OPEN ACCESS

Followness: Revealing Channels of Idea Flow in Research Networks*

연구 네트워크 내 아이디어 전파 경로 분석: 추종성(Followness) 개념

Eunhee Kim (김은희)**
Ji-Hong Park (박지홍)***

ABSTRACT

Finding idea-flow channels in a research network is an important issue because the channels influence the performance of a researcher. We developed a metric called *followness* that measures how closely in time one author gets ideas from another author. Followness is a sum of citation counts each weighted by citation delays. A network of authors constructed from the follow relations reveals fast idea-flow channels among authors. Using the Web of Science data for a prestigious journal in computer science, we show that the followness is more correlated to an author's productivity and the quality of an author's work than the number of coauthors or the citation count. Moreover, follow networks reveal idea-flow channels faster than those of coauthorship networks and citation networks. In addition, authors in a follow relation have a stronger relation than authors in a citation relation in terms of the degree, the connectivity, and the mutuality of the relation.

초 록

연구 네트워크에서 아이디어 흐름 경로를 찾는 것은 연구자 성과에 영향을 미치기 때문에 중요한 문제이다. 본 연구에서는 한 저자가 다른 저자로부터 아이디어를 시간적으로 얼마나 가까운 시점에 받아들이는지를 측정하는 방법으로 followness라는 지표를 개발했다. Followness는 인용 지연에 따른 가중치를 적용한 인용 수의 합이다. 이러한 '팔로우 관계'를 바탕으로 구성된 저자 네트워크는 저자 간의 빠른 아이디어 흐름 경로를 드러낸다. 컴퓨터 과학 분야의 권위 있는 학술지의 Web of Science 데이터를 활용하여 followness가 공동 저자 수나 인용 수보다 연구자의 생산성과 연구의 질과 더 높은 상관관계를 가짐을 보였다. 또한 follow 네트워크는 공동저자 네트워크나 인용 네트워크보다 아이디어 흐름 경로를 더 신속히 보여주며 '팔로우 관계'에 있는 저자들은 '인용 관계'에 있는 저자들보다 관계의 양, 연결성, 및 상호성 측면에서 더 강한 관계를 가짐을 보여준다.

Keywords: followness, follow network, research network, research idea flow, network analysis
팔로우성, 추종성, 추종 네트워크, 연구 네트워크, 연구아이디어 전파, 네트워크 분석

---

* 이 논문은 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5C2A03093597).
** 연세대학교 문헌정보학과 박사과정(eunhee.kim@yonsei.ac.kr) (제1저자)
*** 연세대학교 문헌정보학과 교수(jihongpark@yonsei.ac.kr) (교신저자)
▪ 논문접수일자 : 2025년 5월 19일 ▪ 최초심사일자 : 2025년 6월 6일 ▪ 게재확정일자 : 2025년 6월 11일
▪ 정보관리학회지, 42(2), 127-146, 2025. http://dx.doi.org/10.3743/KOSIM.2025.42.2.127

1. Introduction

Researchers are extending and deepening their research areas while incorporating and developing ideas from others. Exchanging ideas with others is an important factor that significantly affects the productivity of researchers and the quality of their work. For example, the performance of a researcher is affected by the colleagues with whom the researcher exchanges ideas (Azoulay et al., 2010; Borjas & Doran, 2015; Oettl, 2012). Those groups of researchers bound together by their research interests and ideas are called an *invisible college* (Crane, 1972; Wang & Barabási, 2021). Because idea-flows play such a central role in researchers' productivity and the quality of their work, identifying such a cohesive community has been a main research topic (Palla et al., 2005; Palla et al., 2007; Perianes-Rodríguez et al., 2010).

In a research network, ideas flow mainly by means of publishing papers and reading them. We develop a metric called *followness* between authors. We say that an author follows another author when the former keeps track of the latter's publications closely. This follow activity can be observed in the form of a citation. Specifically, we use the term *follow* to indicate the behavior of a researcher who keeps citing another researcher's papers soon after they are published. *Followness* is the weighted sum of citation counts whose weights are depreciated over the citation delay. The network built from the followness reveals fast idea flow channels and strong inter-author relations. Because the followness is a citation-based metric, there are inherent positive correlations between an author's followness and the author's publication counts as well as the citation counts.

Social network analysis reveals the opportunities and constraints an actor will encounter depending on the actor's position in the network and hence, finding the positions of an actor in a network is essential to predict the actor's performance (Borgatti et al., 2018). Centrality metrics for a social network can reveal the positional benefits of an actor regarding how ideas flow among actors (Borgatti, 2005). For example, the closeness centrality shows how fast one can hear ideas from the network, the betweenness centrality measures how often one is in the fastest idea-flow paths, the degree centrality indicates how many idea-flow channels one has, and so on. Varied by the underlying relations between actors, different research networks such as coauthorship networks, co-citation networks, citation networks, etc. (Egghe & Rousseau, 1990; Newman, 2001; 2004; Small, 1973; Price, 1965) can be formed. In this paper, we build a *follow network* based on the follow relation. Because the followness wanes with the citation delay, the metric makes the fast idea-flow channels prominent. As a result, the follow network reveals fast idea-flow channels among authors.

To illustrate the usefulness of the follow network, we compare the network with other networks, namely the coauthorship network and the citation network. From the perspective of coverage and the coherency of a relation, coauthoring a paper is direct evidence showing who exchanges ideas with whom through an explicit collaboration. However, coauthorship is not an easy relation to make in the sense that some physical, topical, operational, and more conditions among authors need to be satisfied. As a result, a coauthorship network reveals only the restricted channels of direct idea-flows. On the other hand, citing a paper is an easier activity: one may freely reference anyone's paper as long as the contents are relevant. As a result, a citation network shows diverse but possibly weak idea-flow channels between papers and hence between their authors. Moreover, its main driving force to flow ideas is the influence of papers rather than the merits of authors. Compared to these relations, following an author is not as restrictive as coauthoring a paper because one does not need to get the consent of the other, and at the same time it is not a weak relation because following someone is a targeted activity that requires nontrivial efforts. Hence, the network constructed from the follow relation reveals broad and consistent idea-flow channels between authors. Another favorable characteristic of the follow network is the association of an idea-flow speed with the network structure. Because the follow network discards slow idea-flow channels, the edges of the network are composed of fast channels. Hence, such metrics as the closeness centrality and the betweenness centrality that reveal how fast one can receive new ideas from the network or conversely disseminate her ideas to the network, and how many times one is in the fastest information flow paths respectively (Borgatti, 2005) can assess the authors' positional benefits in the network more accurately.

Using the *Web of Science* (WoS) data for *IEEE Transactions on Software Engineering* (TSE), a prestigious journal in Computer Science, we show that the followness is more correlated to the productivity of an author and the collective quality of an author's papers than the coauthor counts or the citation counts of the author. Moreover, a follow network reveals faster idea-flow channels compared to those based on the citation counts or the coauthorship counts.

2. Literature Review

A following behavior is often observed in Social Networking Services (SNS) and there are some analogies between an SNS and a research community: in an SNS, a user with many share-worthy postings has many followers and an article becomes a "hot" or popular topic if it is viewed by many users in a narrow-time window (Doerr et al., 2012; Kane, 2018); similarly, an author who published cite-worthy results would have many followers and such results would be cited promptly.

A citation network has been used as an analysis tool for the information flow in a research network (Garfield, 2004; Price, 1965). A citation network is a directed graph whose nodes are documents and whose edges represent the citation relationships between the documents. Ideally, due to the chronological ordering of citations, a citation network should not have cycles and form a Direct Acyclic Graph (DAG). Applying the ideas from the Strahler number, a numerical measure of the branching complexity of a tree that was first developed in hydrology to measure the complexity of rivers and streams (Strahler, 1957), a flow metric has been developed for the DAG of a citation network to measure the flow of knowledge (Herman et al., 1999). Recently, an efficient dynamic algorithm was developed to compute the flow of the influence of a document while ascending the flow, i.e., giving credits to the cited documents (Renoust et al., 2016).

A coauthorship network is a tool to analyze the collaboration relation among authors in a research network (Palla et al., 2007; Wang & Barabási, 2021). Mathematically, it is a graph whose nodes are authors and whose edges are the coauthorship relation between authors, i.e., whether two authors coauthored any documents. One of the utilities of a coauthorship network is to analyze the flow of topics: in the process of collaboration the expertise of an author is transferred to another and vice versa, and hence, topics of one author's networks can be propagated to the networks of the other author. A graph theoretical temporal topic model that identifies and tracks topics from their emergence has been employed by (Churchill et al., 2018) and a hierarchical topic model has been proposed by (Jiang & Zhang, 2016) to capture the topic evolution over time. Another research direction involving a coauthorship network is to utilize the network for the tasks of topic models, for example, to capture the temporal patterns of research interests of authors over time (Jeong et al., 2020), to enhance the performance of the Latent Dirichlet Allocation (LDA) topic model identifying hidden topics (Tran et al., 2012), and so on. To analyze research topic flows between authors, Schäfermeier et.al. (Schäfermeier et al., 2023) proposed a Topic Flow Network (TFN). A TFN is a coauthorship network whose edges are labeled by a year and a topic, discovered by the Non-negative Matrix Factorization topic modeling technique. The authors analyzed the intertopic and intratopic flows using the TFN.

3. Followness

The followness is a measure of how closely in time one author gets ideas from another author. In this section, we formally define the followness and other terms used in this paper.

Let $A=\{a_1, a_2, \cdots, a_m\}$ be the set of all authors and $P=\{p_1, p_2, \cdots, p_n\}$ be the set of all papers. We define the following functions to precisely specify the relation between them:

- $paper(a)$ is a function that takes an author a and returns the set of papers authored by a. We also define a function $PC(a)$ for the number of papers an author a authored, i.e. $PC(a)=|paper(a)|$. For example, if an author a wrote papers p and q and an author b wrote papers q and r, then, $paper(a)=\{p,q\}$, $paper(b)=\{q,r\}$, $PC(a)=|\{p,q\}|=2$ and $PC(b)=|\{q,r\}|=2$.
- $author(p)$ is a function that takes a paper p and returns the set of authors who authored p. For example, if authors a and b coauthored a paper p, then $author(p)=\{a,b\}$.
- $reference(p)$ is a function that takes a paper

p and returns the set of papers referenced by p. For example, if a paper p references papers q and r, then $reference(p) = \{q, r\}$.

- We represent the publication date of a paper as the number of months passed since January of year 0 to the publication date of the paper. $date(p)$ is a function that takes a paper p and returns the number of months from January of year 0 to the publication date of p. For example, if a paper p is published in March 2023, $date(p)$ is equal to $2023 \times 12 + 3 - 1$. In addition, if p references q, then $date(p) - date(q)$ computes the citation delay, the number of months from the publication date of q to the publication date of p. As an example, if p is published in March 2023 and q is published in May 2021, then the citation delay is: $date(p) - date(q) = 2023 \times 12 + 2 - 2021 \times 12 - 4 = 22$ months.

With the definitions above, let us define several related terms used in this paper: the *paper followness*, the *followness*, the *received followness*, and the *induced followness*. The *paper followness* between two papers is a metric, measuring how quickly a published paper is referenced by the other paper. Let a paper p reference a paper q, then the paper followness $PF_{\alpha}(p, q)$ between the paper is:

$$PF_{\alpha}(p, q) = \alpha^{date(p) - date(q)}, \tag{1}$$

where $date(p) - date(q)$ is the citation delay and α is a depreciation constant, having a value in $0 < \alpha \leq 1$. The paper followness, representing how tightly in time p references q, is a function monotonically decreasing with the citation delay between p and q. The depreciation constant decides how quickly the paper followness shrinks for each passing month.

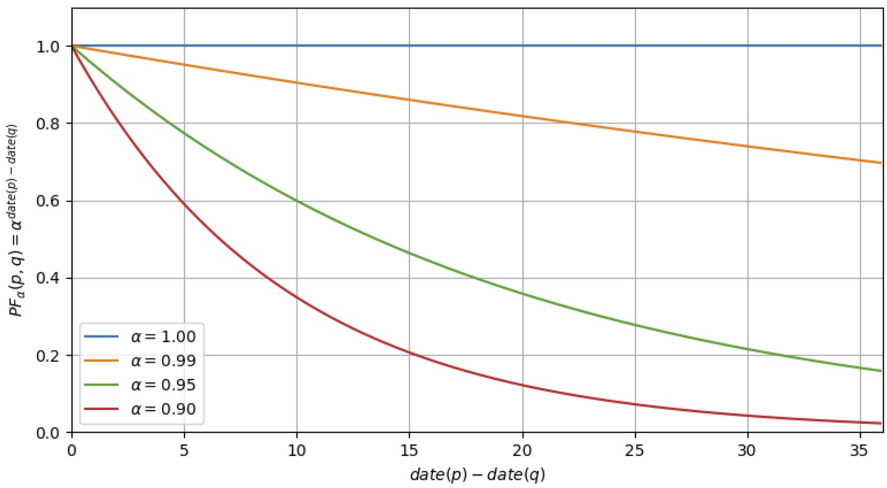


Fig. 1. The paper followness (PF_{α}) over citation delays for different depreciation constant values (α): the smaller the constant α is, the sharper the depreciation is over citation delay

Fig. 1 shows how $PF_{\alpha}(p,q)$ changes over citation delays for different depreciation constant values: when $\alpha=1$ there is no depreciation and when $\alpha=0.9$, 10% of the paper followness is depreciated for each passing month. We believe different values of α would be necessary for different fields and topics as the dynamics of each research area would be different. Finding an optimal value of α is beyond the scope of this paper. However, in general, smaller α would reveal faster idea flow channels and stronger relations between authors at the cost of making the network sparse. In the case study of this paper, we chose $\alpha=0.9$ as it is small enough to make the resulting network distinguishable from the direct citation network but large enough to make the network not sparse.

The *followness* between two authors is a metric that measures how quickly an author's papers cite the other author's papers. To measure the followness, we need to build the set of referencing-referenced pairs of papers between two authors. Particularly, let us call an author a a *follower* and an author b a *followed*, then $R(a,b)$ is a function that takes two authors a and b, and returns the set of referencing-referenced pairs of papers (p,q) such that p is authored by a and q is authored by b. In other words,

$$R(a,b)=\{(p,q)\mid p\in paper(a),\ q\in paper(b),\ q\in reference(p)\} \quad (2)$$

For example, suppose that an author a wrote papers p and q, i.e. $paper(a)=\{p,q\}$; an author b wrote papers p, q and r, i.e. $paper(b)=\{p,q,r\}$; and p references q and q references r, i.e. $reference(p)=\{q\}$ and $reference(q)=\{r\}$. Then, $R(a,b)=\{(p,q),(q,r)\}$.

The followness $F_{\alpha}(a,b)$ between a follower a and a followed b is:

$$F_{\alpha}(a,b)=\sum_{(p,q)\in R(a,b)} PF_{\alpha}(p,q) \quad (3)$$

Example 1 Fig. 2 shows relations about who authored which papers and which paper references which papers. In the diagram, author a wrote paper p and q, author b wrote papers q, r, and s, and author c wrote a paper t. That is, $paper(a)=\{p,q\}$, $paper(b)=\{q,r,s\}$ and $paper(c)=\{t\}$. In addition, paper p references q and r, i.e. $reference(p)=\{q,r\}$; paper q and r reference r and s respectively, i.e. $reference(q)=\{r\}$ and $reference(r)=\{s\}$; and paper t references q and s, i.e. $reference(t)=\{q,s\}$. The numbers in a red font of the form *year.month* in the diagram represent the publication date (year and month) of the papers. For example, paper p is published in April 2023. Hence, $date(p)=2023\times 12+4-1$ and $date(p)-date(q)=1$. With this information, let us consider how much a follows b when the depreciation constant $\alpha=0.9$.

First, let us find which of a's papers reference b's papers, i.e. compute $R(a,b)$. Because $author(a)=\{p,q\}$, $author(b)=\{q,r,s\}$, $reference(p)=\{q,r\}$, and $reference(q)=\{r\}$, $R(a,b)=\{(p,q),(p,r),(q,r)\}$.

As indicated in Fig. 2, suppose that p, q, r, s, and t were published in April, March, February, January, and April of 2023 respectively. Then how much a follows b can be computed as

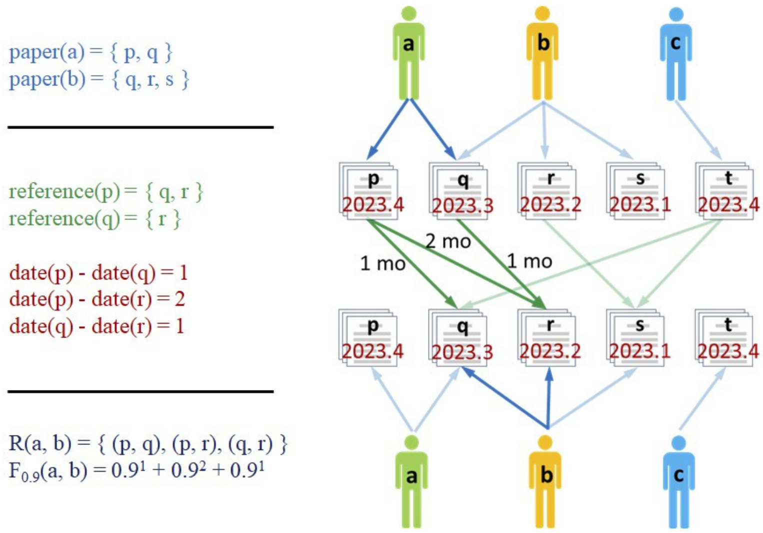


Fig. 2. An example of computing the followness

$$F_{0.9}(a,b) = PF_{0.9}(p,q) + PF_{0.9}(p,r) + PF_{0.9}(q,r) = 0.9^{1} + 0.9^{2} + 0.9^{1} \quad \square$$

The *received followness* $RF_{\alpha}(a)$ of an author a is the sum of all followness the author received from the author's followers and the *induced followness* $IF_{\alpha}(a)$ of an author a is the sum of all followness the author induced while following other authors. In other words:

$$RF_{\alpha}(a) = \sum_{b \in A} F_{\alpha}(b,a), \qquad IF_{\alpha}(a) = \sum_{b \in A} F_{\alpha}(a,b) \quad (4)$$

4. Case Study

4.1 Data Collection

We collected the *Web of Science* (WoS) records for *IEEE Transactions on Software Engineering* (TSE) journal, a prestigious journal in Computer Science, from September 1994 to August 2023. We examined earlier WoS records for the journal from 1977. However, in the earlier records, the author email field, used as an identifier in this paper, does not exist. From the records, we parsed 1950 article records and 5224 author records who authored those articles. For the stability of the analysis, we removed the authors who have less than three article records. After data cleaning and removing those author records, 630 author records remained. We performed our analysis based on the remaining 630 author records and their article records.

To compare the characteristics of the followness with other metrics, let us define the following functions. $citing\_docs(a,b)$ is a function that takes two authors a and b and returns the set of a's papers that reference some of b's papers. That is,

$$citing\_docs(a,b) = \{ p \in P \mid p \in paper(a), reference(p) \cap paper(b) \neq \emptyset \} \quad (5)$$

$CC_{pair}(a, b)$ is a function that returns the number of times a cites b's papers, i.e. $CC_{pair}(a, b) = |citing\_docs(a, b)|$ and $CC(a)$ is a function that returns the total number of citations a received, i.e. $CC(a) = \sum_{b \in A} CC_{pair}(b, a)$.

Similarly, $coauthored\_docs(a, b)$ is a function that returns the set of papers a and b coauthored and $coauthor(a)$ is a function that returns the set of co-authors of a. Specifically,

$$coauthored\_docs(a, b) = paper(a) \cap paper(b)$$
$$coauthor(a) = \{ b \in A \mid coauthored\_docs(a, b) \neq \emptyset \} \quad (6)$$

$AC_{pair}(a, b)$ is a function that returns the number of papers a and b coauthored, i.e. $AC_{pair}(a, b) = |coauthored\_doc(a, b)|$ and $AC(a)$ is a function that returns the number of coauthors of a, i.e. $AC(a) = |coauthor(a)|$.

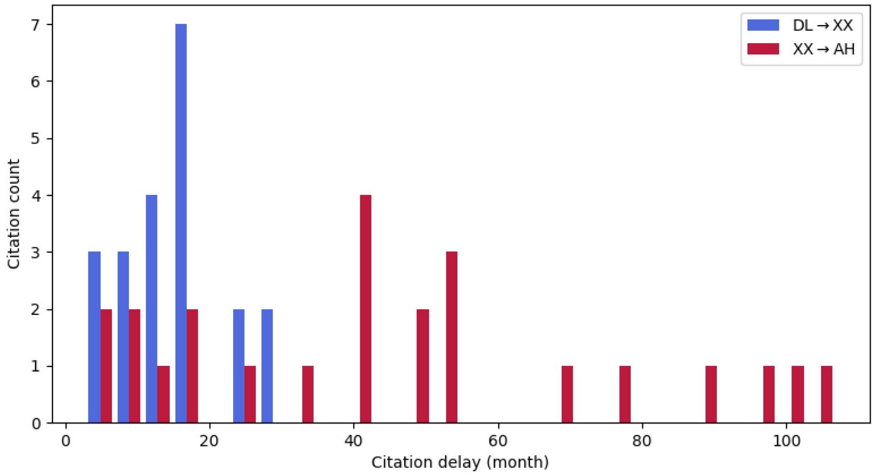
Table 1 shows the top 20 follower-followed pairs sorted by their followness $F_{0.9}$, i.e. the followness measured with the depreciation constant of $\alpha = 0.9$. In the rest of the paper we use the depreciation constant of $\alpha = 0.9$ unless it is specified. In the table, CC_{pair} and AC_{pair} columns show how many times the follower cited some of the followed's papers and how many papers they have coauthored respectively. In the table, the top follower-followed pair is AH and AH. It means that AH referenced many of AH's own papers in a short duration of time. The top non-self-following pair is DL and XX: DL follows XX with the followness of 5.15 by referencing 21 papers of XX. Conversely, XX followed DL with the followness of 3.06, by citing 14 papers of DL. They have coauthored 19 papers.

To bring insight into the followness metric, let us compare the two follower-followed pairs that have similar citation counts, but have somewhat different followness values. In Table 1, DL follows XX with the followness of 5.15 by citing 21 papers of XX and XX follows AH with the followness of 2.80 by citing 24 papers of AH. To see what does the followness measure, we compared their citation patterns over time. Fig. 3 shows the histograms of the citation delays. As the histogram shows, DL referenced XX's new papers soon after they are published,

Table 1. Top 20 follower-followed pairs ordered by their followness $F_{0.9}$. CC_{pair} columns show how many times follower cited followed's papers and AC_{pair} columns show how many papers they have coauthored

Follower	Followed	$F_{0.9}$	CC_{pair}	AC_{pair}	Follower	Followed	$F_{0.9}$	CC_{pair}	AC_{pair}
AH	AH	8.80	68	–	SW	AH	3.21	13	4
DL	DL	5.59	30	–	AH	TC	3.10	9	4
DL	XX	5.15	21	19	XX	DL	3.06	14	19
DL	AH	4.89	36	7	XX	XX	2.96	8	–
CT	CT	4.56	9	–	XX	AH	2.80	24	5
AH	KM	3.52	29	4	XL	XL	2.63	10	–
CC	CC	3.44	6	–	RH	RH	2.53	10	–
CT	JJ	3.27	7	2	JG	JG	2.44	5	–
CT	JG	3.27	7	2	CT	HD	2.43	8	1
TM	TM	3.23	11	–	AH	DL	2.40	11	7

Fig. 3. Histograms of the citation delays between DL and XX (blue) and between XX and AH (red). In the legend $a \rightarrow b$ means the papers of a that reference some of b's papers, i.e. $citing\_docs(a, b)$



whereas XX referenced AH's new and old papers regardless of their ages. Hence, we say that DL follows XX more closely than XX follows AH.

4.2 Comparison Between Metrics

In this section, we compare the three metrics of the followness (F_α), the citation count (CC), and the coauthorship count (AC) with regard to how well they represent the productivity of an author and the quality of an author's work. In general, the number of published papers of an author represents the author's productivity and the number of citations a paper received represents the quality of the paper (Wang & Barabási, 2021). In this paper, we regard the number of papers an author authored as a measure of the productivity of the author and the total number of citations an author received as a measure of the overall quality of the author's papers (Bornmann & Daniel, 2009).

Let us first compare the metrics with regard to the productivity of authors. Fig. 4 shows the scatter plots between the paper count $PC(a)$ of each author a and (a) the received followness $RF_{0.9}(a)$; (b) the induced followness $IF_{0.9}(a)$; (c) the citation count $CC(a)$; and (d) the coauthor count $AC(a)$.

The correlation coefficients between the metrics and the paper count can be found in the second column of Table 2. The table shows the overall correlation between paper count PC and received followness $RF_{0.9}$, induced followness $IF_{0.9}$, citation count CC, and coauthor count AC. As is often acknowledged (Dong et al., 2017; Wang & Barabási, 2021), coauthor count has a strong correlation with an author's productivity. However, as the table shows, the received followness has a marginally but larger correlation with an author's productivity than the coauthor count. Interestingly, the induced followness

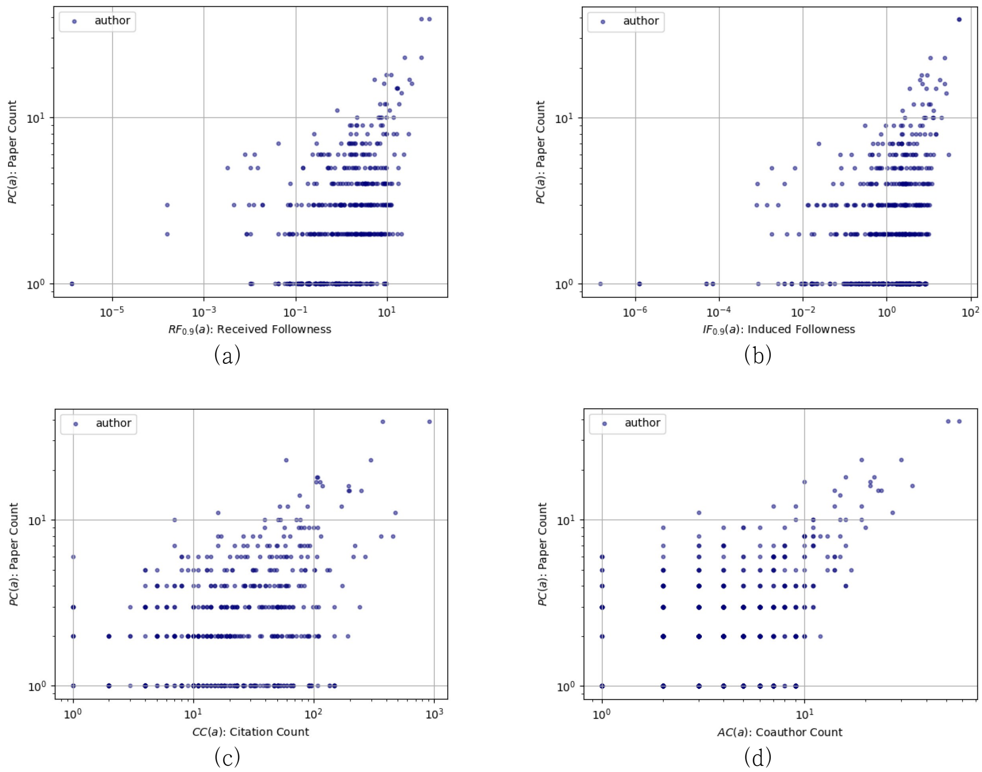


Fig. 4. Scatter plots between the Paper Count $PC(a)$ of each author a and (a) the Received Followness $RF_{0.9}(a)$ (corr: 0.739); (b) the Induced Followness $IF_{0.9}(a)$ (corr: 0.671); (c) the Citation Count $CC(a)$ (corr: 0.625); (d) the Coauthor Count $AC(a)$ (corr: 0.736)

Table 2. Correlations between evaluation metrics and PC and between the metrics and CC

Metric/Range	Correlation		Sample size
	PC: Paper Count	CC: Citation Count	
$RF_{0.9}$: Received Followness (overall)	0.739	0.695	630
84.56 – 2.13	0.757	0.724	210
2.13 – 0.20	0.274	0.188	210
0.20 – 0.00	0.022	0.583	210
$IF_{0.9}$: Induced Followness (overall)	0.671	0.469	630
53.97 – 2.76	0.734	0.618	210
2.74 – 0.61	0.126	-0.027	210
0.60 – 0.00	0.260	0.035	210
CC: Citation Count (overall)	0.625	–	630
971 – 32	0.582	–	210
31 – 8	0.185	–	210
8 – 0	0.478	–	210
AC: Coauthor Count (overall)	0.736	0.601	630
58 – 6	0.843	0.720	210
6 – 4	-0.090	0.079	210
4 – 0	-0.080	0.060	210

takes the third place, not far behind from the first two. It indicates that taking ideas by following others can improve the productivity of an author. In addition, the citation count does not have as strong correlation with the productivity of an author as the other metrics have.

In order to see from where the differences in the metrics come, we further divide the authors into three groups based on their metric values and compute the correlations for the authors in each group. Examining the divided groups, the coauthor count has a strong correlation with an author's productivity when the author has many coauthors. Compared to the coauthor count, the received followness show a larger correlation in the second group as well. The difference might be caused by the fact that the coauthorship includes only the explicit idea-flows, but the follow relation includes implicit idea-flows as well. Compared to other metrics, the citation count has a larger correlation in the third group although it does not have as large correlation as the other metrics in the first group.

Next, we compared the metrics with regard to an overall quality of the authors' papers. Fig. 5 shows the scatter plots between the citation count $CC(a)$ of each author a and (a) the received followness $RF_{0.9}(a)$; (b) the induced followness $IF_{0.9}(a)$; and (c) the coauthor count $AC(a)$. Plotting the relation between $CC(a)$ and $CC(a)$ is skipped as the dots are simply on the diagonal line with the correlation coefficient of 1. The correlation between the citation count CC of an author and the other metrics for the author are summarized in the third column of Table 2. Comparing the overall correlation, the received followness has the strongest correlation with the collective quality of an author's papers, followed by the number of coauthors. When divided into three groups by the value of each metric, the coauthor count shows the relevance only for the first group, but the received followness maintains larger correlations in all three groups. In general, the received followness shows a more stable relevance for the productivity of an author and the overall quality of an author's work than the coauthor count.

Because the received followness and the coauthor count have similar correlation coefficients with the paper count and the citation count, we further examined the relation between $RF_{0.9}(a)$ and $AC(a)$. Fig. 5 (d) shows a scatter plot between $RF_{0.9}(a)$ and $AC(a)$. The correlation between them are 0.729.

4.3 Network Analysis

An author's position in a network affects the performances of the author because some opportunities and constraints are associated with the position of the author in the network (Borgatti et al., 2018). In this section, we construct three networks of authors based on their follow relations, citation relations, and coauthorship relations, and compare them in the view of idea-flow speeds as well as the strength of the relation in terms of the degree of relations, the connectivity, and the mutuality.

A follow network $Fn_{\alpha} = \langle A, E \rangle$ is a labeled directed graph whose nodes are the authors and whose edges represent the follow relation. We write $a \xrightarrow{f} b$

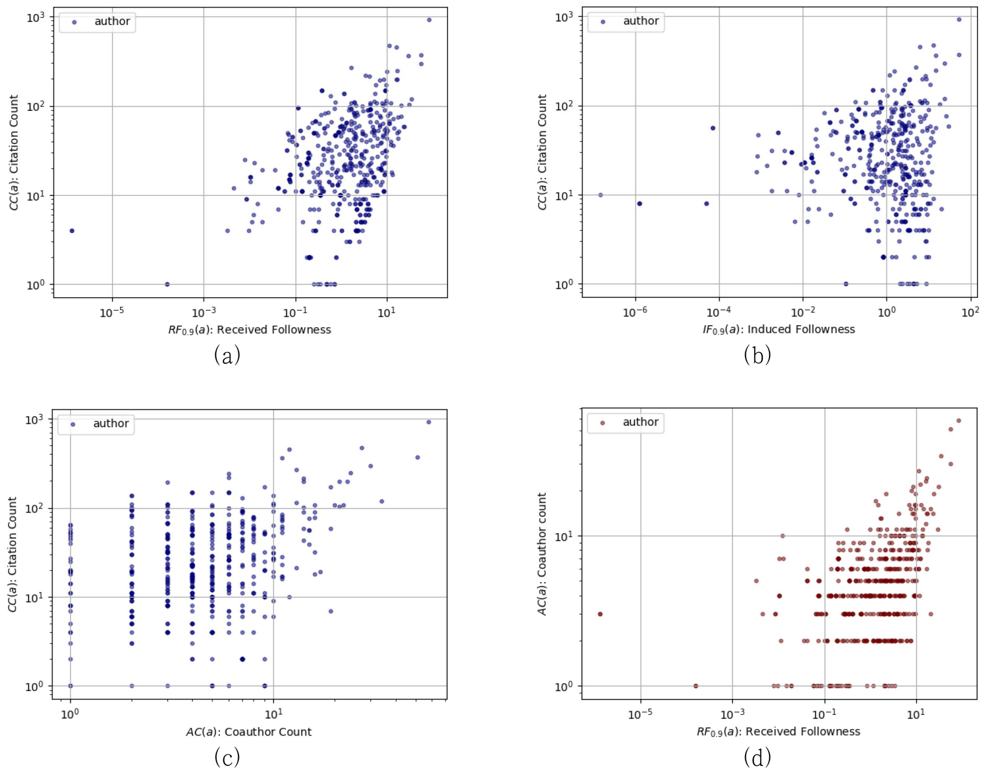


Fig. 5. Scatter plots between the Citation Count $CC(a)$ of each author a and (a) the Received Followness $RF_{0.9}(a)$ (corr: 0.695); (b) the Induced Followness $IF_{0.9}(a)$ (corr: 0.469); and (c) the Coauthor Count $AC(a)$ (corr: 0.601). (d) Scatter plot between $RF_{0.9}(a)$ and $AC(a)$ (corr: 0.729)

for the edge representing the relation that a follows b with the followness $f = F_{a}(a, b)$. Fig. 6 (a) shows a sub-graph of a follow network. Because the whole network has many nodes and densely wired edges connecting them, it is not practical to present the entirety in this paper. Moreover, by reducing the graph to include only the dominant elements of the network, we can identify the key structure more easily. Specifically, we reduce the network to contain only the edges with the largest followness such that 100 authors are included in the network. Such a sub-graph can be built as follows:

- Let $Fn_{a}^{(n)} = \langle A', E' \rangle$ be a subgraph of $Fn_{a} = \langle A, E \rangle$, i.e. $A' \subseteq A$ and $E' \subseteq E$, such that $|A'| = n$ and the followness of edges in E' are larger than or equal to those in the rest, i.e. for all $a' \xrightarrow{f} b' \in E'$ and for all $a \xrightarrow{f} b \in E - E'$, $f \geq f$ holds. For simplicity, we also call $Fn_{a}^{(n+1)}$ as $Fn_{a}^{(n)}$ when $Fn_{a}^{(n)}$ does not exist.
- Starting from $Fn_{a}^{(0)} = \langle \varnothing, \varnothing \rangle$, we construct $Fn_{a}^{(n)}$ from $Fn_{a}^{(n-1)}$ by adding the edge with

the largest followness in the remaining edges until one or two new authors are added to $Fn_{\alpha}^{(n)}$ - two authors can be added together when an edge added has two new authors not in $Fn_{\alpha}^{(n-1)}$.

- Stop the incremental construction when $Fn_{\alpha}^{(100)}$ is built.

The network in Fig. 6 (a) shows $Fn_{0.9}^{(100)}$ constructed this way. To visualize the differences between the networks explicitly, the nodes in all three networks of Fig. 6 are drawn such that the size of a node is proportional to the node's eigenvector centrality and its color indicates the community the node belongs to. The eigenvector centrality of a node reflects the influence of the node in the network (Freeman, 1978). The nodes in a network are grouped by the Louvain algorithm (Blondel et al., 2008) which partitions the network into communities based on the ratio of each node's inbound edges and outbound edges.

As we construct a follow network, we can build a *citation network* as $Cn = \langle A, E \rangle$, where A is the set of authors and E represents the citation relation: an edge $a \xrightarrow{c} b$ means that a cited b's papers $c = CC_{pair}(a, b)$ times. To compare the strength of inter-author relations, we construct a coauthorship network. A *coauthorship network* $An = \langle A, E \rangle$ is an undirected graph, where A is the set of authors and E represents the coauthorship relation: $a \xrightarrow{c} b$ means that a and b coauthored $c = AC_{pair}(a, b)$ papers. Because An is an undirected graph, if $a \xrightarrow{c} b$ is in E, then $b \xrightarrow{c} a$ is in E as well by the symmetry. We construct $Cn^{(n)}$ and $An^{(n)}$ by the same procedure as we build $Fn_{\alpha}^{(n)}$. The networks of Fig. 6 (b) and Fig. 6 (c) are $Cn^{(100)}$ and $An^{(100)}$ respectively. In the rest of the section, we will explain the differences between the three networks.

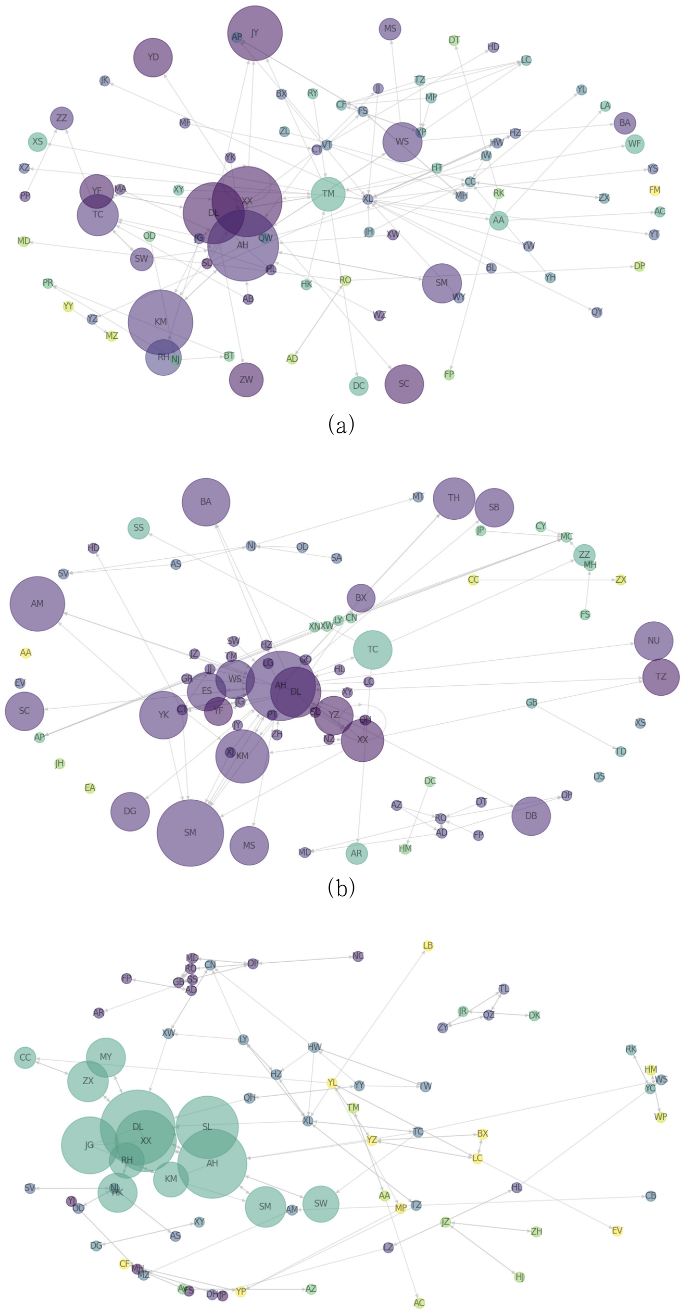
Let us first compare how fast ideas are flowing through the edges identified by the three networks. To compare the *idea-flow speed* of each network, we collect the *citation delays* among the paper written by authors incident with the edges in the network. In other words, given a network $N = \langle A, E \rangle$, we construct the multi-set of citation delays D_N as the union of citation delays between the papers of all authors. Specifically, D_N is defined as below:

$$D_N = \bigcup_{a \xrightarrow{c} b \in E} \{ date(p) - date(q) \mid (p, q) \in R(a, b) \}$$

Observe that because D_N uses only the publication dates of papers, it is not specific to the follow network but applicable to all networks. As an illustration, let us consider the example of Fig. 2. Suppose that a simple network $N = \langle A, E \rangle$ is composed of two authors $A = \{a, b\}$ and a singleton edge $E = \{a \xrightarrow{c} b\}$, where the weight c may differ by the kind of the network. Then, the set of citation delays D_N is

$$D_N = \{ date(p) - date(q),\ date(p) - date(r),\ date(q) - date(r) \} = \{1, 2, 1\}$$

The overall idea-flow speed between authors in a network is inversely proportional to the mean of the collected citation delays, i.e. the idea-flow speed is proportional to $\frac{1}{mean(D_N)}$.



(a) (b) (c)

Fig. 6. (a) a follow network, (b) a citation network, and (c) a coauthorship network. The node colors represent the communities grouped by the Louvain algorithm and the node sizes represent the influence of the node measured by the eigenvector centrality

Table 3. Citation delays of different networks

Network	Citation delay (D_N)	
	Mean (month)	Std. dev.
Follow network	24.11	24.85
Citation network	59.10	47.76
Coauthorship network	30.46	26.10

Table 3 shows the mean and the standard deviation of the citation delays of the three networks. Based on the table, ideas flow fastest in the follow network: on average, a paper is cited 24.11 months later after it is published. The coauthorship network is the next: it takes 30.46 months on average before a paper is cited since it is published. The citation network is the slowest of the three: the citation delay is 59.10 months on average. The reason the follow network is composed of fast idea-flow channels is because the underlying follow relation accounts for the citation delays negatively, i.e. each citation relation is depreciated by its citation delay.

The different idea-flow speeds in the networks have implications in the centralities of the networks. For example, the closeness centrality of an author shows how close the author is from the center of the network, commonly measured by the edge counts in the shortest path between nodes, and it indicates how fast an author can disseminate her ideas (Freeman, 1978). Because ideas flow fast through the edges comprising a follow network, the authors with a high closeness centrality in this network are more likely to disseminate her ideas faster in physical time than the authors with a high closeness centrality in other networks. Similarly, the betweenness centrality that measures how many times an author is in the shortest idea-flow paths between two authors is also affected by the citation delay. The smaller the citation delay of a network is, the more the betweenness centrality accounts for the fastest idea-flow paths in time, not just the smallest hop-count paths.

Let us examine the networks from the perspective of the ease of making the underlying relations, an indirect measure of the strength of the inter-author relations, with regard to the degree of relations, the connectivity, and the mutuality. We assume that following an author requires more time and effort than citing a paper of an author and authoring a paper together with other authors is the most difficult relation of the three as it requires a fluke as well as committed time and effort. As an illustration, let us simply assume that the number of relations an author has is uniformly distributed with a mean μ, then the probability that a randomly chosen author has k relations follows the Poisson distribution (Grimmett & Stirzaker, 2001).

Fig. 7 shows how the *Probability Mass Function* (PMF) changes with μ: as the relation becomes difficult to make (small μ), the PMF graph peaks at a smaller k and becomes narrower, i.e. all authors will have similar number of relations, but as the relation becomes easy to make (large μ), the PMF curve peaks at a larger k and becomes wider, i.e., some authors may have much larger number of relations than some other authors.

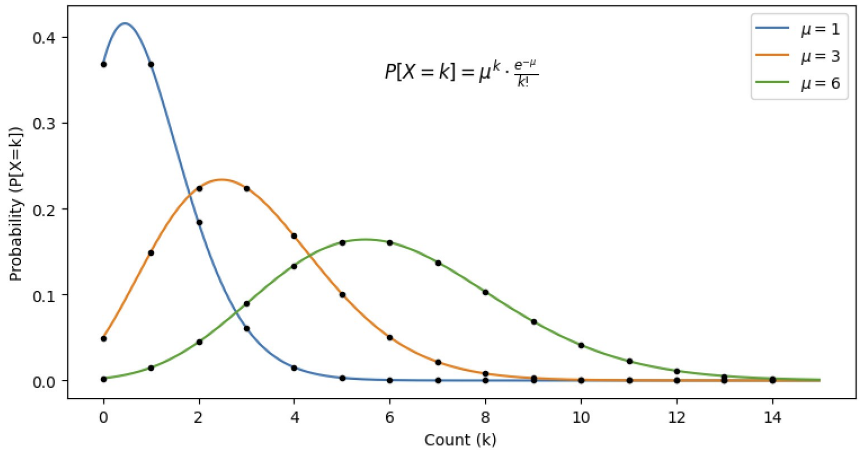


Fig. 7. Probability Mass Function of Poisson distributions with different mean

Table 4 summarizes the means and the standard deviations of the number of relations (edges) each author has in the three networks of Fig. 6.

In accordance with the assumed difficulties of the three relations, the coauthorship network has the smallest mean and the smallest standard deviation, followed by the follow network and the citation network has the largest mean and the standard deviation values.

Now, let us examine how the difficulty of making the underlying relations affects the connectivity of the network.

The Count column of Table 5 shows the number of *Strongly Connected Components* (SCC) (Cormen et al., 2009) in the follow network and in the citation network and the number of *Connected Components* (CC) in the coauthorship network of Fig. 6. Because the coauthorship is the most difficult relation of the three, each author has a small number of relations with others and hence the network is partitioned into many small connected components. By the same reasoning, the follow network takes the next position in the number of SCCs and the citation network has the smallest number of SCCs. The maximum number of authors in the (S)CCs and the mode of those would be in the reverse order of the (S)CC counts as the table shows. Fig. 8 shows a histogram of the size of (S)CCs: the harder a relation is to make, the narrower its distribution is and the smaller value its peak is at.

Table 4. In-degree and out-degree of different networks

Network	In-degree		Out-degree	
	Mean	Std. dev.	Mean	Std. dev.
Follow network	2.09	2.28	2.09	1.93
Citation network	2.76	4.57	2.14	3.02
Coauthorship network	1.86	1.47	1.86	1.47

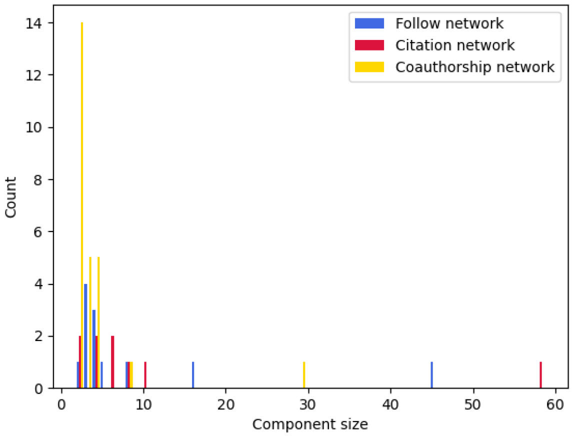


Fig. 8. Histograms of the size of (S)CCs

Table 5. Statistics of (S)CCs

Network	(S)CC Count	(S)CC Max size	(S)CC Mode size
Follow net.	12	45	3
Citation net.	9	58	4
Coauthorship net.	26	29	2

Table 6. The compositions of relation types of different networks

	Follow net.	Citation net.	Coauthorship net.
One-way rel. count	100 (74.63%)	127 (90.07%)	0 (0%)
Mutual rel. count	34 (25.37%)	14 (9.93%)	186 (100%)
Total rel. count	134	141	186

Finally, let us consider how the difficulty of a relation affects the mutuality of the relation. Table 6 shows the number of one-way relations and the number of mutual relations in the three networks of Fig. 6.

In an SNS, we may follow an influential person easily even though the relation is not mutual. However, many SNS users expect reciprocation by default (Consalvo et al., 2006). Similarly, in a research network, we postulate that more relations will be mutual if the relations are hard to make. Citing a paper is the easiest relation to make and hence the relations between authors in a citation network will have more portions of one-way relations than other networks. As indicated in Table 6, Fig. 6 (b) is composed of 90.07% of one-way relations and 9.93% of mutual relations. Because the follow relation is more difficult to make than the citation relation, Fig. 6 (a) will have more mutual relations than Fig. 6 (b). As confirmed in Table 6, the former is composed of 74.63% of one-way relations and 25.37% of mutual relations. The coauthorship relation, the most difficult relation of the three, is by definition a mutual relation.

5. Conclusion

In this paper we designed a new metric called followness to measure how closely an author gets ideas from another author. The followness has a strong positive correlation with an author's productivity, measured by the publication counts, and the overall quality of the papers the author published, measured by the citation counts.

A follow network, a graph of authors constructed based on the follow relation between authors, shows idea-flow channels among authors. Compared to a coauthorship network and a citation network, a follow network reveals fast idea-flow channels because its underlying follow relation accounts for an idea transfer delay. In addition, while the edges in a coauthorship network show only explicit idea exchange channels, the edges in a follow network include implicit idea-flow channels as well.

Relations between authors are evolving dynamically over time (Palla et al., 2007). As a future research direction, we are studying how the followness metric can be utilize to uncover the dynamics of collaboration patterns among authors, and to predict the changes of the productivity of an author and the quality of an author's work over time.

With the rapid advance of technologies, researchers obtain ideas from others faster than ever and identifying fast idea-flow channels is essential to predict the collaboration dynamics among authors and their performances. We believe the followness and the follow network will be a useful tool to understand the evolution of research networks.

References

Azoulay, P., Graff Zivin, J. S., & Wang, J. (2010). Superstar extinction. The Quarterly Journal of Economics, 125(2), 549-589. https://doi.org/10.1162/qjec.2010.125.2.549

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Borgatti, S. P. (2005). Centrality and network flow. Social Network, 27, 55-71. https://doi.org/10.1016/j.socnet.2004.11.008

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). Analyzing Social Networks (2nd ed.). SAGE Publications.

Borjas, G. J. & Doran, K. B. (2015). Which peer matter? The relative impacts of collaborators, colleagues, and competitors. The Review of Economics and Statistics, 97(5), 1104-1117. https://doi.org/10.1162/REST_a_00472

Bornmann, L. & Daniel, H. D. (2009). The state of h index research: Is the h index the ideal way to measure research performance? EMBO Reports, 10(1), 2-6. https://doi.org/10.1038/embor.2008.233

Churchill, R., Singh, L., & Kirov, C. (2018). A temporal topic model for noisy mediums. In Lecture Notes in Computer Science, 42-53. https://doi.org/10.1007/978-3-319-93037-4_4

Consalvo, M., Haythornthwaite, C., & Aoir (2006). Fono, David, and raynes-goldie, kate. Hyperfriends and beyond: Friendship and social norms on live journal. Available: https://api.semanticscholar.org/CorpusID:149201466

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to Algorithms (3rd ed.). Massachusetts: MIT Press.

Crane, D. (1972). Invisible Colleges: Diffusion of Knowledge in Scientific communities. Illinois: University of Chicago Press.

Doerr, C., Tang, S., Blenn, N., & Mieghem, P. V. (2012). Are friends overrated? A study for the social aggregator dig.com. Computer Communications, 35(7), 796-809. https://doi.org/10.1016/j.comcom.2012.02.001

Dong, Y., Ma, H., Shen, Z., & Wang, K. (2017). A century of science: Globalization of scientific collaborations, citations, and innovation. In: Proccedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, KDD ’17, 1437-1446. https://doi.org/10.1145/3097983.3098016

Egghe, L. & Rousseau, R. (1990). Introduction to Informetrics. Elsevier Science Publishers.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215-239. https://doi.org/10.1016/0378-8733(78)90021-7

Garfield, E. (2004). Historiographic Mapping of Knowledge Domains Literature. Journal of Information Science, 30(2), 119-145. https://doi.org/10.1177/0165551504042802

Grimmett, G. R. & Stirzaker, D. R. (2001). Probability and Random Processes (3rd ed.). Oxford: Oxford University Press.

Herman, I., Marshall, M. S., Melançon, G., Duke, D. J., Delest, M., & Domenger, J. (1999). Skeletal images as visual cues in graph visualization. In Eurographics, 13-22. https://doi.org/10.1007/978-3-7091-6803-5_2

Jeong, Y., Lee, S., Gweon, G., & Choi, H. (2020). Discovery of topic flows of authors. The Journal of Supercomputing, 76(10), 7858-7882. https://doi.org/10.1007/s11227-017-2065-z

Jiang, X. & Zhang, J. (2016). A text visualization method for cross-domain research topic mining. Journal of Visualization, 19(3), 561-576. https://doi.org/10.1007/s12650-015-0323-9

Kane, B. (2018). One Million Followers: How I Built a Massive Social Following in 30 Days. BenBella Books.

Newman, M. E. J. (2001) The structure of scientific collaboration networks. Proceeding of the National Academy of Sciences, 98(2), 404-409. https://doi.org/10.1073/pnas.98.2.404

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences, 101(suppl_1), 5200-5205. https://doi.org/10.1073/pnas.0307545100

Oettl, A. (2012). Honour the helpful. Nature, 489, 496-497. https://doi.org/10.1038/489496a

Palla, G., Barabási, A. L., & Vicsek, T. (2007). Quantifying social group evolution. Nature, 446, 664-667. https://doi.org/10.1038/nature05670

Palla, G., Derenyi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435, 814-818. https://doi.org/10.1038/nature03607

Perianes-Rodríguez, A., Olmeda-Gomez, C., & Moya-Anegon, F. (2010). Detecting, identifying and visualizing research groups in co-authorship networks. Scientometrics, 82, 307-319. https://doi.org/10.1007/s11192-009-0040-z

Price, D. J. D. S. (1965). Networks of scientific papers. Science, 149, 510-515. https://doi.org/10.1126/science.149.3683.510

Renoust, B., Claver, V., & Baffier, J. (2016). Flows of knowledge in citation networks. In Studies in Computational Intelligence, 159-170. https://doi.org/10.1007/978-3-319-50901-3_13

Schäfermeier, B., Hirth, J., & Hanika, T. (2023). Research topic flows in co-authorship networks. Scientometrics, 128(9), 5051-5078. https://doi.org/10.1007/s11192-022-04529-w

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science, 24(4), 265-269. https://doi.org/10.1002/asi.4630240406

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. Transactions American Geophysical Union, 38(6), 913-920. https://doi.org/10.1029/TR038i006p00913

Tran, D., Takeda, H., Kurakawa, K., & Tran, M. (2012). Combining topic model and co-author network for KAKEN and DBLP linking. In Lecture Notes in Computer Science, 396-404. https://doi.org/10.1007/978-3-642-28493-9_42

Wang, D. & Barabási, A. L. (2021). The Science of Science. Cambridge: Cambridge University Press.