ferrara, Radicchi, & Flammini, 2015; Van Raan, 2004) and "early rise - rapid decline" (Aversa, 1985) have been discovered, which led to discussions on choices of citation time windows for research evaluation (J. Wang, 2013). Finally, as Goodhart's law states that "When a measure becomes a target, it ceases to be a good measure." (Strathern, 1997), Biagioli (2016) wrote "All metrics of scientific evaluation are bound to be abused." Indeed, there exists manipulations such as self-citations (Fong & Wilhite, 2017; Seeber, Cattaneo, Meoli, & Malighetti, 2019), honorary authorship (Flanagin et al., 1998; Katz & Martin, 1997), coercive citations (Wilhite & Fong, 2012), and even academic misconduct (Martinson, Crain, Anderson, & De Vries, 2009), because of the fierce competition of *publish or perish*. As a result, limitations and cautions should always be considered when using citation-based metrics.

2.2 | Predicting future citation impact

As mentioned earlier, the evaluation of research as well as scientific impact, while often considered as multidimensional, is mostly based on citations received by publications (Radicchi et al., 2017; Waltman, 2016). The vast literature on predicting citation impact can be classified into two categories: (a) future impact of papers (Abramo, D'Angelo, & Felici, 2019; Cao, Chen, & Liu, 2016; Chakraborty, Kumar, Goyal, Ganguly, & Mukherjee, 2014; Sarigöl, Pfitzner, Scholtes, Garas, & Schweitzer, 2014; Stegehuis, Litvak, & Waltman, 2015; D. Wang, Song, & Barabási, 2013; Xiao et al., 2016; Yan, Huang, Tang, Zhang, & Li, 2012; Yan, Tang, Liu, Shan, & Li, 2011; Yu, Yu, Li, & Wang, 2014) and (b) of researchers (Acuna et al., 2012; Ayaz, Masood, & Islam, 2017; Bertsimas et al., 2015; Bütün, Kaya, & Alhajj, 2017; Dong, Johnson, & Chawla, 2016; Mazloumian, 2012; Nezhadbiglari, Gonçalves, & Almeida, 2016; Penner et al., 2013; Sinatra, Wang, Deville, Song, & Barabasi, 2016; Weihs & Etzioni, 2017) both of which are based on bibliometric indicators. Both types of predictions, especially the accurate prediction of researchers' future citation impact, are of great interest to research institutions and funding agencies. Previous studies have claimed to produce highly accurate predictions of researchers' h-index (Weihs & Etzioni, 2017), even with simple models such as linear regressions (Acuna et al., 2012; Ayaz et al., 2017).

However, the consistency of h-index predictions in different contexts remains questionable (García-Pérez, 2013). In particular, there is little research on the effects of collaboration on the prediction of future research impact at the author level. One's papers may receive broader attention and gain wider exposure thanks to (weak) ties with prestigious researchers in the past. Indeed, Sarigöl et al. (2014) find that centrality in coauthorship networks can be used to precisely identify highly cited papers. At the researcher level, many studies (Daud, Abbasi, & Muhammad, 2013; Daud, Ahmad, Malik, & Che, 2014; X. L. Li, Foo, Tew, & Ng, 2009) have aimed at finding rising stars (i.e., young academics with huge potentials) utilizing coauthorship networks. Recently, Amjad et al. (2017) provide empirical evidence that with the opportunity to work with authoritative seniors, junior scholars can benefit from their experience and therefore become rising stars early in their careers. A recent work, Dong et al. (2016), managed to classify whether a paper can contribute to its author's future h-index, and explored the impact of various social factors, such as social network centralities and h-indices of coauthors. The study does not find these social factors as contributors to future h-indices and has several limitations. First, their data set only included authors with h-index of at least 10. In fact, h-index of 10 is a relatively high number, especially for a junior researcher, therefore leading to potentially biased predictive model. In addition, it suffers from the problem of failing to separate researchers at different career stages. Finally, the implications are unclear. Even with an accurate prediction of an individual paper's contribution to a future h-index, there is no easy way to utilize this prediction model to facilitate real-world managerial decision-making.

3 | HYPOTHESIS DEVELOPMENT

While there is abundant literature on various approaches of research impact prediction, there is no systematic study on understanding and predicting author-level future impact across different career stages. Investigations of this problem can provide implications for *people analytics* from the perspectives of both individuals and organizations. As Penner et al. (2013) pointed out, however, it is problematic to use h-index as prediction targets at researcher level even though this is a common practice. The major issue is the highly autocorrelated nature of h-indices—as a nondecreasing measure, a predictive model that takes earlier h-indices into consideration inevitably inflated the explained variances in future impact. In fact, any type of cumulative measures suffer from similar complications. It is therefore important to predict researchers' *future impact generated by their future work* (future impact hereafter). Indeed, what hiring and grant committees look for is not only cumulative impact from a candidate's past work. Rather, they are also interested in predicting the impact of her future work based on what she has achieved. While Mazloumian (2012) concluded that future impact is hardly predictable based on past work, Penner et al. (2013) showed moderate correlations between past and future impact using data of highly cited physicists and biologists. Further, predictions should be made for researchers with similar career ages. Intuitively, there are contextual factors impacting how researchers at different career stages behave and choose differently due to the different goals they have as well as the various expectations from others (Bu et al., 2018; Packalen & Bhattacharya, 2019). A senior research scientist may be, for instance, taking over an administrative role which reduce the intensity of research activities as before. As a result, her productivity may be highly dependent on the collaborative activities (Gingras, Larivière, Macaluso, & Robitaille, 2008). A junior faculty member just entering a research intensive university, on the other hand, is going to be evaluated on her recognition of research output in both quantity and quality. The department, at the same time, would help reduce her service jobs enabling her to spend more time on research with a high expectation (Taylor, Fender, & Burke, 2006). Lumping researchers with different career ages together, as a result, will degrade both predictive powers and practical values.

Additionally, the literature on prediction of individual research impact often took a bottom up approach as feature engineering is mainly based on intuition, instead of being guided by theoretical foundations (García-Pérez,- 2013). Such deficiency in turn leads to an upsurge in applications and developments of methodologies with an ultimate emphasis on the prediction power rather than digging into what constructs are useful for the task or why these work. Weihs and Etzioni (2017), for example, discovered that nonlinear machine learning models contribute to more accurate predictions of future citations at both author and paper levels. While this type of work enables new opportunities for citation predictions especially when computational methods improve at a fast pace and larger amounts of data becomes available, we reiterate and emphasize the call by García-Pérez (2013) to pursue theoretically more sound ways to predict individual future impact from two aspects. First of all, exhausting various choices for models and constructs, while achievable, are accumulating facts instead of knowledge. In order to achieve a deep understanding of one's future research impact dynamics, it is necessary to dive deep into what these constructs really mean and, moreover, why they could potentially affect the target variable of interest. To this end, theories, as abstract and generalizable frameworks, help direct us to sort potentially helpful constructs out, leading to justified and interpretable solutions to the question in hand. More interestingly, as Parsons (1938) noted, researchers are always guided by "logical structures of theoretical schemes," be it implicitly or explicitly. Take a variable capturing authors' cumulative citation count used in Weihs and Etzioni (2017) as an example. While this is an intuitive choice among many others, it implicitly stems from a famous complex network theory, preferential attachment (Barabási & Albert, 1999). In fact, this has been studied by Price (1976) in proposing a theory of the cumulative advantage characteristics of citations. We therefore believe that with more explicit reference to and foundations based on theories, we are able to not only produce precise predictions, but also obtain and summarize generalizable patterns among the constructs we test.

Furthermore, the academic research community is a miniature of the larger social systems. The interaction among researchers, in fact, forms a large-scale social network in both formal and informal ways. The resulting social connections provide convenient information flows across network members to share resources such as knowledge and expertise, workloads, as well as equipment (E. Y. Li, Liao, & Yen, 2013). Collectively, this special type of bonds bring forth the community-owned capital (Nahapiet & Ghoshal, 1998), motivating our use of social capital theory as the framework to investigate future research impact at individual level. While E. Y. Li et al. (2013) is similar in that they study future citation impact using the social capital theory, it is different from the current study in the following facets: (a) We take a prospective approach to help stakeholders better understand how established social capital contributes to future research impact, while Li et al. took a retrospective view to study the relationships between what has happened during the time period of interest (i.e., from 1999 to 2003). Specifically, both study and target variables were measured by data within the 5-year window in E. Y. Li et al. (2013). Therefore, they are in fact studying the coevolution of social capital and research output. Instead, we measure study variables based on already established publication records and the resulting coauthorship networks, while target variables are one's future citations that will be received by one's future research output. (b) We consider the effect of career age. Pooling junior and senior researchers into the same sample set as in E. Y. Li et al. (2013) may, as mentioned earlier, result in misleading findings due to the potentially different patterns for researchers from different cohorts; (c) We used a much larger data set with a broader coverage for the discipline, while Li et al. focused only on 137 scholars during the 5-year period.

In the context of research impact, social capital such as a central position in a coauthorship network has been shown to help citation impact accumulations. Chakraborty et al. (2014), using millions of publication records from Microsoft Academic Graph, found that a paper with productive authors attracts more citations. Likewise, Sarigöl et al. (2014) showed that papers benefit from their authors with high centrality in the coauthorship network with respect to their future citations in 5 years. At researcher level, future success in different dimensions is found to be related to who one has worked with in the past. Particularly, collaboration opportunities with reputable authors enhance a junior researcher's future citation and productivity (Amjad et al., 2017). The sheer number of coauthors, additionally, has been shown to be positively correlated with one's future citation impact (Ayaz et al., 2017; Dong et al., 2016). Incorporating social network analysis, Bertsimas et al. (2015) identified researchers' centrality in coauthorship networks improved predictions of future citations as well as tenure decisions. All these point to the significant role of a researcher's social network via coauthorship in her future research impact. Accordingly, we hypothesize that social capital accumulated through coauthorship activities is positively associated with researchers' future impact. Following Nahapiet and Ghoshal (1998) and E. Y. Li et al. (2013), specifically, we measure social capital from three dimensions: structural, relational, and cognitive:

- *Structural social capital as network centrality*. The structural dimension of social capital refers to the linkage among entities within networks and can be described as connectivity measures. In particular, centrality serves as fundamental metrics to characterize the importance of an entity in various aspects. Depending on the operationalization of centrality measures, a highly central researcher may enjoy various benefits. For example, a researcher connecting otherwise isolated research groups possess a broker position that enhances her access to information flow in the network (Newman, 2004). From a different view, experience working with top-performing researchers implies the accessibility to one's rich resources as well as her own (high) status in the academic community. We therefore hypothesize that

H1 Researcher's centrality in the coauthorship network is critical for her to gain information, therefore positively correlated with a greater future impact.

- *Relational social capital as collaborators' network centrality and research impact*. Relational capital is the personal relationships built from past interactions. In contrast to structural capital, the relational dimension emphasizes the trust, commitment, and reciprocity instead of the topological similarity among the collective (E. Y. Li et al., 2013). Within an academic collaboration network, two researchers coauthor and commit to pursuing of a common goal of research and publication based on the mutual trust. Hence a natural operationalization of relational capital for researchers is to leverage the characteristics of their past coauthors. Specifically, coauthors' statuses, including their research reputation and structural significance in the social network (in our context, coauthorship network), are relevant to a focal researcher's future research impact. For one thing, a researcher can learn from a prolific coauthor and polish her perception of captivating topics for producing papers that easily attract more citations. For another, one can benefit from the copious resources possessed by her coauthors who have built network ties. A prestigious researcher may, for example, expose the focal one to another well-respected researcher who shares similar research interest, which in turn lead to improvement of the focal one's future research. Accordingly, we posit that

H2 Centrality (*H2a*) as well as research impact (*H2b*) of collaborators are both positively correlated with focal researchers' future impact.

- *Cognitive social capital as established research impact and network degree*. Last but not least, cognitive capital concerns resources that can provide common knowledge and understanding within the network. On publishing more papers and attracting more citations, researchers are learning more about their field and community. Therefore, their past research impact serves as an asset to provide them with more access to attaining the shared knowledge. However, one's previous research performance has been a well-studied variable for future research impact (Acuna et al., 2012; Ayaz et al., 2017; Mazloumian, 2012; Penner et al., 2013). As a result, we treat it as a control variable. Building more ties in the network (i.e., degree), in addition, is an efficient way to acquire suggestions and learn about experiences within the area from peer academics via both formal and informal communications. Indeed, the number of coauthors is often used as a strong predictor for one's future h-index (Ayaz et al., 2017; Dong et al., 2016). Therefore, we posit that

H3 Established research impact (*control variable*) and degree in the social network (*study variable*) are positively correlated with researchers' future impact.

Lastly, research output can vary due to the different roles one play at different career stages. In the course of research careers, newcomers are gradually assimilated and accumulating their interpretations of the community. Therefore, a longer publication tenure is often a legitimate indicator for a higher level of cognitive capital (E. Y. Li et al., 2013). Additionally, it is statistically inappropriate to put subjects with different research experience into the same statistical mode given the unfair comparison and their heterogeneous backgrounds. We correspondingly hypothesize that the relationship between social capital and future research impact is *moderated* by their career ages. A subgroup analysis is conducted to validate such hypothesis, where researchers are separated into groups based on their career ages.

4 | METHODS

4.1 | Data description

The data collected for this study concerns researchers in the field of information systems (hereafter IS). The benefits of taking IS as the research context are twofold. First and foremost, IS is a mature academic field and has established and coherent culture and norm (Rai, 2018), with an exclusive focus on the managerial, organizational, as well as societal aspects of information technology (IT; ISR, 2020; MISQ, 2020). By contrast, emerging interdisciplinary fields, such as information (Zuo & Zhao, 2018), usually span a wide spectrum of research directions. As a result, collaborative patterns of interdisciplinary research are usually different from those of more established and disciplinary ones such as IS due to differences in academic promotion assessment (Van Rijnsoever & Hessels, 2011) and funding opportunities (Bromham, Dinnage, & Hua, 2016). Moreover, albeit being a developed field, IS is continuously seeking novel approaches to increase the diversity of its traditional research paradigms, and contributes to collaborative research that provides novel solutions to IT-related problems (Rai, 2018). Second, the outlet of IS research is usually in the form of journal publications, along with a list of well-accepted journals. Other disciplines related to IS, such as computer science (Dong et al., 2016; Weihs & Etzioni, 2017), have more diverse publication venues and different citation patterns even within the same discipline (Meyer, Choppy, Staunstrup, & van Leeuwen, 2009). Therefore, IS serves as an appropriate context for the purpose of this study.

In this regard, we first collected 6,401 authors who have published papers between 1980 and 2017 in any of the eight top IS journals listed by the *Association for Information Systems*[2]: (i) European Journal of Information Systems; (ii) Information Systems Journal; (iii) Information Systems Research; (iv) Journal of the Association for Information Systems; (v) Journal of Information Technology; (vi) Journal of Management Information Systems; (vii) Journal of Strategic Information Systems; (viii) Management Information Systems Quarterly. Scopus APIs[3] were used to retrieve each paper along with its authors in March 2018. To filter out "cameo authors," an author is considered as a valid IS researcher if she has at least two papers in any of the aforementioned eight journals. Two thousand, one hundred and ten out of 6,401 were selected and their publication profiles were retrieved. A total number of 68,032 unique papers were retrieved, along with their annual citation counts. It is worth noting that while we started with a narrow set of eight journals, they are only used as seeds for identifying researchers in this area. The retrieval of papers from these researchers is not limited by these eight journals. In fact, papers we collected for this study appear in a much larger number of venues, including Management Science, European Journal of Operations Research, Decision Sciences, etc. Given the publication and citation records, we calculate productivity (i.e., the number of publications), citation counts, as well as h-index for each author and their collaborator in each year as a measure of research impact.

4.2 | Social network and social capital features

For these 2,110 authors, we construct a dynamic collaboration network based on their coauthorship relationships in each year from 1980 to 2017 as a proxy for their social networks. In the collaboration network G_t at time t, a node is an author who has ever published any paper up till t. There is an undirected edge between two authors if they have ever coauthored before, with the weight being the number of coauthored papers. Several network centrality measures for each node across all times are calculated to capture various dimensions of social capital:

- Nodal degrees, that is, the numbers of connections, are counted as part of the cognitive capital measures. Note that edge weight is not applied since node degree is an approximate of the size of one's social circle by counting the number of distinct coauthors. These are normalized by dividing the maximum possible degree a researcher could have (i.e., the total number of researchers minus one).
- Betweenness centrality is defined as the ratio between the number of shortest paths passing an author and the total number of shortest paths (Freeman, 1977). It captures the brokerage position of a node–the higher an author's betweenness centrality is, the more power she has over the network since information has to pass through her before it can flow from one cluster to another.
- Closeness centrality is the reciprocal of the sum of the length of all shortest paths between the focal and all other nodes in the network (Sabidussi, 1966). A higher closeness centrality indicates that the focal researcher is easily reachable from anywhere in the network. Edge weight is not applied in this context since closeness is used to measure the reachability.
- Eigenvector centrality assumes that the more central a node is, the more influence it has (Bonacich, 1987). Specifically, a researcher has higher influence scores if she is connected with other high influencing researchers.
- PageRank score (Page, Brin, Motwani, & Winograd, 1998) is a variant of eigenvector centrality in that it includes a scaling factor when aggregating collaborators' influence scores to the focal researcher.

As discussed previously, relational social capital is then operationalized as the network centrality scores and research impact of focal researcher's collaborators. Since one can have multiple collaborators at the same time, we aggregate these statistics in three ways to summarize the distribution of collaborator features: mean, median, and maximum values.

4.3 | Model setup

Our first set of analysis builds explanatory models to understand the effects of various social capital measures on researchers' future impact. Future impact is measured by the total number of citations received by future work. Formally, for an author a whose first paper was published in year t_0^a, we define a *cutoff* year t_c^a such that all papers published up till this year (i.e., $[t_0^a, t_c^a]$) are *past work* p_c^a. A *future* year t_f^a is defined so that all publications fall within the window of $(t_c^a, t_f^a]$ are *future work* p_f^a. Career age is therefore $c_a = t_c^a - t_0^a$ and future time window is $f_a = t_f^a - t_c^a$. Correspondingly, the target variable is defined as the citation counts received by p_f^a, denoted as $cite_{p_f^a}$. It is worth noting that those published papers (close to t_f^a) may not have enough time to receive any citation, although they may become influential later. To address this problem, we also try extending the evaluation time windows to $t_f^a + i; i \in \{1..5\}$ for citation accumulation especially for new papers. It turns out the correlation between $cite_{p_f^a}$'s with and without evaluation window extension is very high ($r > 0.8$; $p < 0.001$; see Appendix D for more details). In addition, since every instance in the analysis faces the same problem, this will not lead to unfair comparisons. As a result, we only present results without extending the evaluation time window for simplicity and better interpretability.

In this study, we clearly distinguish researchers at different career stages utilizing publication tenure, that is, the number of years between their first and last publications, with specific selections of early-stage (i.e., $c_a = 3$), mid-career (i.e., $c_a = 5$), and senior researchers (i.e., $c_a = 10$). While such choice may seem arbitrary at first sight, these three different publication tenure in fact manifest researchers at three various career stages. In the field of information systems, the typical peer review process takes about 2 to 3 years (see Appendix A). Therefore a 3-year publication tenure is a reasonable approximation of a fresh junior faculty member just staring her appointment. A 5-year history implies a mid-career whereas 10 years of publication record refer to an already well-established senior researcher in the field. Figure 1 provides an illustrative definition of these three types of career stages.

Based on the categorization, we investigate factors related to their future impact in both short and long

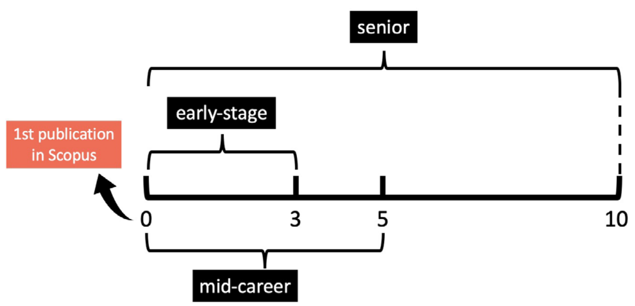

FIGURE 1 Definition of early-stage, mid-career, and senior researchers, with the start (0) being the first publication indexed by Scopus database [Color figure can be viewed at wileyonlinelibrary.com]

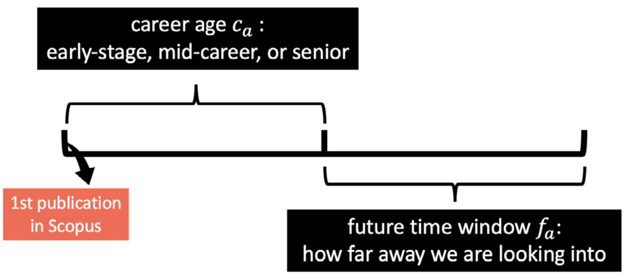

FIGURE 2 Problem setup: the future time window f_a is dependent on career age c_a, starting from the first publication [Color figure can be viewed at wileyonlinelibrary.com]

terms (i.e., $f_a \in \{3, .., 10\}$; Figure 2). Study variables are the three dimensions of social capital, controlling for individuals' own research impact, all calculated up till t_c^a. Regression analysis is conducted to analyze the roles of various social capital measures in future impact at different career stages using different sets of c_a and f_a. It is noteworthy that f_a less than 3 and above 10 are not included because (a) predictions of only 1 or 2 years ahead have low pragmatic values due to time lags between publications and citations and (b) predicting over 10 years is quite challenging and may produce unreliable results.

For senior researchers (i.e., $c_a = 10$), moreover, we further categorize their collaborators into three categories: (a) earlier collaborator with whom they coauthored a while ago but did not collaborate in recent years; (b) recent collaborator with whom they published papers recently; (c) "life partner" collaborators with whom they have continuing collaboration relationship. The threshold for such categorization is based on the median of the years a specific senior researcher collaborated with her coauthors. For example, a senior researcher X has four coauthored papers with D in 1990, with B in 1997, with C in 2005, and with D again in 2010. The threshold is 2001.5 such that B is X's earlier collaborator, C is a recently built tie, and D is a life partner collaborator. Similar to the rationale of investigating the future impact for different cohorts, collaborators who worked with a researcher at various stages of the researcher's long career history may play different roles in contributing to her future impact. The finalized variables along with their pairwise zero-order Pearson correlations can be found in Figure 3, while a detailed description of variable selection can be found in Appendix B.

Our second step is to construct predictive models to forecast whether a researcher will become a "star" in the future. Instead of predicting the actual future citations, a binary classification task may be more practical and valuable for decision-making. In particular, for each pair of career age c_a and future window f_a, we categorize an author to be a *star* if her future impact, $cite_{p_f^a}$, is no less than the 80th percentile of all authors whose statistics can be calculated, in accordance with the *Pareto Principle* (a.k.a., 80/20 rule.) Otherwise, she is considered as a non-star. Seventieth percentile is also used as another threshold for robustness check. Logistic regression with L2-norm regularization (a.k.a., Ridge regression; Hoerl & Kennard, 1970) implemented by Scikit-Learn (Pedregosa et al., 2011) is used as the classification model. L2-norm regularization is a technique to avoid overfitting by adding a penalty term to models' cost function in the form of the sum of squares of all coefficients (Bishop, 2006). While simple, logistic regression has been proved to be comparable to the sophisticated machine learning algorithms such as artificial neural networks (Ayer et al., 2010; Rajkomar et al., 2018). We investigate the predictive power of social capital features by comparing the classification performance of models with and without these features, while individuals' past research impact manifested by h-index is used as a baseline feature always present in all settings.

We applied nested cross validation (CV; Cawley & Talbot, 2010) to search for an (sub)optimal regularization strength. Specifically, we split each data set into 10 equal-size stratified folds (a.k.a., outer folds), where class distribution within onefold is similar to that in the whole data set. Each of the 10 is picked as an outer test set (i.e., held out from the training and validation process). Given each outer test set, we conduct inner CV to search for the best hyperparameters for each predictive model based on the average performance across all folds. During inner CV,

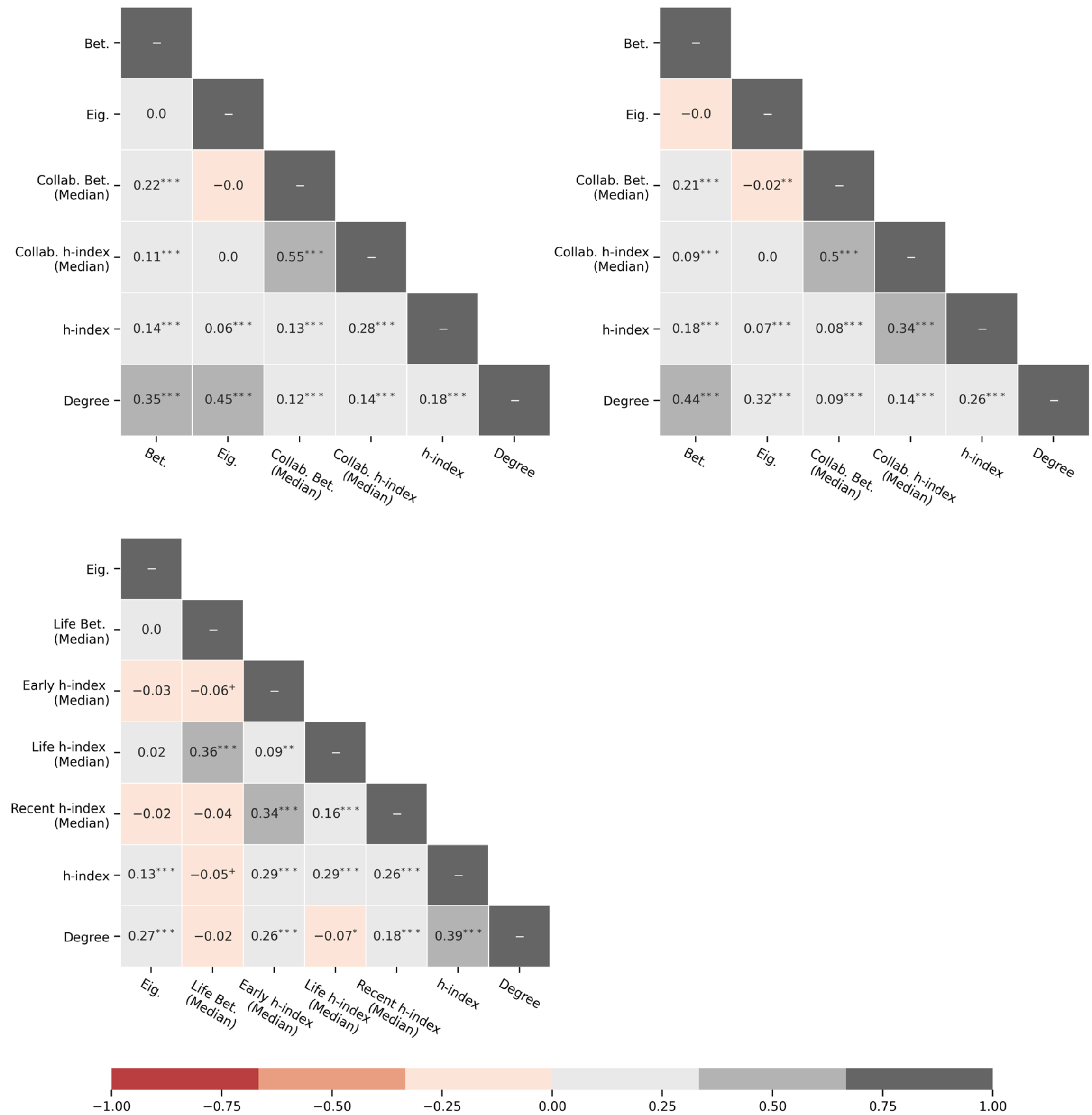

FIGURE 3 Bivariate correlation between pairs of covariates. + : $p < 0.1$; *: $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$ [Color figure can be viewed at wileyonlinelibrary.com]

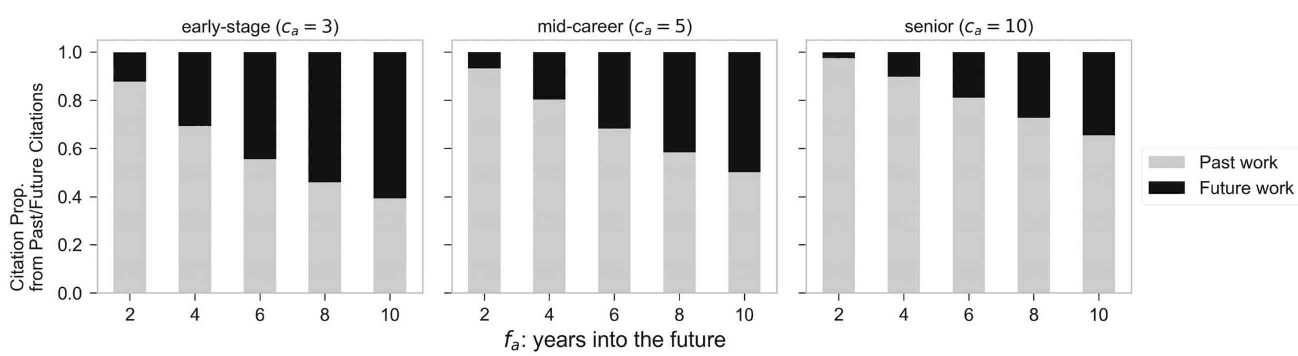

FIGURE 4 The relative contributions between past papers and future papers to the total number of citations at a certain future time point

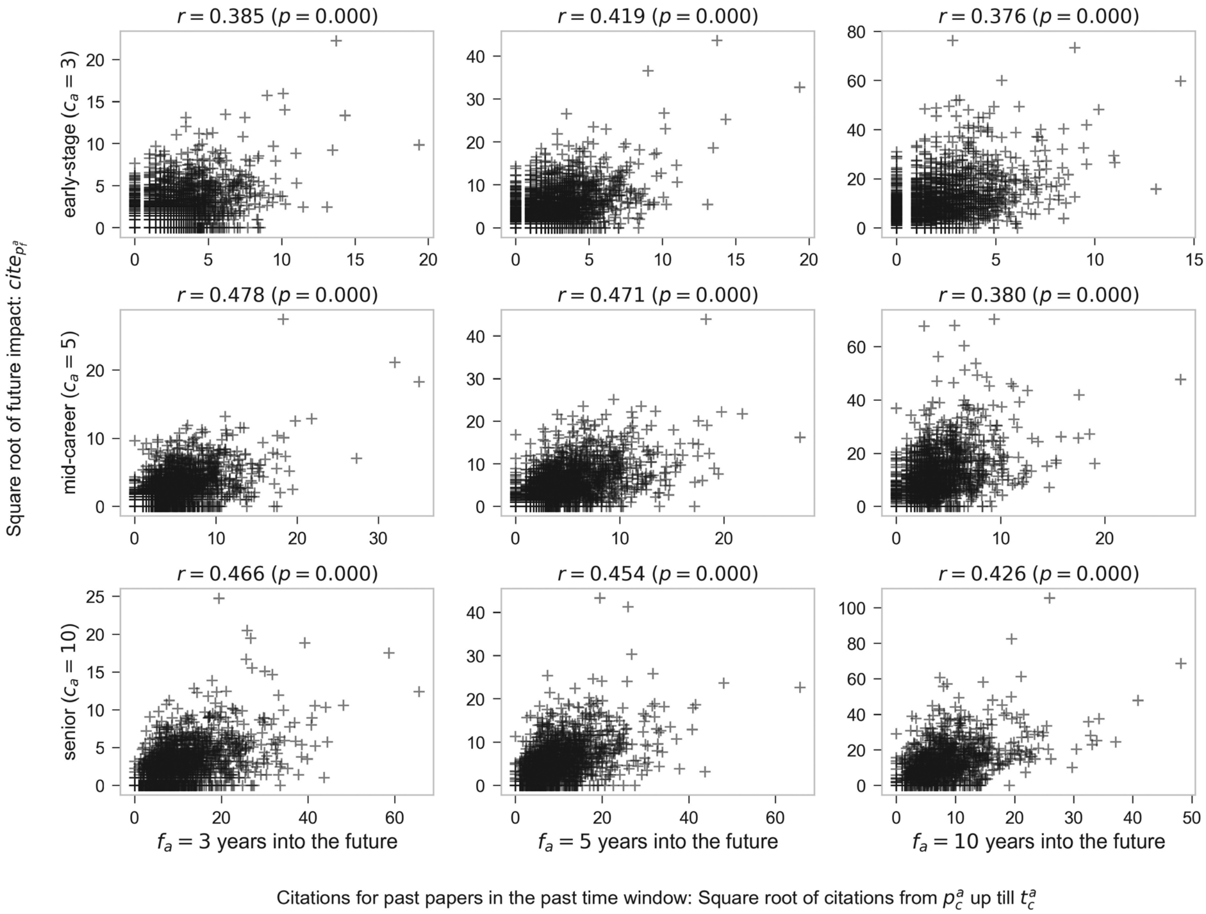

FIGURE 5 Correlations between citations to past work before future time window and future work

TABLE 1 Selected variables for each hypothesis for early-stage and mid-career researchers

Social capital	Variable	Hypothesis
Structural	Betweenness eigenvector	H1
Relational	Collab. Betweenness (median)	H2a
	Collab. h-index (median)	H2b
Cognitive	h-index	Control
	Network degree	H3

we conduct 10-fold CV, similar to the outer CV. Different sets of outer CV may select different hyperparameters. The final prediction performance is the average value of those on each test set. The prediction performance is evaluated using area under the receiver operating characteristics curve (AUC). AUC is a single-number metric (with values between 0 and 1) that captures the extent to which the two classes (in this case, star and nonstar researcher) are separated by the model. The higher it is, the better the model is in classification performance.

TABLE 2 Selected variables for each hypothesis for senior researchers

Social capital	Variable	Hypothesis
Structural	Eigenvector	H1
Relational	Life partner collab. Betweenness (median)	H2a
	Earlier collab. h-index (median)	H2b
	Life partner collab. h-index (median)	
	Recent collab. h-index (median)	
Cognitive	h-index	Control
	Network degree	H3

5 | RESULTS

5.1 | Why citations of future work?

We first present the importance of studying citations of future work by quantitatively depicting the

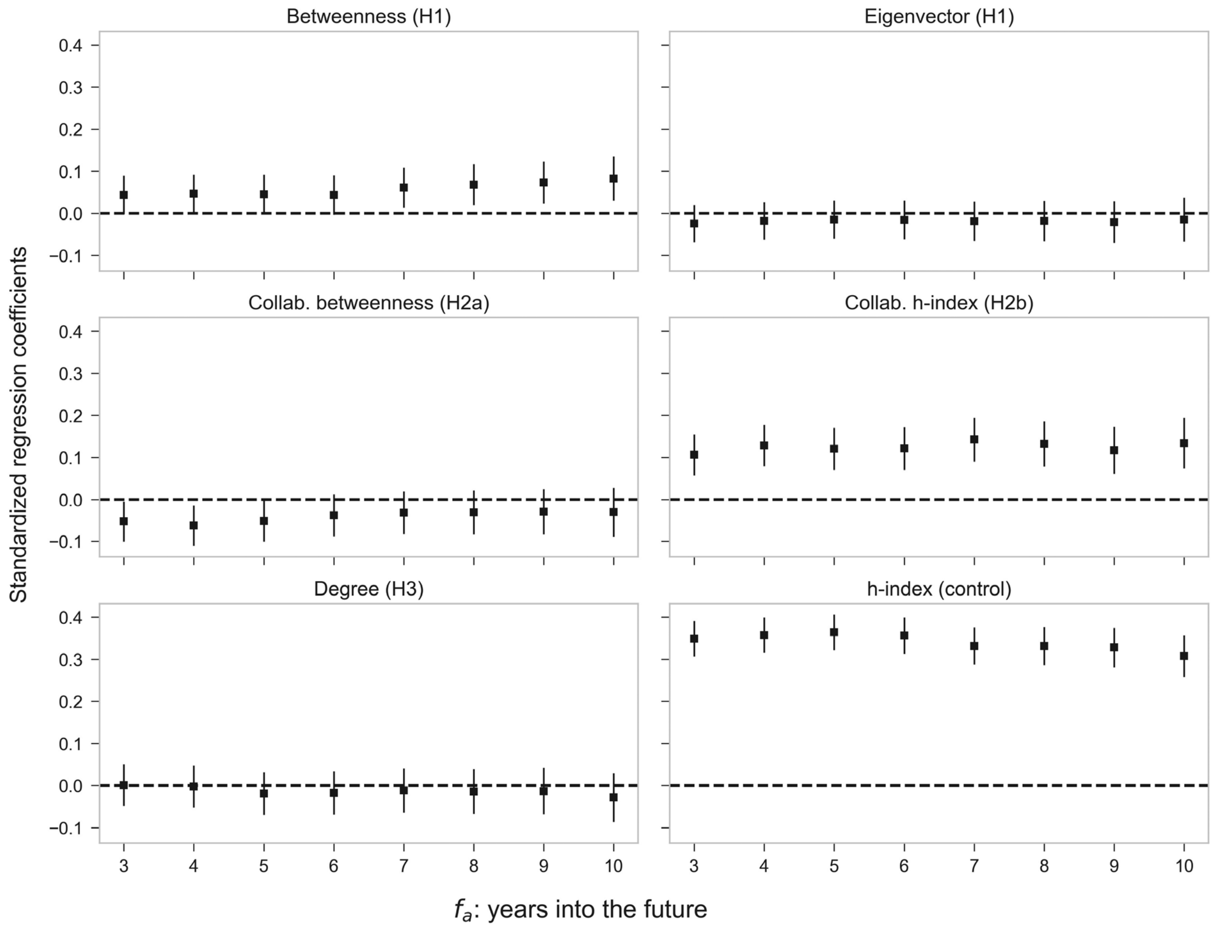

FIGURE 6 The effects of covariates on future impact and their changes over different f_a for early-stage researchers. The marker points are the point estimates of the standardized regression coefficients for covariates on the horizontal axis. The error bars are 95% confidence intervals. The reference dashed lines are zero, indicating the significance of the regression coefficients

cumulative nature of citation-based impact measures. In particularly, for $c_a \in \{3,5,10\}$, we visualize how p_c^a and p_f^a on average contribute to an author a's citation counts at various f_a across all authors (Figure 4). Indeed, the majority of one's citations are attributed to their past papers. Even if we focus on long-term impact using relatively short history with $c_a \in \{3, 5\}$ and $f_a = 10$ (i.e., the last bars in the first two subplots in Figure 4), citations received by past papers still make up 40% and 50% of the total amounts. Thus it is easy to predict future impact if it is operationalized as a cumulative measure such as h-index.

Although a noncumulative measure, we confirm the predictability of $cite_{p_f^a}$ based on previous research impact (Figure 5). On one hand, in particular, the moderate correlations across different career stages and various future time windows imply the association between past and future citation impact. On the other hand, it shows that past citation impact is not necessarily perfectly correlated with the future impact, suggesting the need to seek additional signals. Compared to Penner et al. (2013) where a limited set of early-stage and mid-career prominent researchers are considered, we further show that historical achievements are consistently informative of future impact regardless of career stages and past attainment.

5.2 | The role of social Capital in Future Impact across Career Stages

To analyze the effect of various dimensions of social capital on future research impact over time, we run regression analysis using variables listed in Tables 1 and 2 with $c_a \in \{3,5,10\}$ and $f_a \in \{3, .., 10\}$, resulting in 24 different models (see Appendix C for detailed regression results including point estimates, 95% confidence intervals, and variance inflation factors in tabular forms). All variables are standardized such that, given one c_a, we can compare the effect of a covariate across different f_a's. Considering $c_a = 3$ and $f_a = 5$ (i.e., early-stage researchers with their fifth year into the future), for example, every standard

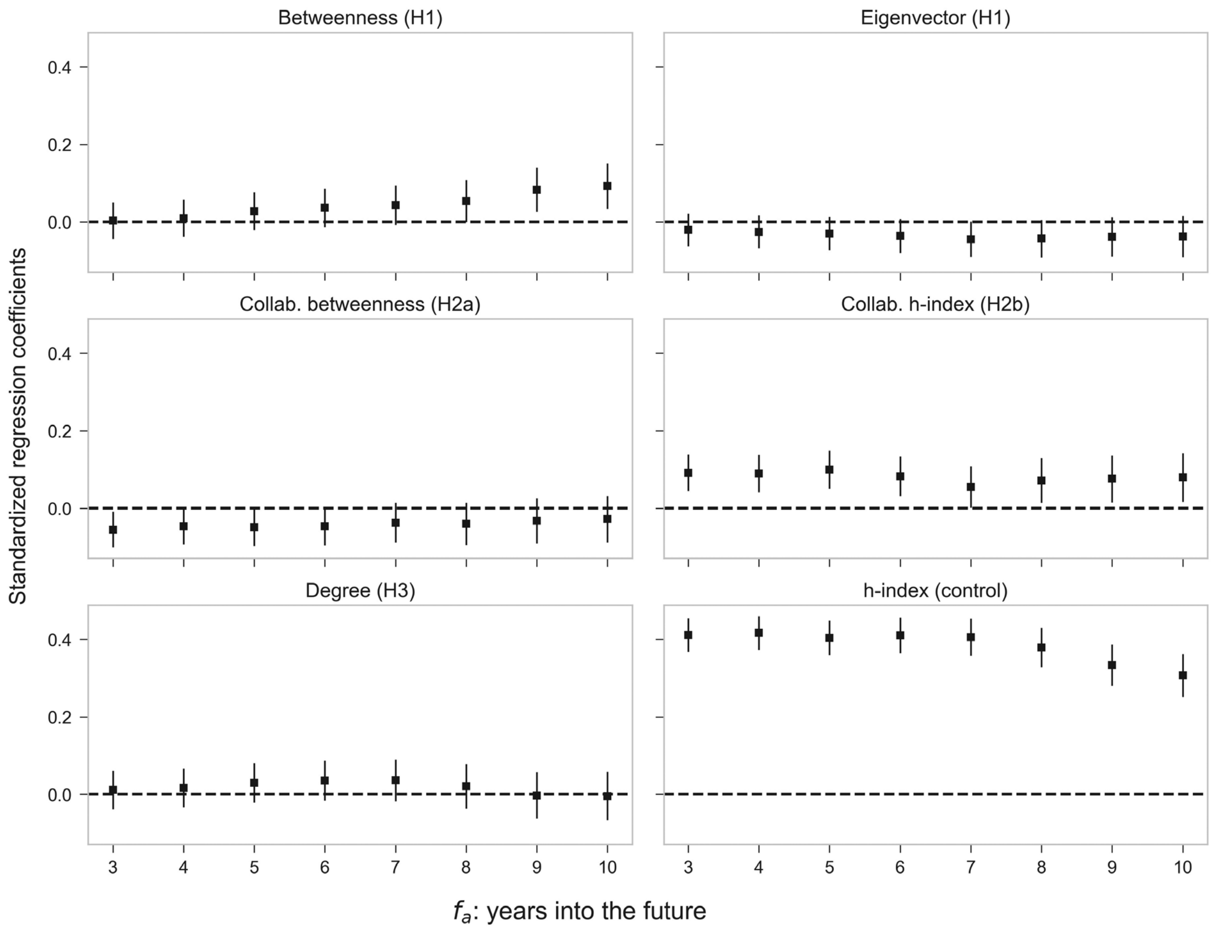

FIGURE 7 The effects of covariates on future impact and their changes over different f_a for mid-career researchers. Figure aesthetics similar to Figure 6

deviation increase in betweenness centrality in the coauthorship network is correlated with 0.045 standard deviation increase in future citation impact. Such an effect of betweenness centrality on future citation impact grows into 0.074 standard deviation increase when we are looking at 9 years into the future for the same cohort.

Results are similar between early-stage and mid-career researchers (Figures 6 and 7). Our hypothesis that researchers' own network centrality is positively correlated with future impact (H1) is partially supported—betweenness centrality shows its strength when it comes to long-term impact. However, eigenvector centrality which captures the influence over the collaboration network is insignificant. The research impact of current collaborators' has a positive and significant effect on future impact and therefore supports H2b. The remaining hypotheses are not supported by the data as collaborators' bridging role (H2a) and network degree (H3) are not significant predictors of future impact. Finally, the control variable, h-index, which captures part of the cognitive social capital, remains a positive predictor across all models.

For senior researchers (Figure 8), H2b is partially supported. In particular, the results show that "not all collaborators are equal"—the role of recently built collaboration ties stand out, whereas the earlier and life partner ones are insignificant. The effects of researchers' own academic impact are similar with early-stage and mid-career researchers, explaining a large portion of variance in future impact. However, the remaining hypotheses are not supported.

5.3 | Stargazing: Predicting prominent researchers

In the previous section, we discover that various aspects of social capital indeed explain variances in the noncumulative measure of future impact after adjusting for researchers' established impact. A natural follow-up question comes up—can we predict who will stand out among colleagues in the same field, given our current knowledge of factors driving future impact? Specifically, with individuals' established impact as the baseline

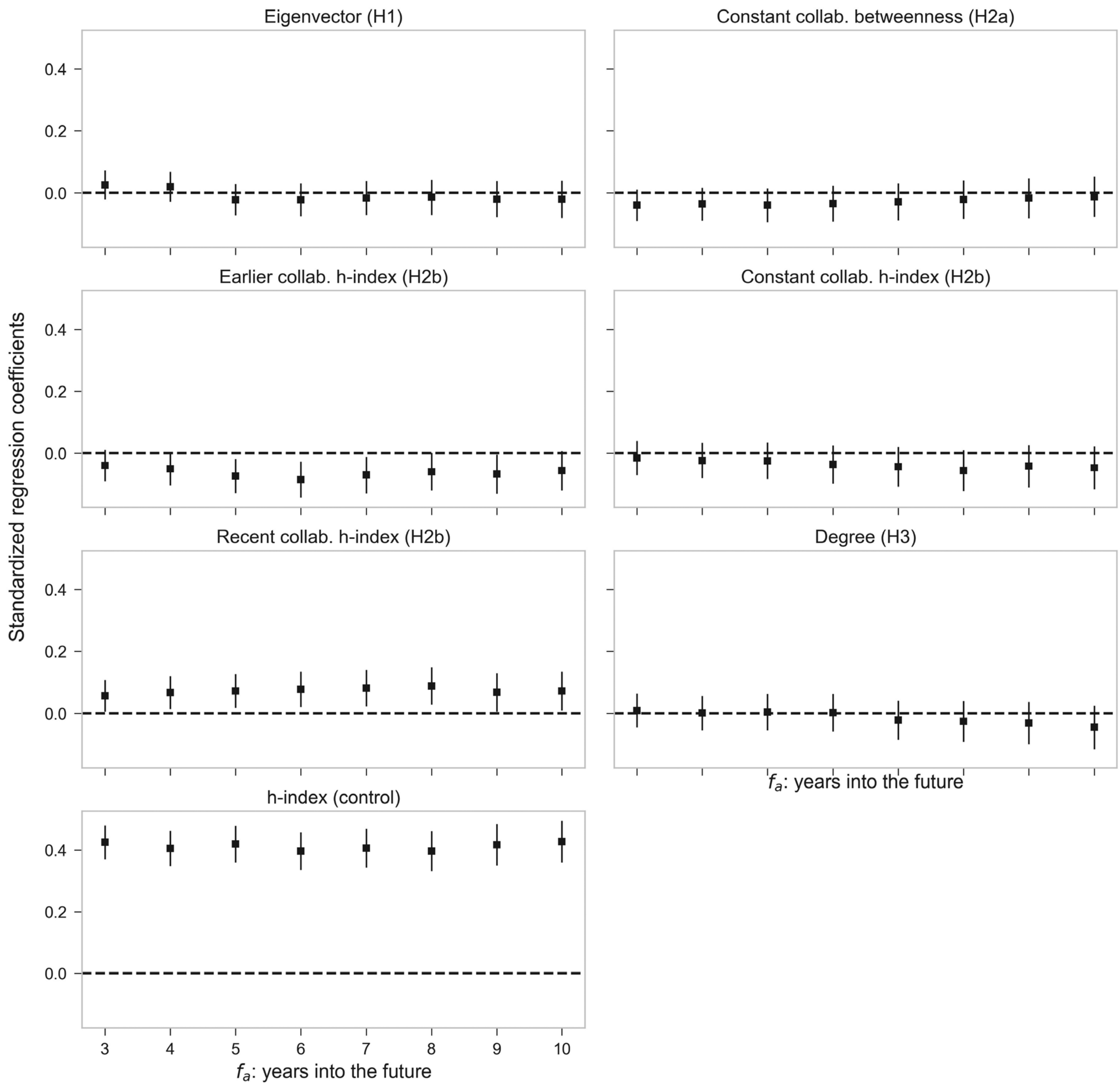

FIGURE 8 The effects of covariates on future impact and their changes over different f_a for senior researchers. Figure aesthetics similar to Figure 6

feature, we investigate to what extent their social network features manifesting social capital contribute to predicting whether their future impact will be among the top 20% or 30%. For early-stage and mid-career scholars, social capital indeed boosts the prediction performance significantly (Figures 9 and 10). However, this is not the case for senior scholars (Figure 11). There is no improvement (and even declines, though statistically insignificant) by adding social capital features. Despite a significantly positive effect of recently built collaboration ties on explaining future impact by future work, its contribution is negligible when it comes to stargazing.

In addition, as one would expect, it would be more difficult to predict future impact that is further away—compared to predicting the next 3 to 5 years, there are some minor declines in AUC scores for mid-career and senior scholars when we are trying to predict their impact for the next 6 to 10 years. Nevertheless, on average, the models are able to predict both short- and long-term impact for scholars at different career stages, with AUC scores ranging from 0.7 to 0.8. Finally, we find consistent results between 30/70 and 20/80 split on the classification label, indicating the robustness of the predictive power from social capital features.

6 | CONCLUSIONS

In this paper, we investigate the role of social network in future research impact of researchers across career stages

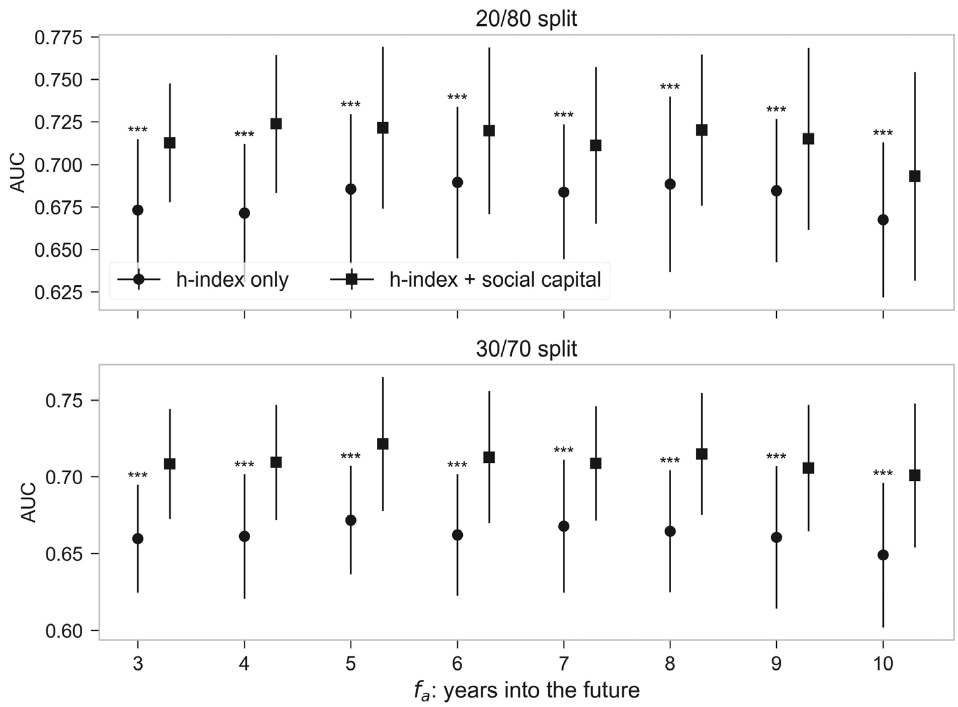

FIGURE 9 Average area under the receiver operating characteristics curve (AUC) scores (markers) across different f_a for early-stage researchers. Error bar indicates one standard deviation. Star notations are based on t tests comparing classification performance with and without social capital features: + : $p < 0.1$; *: $p < 0.05$; **: $p < 0.01$; *** : $p < 0.001$

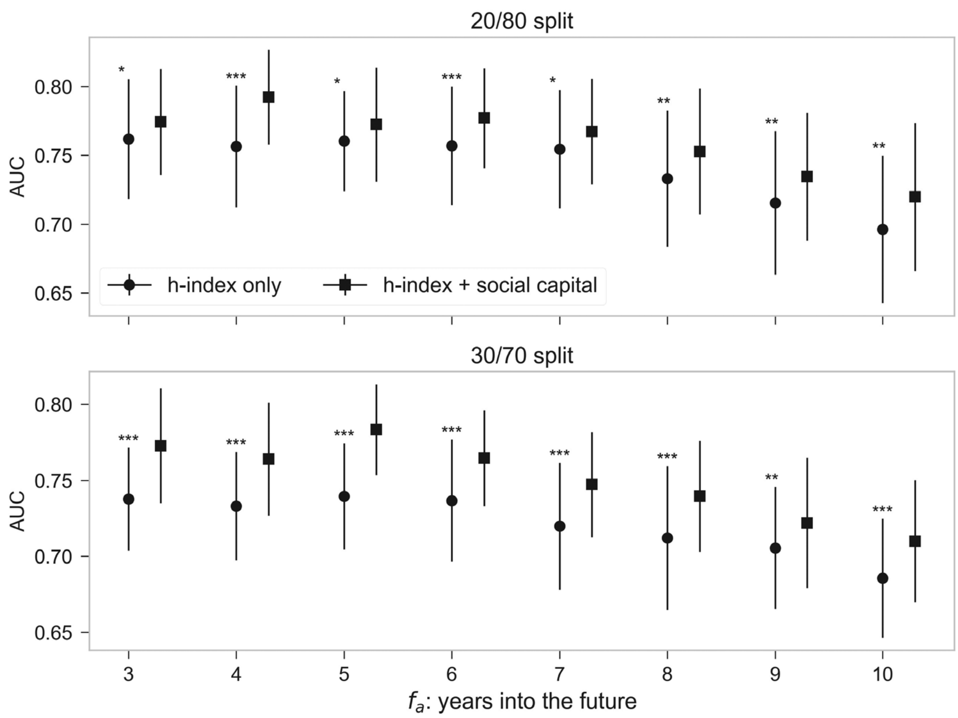

FIGURE 10 Average area under the receiver operating characteristics curve (AUC) scores (markers) across different f_a for mid-career researchers. Figure aesthetics similar to Figure 9

leveraging social capital theory as the theoretical background. In particular, we propose the operationalization of the three dimensions of social capital based on coauthorship network statistics, including network degree and centrality measures. Extending Penner et al. (2013), a novelty of this study is to understand and predict the dynamics of a noncumulative impact index, citation counts received by future papers, for early-stage, mid-career, and senior researchers. The results show that social capital characteristics are indeed significant predictors for future impact, but it plays different roles for different career stages. In particular, for early-stage and mid-career researchers, it is beneficial to work with high-standing collaborators. While this is helpful, not everyone has the access to collaborating with "celebrities"—such relational capital is usually unevenly distributed.

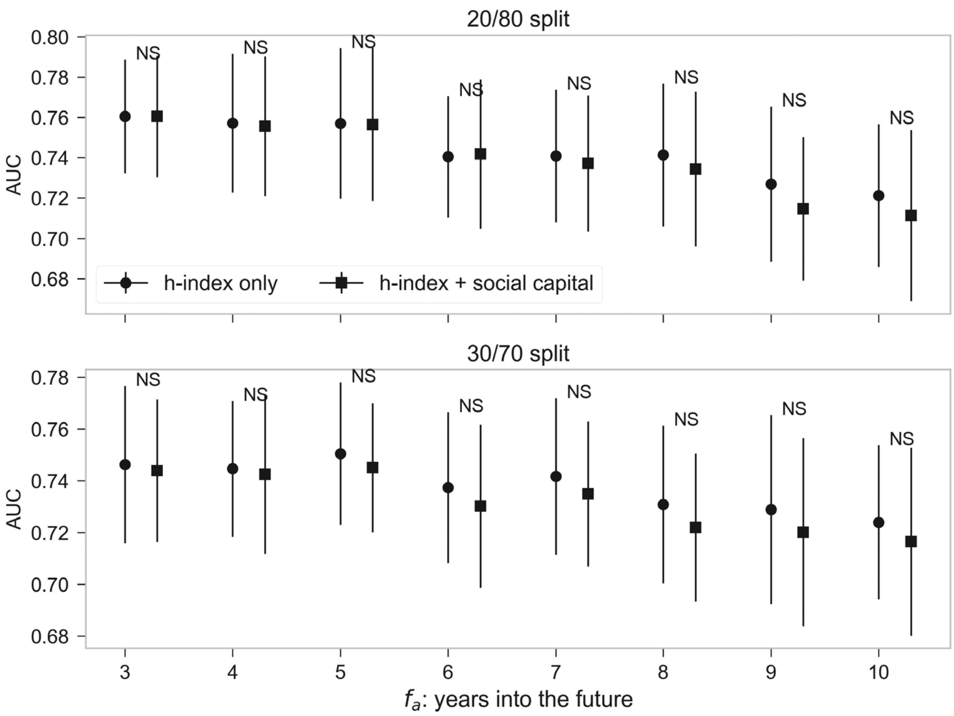

FIGURE 11 Average area under the receiver operating characteristics curve (AUC) scores (markers) across different f_a for senior researchers. Figure aesthetics similar to Figure 9. NS stands for no significant difference is identified

Nonetheless, our findings suggest that besides standing on the shoulders of giants, every tie researchers have ever built can also contribute to their future impact. In fact, by keeping building ties to become a broker that connects various subgroups, one can accumulate her structural capital. Although the resulting benefit may not be immediate, researchers are able to enjoy the advantages of having valuable information flow eventually to boost their future impact in the long run. As for senior researchers, past achievements are the major drivers to their future impact, followed by the reputations of those whom they recently collaborated with. For hiring organizations, as a result, the findings imply that there should be different emphases when recruiting talents at different career stages.

We also achieve decent predictive performance in identifying scholars of great potentials across different career stages. While long-term impact may seem harder to predict due to the stochastic nature of (scientific) careers at first sight, the consistent prediction performance of our models for early-stage, mid-career, or senior researchers on their short- or long-term impact, indeed shows the existence of regularity in researchers' career trajectories. For both individuals and organizations, further exploration of predictive models can be applied in multiple occasions such as hiring and grant decisions. One caveat here is that senior researchers, while seeing the addition of new and high-standing ties may be evidence of attracting more citations, already have well-established identity and network. As such, a binary classification of star or not is unable to benefit from the relational capital. In brief, if a senior researcher has already shown scholarly excellence, it is very likely that she is able to keep the "hot hand".

There are several limitations of this study. First, we only use researchers from the discipline of information systems. Other disciplines may manifest different patterns. Second, collaboration through coauthorship is only part of one's social network and social capital. One can accumulate different types of social capital via various venues, such as informal conversation in conferences. When teams get bigger (e.g., in high energy physics or biomedicine), furthermore, it is also likely that coauthorship no longer captures collaborations accurately. Nonetheless, this is not an issue in this study given the small coauthor team sizes—99% of the papers have no more than three coauthors. Third, the current definition of career age is based on the first publication in Scopus, along with the choice of 3, 5, and 10 years being three various career stages. Even though our categorization of career ages is based on the peer review cycle, this is coarse grained and more data may be needed to more clearly define the starting point of one's research career (e.g., ProQuest dissertation and theses database.) To mitigate biases brought by potential arbitrariness of such categorization, we conducted regression analyses for $c_a \in \{3, ..., 15\}$ along with $f_a \in \{3, ..., 10\}$ with detailed discussions in Appendix D. Career trajectories may, in addition, be affected by many different factors such as family reasons. For example, one may prefer a job offer that can address two body problems, even if another offer which is unable to do this provides better platforms for publications in top tier journals. Therefore, our analysis-based solely on bibliometric databases only reflect part of underlying mechanism of citation impact dynamics. Finally, our study is limited by name ambiguity of researchers, a common problem for research based on scholar data from bibliography databases. Although Scopus has done name disambiguation for authors in its database, there is still room for improvement.

As a pioneer work in understanding and predicting *noncumulative* future research impact, we hope this study can inspire more studies that systematically understand the academic career of individuals. An interesting direction is to focus more on the prediction model—not only individuals care about their future impact, organizations and committees are always looking for accurate estimates of one's future potential. We currently only utilized logistic regression model, where social connections are approximated by one-hop neighbors (i.e., collaborators) in the coauthorship network. Higher orders of connections may also provide subtle signals that contribute to better prediction performance. Further explorations may not only benefit the area on people analytics but also provide unique opportunities to further enrich social capital theory. Moreover, the employed strategy in this study is to aggregate historical information into a single number (e.g., h-index calculated based on all past papers). However, it is possible that not all past information is helpful. Instead, learning what to keep and what to forget may be helpful to improve the predictive performance, implying critical points largely affecting one's career. Besides, the current model setup considers all researchers regarding their career ages. While this mitigates the potentially mixed patterns among those in different career phases, it is also likely that citations may exhibit various patterns from year to year. Therefore, a more fine-grained analysis taking time into consideration may produce more insights into how the prediction of future impact may differ over time, because a discipline may experience changes over years. Finally, the methodology, while based on scholarly data, can in fact be applied beyond academia. In particular, the advancement of IT has enabled social coding and online open collaboration (e.g., GitHub) where people self-organize to contribute to various open source projects. The proposed method can be applied to, for example, understand how social capital affects GitHub users' activity as well as project impact with minor modifications.

ORCID

Zhiya Zuo https://orcid.org/0000-0002-3771-4731
Kang Zhao https://orcid.org/0000-0002-8321-2804

ENDNOTES

[1] The use of this phrase here is being allowed as it is exclusively an example. Current JASIST guidelines and an article published in JASIST discourage use of this phrase as it is considered a pejorative term.

[2] https://aisnet.org/?SeniorScholarBasket.

[3] https://dev.elsevier.com/.

REFERENCES

Abramo, G., D'Angelo, C. A., & Felici, G. (2019). Predicting publication long-term impact through a combination of early citations and journal impact factor. *Journal of Informetrics*, *13*(1), 32–49. https://doi.org/10.1016/j.joi.2018.11.003.

Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Future impact: Predicting scientific success. *Nature*, *489*(7415), 201–202. https://doi.org/10.1038/489201a.

Amjad, T., Ding, Y., Xu, J., Zhang, C., Daud, A., Tang, J., & Song, M. (2017). Standing on the shoulders of giants. *Journal of Informetrics*, *11*, 307–323. https://doi.org/10.1016/j.joi.2017.01.004.

Arellano, C., DiLeonardo, A., & Felix, I. (2017). Using people analytics to drive business performance: A case study. *The McKinsey Quarterly*, *6*.

Aversa, E. S. (1985). Citation patterns of highly cited papers and their relationship to literature aging: A study of the working literature. *Scientometrics*, *7*(3–6), 383–389. https://doi.org/10.1007/BF02017156.

Ayaz, S., Masood, N., & Islam, M. A. (2017). Predicting scientific impact based on h-index. *Scientometrics*, *114*, 1–18. https://doi.org/10.1007/s11192-017-2618-1.

Ayer, T., Chhatwal, J., Alagoz, O., Kahn, C. E., Woods, R. W., & Burnside, E. S. (2010). Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics*, *30*(1), 13–22. https://doi.org/10.1148/rg.301095057.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Bertin, M., Lariviere, V., & Sugimoto, C. R. (2016). The linguistic patterns and rhetorical structure of citation context: An approach using n-grams. *Scientometrics*, *109*, 1417–1434. https://doi.org/10.1007/s11192-016-2134-8.

Bertsimas, D., Brynjolfsson, E., Reichman, S., & Silberholz, J. (2015). OR forum—Tenure analytics: Models for predicting research impact. *Operations Research*, *63*(6), 1246–1261. https://doi.org/10.1287/opre.2015.1447.

Biagioli, M. (2016). Watch out for cheats in citation game. *Nature*, *535*(7611), 201–201. https://doi.org/10.1038/535201a.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (pp. 4–10). New York, NY: Springer.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*(5), 1170–1182. https://doi.org/10.1086/228631.

Bornmann, L., & Daniel, H.-D. (2006). Selecting scientific excellence through committee peer review—A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics*, *68*(3), 427–440. https://doi.org/10.1007/s11192-006-0121-1.

Bornmann, L., Mutz, R., & Daniel, H. D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, *59*(5), 830–837. https://doi.org/10.1002/asi.20806.

Bornmann, L., & Tekles, A. (2019). Productivity does not equal usefulness. *Scientometrics*, *118*(2), 705–707. https://doi.org/10.1007/s11192-018-2982-5.

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534* (7609), 684–687.

Bu, Y., Murray, D. S., Xu, J., Ding, Y., Ai, P., Shen, J., & Yang, F. (2018). Analyzing scientific collaboration with "giants" based on the milestones of career. *Proceedings of the Association for Information Science and Technology*, *55*(1), 29–38.

Bütün, E., Kaya, M., & Alhajj, R. (2017). A supervised learning method for prediction citation count of scientists in citation networks. *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2017*. Sydney, Australia. https://doi.org/10.1145/3110025.3110160.

Cao, X., Chen, Y., & Liu, K. J. R. (2016). A data analytic approach to quantifying scientific impact. *Journal of Informetrics*, *10*, 471–484. https://doi.org/10.1016/j.joi.2016.02.006.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*, 2079–2107.

Chakraborty, T., Kumar, S., Goyal, P., Ganguly, N., & Mukherjee, A. (2014). Towards a stratified learning approach to predict future citation counts. In *Proceedings of the acm/ieee joint conference on digital libraries*. London, UK. (pp. 351–360). https://doi.org/10.1109/JCDL.2014.6970190.

Daud, A., Abbasi, R., & Muhammad, F. (2013). Finding rising stars in social networks. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 7825, pp. 13–24). Berlin/Heidelberg, Germany: Springer Science+Business Media, *PART 1*. https://doi.org/10.1007/978-3-642-37487-6_4.

Daud, A., Ahmad, M., Malik, M. S., & Che, D. (2014). Using machine learning techniques for rising star prediction in co-author network. *Scientometrics*, *102*(2), 1687–1711. https://doi.org/10.1007/s11192-014-1455-8.

Dong, Y., Johnson, R. A., & Chawla, N. V. (2016). Can scientific impact be predicted? *IEEE Transactions on Big Data*, *2*(1), 18–30.

Dorogovtsev, S. N., & Mendes, J. F. (2015). Ranking scientists. *Nature Physics*, *11*(11), 882–883.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152.

Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D., & Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *Journal of the American Medical Association*, *280*(3), 222–224. https://doi.org/10.1001/jama.280.3.222.

Fong, E. A., & Wilhite, A. W. (2017). Authorship and citation manipulation in academic research. *PLoS One*, *12*(12), e0187394. https://doi.org/10.1371/journal.pone.0187394.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35. https://doi.org/10.2307/3033543.

García-Pérez, M. A. (2013). Limited validity of equations to predict the future h index. *Scientometrics*, *96*(3), 901–909. https://doi.org/10.1007/s11192-013-0979-7.

Garfield, E. (1955). Citation indexes for science; a new dimension in documentation through association of ideas. *Science*, *122* (3159), 108–111.

Garfield, E. (1999). Journal impact factor: A brief review. *CMAJ: Canadian Medical Association Journal*, *161*(8), 979–980.

Gingras, Y., Larivière, V., Macaluso, B., & Robitaille, J.-P. (2008). The effects of aging on researchers' publication and citation patterns. *PLoS One*, *3*(12), e4048.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hornbostel, S., Böhmer, S., Klingsporn, B., Neufeld, J., & von Ins, M. (2009). Funding of young scientist and scientific excellence. *Scientometrics*, *79*(1), 171–190. https://doi.org/10.1007/s11192-009-0411-5.

ISR. (2020). *Editorial statement | information systems research*. Retrieved from https://pubsonline.informs.org/page/isre/editorial-statement.

Katz, J., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1.

Kaur, J., Ferrara, E., Menczer, F., Flammini, A., & Radicchi, F. (2015). Quality versus quantity in scientific impact. *Journal of Informetrics*, *9*(4), 800–808. https://doi.org/10.1016/j.joi.2015.07.008.

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7426–7431. https://doi.org/10.1073/pnas.1424329112.

Kelly C., Jennions M (2006). The h index and career assessment by numbers. *Trends in Ecology & Evolution*, *21*(4), 167–170. http://dx.doi.org/10.1016/j.tree.2006.01.005.

Kosmulski, M. (2006). A new hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, *2* (3), 4–6.

Leonardi, P., & Contractor, N. (2018). Better people analytics. *Harvard Business Review*, (November–December 2018), 70–81. https://hbr.org/2018/11/better-people-analytics.

Li, E. Y., Liao, C. H., & Yen, H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy*, *42*(9), 1515–1530. https://doi.org/10.1016/j.respol.2013.06.012.

Li, X. L., Foo, C. S., Tew, K. L., & Ng, S. K. (2009). Searching for rising stars in bibliography networks. In: X. Zhou, H. Yokota, K. Deng, & Q. Liu, (eds.). *Database Systems for Advanced Applications. DASFAA 2009. Lecture Notes in Computer Science*, vol 5463. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00887-0_25.

Martinson, B. C., Crain, A. L., Anderson, M. S., & De Vries, R. (2009). Institutions' expectations for researchers' self-funding, federal grant holding, and private industry involvement: Manifold drivers of self-interest and researcher behavior. *Academic*

Medicine, *84*(11), 1491–1499. https://doi.org/10.1097/ACM.0b013e3181bb2ca6.

Mazloumian, A. (2012). Predicting scholars' scientific impact. *PLoS One*, *7*(11), e49246–e49246. https://doi.org/10.1371/journal.pone.0049246.

McNutt, M. (2014). The measure of research merit. *Science*, *346*(6214), 1155. https://doi.org/10.1126/science.aaa3796.

Meyer, B., Choppy, C., Staunstrup, J., & van Leeuwen, J. (2009). Viewpoint research evaluation for computer science. *Communications of the ACM*, *52*(4), 31–34.

MISQ. (2020). *About mis quarterly*. Retrieved from https://www.misq.org/about/.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242–266. https://doi.org/10.5465/amr.1998.533225.

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5200–5205.

Nezhadbiglari, M., Gonçalves, M. A., & Almeida, J. M. (2016). Early prediction of scholar popularity. *Proceedings of the 16th ACM/-IEEE-CS on joint conference on digital libraries*, pp. 181–190. Newark, NJ: Rutgers University. https://doi.org/10.1145/2910896.2910905.

Packalen, M., & Bhattacharya, J. (2019). Age and the trying out of new ideas. *Journal of Human Capital*, *13*(2), 341–373.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web. *World Wide Web Internet and Web Information Systems*, *54* (1999–66), 1–17.

Parsons, T. (1938). The role of theory in social research. *American Sociological Review*, *3*(1), 13–20.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Penner, O., Pan, R. K., Petersen, A. M., Kaski, K., & Fortunato, S. (2013). On the predictability of future impact in science. *Scientific Reports*, *3*(1), 3052. https://doi.org/10.1038/srep03052.

Ponomariov, B., & Toivanen, H. (2014). Knowledge flows and bases in emerging economy innovation systems: Brazilian research 2005–2009. *Research Policy*, *43*(3), 588–596.

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306.

Radicchi, F., Weissman, A., & Bollen, J. (2017). Quantifying perceived impact of scientific publications. *Journal of Informetrics*, *11*(3), 704–712. https://doi.org/10.1016/J.JOI.2017.05.010.

Rai, A. (2018). Beyond Outdated Labels: The Blending of IS Research Traditions. *MIS Quarterly*, *42*(1), iii–vi. https://aisel.aisnet.org/misq/vol42/iss1/2/.

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., … Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, *1*(1), 18. https://doi.org/10.1038/s41746-018-0029-1.

Sabidussi, G. (1966). The centrality of a graph. *Psychometrika*, *31*(4), 581–603.

Sarigöl E., Pfitzner R., Scholtes I., Garas A., Schweitzer F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, *3*(1), 9. http://dx.doi.org/10.1140/epjds/s13688-014-0009-x.

Seeber, M., Cattaneo, M., Meoli, M., & Malighetti, P. (2019). Self-citations as strategic response to the use of metrics for career decisions. *Research Policy*, *48*(2), 478–491. https://doi.org/10.1016/j.respol.2017.12.004.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabasi, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*(6312) aaf5239–aaf5239, aaf5239. https://doi.org/10.1126/science.aaf5239.

Stegehuis, C., Litvak, N., & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of Informetrics*, *9*(3), 642–657. https://doi.org/10.1016/J.JOI.2015.06.005.

Strathern, M. (1997). Improving ratings': Audit in the British university system. *European Review*, *5*, 305–321. https://doi.org/10.1002/(SICI)1234-981X(199707)5:33.0.CO;2-4.

Taylor, S. W., Fender, B. F., & Burke, K. G. (2006). Unraveling the academic productivity of economists: The opportunity costs of teaching and service. *Southern Economic Journal*, *72*, 846–859.

Van Raan, A. F. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467–472. https://doi.org/10.1023/B:SCIE.0000018543.82441.f1.

Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, *40*(3), 463–472.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391.

Wang, D., Song, C., & Barabási, A.-L. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127–132. https://doi.org/10.1126/science.1237825.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872. https://doi.org/10.1007/s11192-012-0775-9.

Way, S. F., Larremore, D. B., & Clauset, A. (2016). Gender, productivity, and prestige in computer science faculty hiring networks. In *Proceedings of the 25th international conference on world wide web - www '16*, (pp. 1169–1179). Montreal, Canada. https://doi.org/10.1145/2872427.2883073.

Weihs, L., & Etzioni, O. (2017). Learning to predict citation-based impact measures. In *2017 acm/ieee joint conference on digital libraries (jcdl)* (pp. 1–10). Location: Toronto, Ontario, Canada. https://doi.org/10.1109/JCDL.2017.7991559.

Wilhite, A. W., & Fong, E. A. (2012). Coercive citation in academic publishing. *Science*, *335*(6068), 542–543. https://doi.org/10.1126/science.1212540.

Xiao, S., Yan, J., Li, C., Jin, B., Xiangfeng, W., Yang, X., …Zha, H. (2016). On modeling and predicting individual paper citation count over time. *IJCAI international joint conference on artificial intelligence*, *2016-Janua*, (pp. 2676–2682). New York, NY, USA.

Yan, R., Huang, C., Tang, J., Zhang, Y., & Li, X. (2012). To better stand on the shoulder of giants. In *Proceedings of the 12th acm/-ieee-cs joint conference on digital libraries - jcdl '12*, p. 51. Washington DC: The George Washington University. https://doi.org/10.1145/2232817.2232831.

Yan, R., Tang, J., Liu, X., Shan, D., & Li, X. (2011). Citation count prediction. *Proceedings of the 20th ACM international conference on information and knowledge management - CIKM '11*,

p. 1247. Glasgow, Scotland, UK. https://doi.org/10.1145/2063576.2063757.

Yu, T., Yu, G., Li, P.-Y., & Wang, L. (2014). Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics*, *101*(2), 1233–1252. https://doi.org/10.1007/s11192-014-1279-6.

Zuo, Z., Qian, H., & Zhao, K. (2019). Understanding the field of public affairs through the lens of ranked Ph.D. programs in the United States. *Policy Studies Journal*, *47*(S1), S159–S180. https://doi.org/10.1111/psj.12322.

Zuo, Z., & Zhao, K. (2018). The more multidisciplinary the better? The prevalence and interdisciplinarity of research collaborations in multidisciplinary institutions. *Journal of Informetrics*, *12*(3), 736–756.

Zuo, Z., Zhao, K., & Eichmann, D. (2017). The state and evolution of U.S. iSchools: From talent acquisitions to research outcome. *Journal of the Association for Information Science and Technology*, *68*(5), 1266–1277. https://doi.org/10.1002/asi.23751.

Zuo, Z., Zhao, K., & Ni, C. (2019). Standing on the shoulders of giants? Faculty hiring in information schools. *Journal of Informetrics*, *13*(1), 341–353.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zuo Z, Zhao K. Understanding and predicting future research impact at different career stages—A social network perspective. *J Assoc Inf Sci Technol*. 2021;72:454–472. https://doi.org/10.1002/asi.24415