IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 28, 2019, accepted July 24, 2019, date of publication August 23, 2019, date of current version September 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937220

A New Metric for the Analysis of the Scientific Article Citation Network

LIVIA LIN-HSUAN CHANG[1], **FREDERICK KIN HING PHOA**[2], **AND JUNJI NAKANO**[3]
[1]Department of Statistical Science, SOKENDAI (The Graduate University for Advanced Studies), Tokyo 1908562, Japan
[2]Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan
[3]Department of Global Management, Chuo University, Tokyo 1920393, Japan

Corresponding author: Livia Lin-Hsuan Chang (livia@ism.ac.jp)

This work was supported in part by the Career Development Award of Academia Sinica, Taiwan, under Grant 103-CDA-M04, in part by the Ministry of Science and Technology, Taiwan, under Grant 107-2118-M-001-011-MY3 and Grant 107-2321-B-001-038, in part by the ISM Cooperative Research Program under Grant 2016-ISMCRP-4408, Grant 2017-ISMCRP-4206, and Grant 2017-ISMCRP-4210, and in part by SOKENDAI Publication Grant for Research Papers.

ABSTRACT Citation plays an important role in the bibliometrics analysis since the introduction of the impact factors, but traditional measures mainly focused on the direct citations between articles. In this work, we introduce a new metric, namely Article Network Influence (ANI), to measure the influence of an article by using broader citation relationships quantitatively. We prepare our article citation networks from one of the largest citation databases called the Web of Science, and we demonstrate the use of ANI on the analysis of these networks in the statistics research community. These analyses appear in the top-20 influential articles in statistics within every 11 years during 1981-2016. We consider differences between the new metric and several traditional measures, including the impact factor, PageRank, and Field-Weighted Citation Impact (FWCI).

INDEX TERMS Article Network Influence, bibliometrics, citation analysis, Web of Science.

I. INTRODUCTION

Bibliometrics [1] is an analysis of research performance that governments and research institutes advocate and emphasize nowadays. Among all bibliometrics, citation plays an important role in the performance analysis. In the regime of big data, citation information is gathered from large-scale databases and represented in network forms. A citation network analysis is a quantitative method to identify important and impacted literature of a field based on how often a publication is cited in other publications. This analysis has recently become an essential tool to evaluate scientific achievements in different entities, including but not limited to, research articles, individual researchers, scientific journals, international conferences, universities and research institutes, or even countries. Governments and funding agencies make decisions on the allocation of their resources to these entities according to their research performances. For example, an education department can allocate its educational funds to schools based on the quantitative reports of their performances.

Impact Factor [2] was perhaps the pioneer to introduce citation indexes, and it is one of the most well-known indicators for evaluating the quality of articles in scientific journals. Since then, many metrics were introduced to measure research performance and quality from the different perspectives of citation analysis, such as Eigenfactor [3], h-index [4], Immediacy Index [5], Raw Impact per Publication (RIP) [6], SCImago Journal Rank (SJR) [6], Source Normalized Impact per Paper (SNIP) [7], Field-Weighted Citation Impact (FWCI) [6], PageRank [8], and many others. Recently, [9] measured the performance of statistics journals via a sophisticated Stigler model-based approach.

Most conventional methods listed above only consider "direct citations" and ignore "indirect citations". The formal definitions of the citation directness will be given in the next section, and we use an example to describe the scenario here. Assume we have a small citation network with three articles and two citations, where Article A was cited by Article B and Article B was cited by Article C. If we measure the influence of paper A in this network, conventional methods only take direct citations into consideration and conclude that Article A has influence only on Article B. However, likely, Article A inspired Article C through Article B. This scenario might happen when one of the above two citations (or both) was interdisciplinary, or Article B was a review-type article. For the past hundred years, scientific advancement was built on the continuous propagation of one's research

The associate editor coordinating the review of this article and approving it for publication was Donghyun Kim.

results to others, and conventional methods are limited to quantify this phenomenon.

This work aims at analyzing the citation network from the network perspective. In fact, there is much research investigates the influence from the network perspective [10]. Influence is a capacity or a power of things to be a compelling force or to produce effects on the actions or opinions of others. In the citation network, when a researcher published a scientific article, others read it and followed his step. When these followers completed their own research that was influenced or inspired by the original article, they cited it in the reference when they wrote their scientific articles. It led to a citation, which is officially defined as an abbreviated alphanumeric expression embedded in the body of intellectual work. It denotes an entry in the bibliographic reference section to acknowledge the relevance of other literature in the discussion at the spot where the citation appears. After years of progress, a citation network has resulted. We introduce a new influence-based metric for the importance of an article, and it is in terms of the influence of an article towards all other articles in its associated field or the whole citation network.

The structure of this paper is as follows. Section 2 provides the definition and theoretical background of our proposed metric. Section 3 demonstrates the use of our proposed metric to the analysis of citation networks in the statistics research community and interprets the analytical results. Section 4 compares several existing metrics with our proposed metric. A conclusion is given in the last section.

II. ARTICLE NETWORK INFLUENCE (ANI)

Let $G = (V, E)$ be a citation network, where $V = \{v_1, \ldots, v_m\}$ is a set of m articles (nodes), and E is a set of citations (edges). In specific, we denote $v_t \rightarrow v_s$ in G when Paper v_t cites Paper v_s. Here are some special properties of the citation network:

1. All edges in G are directed at only one direction. Since an edge between two nodes in G represents a citation relationship between two articles, it is obvious that the direction of the edge indicates which article cites one another. We do not expect to see a two-way edge in G, because it is unlikely to have a situation of "*A* cites *B* while *B* cites *A*".
2. There is no self-connection on any nodes in G. The self-connection of a node in G represents the self-citation of an article which is unlawful in nature. We ignore the existence of self-connection in our network and treat them as incorrect entries.

We define a *citation path* between two articles as a finite sequence of citations that connect a set of distinct articles. Let $S \subseteq G$ be an ordered sequence of articles appearing along the path, i.e. the first and last articles in S are the articles of interest, and the length of a citation path between two articles, or *pathlength* in short, is defined as $l_S = |S| - 1$. Note that the citation path between any two articles is not necessarily unique. Assume that there are k different citation paths between two articles, denoted as $l_{S_1}, \ldots, l_{S_k}$, we define the *influence range* between two articles as $r = \min(l_{S_1}, \ldots, l_{S_k})$.

Let v_t be an article of interest, where $1 \leq t \leq m$. Then the Article Network Influence (ANI) of v_t, denoted as ANI_t, is

$$ANI_t = \sum_{r=1}^{r_M} g(r) o_t(r), \tag{1}$$

where r is the influence range from v_t to all other articles in G, r_M is the maximum influence range of interest from v_t, with $1 \leq r_M \leq m$, $o_t(r)$ is the observed number of articles with the influence range r from v_t, and $g(r)$ is a weighted function to represent the decay effect of article citations as the influence range increases.

$o_t(r)$ can be obtained from the citation database easily via calculating the number of citations of v_t received from articles with the influence range r. To obtain $g(r)$, we first denote $a_t(r)$ as the number of articles which pathlengths from v_t are exactly r for $t = 1, \ldots, m$. The estimated average number of citations of all m articles within the pathlength r over the whole citation network G is

$$e(r) = \frac{\sum_{t=1}^{m} a_t(r)}{m}. \tag{2}$$

Then we define the weighting function $g(r)$ as the normalization of a function $g'(r)$ below. If $e(r-1) \leq e(r)$ for $r = 1, \ldots, r_X$ and $e(r_X) > e(r_X + 1)$, then $g'(r) = e(r)^{-1}$ for $r = 1, \ldots, r_X$ and $g'(r) = e(r_X)^{-1}$ for $r = r_X + 1, \ldots, r_M$. Note that it is possible that $r_X = r_M$, i.e. $g(r)$ is a monotonically decreasing function within $\{1, \ldots, r_M\}$.

Generally speaking, the number of citations with the influence range r will increase as r increases. Therefore, the value of a single citation, which is proportional to the reciprocal of the average number of citations, becomes less important as the influence range r increases. However, this phenomenon does not always occur because a paper may be cited by minor papers that do not receive any other citations. It leads to a stop in the increase in the average number as the influence range r increases. The boundary indicates that the paper becomes less important at the point r_X, so we fix $g'(r)$ as $g'(r_X)$ when the value starts to increase.

III. AN ANALYSIS OF THE CITATION NETWORK OF STATISTICS RESEARCH ARTICLES

A. DATA PREPARATION

The Web of Science (WoS) [11] is one of the finest bibliographic databases in the world. It contains relevant attributes of published scientific articles, such as journals where the articles published, their publication years, their authors, their reference lists, and many others. In the current framework of the WoS database, we can access the information of scientific articles during 1981-2016.

Technically, the accessible WoS database is stored in the neo4j graph database. Details about neo4j can refer to [12]. We use the Cypher query language to extract the required data from the database. Given a targeted subject, we first

TABLE 1. The median network influence of top 10 statistics journals.

Journal	Influence	Impact Factor
JRSS SERIES B	29.4321	4.610
ANNALS OF STATISTICS	22.7762	3.023
BIOSTATISTICS	17.4442	1.798
JASA	16.0050	2.016
BIOMETRIKA	14.5430	1.448
JCGS	13.9627	1.735
BERNOULLI	12.2067	1.070
ECONOMETRIC REVIEWS	11.6636	1.333
BIOMETRICS	8.7595	1.329
STATISTICA SINICA	8.4241	0.899

extract from the database that all articles are labeled as the targeted subject. The data subset is then downloaded, and the resulting file is in the CSV file that is structured as a table of two columns. Each row of the table represents a citation between two articles. In specific, the articles listed in the first and second columns are the articles cited and the articles being cited. The rest of the analyses are conducted using R [13], especially the R package "igraph" [14].

B. INTERPRETATION AND ANALYSIS

We applied our method to the citation network of statistics research articles, which can be obtained via the extraction of data with labeled "Statistics" in the whole WoS database. All the data are first divided into the 11-year time span, which is equivalent to a 10-year period after publication. By dividing the total data into small segments, the historical trend might be able to be observed, and we can also avoid the situation that all the early published papers are on the top and the rest of papers are invisible from the list. Therefore, we have 26 networks from the year 1981-1991, the year 1982-1992,..., until the year 2006-2016. ANI is applied to each network, and the results are listed in the appendix. There are totally 26 tables in the supplementary material, each table lists the top-20 influential articles from each network. Below are some observations from these tables.

First, most of the top-20 articles were published at the early stage when compared to the range of years being considered. For example, all 20 articles in Table 1 (1981-1991) were published in either 1981, 1982, or 1983. The most extreme of this happened in Table 15 (1995-2005), where only one article was published in 1996 and the rest were published in 1995. Similar extreme tables include Tables 16 (1996-2006) and 17 (1997-2007). This phenomenon is reasonable because the influence of an article, especially in statistics or most mathematical sciences, needs time to accumulate. It is highly unlikely to have an article that its influence is high enough to jump into the top-20 list in its first year of publication. From this perspective, we can actually identify some exceptional articles from these tables. For example, "Sampling-based approaches to calculating marginal densities" [15], published in 1990, was the only article being on the list six times in all possible eleventh years. It entered into the top-20 list (ranked the 17th) in its fifth year (Table 5), and its rank rose to the 7th in its sixth year (Table 6) and became the 1st in the next four years (Tables 7-10). Several excellent articles entered the list 4-5 times with consecutive first ranks before their eleventh year of publication.

Second, some articles with a smaller number of direct citations are ranked higher than others with a larger number of direct citations. For example, in Table 25 (2005-2015), the first-rank article received 552 direct citations from other articles while the second-rank article received 691 direct citations. For most research metrics that consider only direct citations, the latter article should rank on top of the former article. Oppositely, ANI considers indirect citations with various pathlengths. It reverses the ranks between these two articles because of the much larger number of indirect citations of the first-rank article, which can be interpreted as an implicit spread of its idea to a broader group of audiences or authors in the statistics community. In fact, such phenomena appear quite often in these 26 tables.

For a deeper insight, the topics of these top-20 articles listed in these 26 tables can be viewed as the historical development of statistical research in the past 36 years. For example, the first article related to lasso that reached the top-20 lists was published in 2006, and it appeared early in Table 22 (2002-2012). It was its fourth year after publication, and it indicated the importance and attractiveness of lasso in the statistics community. We found many articles related to lasso appeared in the top-20 lists in the latter years. For example, there are seven lasso-related articles in Table 26 (2006-2016). Similar hot and representative topics, including regression, the Bayesian method, and many others, governed the recent developments of statistics.

In opposite, the first article related to the Gibbs sampling that reached the top-20 list appears in Table 7 (1987-1997). It has received numerous attention since then and reached its peak in Tables 10 (1990-2000) to 12 (1992-2002). We observed a diminishing interest in this topic thereafter, and this topic disappeared completely from the top-20 list after Table 19 (1999-2009). A similar phenomenon also appears in articles related to microarray. The first article appeared in the top-20 list is in Table 18 (1998-2008), and it reached its peak in Table 20 (2000-2010), then the interest diminishes until its disappearance from the top-20 list in Table 25 (2005-2015). It is not trivial to explain in deep why such phenomena of diminishing interests appear, especially in these two topics that might still be popular until now. Some believe that their attractions remain constant while new and hot topics arise and attract more attention, but this guess requires careful evaluations by experts in related fields.

As a final note, we consider the influence range from 1 to 6 in these tables, but one can select the maximum number of ranges suggested by experts. The selection of 6 is simply based on observation in these tables that the number of articles cited an article indirectly reach its maximum when the influence range is 3 or 4, and the delay afterward is steep. Combining with the effect of weighted function $g(r)$, the effects of indirect citations from articles with influence range greater than 4 are small. We do not know if a similar phenomenon appears in other networks, so the selection of this maximum influence range requires expert knowledge or adequate observations.

IV. EXTENSION OF ARTICLE NETWORK INFLUENCE AND COMPARISON TO SOME EXISTING MEASURES OF RESEARCH METRICS

We have demonstrated how the ANI can be used to quantitatively measure the influence of a research article in its scientific field. In this section, we extend the use of network influence on other levels of citations and discuss the differences between ANI and three measures of research metrics found in the literature.

A. A COMPARISON TO IMPACT FACTOR IN JOURNAL-LEVEL CITATION NETWORK

Reference [2] introduced a pioneered concept to quantitatively measure the quality of a scientific journal using a single index about 60 years ago, and it is now well-known to be the Impact Factor among researchers in scientific fields. The formal definition of an impact factor is

$$IF_y = \frac{Citations_{y-1} + Citations_{y-2}}{Publications_{y-1} + Publications_{y-2}},$$

where y is the year of measurement, $Citations_t$ and $Publications_t$ are the numbers of total citations and the number of published papers in a journal respectively for $t = y-1$ and $y-2$.

Following the spirit of the impact factor, we extend the use of ANI to the journal-based level via considering the location of ANI of the articles published in the journal, namely Journal Network Influence (JNI). Since the distributions of ANI in every journal are mostly skewed, the considered location of distribution is the median value instead of the mean value. The following table lists the top ten journals with the largest median network influence, and we compare them with the values of their impact factors. The network influence is calculated during 2006-2016, and the impact factor is taken in the year 2016.

Although the top-10 order follows quite similarly, it is obvious that some ranks are out of the usual expectation. Here are two observations: (1) Journal of Biostatistics ranked the third and had a higher impact than two of the traditional top-4 journals. It represents that 2006-2016 is a period that this subfield develops quickly and expands its influence towards the whole statistics community over the traditional statistics subfield. (2) Except for Journal of Biostatistics, Biometrika ranked back to top-4 statistics journals with JRSSB, Annals of Statistics, and JASA. This prestigious journal has its impact factors dropped to the sixth rank due to the rise of biostatistics and data visualization. However, if we examine its influence over the whole statistics community, it still remains as top 5. Similar situations appear in other two prestigious journals featuring theoretical statistics, Bernoulli and Statistica Sinica.

The main cause of these discrepancies comes from the conceptual difference between the network influence and the impact factor summarized as follows.

First, as mentioned in previous sections, an impact factor considers only an entity's direct citations and ignores its indirect citations. It is equivalent to set $r_M = 1$ or define weighting functions $g(1) = 1$ and $g(r) = 0$ for all $1 < r < r_M$ in the definition of ANI_t. Such setting completely ignores the indirect citations, so the influence of a potentially highly-influential article may be underestimated if one of its followers wrote another excellent article and received many citations after published.

Second, the definition of the impact factor reveals that it considers a subset of the citation database dates back for only two years. It is a major source of unfair comparisons between articles or journals in different subjects because different subjects may have different citation habitats for their distinct natures. For example, the journals in biological sciences and computer sciences enjoy high impact factors as their advancements are straightforward and time applicable, while those in mathematical sciences always have relatively low impact factors as the contents require a deep understanding on the background and adequate time for full digestion. Instead of imposing a year limit, network influence considers the complete citation database, and it is ready to investigate the long-time influence accumulations across the years.

B. A COMPARISON TO PAGERANK ALGORITHM IN ARTICLE-LEVEL CITATION NETWORK

Reference [8] introduced a state-of-the-art algorithm called PageRank to the rank website in Google search engine. It provides a measure of the importance of a web page by counting the number and quality of links to that web page under the assumption of preferential attachments. As a simple definition, the PageRank of a node, denoted as $R(v)$, is defined implicitly as the solution of the following equation

$$R(v) = c \sum_{u \in M(v)} \frac{Ru}{N_u} + \frac{1-c}{n},$$

where c is a damping factor between 0 and 1, $M(v)$ is a set of nodes that link to v, N_u is the out-degree of node u, and n is the number of nodes in the graph.

If we consider a node as an article, it is possible to implement PageRank to measure the importance of an article. In fact, Eigenfactor is a journal ranking method similar to PageRank. Since the calculation of PageRank of a node considers the quality of its neighboring nodes, it implicitly aggregates information from all nodes in the whole network rather than considering direct citations only. However, there are still several differences between PageRank and network influence as stated below.

First, PageRank is an algorithm, and it requires multi-step propagations towards convergence. The PageRank values of nodes are easily altered, and the propagation process is required to run again when a new node is inserted to the network. It may take an unexpectedly large amount of computational resources to obtain the converged values for all nodes if the node addition is very frequent, and it is the reality of the citation network in the Web of Science. Therefore, we suggest using network influence as it can be obtained via a mathematical function that depends only on the number of citations of a target node and its neighboring nodes of interest. When a new node is added, it is not necessary to recalculate the values of all nodes, and only a subset of nodes requires recalculations. Besides, it is not necessary to update the weighting function whenever a new node is added because this function aims at providing a decreasing weight towards the pathlengths between two nodes. We suggest updating annually whenever an annual report on the network influence of the Web of Science is published.

Second, in the iterative formula of PageRank, there is a damping factor (c) that is a user-defined parameter. Its complement $(1 - c)$ is technically an adjustment or a weight on the probability of average crediting to all nodes in the network. Conventional wisdom suggests $c = 0.85$ as it provides satisfactory results, but this suggestion is arbitrary and without statistical justification. Even worse, a different setting of c results in different PageRank. To the extreme, if we set $c = 0$, all nodes will enjoy equivalent importance. It has been questionable on how to adjust this parameter properly. ANI has a user-defined parameter r_M, but it has a clear and interpretable meaning for users. Its weighting function is composed of terms with statistical meaning.

C. A COMPARISON TO FIELD-WEIGHTED CITATION IMPACT (FWCI) IN ARTICLE-LEVEL CITATION NETWORK

Field-Weighted Citation Impact (FWCI) is a new measure introduced by [6] to evaluate the entity's impacts in Scopus, the database of Elsevier since 1996. It indicates how the number of citations received by an entity's publications compared to the average number of citations received by all other similar ones. Mathematically, the FWCI of an entity is defined as

$$FWCI = \frac{1}{N}\sum_{i=1}^{N}\frac{c_i}{e_i},$$

where N is the number of publications by an entity, c_i is the citations received by publication i, and e_i is the expected number of citations received by all similar ones in the publication year plus the following three years. When a similar publication is allocated to more than one discipline, the harmonic mean is used to calculate e_i.

FWCI can be viewed as a simple modification from the impact factor. It differs in two places: (1) the inclusion of the number of citations changes from two backward years backward to three forward years; (2) instead of the number of publications as the denominator, FWCI considers the average number of citations of similar entities. Thus, two differences between ANI and the impact factor also inherit in the difference to FWCI, and they are indirect citations and subsets in terms of years.

Moreover, when one considers an aggregated ANI from several articles, FWCI takes the arithmetic means on the ratio of actual to the expected citations of each publication while JNI defined in section IV-A takes the median values of ANI of each publication. Notice that it is highly unlikely for both the ratio and the influence of an article in a citation network to be normally distributed, and as we observe in most cases, they are highly right-skewed. Therefore, FWCI tends to provide an inflated value on the average impact of an entity, and it is highly sensitive to the outlying values. In contrast, we propose to consider the median value of the ANIs of several articles instead. The median value is well-known to be resistant to these outliers, and it provides an informative average value towards a non-normal set of values.

V. CONCLUSION

In this work, we introduce a new quantity called network influence to measure the importance of an article when a citation network is given. Through the article citations, the influence of an article is propagated not only to its followers but also the followers' followers and so on. We define the quantity called influence range to describe this hierarchical relationship.

There are several extensions of this work. First, it is desired to investigate in the article citation networks of other subjects. Notice that the citation in statistics mostly represents an advancement of a theory or method, but different interpretations may appear in other scientific fields. For example, in addition to the scientific advancement, we expect the existence of other meanings of citations in physics, including but not limited to, the use of the instruments, the reference of the experimental setup, the use of the specific software or websites, and many others. Thus, the investigation of cross-disciplinary citations becomes an obvious step for future work.

Second, it is obvious that different subjects may have different traditions or habits in publishing research results. This leads to significant non-random biases among subjects. Unlike FWCI, we do not actively provide an adjustment on the current version of ANI to compensate the biases from the subject difference. This also becomes another obvious step for future work.

Third, there are many attributes for all articles in the WoS database, like authors, institutions, etc. Each attribute can form a citation network with different new challenges. For example, it is necessary to deal with self-citations in an author citation network because a researcher is allowed to cite his own past articles. Moreover, not all citations are connected equally because Researcher A may cite Researcher B ten times in ten different articles but only cite Researcher C once. Thus, a normalization scheme is essential for edge standardization.

Lastly, WoS consists of articles in the past thirty years, and it recorded the publication years of all articles.

Therefore, it is possible to create, for the whole network or a subgraph on a specific subject, multiple networks at different times. Then it is highly interested in studying the evolution of some citation networks over the past thirty years from a dynamical point of view. The simplest way to achieve this goal is to study the connection probabilities between articles, which is expected to have strong relations with the articles' in-degrees.

ACKNOWLEDGMENT

The authors would like to thank Clarivate Analytics to provide access to the raw data of the Web of Science database for research investigations. They also thank the URA team of ISM, Dr. Keisuke Honda and Mrs. Hiroka Hamada in specific, for transforming the data into the neo4j database and providing the neo4j database for analysis in this work. Lastly, the authors would like to thank Ms. Ula Tzu-Ning Kung for her editing service to improve the English quality and readability of this manuscript.

REFERENCES

[1] A. Pritchard, "Statistical bibliography or bibliometrics?" *J. Doc.*, vol. 25, no. 4, pp. 348–349, Jan. 1969.

[2] E. Garfield, "Citation indexes for science: A new dimension in documentation through association of ideas," *Science*, vol. 122, no. 3159, pp. 108–111, Jul. 1955. doi: 10.1126/science.122.3159.108.

[3] C. T. Bergstrom, J. D. West, and M. A. Wiseman, "The eigenfactor metrics" *J. Neurosci.*, vol. 28, no. 45, pp. 11433–11434, Nov. 2008. doi: 10.1523/JNEUROSCI.0003-08.2008.

[4] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci.*, vol. 102, no. 46, pp. 16569–16572, Nov. 2005. doi: 10.1073/pnas.0507655102.

[5] K. Kim and Y. Chung, "Overview of journal metrics," *Sci. Ed.*, vol. 5, no. 1, pp. 16–20, Feb. 2018.

[6] *Research Intelligence: Research Metrics Guidebook*, Elsevier, Amsterdam, The Netherlands, 2018. [Online]. Available: https://www.elsevier.com/research-intelligence/resource-library

[7] H. F. Moed, "Measuring contextual citation impact of scientific journals" *J. Informetr.*, vol. 4, no. 3, pp. 265–277, Nov. 2009. doi: 10.1016/j.joi.2010.01.002.

[8] L. Page, S. Brin, R. Motvani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford InfoLab, Stanford, CA, USA, Tech. Rep. 1999-66, 1999.

[9] C. Varin, M. Cattelan, and D. Firth, "Statistical modelling of citation exchange between statistics journals," *J. Roy. Stat. Soc. A*, vol. 179, no. 1, pp. 1–63, Jan. 2016. doi: 10.1111/rssa.12124.

[10] S. Gualdi, M. Medo, and Y.-C. Zhang, "Influence, originality and similarity in directed acyclic graphs," *Europhys Lett.*, vol. 96, no. 1, Sep. 2011, Art. no. 18004. doi: 10.1209/0295-5075/96/18004.

[11] Web of Science. (2018). *Clarivate Analytics*. [Online]. Available: http://www.webofknowledge.com/

[12] J. Baton and R. van Bruggen. (2019). *Learning Neo4j 3.x (2nd Edition)*. [Online]. Available: https://www.packtpub.com/big-data-and-business-intelligence/ learning-neo4j-3x-second-edition

[13] R Core Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org/

[14] G. Csardi and T. Nepusz, "The Igraph software package for complex network research," *InterJ., Complex Syst.*, vol. 11, no. 1695, pp. 1–9, 2006. [Online]. Available: http://igraph.org

[15] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities" *J. Amer. Stat. Assoc.*, vol. 85, no. 410, pp. 398–409, Jun. 1990. doi: 10.2307/2289776.

LIVIA LIN-HSUAN CHANG received the B.B. degree in business mathematics from Soochow University, Taiwan, and the M.S. degree in applied statistics from California State University, Long Beach, CA, USA. She is currently pursuing the Ph.D. degree in statistical science with SOKENDAI, The Graduate University for Advanced Studies, Tokyo, Japan. She was a Research Assistant with the Institute of Statistical Science, Academia Sinica, from 2014 to 2017. Her research interest includes network data analysis.

FREDERICK KIN HING PHOA received the B.S. degree in physical chemistry, the B.S. degree in applied mathematics, the M.S. and Ph.D. degrees in statistics from the University of California at Los Angeles (UCLA), in 2001, 2002, 2006, and 2009, respectively.

He has been a Research Fellow with the Institute of Statistical Science, Academia Sinica (ISSAS), since 2018. He was an Assistant Research Fellow and an Associate Research Fellow of the ISSAS from 2009 to 2013 and from 2013 to 2018. His research interests include design and analysis of experiments, network data analysis, nature-inspired metaheuristic optimization, big data analytics, techniques in deep learning and artificial intelligence, and statistical practices. From 2009 to 2019, he has published more than 45 scientific articles in international journals, has given more than 110 invited talks in the international conferences, and more than 65 seminar talks in the universities around the world.

Dr. Phoa was a recipient of the Career Development Award (2014) from Academia Sinica, the Ta-You Wu Memorial Award (2014), and the Outstanding Research Award (2017) from the Ministry of Science and Technology (MOST). He received the Best Paper Award in the World Congress of Engineering, in 2015. He conducted the Excellent Young Researcher Research Project supported by MOST, from 2013 to 2016, an International Cost-Share Exchange Scheme funded by MOST and Royal Society, U.K., from 2016 to 2018, and an interdisciplinary project on artificial intelligent production in layer industry funded by MOST, from 2018 to 2021.

JUNJI NAKANO received the B.S. and M.S. degrees in applied mathematics from the University of Tokyo, in 1978 and 1980, respectively, and the Ph.D. degree in statistics from the Tokyo Institute of Technology, in 1986.

He worked as an Assistant Professor and an Associate Professor with Tokushima University, Saitama University, and Hitotsubashi University, Japan, from 1980 to 1999. He joined the Institute of Statistical Mathematics, Tokyo, Japan, as a Professor, in 1999, and worked there, until 2018. He is currently a Professor with the Department of Global Management, Chuo University, Tokyo, Japan, since 2019. His research interests include time series analysis, statistical computing, and big data analysis, especially data visualization and graph structure analysis.

Prof. Nakano is a member of the International Association for Statistical Computing, the International Statistical Institute, the Japan Statistical Society, and the Japanese Society of Computational Statistics. He is a recipient of the Software Development Award of the Japanese Society of Computational Statistics, in 2008, the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology, Japan (Prizes for Science and Technology, Public Understanding Promotion Category), in 2017.