Journal of Informetrics 14 (2020) 101041

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Early-stage reciprocity in sustainable scientific collaboration

Wei Wang[a], Jing Ren[a], Mubarak Alrashoud[b], Feng Xia[c,*], Mengyi Mao[a], Amr Tolba[d,e]

[a] School of Software, Dalian University of Technology, Dalian 116620, China
[b] Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11437, Saudi Arabia
[c] School of Science, Engineering and Information Technology, Federation University Australia, Ballarat, VIC 3353, Australia
[d] Computer Science Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia
[e] Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shebin-El-kom 32511, Egypt

ARTICLE INFO

Article history:
Received 2 August 2019
Received in revised form 29 February 2020
Accepted 1 April 2020
Available online 7 June 2020

Keywords:
Scientific collaboration
Reciprocity
Collaboration sustainability
Science of science

ABSTRACT

Scientific collaboration is of significant importance in tackling grand challenges and breeding innovations. Despite the increasing interest in investigating and promoting scientific collaborations, we know little about the collaboration sustainability as well as mechanisms behind it. In this paper, we set out to study the relationships between early-stage reciprocity and collaboration sustainability. By proposing and defining h-index reciprocity, we give a comprehensive statistical analysis on how reciprocity influences scientific collaboration sustainability, and find that scholars are not altruism and the key to sustainable collaboration is fairness. The unfair h-index reciprocity has an obvious negative impact on collaboration sustainability. The bigger the reciprocity difference, the less sustainable in collaboration. This work facilitates understanding sustainable collaborations and thus will benefit both individual scholar in optimizing collaboration strategies and the whole academic society in improving teamwork efficiency.

1. Introduction

Scientific collaboration is becoming one of the most significant features of scientific activities in modern academia and has evolved gradually to become one of the most important forms of knowledge production. Studies have shown that scholars are becoming more and more collaborative and collaborative papers are gaining more citations (Adams, 2013; Sonnenwald, 2007; Zuo & Zhao, 2018). As scientific collaboration networks are expanding and characterizing the social construction of science (Tang et al., 2008; Xia, Wang, Bekele, & Liu, 2017), many researchers and institutions have a growing interest in explaining the patterns of scientific collaboration and co-authorship (Cohen & Ebel, 2013; Tang, Wu, Sun, & Su, 2012; Tsai & Lin, 2016; Wu, Venkatramanan, & Chiu, 2015).

Although scientific collaboration is dynamic in nature, the majority of existing studies investigate the static collaboration behaviours (Newman, 2001b, 2004; Zhang, Bu, Ding, & Xu, 2018). Tracing and modeling the dynamics of scientific collaboration can improve our understanding of successful scientific collaborations and enhance collaboration efficiency and productivity (Barabâsi et al., 2002; Sinatra, Wang, Deville, Song, & Barabási, 2016; Wang, Cui, et al., 2017). Usually, scientific collaboration is not a one-shot deal where two scholars may collaborate many times. While a variety of feature, patterns,

* Corresponding author.
E-mail address: f.xia@ieee.org (F. Xia).

https://doi.org/10.1016/j.joi.2020.101041

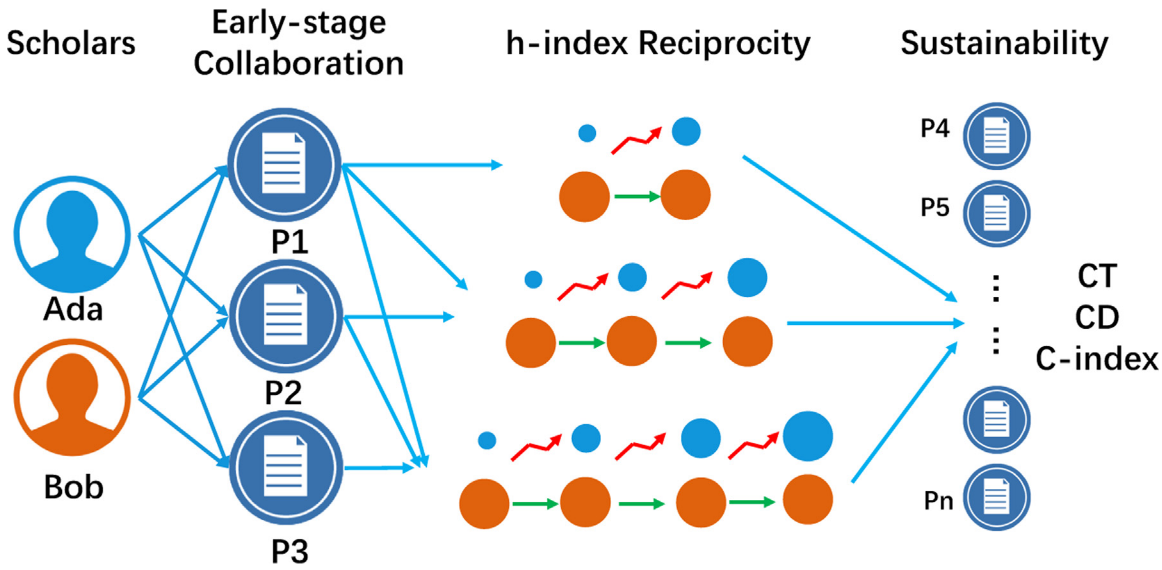

Fig. 1. An illustrative example of early-stage reciprocity in scientific collaboration. It shows an extreme case that each collaboration increases Ada's h-index, while Bob does not benefit in h-index increasing all the time. We want to explore how such reciprocity in h-index influences the collaboration sustainability in terms of collaboration times, duration, and collaboration index.

and strategies of scientific collaboration have been identified in previous work (Wang, Yu, Bekele, Kong, & Xia, 2017; Zhang et al., 2018), there is a lack of systematic understanding of factors that make collaboration sustainable.

In this paper, we aim at leveraging the early-stage reciprocity to reveal sustainable scientific collaborations. Reciprocity is an important factor in human social interaction. Reciprocity means that "If you do something for me, I will do something you" which indicates that people will reward kind actions and punish unkind actions to them (Berg, Dickhaut, & McCabe, 1995; Falk & Fischbacher, 2006; Li, Aste, Caccioli, & Livan, 2019; Romano & Balliet, 2017). Based on the reciprocity, we aim to address the following questions: **Q1** What is the relationship between early-stage reciprocity and scientific collaboration sustainability? **Q2** What kind of early-stage reciprocity may bring more persistent collaboration?

Specifically, we propose and define the h-index reciprocity. We know that scientific collaboration will benefit both scholars in increasing publication quantity. However, the collaboration may have different benefits in impact, e.g., h-index increase (see Fig. 1). To explore **Q1** and **Q2**, we use a comprehensive descriptive analysis of the relationships between early-stage reciprocity (including first, twice, and thrice reciprocity) and collaboration sustainability in terms of collaboration times, duration, and index based on APS dataset. We find that unfair reciprocity accounts for more proportions with short sustainability and less proportion with long sustainability. Moreover, to account for prestige effects, we define two groups including top scholars and the beginner scholars and find that there are fewer unfair reciprocities between top scientist. Furthermore, we compare the h-index reciprocity between persistent collaborations and non-persistent collaborations and find that the precondition of a persistent collaboration is that the collaboration is fair in reciprocity.

2. Related work

2.1. Reciprocity

Reciprocity is a critical factor in human social interaction. The simplest concept of reciprocity is "If you do something for me, I will do something for you", in other words, people are reciprocal if they reward kind actions and punish unkind actions to them (Sandoval, Brandstetter, Obaid, & Bartneck, 2016). It has been studied and proven that reciprocity outperforms conformity in promoting human cooperation (Li et al., 2019; Romano & Balliet, 2017). Enhancing reciprocity by giving people a small gift before requesting a contribution increases the likelihood of a positive contribution (Alpizar, Carlsson, & Johansson-Stenman, 2008). Since reciprocity is highly related to human behaviors, it has been extensively used to explore human cooperation with game theory such as the prisoner's dilemma, and public goods games (Cortés-Berrueco, Gershenson, & Stephens, 2016; Falk & Fischbacher, 2006; Yang & Chen, 2015). However, although scholars have explored the impact of homophily, transitivity, and preferential on promoting new scientific collaboration (Chung, Kwon, & Lee, 2016), little is known about how the reciprocity influences the collaboration dynamics between scientists.

2.2. Scientific collaboration

Previous studies on scientific collaboration analysis mainly fall into two categories including macroscopic collaboration behaviour analysis and microscopic collaboration behaviour analysis (Amjad et al., 2017). In microscopic collaboration behaviour analysis, the patterns of international scientific collaboration have been extensively studied for its impetus in improving research quality, advancing the efficiency of the scientific production, and fostering breakthroughs in a shorter time (Bu, Ding, Liang, & Murray, 2018; Coccia & Wang, 2016; Petersen et al., 2014; Zhang et al., 2018). Coccia and Wang (2016) explore the long-run patterns of international research collaboration across scientific fields as well as their structural changes over time. One key finding therein is that the convergence of collaboration patterns across research fields is possibly one of contributing factors that support the evolution of scientific disciplines. Zhang et al. (2018) propose an approach to analyzing the differences in possibilities that two authors (i.e., scholars) will cooperate with each other from the perspectives of homophily, transitivity, and preferential attachment. The authors find that considering the authors' attributes and homophily effects as well as the transitivity and preferential attachment effects of the co-authorship network in which they are embedded can help better the understanding of scientific collaboration. Petersen (2015) adopts an egocentric view to track research careers from their inception along their longitudinal growth trajectory, revealing that scientific collaboration might be characterized by a high turnover rate juxtaposed with surprisingly frequent "life partners".

Most macroscopic collaboration behaviour analysis is performed based on network approach by constructing a scientific collaboration network based on co-authorship. From a network science perspective, scientific collaborations/co-authorship among scholars will naturally form a social network. It has been widely known that the preferential attachment influences the generation of new scholarly collaborations (Newman, 2001a). Barabâsi et al. (2002) demonstrate the presence of preferential attachment in two collaboration networks in the fields of mathematics and neuroscience for an eight-year period. It is found that a new publication is more likely to emerge among those who already have a large number of co-authors. Newman (Newman, 2001b, 2004) analyzes the preferential attachment in co-authorship networks in the fields of physics and biology, showing that the number of new collaborations one scholar/author gains each year increases with the number of past collaborators. Recently, due to the importance of scientific collaboration, many collaboration-based academic services have been developed, for example, collaborator recommendation (Wang, Liu, Yang, Kong, & Xia, 2019).

2.3. Persistence in science

It is well-known that persistence is important to success. By proposing and analyzing UCP (Uninterrupt and Continuous Presence) behaviors, Ioannidis, Boyack, and Klavans (2014) find that UCP scholars can gain more citations than non-UCP scholars, which indicates that persistence is important to the structure, stability, and vulnerability of a scientific career. However, they merely study publication persistence while the collaboration persistence is not considered. Few research has been done to explore the persistent collaboration. Petersen (2015) studies the impact of weak, strong, and super collaboration on researchers' career from an egocentric perspective. He finds that sustainable collaborations, especially "life-partner" collaborators, have a significant positive impact on productivity and citation. Wang, Cui, et al. (2017) design a model based on structural similarity indices, authorship properties, and research interests to predict collaboration sustainability. Although these research explore the significance as well as the predictability of collaboration sustainability, factors that influence the sustainability of scientific collaboration are not clear.

3. Data and methodology

3.1. Dataset

We adopt the APS (American Physical Society) dataset to investigate the influence of early-stage reciprocity on collaboration sustainability. The APS dataset is a corpus of many high-quality physical journals containing over 45,000 articles since 1893. It can be accessed online by requesting.[1] The dataset contains two subsets. The article meta dataset consists of the basic metadata of all APS journals including author list from which we can extract collaboration pairs and construct collaboration networks. The citing article pairs dataset consists of pairs of APS articles that cite each other from which we can calculate the c_{10} and h_{10}.

We first perform name disambiguation to authors in APS dataset from 1893 to 2015 based on the method in previous work (Sinatra et al., 2016). Then, to eliminate authors who leave research at early stage of their career (Sinatra et al., 2016), we extract scholars who: (i) have published at least one paper every 5 years, (ii) have published at least 10 papers, and (iii) their career lasts at least 20 years, arriving 7284 scholars. These scholars are **S**eed **S**cholars (**SS**). Next, we extract SS's **C**ollaborators (**SC**) based on the limitation that their first collaboration should happen before 2006 to ensure sufficient time for h_{10} and c_{10} calculation, arriving 69,394 scholars and 172,680 collaboration pairs. Note that a scholar may involve in various collaboration pairs with different collaborators. Then, we calculate each collaboration pair's c_{10}, h_{10}, CT, CD, and C-index based on their definitions.

The statistics of the dataset is shown in Fig. 2. We can observe from this figure that: (1) The average CT, CD, and C-index of all collaboration is 5.42, 4.36, and 3.63, respectively. (2) h_0 of SS is higher than that of SC. The reason is that SS is selected with at least 10 papers while their collaborator SC is not selected.

[1] https://journals.aps.org/datasets.

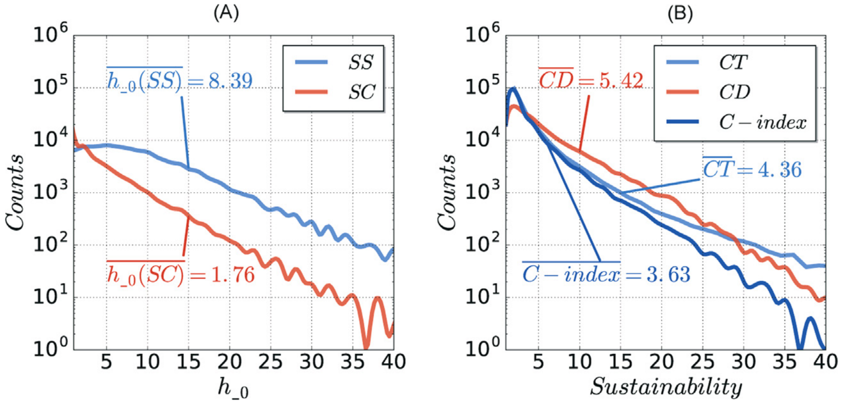

Fig. 2. Statistics of data set. Both h_0 and sustainability follow a power-law distribution.

3.2. Methods

In this study, we explore the problem of early-stage reciprocity in sustainable collaboration as follows: given the early-stage (first three times) collaboration behaviors between scholars, we aim to investigate how the h-index reciprocity influences the collaboration sustainability. We first introduce some commonly-used terms in this work including collaboration sustainability, h-index reciprocity, and h-index reciprocity calculation.

3.2.1. Collaboration sustainability

Scientific collaboration is usually not a one-shot deal. We explore the sustainability of scientific collaboration from three perspectives, including collaboration times (CT), collaboration duration (CD), and collaboration index (C-index).

Definition 1. Collaboration Times. CT is the direct reflection of the collaboration sustainability. The collaboration times shows how many papers two scholars have coauthored with each other. In most cases, a sustainable collaborator will bring a higher collaboration times. However, in some extreme cases where two scholars collaborate many times (e.g., 15 papers) in a short time period (e.g., 2 years), those collaborations are not sustainable.

Definition 2. Collaboration Duration. CD reflects how long two scholars have collaborated with each other. Usually, a sustainable collaboration will last for a long time. In some extreme cases, however, where two scholars collaborate for a long time (e.g., 15 years) with few publications (e.g., 2 publications), those collaborations cannot be regarded as sustainable collaboration either.

Definition 3. Collaboration Index. C-index is proposed to better quantify the collaboration sustainability between two scholars. The C-index is inspired by the idea of h-index which is proposed to quantify scientific output (Hirsch, 2005). Specifically, a collaboration has C-index h if two scholars collaborate with each other at least h times within at least h years. For example, the collaboration between scholars A and B has 10 C-index if they collaborate with each other at least 10 times during 10 years.

3.2.2. h-index reciprocity

Reciprocity is an important factor in human-human interaction. In this article, we explore the scholars' reciprocal behaviors in terms of h-index. There is no doubt that two coauthored scholars (e.g., Ada and Bob) will benefit from the collaboration in publication quantity increase. However, the quality of the collaborated paper may have different influences on the scholars' academic impact, i.e., h-index. For example, scholar Ada may gain h-index increase while scholar Bob may not. We aim to explore the influence of such h-index reciprocity on the collaboration sustainability.

The reasons why we employ the h-index to measure the reciprocity are as follows: (1) h-index is one of the most important measures of scholar influence which can directly show the reputation of a given scholar. Scholars will be more interested in papers that can improve their h-index (Dong, Johnson, & Chawla, 2015); (2) h-index reciprocity is one of the direct outcomes of scientific collaboration. Whether the collaborated paper will increase scholars' h-index can definitely evaluate the success of this collaboration; Instead of an abstract concept of reciprocity, h-index reciprocity can be calculated directly from the data, where the subjective bias can be avoided. Specifically, we define two kinds of h-index reciprocity:

Definition 4. Inexplicit Reciprocity. For this mechanism, scholars merely gain publication quantity increase from the collaboration. Their number of papers increases by one, but this publication will not increase their h-index.

Definition 5. Explicit Reciprocity. For this situation, scholars gain both publication quantity and h-index increase. By h-index increase, we mean that the citation of the co-authored paper meets the minimum requirements of the h-index threshold of the scholar.

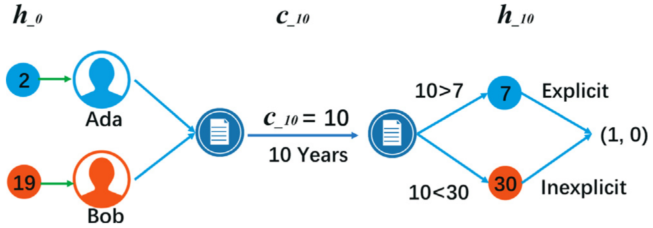

Fig. 3. An illustrative example of explicit and inexplicit reciprocity calculation. The h-index reciprocity for Ada is explicit because the paper's c_{10} is larger than Ada's h_{10} ($10 > 7$). On the other hand, Bob receives the inexplicit reciprocity ($10 < 30$). The reciprocity of this collaboration is (1, 0).

Table 1
Reciprocity categories. In all cases, we do not distinguish (x, y), i.e., (1, 3) with (y, x), i.e., (3, 1) since scientific collaborations are indirect.

Category	First	Twice	Thrice
Fair	(0, 0)(1, 1)	(0, 0)(1, 1) (2, 2)	(0, 0)(1, 1)(2, 2) (3, 3)
Unfair	(1, 0)	(0, 1)(0, 2) (1, 2)	(0, 1)(0, 2)(0, 3) (1, 2)(1, 3)(2, 3)

3.2.3. Reciprocity calculation

We use c_{10} (Sinatra et al., 2016), the number of citations 10 years after publication, to measure the future impact of a collaborative publication. We use the h_{10}, the h-index of the scholar 10 years after collaboration, to measure the future impact of a scholar. If the c_{10} is larger than the h_{10}, this collaboration brings explicit reciprocity to this scholar. Otherwise, this collaboration is inexplicit reciprocity. Fig. 3 shows an example of the explicit and inexplicit reciprocity calculation. Note that we calculate the explicit/inexplicit reciprocity by comparing the c_{10} with h_{10} instead of h_0, the h-index of the scholar when collaborating. The reason is the time delay in getting citations. It takes time for papers to gain citations so that we use c_{10} to measure the paper impact. However, the papers contributed a scholar's h-index when collaborating may not have enough time to gain citations. Therefore, we use h_{10} to ensure sufficient time for h-index calculation.

If a scholar receives an explicit reciprocity, we assign him/her a value 1; if a scholar receives an inexplicit reciprocity, we assign him/her a value 0. Thus, the h-index reciprocity in Fig. 3 is (1, 0). Here, (x, y) denotes the expected reciprocity outcomes of two collaborated scholars. That is, the h-index reciprocity output of two collaborated scholars. If a scholar receive the h-index reciprocity, the value will be 1, otherwise 0. Since collaboration is undirected, we do not distinguish between (0, 1) and (1, 0). For multiple h-index reciprocities, we calculate their cumulative results. Thus, we have six kinds of h-index reciprocity for two times collaborations and ten kinds of h-index reciprocity for three times collaborations. Based on their outcomes, we define them into two categories including fair and unfair, as shown in Table 1. The fair reciprocity means that two scholars receive the same h-index increase from their collaborations. Otherwise, it is unfair reciprocity.

We conduct the analysis with a descriptive overview on the effects of reciprocity on CT, CD, and C-index. We first study the statistics of first collaboration reciprocity and its relationships with CT, CD, and C-index. We aim to explore the relationships between reciprocity and collaboration sustainability for different h_0 collaboration pairs. Similar statistical analysis has been done on multiple reciprocities including first two times collaborations and first three times collaborations. Since we aim to explore the mechanism behind the sustainable collaboration, more times of collaboration would be meaningless. Our goal is to explore what factors in the early stage of scientific collaboration may be related to collaboration sustainability. For early-stage, we intend to explore the first collaboration. For more generality, we further explore the first two times collaborations and the first three times collaborations.

4. Results

In this section, we explore the relationships between reciprocity and sustainability in terms of first collaboration, first two collaborations, and first three collaborations. Meanwhile, to account for the prestige effects, we investigate the early-stage h-index reciprocity for top-cited scholars. Finally, to further reveal the mechanism of reciprocity, the early-stage h-index reciprocity of most sustainable collaborations is investigated.

4.1. Once reciprocity

Table 2 and Fig. 4 show the results between once reciprocity and collaboration sustainability. As shown in Table 1, there are three kinds of reciprocity in first collaboration including (0, 0), (0, 1), and (1, 1). We can see from Fig. 4(a) that (1, 1) reciprocity accounts for the largest proportion (37%). Table 2 shows the average CT, CD, and C-index of each reciprocity. Surprisingly, the (0, 0) reciprocity achieves higher average CT (5.29) and CD (5.63) than those (5.24 and 6.13) of (1, 1). However, the difference is relatively small and (1, 1) achieves higher C-index. We can see that the fair reciprocity including (0, 0) and (1, 1) brings higher sustainability than unfair reciprocity (0, 1).

Table 2
Sustainability of once reciprocity. The blue color indicates the highest number while the red color indicates the lowest number.

Reciprocity		Proportion	CT	CD	C-index
Fair	(0, 0)	27%	**5.29**	**6.53**	4.29
	(1, 1)	37%	5.24	6.13	**4.31**
Unfair	(0, 1)	36%	**4.01**	**5.08**	**3.41**

We further shows the probability density function (PDF) of (0, 0), (0, 1), and (1, 1) in terms of CT, CD, and C-index in Fig. 4(b)–(d), respectively. We can see from these figures that the unfair reciprocity (0, 1), which is presented by the red color scheme, accounts for more proportions with short (i.e., C-index≤4) sustainability and less proportion with long (i.e., C-index >5) sustainability than the fair reciprocity (0, 0) and (1, 1), which is presented by the blue color scheme.

To get a sense for how scholars' h_0 influences collaboration sustainability with different reciprocity, we show the sustainability distribution over different h_0 of (0, 0), (0, 1) and (1, 1) reciprocity in Figs. 5–7, respectively. We can observe from these figures that: (1) In fair reciprocity (0, 0) (Fig. 5(a)) and (1, 1) (Fig. 7(a)), the h_0 of SS is relatively high than that of SC while in unfair reciprocity (0, 1) (Fig. 6(a)), the h_0 of SS is obviously higher than that of SC. This indicates that for unfair reciprocity (0, 1), the author impact difference between two scholars is obvious. In other words, a "big name" collaborates with "no name". (2) In all three figures, with the increase of h_0 the collaboration sustainability will decrease accordingly. This is because that a scholar with a small h_0 may be a student and advisor-advisee relationships will be maintained for a long time.

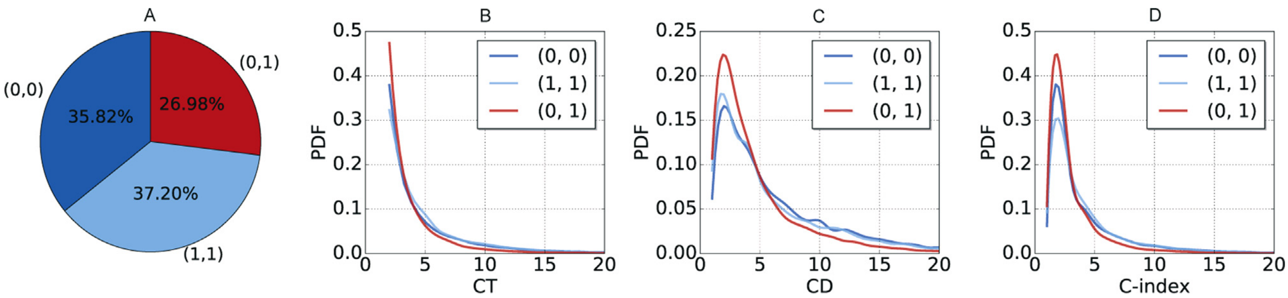

Fig. 4. Proportion and PDF of each reciprocity in the first collaboration. The red color scheme means the unfair reciprocity, (0, 1) while the blue color scheme means the fair reciprocity, (0,0) and (1, 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

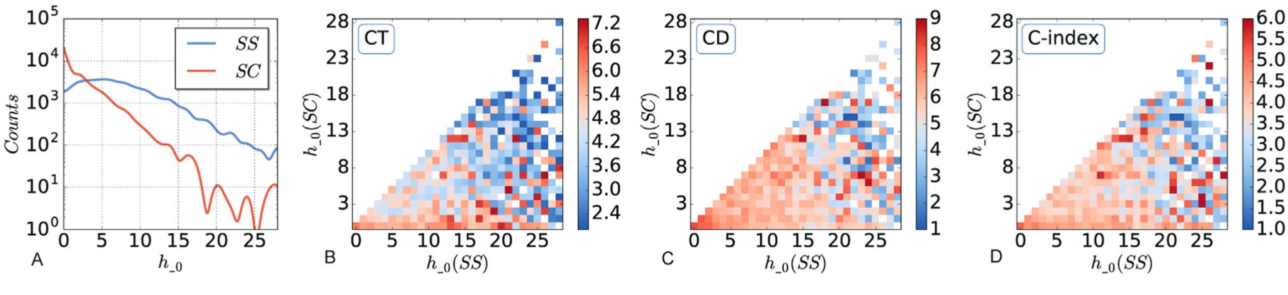

Fig. 5. Sustainability distribution over different h_0 of (0,0) strategy. Subfigure A shows the h_0 distribution of SS and SC. Subfigures B, C, and D show the heat maps of CT, CD, and C-index between different pairs of SS and SC, respectively. The x-axis corresponds to the h_0 of SS and the y-axis to the h_0 of SC.

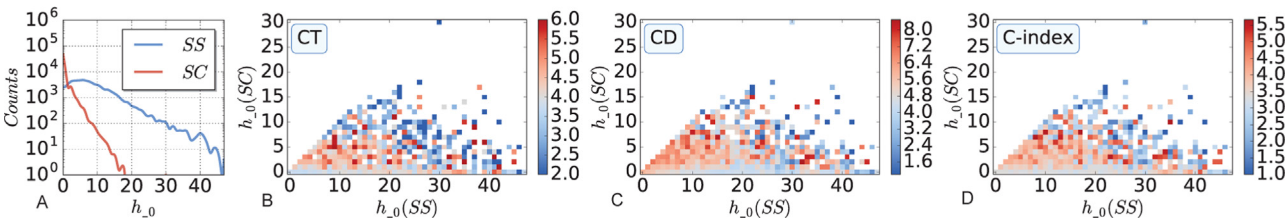

Fig. 6. Sustainability distribution over different h_0 of (0,1) strategy.

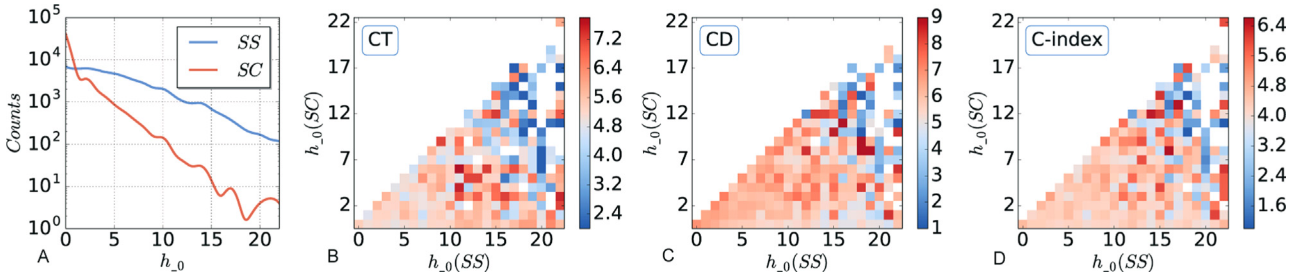

Fig. 7. Sustainability distribution over different h_0 of (1, 1) strategy.

Table 3
Sustainability of twice reciprocity.

Reciprocity		Proportion	CT	CD	C-index
Fair	(0, 0)	13%	**7.66**	**8.68**	5.96
	(1, 1)	19%	7.64	8.41	**6.08**
	(2, 2)	19%	7.13	7.71	5.71
Unfair	(0, 1)	17%	6.45	7.44	5.26
	(0, 2)	14%	**5.30**	**6.23**	**4.45**
	(1, 2)	18%	6.07	6.87	4.98

4.2. Twice reciprocity

In twice reciprocity, there are three kinds of fair reciprocity including (0, 0), (1, 1), and (2, 2), and three kinds of unfair reciprocity including (0, 1), (0, 2) and (0, 3). Note that for twice reciprocity, we screen out these collaborations which last only once so that the CT, CD, and C-index are higher than that of once reciprocity. Table 3 and Fig. 8 show the results between twice reciprocity and collaboration sustainability. It can be seen from Fig. 8(a) that (1, 1) reciprocity accounts for the largest proportion (18.69%). We can see from Table 3 that although the (0, 0) reciprocity reaches the highest CT and CD, the sustainability difference among the three fair reciprocity is small. Meanwhile, the unfair reciprocity (0, 2) has the lowest sustainability.

Similar to once reciprocity, we show the PDF of all fair and unfair reciprocity in terms of CT, CD, and C-index in Fig. 8, where the red color scheme means the unfair reciprocity, (0, 1), (0, 2), and (1, 2) while the blue color scheme means the fair reciprocity, (0,0), (1, 1), and (2, 2). We can observe similar trends with once reciprocity that unfair reciprocity accounts for more proportions with short sustainability and less proportion with long sustainability. Due to the page limitation, we do not show the heat maps of CT, CD, and C-index between different h_0 pairs of SS and SC in twice and trice reciprocity. In fact, the trends are similar to that of once reciprocity.

4.3. Thrice reciprocity

In thrice reciprocity, there are four kinds of fair reciprocity including (0, 0), (1, 1), (2, 2), (3, 3) and six kinds of unfair reciprocity including (0, 1), (0, 2), (0, 3), (1, 2), (1, 3), and (2, 3). For thrice reciprocity, we screen out these collaborations which last only twice so that the CT, CD, and C-index are higher than that of once and twice reciprocity. Table 4 and Fig. 9 show the results between twice reciprocity and collaboration sustainability. We can see from Fig. 9(a) that (2, 2) reciprocity accounts for the largest proportion (13.84%). Similar trends can be seen from Table 4 and Fig. 9 that fair reciprocity results in higher sustainability than unfair reciprocity.

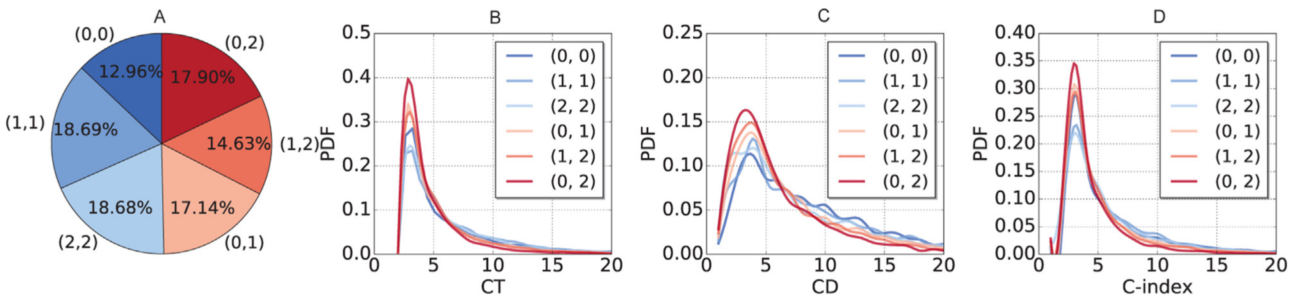

Fig. 8. Proportion and PDF of each reciprocity in twice collaborations. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Table 4
Sustainability of thrice reciprocity.

Reciprocity		Percent	CT	CD	C-index
Fair	(0, 0)	8%	9.84	10.18	7.34
	(1, 1)	11%	9.92	10.07	7.55
	(2, 2)	14%	9.42	9.71	7.41
	(3, 3)	10%	8.95	8.92	6.96
Unfair	(0, 1)	10%	8.75	9.24	6.89
	(0, 2)	11%	7.43	8.03	6.06
	(0, 3)	5%	6.57	7.05	5.36
	(1, 2)	13%	8.45	8.82	6.70
	(1, 3)	8%	7.10	7.49	5.75
	(2, 3)	10%	7.74	8.05	6.12

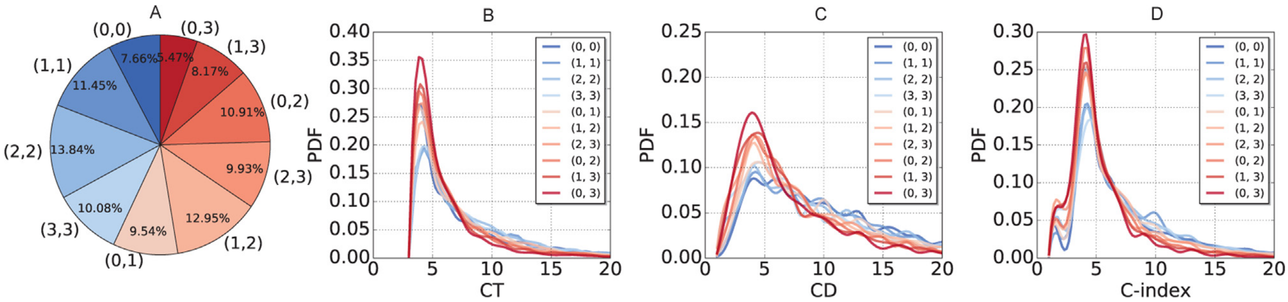

Fig. 9. Proportion and PDF of each reciprocity in thrice collaborations.

Table 5
Reciprocity results between top scientists. The collaboration between top scientists means that the h_0 of SS and SC is no less than 7, compared beginner scholars where the h_0 of SS and SC is less than 1.

		First			Twice						Thrice									
Reciprocity		(0,0)	(0,1)	(1,1)	(0,0)	(0,1)	(0,2)	(1,1)	(1,2)	(2,2)	(0,0)	(0,1)	(0,2)	(0,3)	(1,1)	(1,2)	(1,3)	(2,2)	(2,3)	(3,3)
$h\_0 \geq 7$	Percent(%)	64.5	10.6	24.9	43.2	12.4	1.7	27.3	5.3	10.1	30.1	13.7	2.5	0.3	22.4	9.0	1.3	11.5	4.5	4.7
	CD	5.93	5.71	6.11	8.09	8.22	7.45	8.74	8.06	8.91	9.64	9.25	7.14	5.50	8.89	8.32	7.37	9.02	6.32	12.77
	CT	4.77	4.50	5.26	7.23	6.91	6.95	7.21	7.16	7.50	9.84	9.14	7.70	4.75	8.52	8.39	6.21	9.92	7.67	8.93
	C-index	3.95	4.12	4.32	5.76	5.82	5.60	5.83	5.96	6.06	7.30	7.39	7.73	4.25	6.91	6.70	6.00	7.04	5.29	8.57
$h\_0 \leq 1$	Percent(%)	19.1	23.8	57.1	7.3	9.8	6.0	24.9	18.1	33.9	3.2	4.8	4.6	1.4	12.8	12.3	5.3	24.4	12.3	18.9
	CD	7.88	5.81	6.79	9.99	8.56	7.16	9.42	7.37	8.43	11.78	10.51	9.33	9.23	11.15	9.43	8.26	10.45	8.49	9.90
	CT	5.96	3.81	5.23	8.26	6.17	4.86	7.48	5.54	7.05	10.06	8.16	6.92	6.19	9.80	7.68	6.32	9.07	7.01	8.93
	C-index	4.91	3.38	4.47	6.43	5.36	4.36	6.25	4.75	5.92	8.00	7.04	6.12	5.50	7.64	6.48	5.41	7.54	5.84	7.39

Meanwhile, we show the PDF of thrice reciprocity in Fig. 9. We can observe that at the beginning, all red lines, which indicate unfair reciprocity, are above blue lines, which denote the fair reciprocity and becomes lower in the later. This indicates that unfair reciprocity accounts for more proportions of short sustainability and less proportion of long sustainability.

Then, to account for the prestige effects, we investigate the early-stage h-index reciprocity for top-cited scholars. Finally, to further reveal the mechanism of reciprocity, the early-stage h-index reciprocity of most sustainable collaborations is investigated.

4.4. Reciprocity in top scientists

To account for prestige effects, we define two groups of scholars, facilitating a comparison of top scientists with beginner scholars. Specifically, the top scientists are scholars whose h-index is in the top 10% of all scholars Petersen (2015) whereas the h-index of beginner scholars is 1. After calculation, the h-index threshold of top scientists is 7, arriving 5766 collaboration pairs. Note that the threshold value 7 is small because the data is extracted before 2006 when scholars gained less citations (Yin & Wang, 2017). We are interested in the reciprocity between top scientists' collaborations which means that the h_0 of SS and SC is no less than 7. The reciprocity results are shown in Table 5.

We can observe that: (1) (0, 0) reciprocity accounts for the largest proportion in top scientists' collaboration while (1, 1) and (2, 2) reciprocity accounts for the largest proportions in beginner scholars' collaboration. The reason is that it is easier for beginner scholars to get an h-index increase. (2) There are fewer unfair reciprocities between top scientist. For example, the (0, 3) reciprocity is merely 0.3%. (3) The fair reciprocity i.e., (2, 2) in twice collaboration between top scientists will result in more sustainable collaborations (C-index 6.06) whereas the unfair reciprocity (0, 2) will result in less sustainable collaboration (C-index 5.60).

Table 6
Reciprocity results in persistent collaborations. The persistent collaboration are those collaborations whose c-index is no less than 8, compared with short collaborations whose C-index is less than 3.

Percent	First(%)			Twice(%)						Thrice(%)									
Reciprocity	(0,0)	(0,1)	(1,1)	(0,0)	(0,1)	(0,2)	(1,1)	(1,2)	(2,2)	(0,0)	(0,1)	(0,2)	(0,3)	(1,1)	(1,2)	(1,3)	(2,2)	(2,3)	(3,3)
C-idnex≥ 8	33.0	23.6	43.4	15.8	15.8	8.0	25.0	14.0	21.4	**9.0**	10.0	8.2	2.8	14.9	12.9	5.4	17.6	7.9	11.3
C-idnex≤ 3	36.2	38.3	25.5	14.7	19.0	18.2	15.7	18.9	13.5	9.6	8.6	12.8	13.1	8.9	8.5	11.6	10.1	11.3	5.5

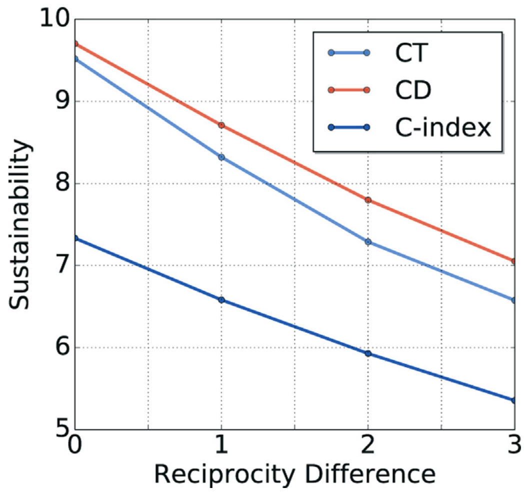

Fig. 10. Sustainability over reciprocity difference. The *x*-axis indicates the reciprocity difference which is the absolute value of two scholars' reciprocity difference, i.e., the difference of $(0, 3)$ is 3.

4.5. *Reciprocity in persistent collaborations*

We further explore the reciprocity in persistent collaboration. Here, a collaboration is persistent if its C-index is in the top 10%, arriving 186,885 collaborations with a threshold 8. Meanwhile, we compare the reciprocity of persistent collaborations with short collaborations whose C-index is less than 3. The results are shown in Table 6.

We can observe from the table that fair reciprocity accounts for largest proportions in persistent collaborations whereas unfair reciprocity accounts for the largest proportions in short collaboration. Take twice reciprocity for example. We can see that there is merely 8.0% unfair reciprocity $(0, 2)$ while the proportion of fair reciprocity $(1, 1)$ and $(2, 2)$ are larger than 20%. This indicates that the precondition of a persistent collaboration is that the collaboration is fair in reciprocity.

We have given a comprehensive analysis on how different reciprocity influences the collaboration sustainability. We find that fair reciprocity will bring about more sustainable collaboration while unfair reciprocity will greatly harm the collaboration reciprocity. To further show the influence of unfair reciprocity on collaboration sustainability, we analyze the relationships between reciprocity difference and collaboration sustainability. The results are shown in Fig. 10. We can see from the figure that with the increase of reciprocity difference, the CT, CD, and C-index will gradually decrease. Such phenomenon indicates that scholars are not altruism which means that if two scholars cannot gain equal benefit (h-index increase), their collaborations will not last for a long time.

5. Discussions

In this paper, we want to explore the mechanism of scientific collaboration from the perspective of sustainability. As an important social interaction between scholars, scientific collaborations have been studied by researchers from diverse disciplines. Some extensive investigated topics include the impact of scientific collaboration on academic reputation (Petersen et al., 2014), the structure of scientific collaboration networks (Newman, 2004), scientific collaboration pattern (Zhang et al., 2018), scientific collaboration recommendation (Cohen & Ebel, 2013), etc. However, in practical, scientific collaboration is not a one-shot deal where two scholars may collaborate with each other many times. Though previous studies (Bu et al., 2018; Petersen, 2015) have shown that sustainable collaborators may have a higher impact on scholars, the mechanism underlying sustainable collaboration is seldom investigated. In other words, no previous quantitative analysis has been performed to explore why the collaboration sustainability between scholars are different. The biggest contribution of this work is provide a new angle for investigating scientific collaboration patterns from the perspective of persistence in collaboration with a newly proposed metric h-index reciprocity.

In our opinion, the key to a sustainable collaborator is the early stage interaction between two collaborators. It is said that a good beginning is half the battle. Therefore, we propose to explore sustainable scientific collaboration based on the early-stage interactions between two scholars. To model the early-stage collaborations between scholars, we propose the h-index reciprocity, as shown in Fig. 1. h-index reciprocity is proposed to measure the impact of the collaborated paper on two collaborators. It can be regarded as an indicator of the relative significance of this collaboration to each collaborator. In other words, two scholars may benefit differently from the collaborated papers. By proposing and defining h-index reciprocity, to some degree, we are able to evaluate the success of the scientific collaboration. The biggest advantage of h-index reciprocity is that it is calculated directly from the dataset so that subjective bias can be avoided. Meanwhile, to quantify the sustainability of scientific collaboration, we define three indicators, including collaboration times, collaboration duration, and collaboration index. By statistical analysis of the h-index reciprocity with 172,680 collaboration records between 69,394 scholars, we find that unfair reciprocity many harm scientific collaboration sustainability. This finding is in line with humanity because people are rational based on the game theory. Asymmetric benefits will break the balance and can now last long. Meanwhile, to account for the prestige effects (Petersen, 2015), we analyze the h-index reciprocity between top scientists and the results indicate that the h-index reciprocity between top scientific is relatively fair. To better differ fairness in persistent and transient scientific collaboration, we explore the h-index reciprocity in persistent collaborations. The observed results further suggest the fact that fairness in h-index reciprocity is important for collaboration sustainability. These results are observed directly from real-world datasets by statistic analysis, which is, to some degree, not biased by man-made selection bias.

Our work may shed new light on understanding scientific collaboration from the perspective of sustainability. Meanwhile, our proposed h-index reciprocity may be used as a novel metric for evaluating scientific collaboration. This is because h-index reciprocity provides a quantitative way to measuring the expected collaboration benefits for two collaborators. The expected outcome will be (0, 0), (0, 1) or (1, 1). Thus, the mutual influence between two collaborators can be quantitatively measured. Moreover, our observation that fairness is vital for sustainable collaboration may help scholars better find his/her lifelong collaborators. In addition, our findings may inspire policymakers in promoting scientific collaboration from the perspective of collaboration sustainability. While our proposed h-index reciprocity metric can help explore the mechanism of collaboration sustainability, earnest and unbiased professional judgment for promoting scientific collaboration should be also considered. For example, scholars may have their own choice for seeking collaborators and their decisions may be affected by diver reasons, i.e., physical distance, topic similarity, collaboration preference, etc.

6. Conclusions

In this paper, we have studied how early-stage reciprocity influences collaboration sustainability. To this end, we define and quantify h-index reciprocity and collaboration sustainability. We present a descriptive analysis of the relationships between h-index reciprocity including fair and unfair reciprocity and collaboration sustainability including CT, CD, and C-index. Results suggest that fair reciprocity will bring about sustainable collaborations and unfair reciprocity has a significant negative impact on collaboration sustainability. Moreover, to account for prestige effects, we define two groups including top scholars and the beginner scholars and find that there are fewer unfair reciprocities between top scientist. Furthermore, we compare the h-index reciprocity between persistent collaborations and non-persistent collaborations and find that the precondition of a persistent collaboration is that the collaboration is fair in reciprocity. These indicate that the key to sustainability is fairness. Our work is relevant for scholars interested in human behaviors and scientific collaborations. Obtained insights may shed light on studying scientific collaboration from the perspective of reciprocity.

Author contributions

Wei Wang: Performed the analysis; Wrote the paper.
Jing Ren: Collected the data; Performed the analysis; Wrote the paper.
Mubarak Alrashoud: Conceived and designed the analysis; Contributed data or analysis tools; Wrote the paper.
Feng Xia: Conceived and designed the analysis; Performed the analysis; Wrote the paper.
Mengyi Mao: Collected the data; Contributed data or analysis tools.
Amr Tolba: Conceived and designed the analysis; Contributed data or analysis tools; Performed the analysis.

Acknowledgement

The authors extend their appreciation to the International Scientific Partnership Program ISPP at King Saud University for funding this research work through ISPP-78. This work is partially supported by China Postdoctoral Science Foundation (2019M651115).

References

Adams, J. (2013). Collaborations: The fourth age of research. *Nature*, *497*, 557–560.
Alpizar, F., Carlsson, F., & Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in costa rica. *Journal of Public Economics*, *92*, 1047–1060.
Amjad, T., Ding, Y., Xu, J., Zhang, C., Daud, A., Tang, J., et al. (2017). Standing on the shoulders of giants. *Journal of Informetrics*, *11*, 307–323.
Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*, 590–614.
Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142.

Bu, Y., Ding, Y., Liang, X., & Murray, D. S. (2018). Understanding persistent scientific collaboration. *Journal of the Association for Information Science and Technology*, *69*, 438–448.

Chung, E., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio-and nanoscientists' motivations, information-sharing and communication practices, and barriers to collaboration. *Journal of the Association for Information Science and Technology*, *67*, 1836–1848.

Coccia, M., & Wang, L. (2016). Evolution and convergence of the patterns of international scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 2057–2061. http://dx.doi.org/10.1073/pnas.1510820113, http://www.pnas.org/content/113/8/2057.full.pdf. http://www.pnas.org/content/113/8/2057.abstract

Cohen, S., & Ebel, L. (2013). Recommending collaborators using keywords. *Proceedings of the 22nd international conference on world wide web*, 959–962.

Cortés-Berrueco, L. E., Gershenson, C., & Stephens, C. R. (2016). Traffic games: Modeling freeway traffic with game theory. *PLOS ONE*, *11*, e0165381.

Dong, Y., Johnson, R. A., & Chawla, N. V. (2015). Will this paper increase your h-index? Scientific impact prediction. *Proceedings of the eighth ACM international conference on web search and data mining*, 149–158.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*, 293–315.

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*, 16569.

Ioannidis, J. P., Boyack, K. W., & Klavans, R. (2014). Estimates of the continuously publishing core in the scientific workforce. *PLOS ONE*, *9*, e101698.

Li, W., Aste, T., Caccioli, F., & Livan, G. (2019). Reciprocity and impact in academic careers. *EPJ Data Science*, *8*, 20.

Newman, M. E. (2001a). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*, 025102.

Newman, M. E. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 404–409.

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

Petersen, A. M. (2015). Quantifying the impact of weak, strong, and super ties in scientific careers. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, E4671–E4680.

Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., Rungi, A., et al. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 15316–15321.

Romano, A., & Balliet, D. (2017). Reciprocity outperforms conformity to promote cooperation. *Psychological Science*, 0956797617714828

Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2016). Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game. *International Journal of Social Robotics*, *8*, 303–317.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*, aaf5239.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*, 643–681.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). Arnetminer: Extraction and mining of academic social networks. *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*, 990–998.

Tsai, C. H., & Lin, Y. R. (2016). Tracing and predicting collaboration for junior scholars. *Proceedings of the 25th international conference companion on world wide web*, 375–380.

Wang, W., Cui, Z., Gao, T., Yu, S., Kong, X., & Xia, F. (2017). Is scientific collaboration sustainability predictable? *Proceedings of the 26th international conference on world wide web companion*, 853–854.

Wang, W., Yu, S., Bekele, T. M., Kong, X., & Xia, F. (2017). Scientific collaboration patterns vary with scholars' academic ages. *Scientometrics*, 1–15.

Wang, W., Liu, J., Yang, Z., Kong, X., & Xia, F. (2019). Sustainable collaborator recommendation based on conference closure. *IEEE Transactions on Computational Social Systems*, *6*, 311–322.

Wu, Y., Venkatramanan, S., & Chiu, D. M. (2015). Research collaboration and topic trends in computer science: An analysis based on ucp authors. *Proceedings of the 24th international conference on world wide web*, 1045–1050.

Xia, F., Wang, W., Bekele, T. M., & Liu, H. (2017). Big scholarly data: A survey. *IEEE Transactions on Big Data*, *3*, 18–35.

Yang, Z., & Chen, W. (2015). A game theoretic model for the formation of navigable small-world networks. *Proceedings of the 24th international conference on world wide web*, 1329–1339.

Yin, Y., & Wang, D. (2017). The time dimension of science: Connecting the past to the future. *Journal of Informetrics*, *11*, 608–621.

Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology*, *69*, 72–86.

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