International Journal of Information Science and Management
Vol. 21, No. 4, 2023, 23-38
DOI: https://doi.org/10.22034/ijism.2023.1977713.0 / DOR: https://dorl.net/dor/20.1001.1.20088302.2023.21.4.2.8

Original Research

Analyzing the Impact of Social Collaborations on Influence Identification in Scientific Literature Analytic: An Analysis on ResearchGate and Academia

Mitali H. Desai
Assistant Prof., Computer Engineering Department, C. K. Pithawala College of Engineering and Technology, Surat-07, Gujarat, India.
Corresponding Author: mitalidesai17@gmail.com
ORCID iD: https://orcid.org/0000-0002-3264-6143

Rupa G. Mehta
Associate Prof., Computer Engineering Department, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat-07, Gujarat, India.
rgm@coed.svnit.ac.in
ORCID iD: https://orcid.org/0000-0002-6260-0488

Dipti P. Rana
Assistant Prof., Computer Engineering Department, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat-07, Gujarat, India.
dpr@coed.svnit.ac.in
ORCID iD: https://orcid.org/0000-0002-5058-1355

Received: 06 February 2022
Accepted: 24 December 2022

Abstract

Influence identification, one of the compelling applications of Social Network Analysis (SNA) is gaining immense attention in scientific literature analytics. Existing influence identification techniques in the scientific domain majorly explore scientific collaborations (co-author and co-citation networks) of researchers. Standard centrality algorithms are widely applied for this purpose. The emergence of digital scholarly platforms allows researchers to build their social community in the scientific environment. Few scholarly platforms maintain social media like follower and following relations apart from co-authorships and co-citations of researchers. This research examines the impact of followers and followings on influence identification in the scientific domain. The real scientific information from widely utilized digital scholarly platforms: ResearchGate (RG) and Academia is extracted. From the collected information, scholarly networks are constructed based on follower-following relations. Standard centrality algorithms are implemented to identify the influence of these networks. The results are compared with i) the researcher's influence scores provided by RG and Academia ii) three legitimate global ranking lists of researchers. The outcome suggested that, like SNA, social collaborations among researchers in terms of followers and followings significantly impact influence identification in the scientific domain.

Keywords: Influence Identification, Scholarly Platforms, Social Collaborations, Centrality Algorithms, Scientific Literature Analytic, Scholarly Networks.

Introduction

Digital scholarly platforms allow researchers to conduct various scientific activities such as scientific research sharing, academic communications, research collaborations, research feedback collection, information propagation, and technical knowledge dissemination (Espinoza Vasquez & Caicedo Bastidas, 2015; Gasparyan, Nurmashev, Yessirkepov, Endovitskiy, Voronov & Kitas, 2017). Such activities have led to extensive growth in digital scientific information. Scientific literature analysis deals with accurately analyzing this information to generate concise outcomes (Desai, Mehta & Rana 2022b).

Prevalent utilization of scholarly platforms such as ResearchGate (RG), Academia.edu, Mendeley, and Publons has opened a novel paradigm of research in the field of scientific influence identification (Wu & Zhang, 2019), which is a compelling application of scientific literature analytic. Scientific influence identification means identifying dominant scientific entities such as scientific articles, researchers, topical experts, domain-specific journals/conferences, etc. from an extensive scientific literature (Savov, Jatowt & Nielek, 2020; Wu, Fan & Yuan, 2021; Liang, Mao, Lu, Ba & Li, 2021). This paper focuses on identifying influential researchers in scientific literature analytics.

Existing techniques to identify influential researchers work with scientific collaboration networks: co-authorships and co-citations. Such networks are generally available for many scholarly platforms, scientific repositories, and bibliometric services such as Mendeley, Academia.edu, Scopus, Google Scholar (GS), Digital Bibliography & Library Project (DBLP) and Association for Computing Machinery (ACM) (Ortega, 2017; Keller, 2019). Besides traditional co-authorship and co-citation networks, scholarly platforms like ResearchGate (RG) and Academia provide researchers social collaboration networks regarding followers and followings (Jordan, 2019). Social Network Analysis (SNA) extensively explores such social collaborations to identify influential users (Malinen & Koivula, 2020; Valsesia, Proserpio & Nunes, 2020). This research analyzes the impact of social collaborations i.e., followers and followings, on influence identification in scientific literature analytics.

Real scientific information from RG and Academia is extracted. From the extracted information, networks based on follower and following relations are constructed respectively for RG and Academia. In the constructed networks, each node represents a researcher, whereas the edge denotes a follower or a following relation between two connected researchers. Standard network centrality algorithms: betweenness, degree, closeness, eigenvector, and PageRank are implemented on constructed networks for influence identification. The notion is to leverage the follower and following relations to see how the influence of one researcher affects the influence of other connected researchers through centrality measures. Each centrality algorithm delivers a list of top-k influential researchers from respective networks.

The results are compared with the official researcher's influence indicator of RG and Academia - RGScore and AcadScore, respectively. The results are also compared with three recognized global ranking lists of researchers: IDEAS, SSRN (Social Science Research Network), and Clarivate provided by the Research Division of the Federal Reserve Bank of St. Loui, Elsevier, and Web of Science (WoS) respectively. The outcome indicated that followers and followings significantly impact influence in the scientific domain. The achieved efficient results open a novel research paradigm concerning social collaborations in scientific literature analytics.

The remainder of this paper is organized as follows: In the section Literature Review, the pioneering work in the domain of scientific influence finding and centrality algorithms are briefly reviewed. In the section Research Objective and Research Contributions, the identified research gaps, the objective of this research, and our research contributions are presented. In the section Research Methodology, the proposed research methodology is discussed. The section Data Analysis and Results suggests the experimentation results with detailed reasoning.

In the section Discussion, the findings of this research are discussed, interpreted, and compared with the existing research. In the Conclusion and Future Work section, we conclude this research along with future expansion possibilities.

Literature Review

In this section, the literature work is described for the existing studies in the domain of influence identification in scholarly platforms and standard centrality algorithms in scholarly networks.

Influence Identification in Scholarly Platforms

Social media platforms are used for social communication, sharing personal information, and collaborations. Accurately analyzing the information on various social media platforms plays a vital role in numerous applications such as influence analysis, community detection, recommendation systems, sentiment analysis, emotion detection, and many more. Inspired by the growing utility and benefits of social network platforms (Khvatova, Dushina & Nikolaenko, 2017), the paradigms of specialized scholarly platforms emerged in 2008.

The embodiment of open accessibility, availability, digitization, and flexibility offered by scholarly platforms allow researchers from diverse research domains and geographically dispersed locations to connect and cooperate (Desai, Mehta & Rana, 2019). Scholarly platforms are assumed to be adaptable to incorporate seamless changes in scientific information shared across the web. The information present on such platforms keeps the global scientific community enlightened of inclining research domains, research articles, and significant scientific contributors and also unveils several research opportunities for emerging researchers (Ansari & Khan, 2020; Hashmi, Kayani, Toor, Mansoor & Raheem, 2020; Desai, Mehta & Rana 2023). The information is available across a wide range of diverse scholarly platforms. It is a potential base for many applications in scientific literature analytics, among which identifying influential researchers is leading.

Existing literature performs influence identification concerning co-author (Ding, 2011; Amjad, Daud, Akram & Muhammed, 2016; Amjad, Daud, Che & Akram, 2016; Amjad, Daud & Aljohani, 2018; Bibi, Khan, Iqbal, Farooq, Mehmood & Nam, 2018; Maia, Lenzi, Rabello & Oliveira, 2019; Alizade Zowj, Ghane & Ehsanifar, 2019; Makkizadeh, Dehghan & Mostafavi, 2020; Wang, Ding, Wei & Long, 2021; Kesht-Karan, Ghane & Danesh, 2021) and co-citation (Zhang, Zhao, Cheng, Cheng & Wang, 2016; Bai, Zhang, Hou, Lee, Kong, Tolba & Xia, 2018; Yaghtin, Sotudeh, Mohammadi, Mirzabeigi & Fakhrahmad, 2019; Wang et al., 2021; Samie, Biranvand, Rahmaniyan & Varnamkhasti, 2022) relations among researchers. Most studies are conducted on readily available ancient co-author and co-citation networks, whereas in some studies, such networks are prepared by collecting recent scientific information. These networks are further explored using various centrality algorithms to identify influence.

Among available digital scholarly platforms, RG and Academia are well-known and currently provide followers-following information of researchers. RG and Academia facilitate the researchers to generate their profiles, share their research and research interests, discover peers with expertise in relevant domains, and get insights into inclining research. Information shared on RG is categorized into researcher’s profile content: research items, citations, reads, recommendations, department, position, and affiliation, as well as researcher-to-researcher links: followers, followings, co-authors, and co-citations. Similarly, the information present on Academia is categorized into researcher’s profile content: research items, citations, reads, recommendations, department, position, affiliation as well as researcher-to-researcher links:

followers, followings, co-authors, and co-citations. RG and Academia are currently the only digital scholarly platforms allowing researchers to connect with other researchers regarding followers and followings. In SNA, followers and followings are widely utilized in influence identification. In today's era of technological advancements and rapid developments, it is essential to consider these social collaborations in the scientific domain while gauging their influence.

Centrality Algorithms in Scholarly Networks

To gasp a comprehensive impression of the collected information, we conducted an analysis to determine which researchers (nodes) are at the 'center' of a network (Wan, Mahajan, Kang, Moore & Cho, 2020). In network analytics, centrality is a significant concept to identify influential (central) nodes (Bhattacharya & Sarkar, 2021). Finding influential nodes is subjective depending on how the 'influence' is defined (Desai, Mehta & Rana 2022a). Different centrality algorithms determine the influence from different perspectives. Our analysis considers standard centrality algorithms, betweenness, degree, closeness, eigenvector, and PageRank. Their theoretical concepts are briefly explained below. For a given scholarly network: S_N (N, R), where N is a set of nodes (researchers) and R is a set of relations (followers/followings):

Betweenness Centrality (BC) measures the importance of a node. $n_i, \forall n_i \in N$ based upon how often it occurs in the shortest path between all the pairs of nodes in a network. In Equation 1, g_{jik} is all geodesics connecting nodes *j* and *k* passing through a node *I* and g_{jk} is the geodesic distance between nodes j and i. A node being in between (having a greater Bc score) in a network can control the information flow between others.

$$Bc(n_i) = \sum_{j,k \neq i} \frac{g_{jik}}{g_{jk}} \tag{1}$$

Degree Centrality (DC) of a node n_i, $\forall n_i \in N$ is the number of ties that a node n_i has with other nodes. DC of any node can be calculated as represented in Equation 2. Here, $d(n)$ is the degree (number of ties) of the node n_i.

$$Dc(n_i) = d(n_i) \tag{2}$$

In a network with non-directed relations, the degree of a node is identified as the number of direct relations a node has with other nodes. In a directed network, the degree of a node can be In-Degree or Out-Degree. In-Degree Centrality (IDC) of a node n_i, $\forall n_i \in N$ refers to the number of relations incidents and is calculated as displayed in Equation 3. Here, $d_i(n_i)$ is in-degree of a node n_i. Out-Degree Centrality (ODC) of a node n_i, $\forall n_i \in N$ refers to the number of relations from it to other nodes and is calculated as mentioned in Equation 4. Here, $ODc(n_i)$ is out-degree of node n_i.

$$IDc(n_i) = d_i(n_i) \tag{3}$$

$$ODc(n_i) = d_o(n_i) \tag{4}$$

In a connected network, the Closeness Centrality (CC) of a node $n_i, \forall n_i \in N$ is the average length of the shortest path between n_i and all other nodes. The notion is that if a node is more central in a network, it is geodetically closer to all other nodes. Closeness centrality concentrates on the extensivity of influence over the entire network. In Equation 5, $Cc(n_i)$ is the closeness centrality of a node n_i and $d(n_i, n_j)$ is the distance between two nodes n_i and n_j in a network.

$$Cc(n_i) = \sum_{i=1}^{N} \frac{1}{d(n_i, n_j)} \tag{5}$$

Eigenvector Centrality (EVC) measures the importance of a node $n_i, \forall n_i \in N$ as a function of the importance of its neighbors. If a node n_i is connected to highly important nodes, it has a higher EvC score than others. Considering a weighted relation with weight (ω_{ij}) between two researchers (say n_i and n_j); EvC score of a node n_j is calculated as Equation 6. Here, σ is a constant.

$$Evc(n_j) = \frac{1}{\sigma} \sum_{n_i \in N(n_j)} \omega_{ij} \times Evc(n_j) \tag{6}$$

PageRank Centrality (PRC) calculates the number and the quality of connections to a node $n_i, \forall n_i \in N$ to estimate the importance of n_i in a network. The important nodes are likely to receive more connections from other nodes. A node has a high rank if the aggregated sum of the ranks of its backlinks is high. PRC of a node n_i is calculated as Equation 7.

$$PRc(n_i) = (1 - d_f) + d_f \times \sum \frac{PRc(n_j)}{N(n)} \tag{7}$$

Research Objective and Research Contributions

From the literature, it is identified that in the scientific domain, influence identification is widely performed in terms of co-authorship and co-citation networks. It is essential to anatomize the impact of SNA-inspired social collaborations on influence. Also, the literature has majorly explored readily available co-author and co-citation networks. Identification of influence in the scientific domain by fetching the real-time linked information serves as an establishment of developing potential research.

As it is proven in ingrained social theories that the followers and followings have a significant impact in influence prorogation, this research aims at examining the same in the scientific domain.

This research aims to analyze the impact of social collaborations i.e., followers and followings on influence identification in scientific literature analytics. Real-time linked information from scholarly platforms is extracted, and respective networks are constructed. The well-established network-based centrality algorithms are implemented on constructed networks for influence identification.

The significant contributions of this research are to:

- Render the real-time linked scientific information from digital scholarly platforms
- Model the rendered information into scholarly networks
- Propose a methodology to identify the relevance of social collaborations (followers and followings) on scientific influence identification
- Identify the applicability and legitimacy of standard centrality algorithms on influence identification in the scientific domain concerning social collaborations

Materials and Methods

The proposed research methodology is an initial step to generate and utilize a network-based model of scientific information on RG and Academia to identify influence considering researchers' followers and following relations. The framework of the proposed methodology constitutes three sub-procedures: data extraction, scholarly network construction, and implementation of centrality algorithms, as demonstrated in Figure 1.

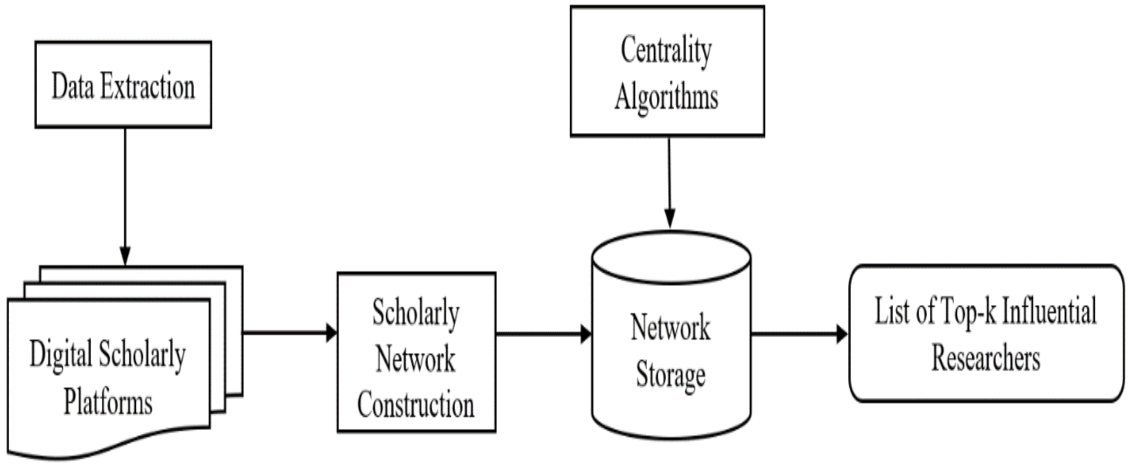


Figure 1: Framework of the Proposed Methodology

The pseudo-code of the proposed methodology is presented in the below Algorithm.

Algorithm 1

Input: Target Researcher T
Output: List of Top-k Researchers

```
Step 1: Data Extraction
        For i=1 to d                          // d is the depth of the network
           Render FL of T                     // T is Target researcher
           For i=1 to n                       // FL is the set of Follower Links
                T = FL[i]                     // n is size of FL
           End For
        End For
        For i=1 to d
           Render FW of T                     // FW is the Following Links
           For i=1 to n
                T = FL[i]
           End For
        End For
Step 2: Scholarly Network Construction
        Initiate Neo4j instance
        CQL load(N)                           // N is set of Nodes
        CQL load(R)                           // R is set of Relations
Step 3: Implementing Centrality Algorithms
```

Calculate $Bc(n_i)$ as per Equation 1; $\forall n_i \in N$
Calculate $IDc(n_i)$ as per Equation 3; $\forall n_i \in N$
Calculate $ODc(n_i)$ as per Equation 4; $\forall n_i \in N$
Calculate $Cc(n_i)$ as per Equation 5; $\forall n_i \in N$
Calculate $Evc(n_i)$ as per Equation 6; $\forall n_i \in N$
Calculate $PRc(n_i)$ as per Equation 7; $\forall n_i \in N$

While for many scholarly platforms, the co-authorship and co-citation networks of researchers are available, it is challenging to collect real linked information from RG and Academia. The required information is extracted using a rendering process (Desai, Mehta & Rana, 2021) that renders publicly accessible researchers' information from both platforms. One known influential researcher who is active in RG and Academia is considered a target researcher (*T*). For any selected target researcher *T*, all the followers (*FL*) and followings (*FW*) up to the required depth (*d*) are extracted as shown in Step 1 of Algorithm 1.

RG and Academia networks are constructed from the rendered information, as shown in Step 2 of Algorithm 1. Each 'node' in both networks represents a researcher from RG and Academia, respectively. Their relations are modeled as 'directed edges' with labels 'Following' and 'Follower'. *N* nodes and R relations are created among *N* researchers based on the extracted information. The networks are stored in Neo4j, the leading network database. Cypher Query Language (CQL) interacts with networks stored in Neo4j.

The centrality algorithms are used to discover the influential nodes in a network. In this research, four standard centrality algorithms - betweenness, degree, closeness, eigenvector, and PageRank are implemented iteratively on RG and Academia networks to find the influential researchers, as shown in Step 3 of Algorithm 1.

The identified list of top-k researchers is compared with considered benchmarks by implementing each centrality algorithm.

Results

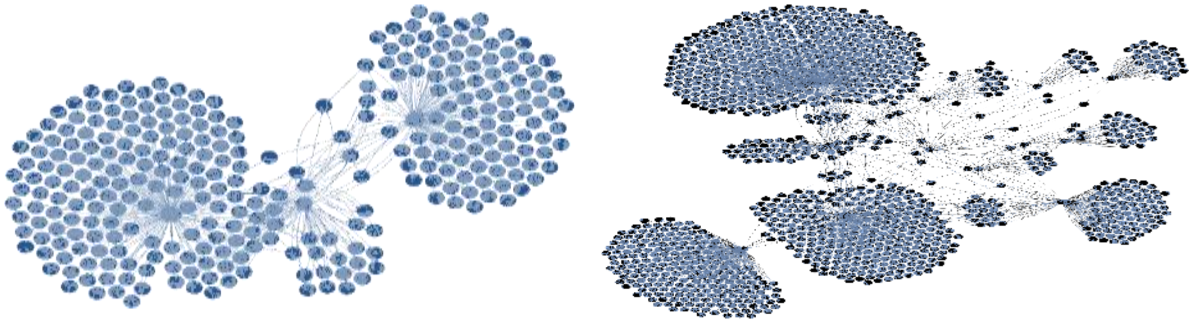
The proposed methodology is implemented on a machine with Ubuntu 18.04 LTS (64-bit) operating system, 8 GB RAM, and an Intel Core i7-7700 processor using Python language (version 3.3). For data collection, web rendering and Xpath are used. Graphml and Neo4j (version 3.5.8) are used for network generation and storage. To visualize network layout, Neo4j is used.

A known influential researcher, Vernon Smith, is targeted in data extraction. The information of his followers and followings is extracted from RG and Academia, respectively. The extracted information is modeled into networks, which are stored in Neo4j. The statistics of RG and Academia networks are displayed in Table 1.

Table 1
Dataset Statistics

	No. of Nodes	No. of Relations
RG	12172	11799
Academia	2171	2999

The visual layouts of both networks are displayed in Figure 2. In the Figure, the partial RG network is shown due to the restriction of Neo4j in showing a huge network.



(a) RG (Partial) (b) Academia (Complete)

Figure 2: Researchers' Follower-Following Networks

After implementing four centrality algorithms, each algorithm computes the list of top-k influential researchers for each network. Table 2 and 3 show the sample list for k=25 (top-25) identified influential researchers from RG and Academia.

Table 2
List of Identified Top-25 RG Researchers Using Centrality Algorithms

Top-k Ranks	BC	IDC	ODC	CC	EVC	PRC
1	Hernan Bejarano	Dariusz Prokopowicz	Federico Del Giorgio Solfa	Vernon L. Smith	Eric Schniter	David Sloan Wilson
2	Eric Schniter	Edward L Deci	Dariusz Prokopowicz	Hernan Bejarano	Michael Gurven	Ernest Aryeetey
3	Vernon L. Smith	Federico Del Giorgio Solfa	Даниил Ковалев	Charles Plott	David P Porter	Raymond Riezman
4	Erik O. Kimbrough	Ian Gilligan	Алена Бычкова	Mary L. Rigdon	Hernan Bejarano	George Akerlof
5	C. Monica Capra	Stephen I. Ternyik	Ilgar Gurbat oglu Mamedov	George Akerlof	Brice Corgnet	Rosemarie Nagel
6	Mary L. Rigdon	Krishnan Umachandran	Алексей Бычков	Praveen Kujal	Stephen J. Rassenti	Daniel Kahneman
7	George Akerlof	Shalom H Schwartz	Krishnan Umachandran	Charles Noussair	Erik O. Kimbrough	Robert W. Crandall
8	Michael Gurven	H. Eugene Stanley	Riccardo Vecellio Segate	David P Porter	Charles Plott	Michael Gurven
9	Ernest Aryeetey	Uta Frith	Yichuan Zhao	Erik O. Kimbrough	Charles Noussair	Erik O. Kimbrough
10	Albert Schram	Hashem Pesaran	Ali Mohammadi	Eric Schniter	Praveen Kujal	Eric Schniter
11	David Sloan Wilson	Ross Levine	Pelayo Munhoz Olea	Brice Corgnet	Rachel Anneliese Bodsky	Lawrence Weiskrantz
12	Raymond Riezman	David J. Teece	Mohammad Ismail Al-berfkani	Stephen J. Rassenti	Taylor Jaworski	Mirac Yazici
13	Juliette Milgram	Michael Posner	Aieman Ahmad Al-Omari	Daniel Mcfadden	Mary L. Rigdon	James Marvin Walker
14	David P Porter	Cris Oprea	Ricardo Candea Sa Barreto	Daniel Kahneman	Verena Utikal	Philip J. Grossman
15	Brian Krauth	Richard H. Thaler	Benedikt Herz	Taylor Jaworski	Raymond Riezman	D. Wade Hands
16	Ramzi Suleiman	Sundarapandian Vaidyanathan	Pierluigi Siano	Mirac Yazici	Rosemarie Nagel	David P Porter

Top-k Ranks	BC	IDC	ODC	CC	EVC	PRC
17	Mirac Yazici	Sofia D. Wechsler	Philip Govule	Don Coursey	Ramzi Suleiman	Robert J. Shiller
18	John Hey	John T Cacioppo	H Gin Chong	Henry DANIEL Vera Ramirez	Jeffrey Lindstrom	Charles Noussair
19	Daniel Mcfadden	James Heckman	Volodymyr Saienko	C. Monica Capra	Arlington W. Williams	Albert Schram
20	Thomas Hazlett	Aieman Ahmad Al-Omari	Hanna Tolchieva	Ernest Aryeetey	C. Monica Capra	Hernan Bejarano
21	Charles Plott	David M Buss	Sundarapandian Vaidyanathan	David Sloan Wilson	Eric Alden Smith	Charles Plott
22	Don Coursey	Jay Belsky	Alexander Brem	Ramzi Suleiman	John Hey	Praveen Kujal
23	Praveen Kujal	Norbert Schwarz	Stephen I. Ternyik	Michael Gurven	James Marvin Walker	Thomas Hazlett
24	Herbert Gintis	Martin A Nowak	Arup Barman	Rosemarie Nagel	Herbert Gintis	Taylor Jaworski
25	Arthur T. Denzau	Elliot Bendoly	Saeed Ur Rehman	Arlington W. Williams	Mark Schneider	Daniel Mcfadden

The top-k list received by each centrality algorithm for RG is compared with RGScore, IDEAS, SSRN, and Clarivate. Similarly, the top-k list received by each centrality algorithm for Academia is compared with AcadScore, IDEAS, SSRN, and Clarivate. ResearchGate score (RGScore) and AuthorRank Score (AcadScore) are the researcher’s influence scores provided by RG and Academia, respectively. We have also considered three recognized global ranking platforms for thorough comparison. The notion is to analyze the relevance of follower-following networks on influence for scholarly platforms’ official influence indicators and the global landscape.

Table 3
List of Identified Top-25 Academia Researchers Using Centrality Algorithms

Top-k Ranks	BC	IDC	ODC	CC	EVC	PRC
1	Vernon Smith	Maurizio Forte	Raj Kishor	Vernon Smith	Hernan D Bejarano	Hernan D Bejarano
2	Hernan D Bejarano	Steven Pinker	Muhammad Asghari	John Thrasher	Stephen Rassenti	Lynne Kiesling
3	John Thrasher	Matthew O. Jackson	Timbul Widodo	Hernan D Bejarano	Carl Johnston	Rebecca Storey
4	Lynne Kiesling	Gary Feinman	Hadrien J Rambach	Steven Pinker	Erik O Kimbrough	John Thrasher
5	Rebecca Storey	Jeffrey Wooldridge	Gary Feinman	Erik O Kimbrough	Praveen Kujal	Philip Tetlock
6	Erik O Kimbrough	Richard Price	Henning Papendorf	Cristina Bicchieri	Lance Clifner	Erik O Kimbrough
7	Philip Tetlock	Jonathan Zittrain	Krishnamurthy Prabhakar	William Easterly	Vernon Smith	Robert Litan
8	Robert Litan	Asian Business Review	Matthieu Queloz	Eric C. Ip	Yoanna Ganeva	R. Shiller
9	R. Shiller	Philip N Pettit	Bernardo Dainese	Philip Tetlock	Rimvydas Baltaduonis	Praveen Kujal
10	Praveen Kujal	Angel Versetti	Luiz Carlos Montans Braga	Carl Johnston	Rebecca Morton	Vernon Smith
11	Rimvydas Baltaduonis	Luiz Guilherme Marinoni	Ben Fulton	Peter Boettke	Alonso Alfaro	James Cox

Top-k Ranks	BC	IDC	ODC	CC	EVC	PRC
12	James Cox	Joseph Raz	Jose Miguel Camacho-Castro	Praveen Kujal	Alessandro Tavoni	Forrest Nelson
13	Laurence Iannaccone	Thom Brooks	Naveen Rajput	Stephen Rassenti	Marcelo Olivera	Stephen Rassenti
14	Stephen Rassenti	Richard Bellamy	Ruslan Spevakov	Lance Clifner	Vinayak Dixit	Carl Johnston
15	Kevin Vallier	Martin O'Neill	Siddharta Legale	Bart J Wilson	Pradeep Kumar	Lance Clifner
16	Peter Boettke	Tim Crane	Professor.Krishan Bir Singh	Roman Sheremeta	Ashley Sitar	Wesley D Stoner
17	Andrew J Cohen	Timothy Williamson	Marcin Paprzycki	Laurence Iannaccone	Juan Manuel Puerta	Michael Spence
18	Dan Shahar	Krishnamurthy Prabhakar	James Kierstead	Lynne Kiesling	Milagro Saborio-Rodriguez	Fred Longstaffe
19	Ryan Patrick Hanley	Cristina Bicchieri	Victor Ricciardi	Caleb Sturges	Robin Dillaway	Gina M . Buckley
20	A. Williams	Moris Polanco	Steven Bonacorsi	Miles Zimmerman	Mark Ellyne	Linda Rosa Manzanilla
21	Caleb Sturges	Carla Bagnoli	Manuel Pulido Mendoza	Erte Xiao	Shubhro Sarkar	Matthew Piscitelli
22	Miles Zimmerman	Andrew Benjamin	Juan Pablo Barros	Maximo Rossi	Stephen Smith	Katherine A . Miller Wolf
23	Erte Xiao	Susan Haack	Mirac YAZICI	Diego Gambetta	Ashwin Samant	Richard Sutter
24	Bart J Wilson	Jonathan Zittrain	Martin O'Neill	Nicolette (Nikki) Sullivan	Cristina Bicchieri	Christopher Pool
25	Maximo Rossi	Brad Hooker	Jorge L Fabra-Zamora	Virgil Storr	Laurence Iannaccone	Kari A. Zobler

For result comparison, the standard correlation measures Spearman and Kendall's correlations are considered. Table 4 shows the correlation tuples created for top-k comparisons. Here, *i* varies from *i=1 to k* for top-k results. RG and Acad denote RGScore and AcadScore, respectively. RI, RS, and RC denote RG top-k tuples with three global top-k lists for comparison with global ranks. Similarly, AI, AS, and AC denote Academia top-k tuples with three global top-k lists.

Table 4
Ranking Tuples Notations for Top-k

	RG				Academia			
	RGScore Tuple	Global Rank Tuple			AcadScore Tuple	Global Rank Tuple		
		IDEAS	SSRN	Clarivate		IDEAS	SSRN	Clarivate
BC	$<BC_i, RG_i>$	$<BC_i, RI_i>$	$<BC_i, RS_i>$	$<BC_i, RC_i>$	$<BC_i, Acad_i>$	$<BC_i, AI_i>$	$<BC_i, AS_i>$	$<BC_i, AC_i>$
IDC	$<IDC_i, RG_i>$	$<IDC_i, RI_i>$	$<IDC_i, RS_i>$	$<IDC_i, RC_i>$	$<IDC_i, Acad_i>$	$<IDC_i, AI_i>$	$<IDC_i, AS_i>$	$<IDC_i, AC_i>$
ODC	$<ODC_i, RG_i>$	$<ODC_i, RI_i>$	$<ODC_i, RS_i>$	$<ODC_i, RC_i>$	$<ODC_i, Acad_i>$	$<ODC_i, AI_i>$	$<ODC_i, AS_i>$	$<ODC_i, AC_i>$
CC	$<CC_i, RG_i>$	$<CC_i, RI_i>$	$<CC_i, RS_i>$	$<CC_i, RC_i>$	$<CC_i, Acad_i>$	$<CC_i, AI_i>$	$<CC_i, AS_i>$	$<CC_i, AC_i>$
EVC	$<EVC_i, RG_i>$	$<EVC_i, RI_i>$	$<EVC_i, RS_i>$	$<EVC_i, RC_i>$	$<EVC_i, Acad_i>$	$<EVC_i, AI_i>$	$<EVC_i, AS_i>$	$<EVC_i, AC_i>$
PRC	$<PRC_i, RG_i>$	$<PRC_i, RI_i>$	$<PRC_i, RS_i>$	$<PRC_i, RC_i>$	$<PRC_i, Acad_i>$	$<PRC_i, AI_i>$	$<PRC_i, AS_i>$	$<PRC_i, AC_i>$

To identify the correlations among the generated ranking tuples, Spearman and Kendall correlations are used. Figure 3 and Figure 4, respectively, display the Spearman and Kendall correlations for RGScore. Similarly, Figure 4 and Figure 5, respectively, display the Spearman and Kendall correlations for AcadScore. It is evident that for RG and Academia, the correlations achieved from highest to lowest are by IDC, PRC, BC, ODC, EVC, and CC. For Academia, IDC and PRC yield the exact Spearman correlation. Compared to the RGScore and AcadScore, the highest correlation achieved is 0.89 and 0.92. This denotes that the centrality top-k lists received from follower-following networks of RG and Academia match their own researchers' influence indicators by 89% and 92%, respectively.

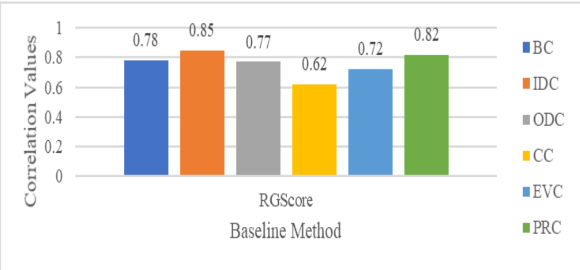


Figure 3: Spearman Correlation for RGScore

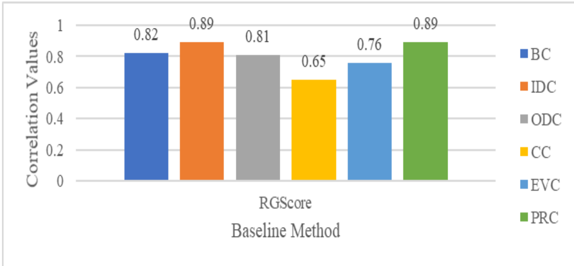


Figure 4: Kendall Correlation for RGScore

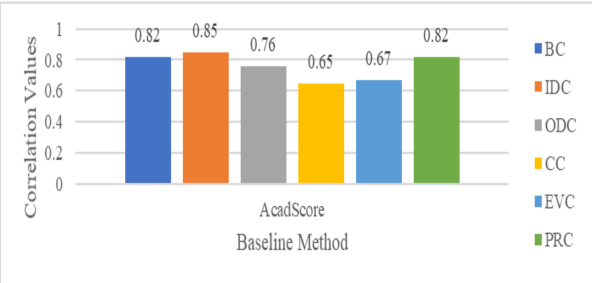


Figure 5: Spearman Correlation for AcadScore

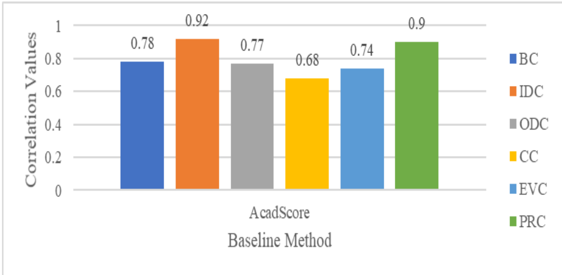


Figure 6: Kendall Correlation for AcadScore

For RG, Figure 7 and Figure 8 respectively display the Spearman and Kendall correlations for considered three global ranking lists. Similarly, For Academia, Figure 9 and Figure 10, respectively, indicate the Spearman and Kendall correlations for considered three global ranking lists. It is evident that for RG and Academia, the correlations achieved from highest to lowest in order by IDC, PRC, BC, ODC, EVC, and CC. IDC and PRC yield the same Spearman correlations for RG compared to Clarivate. IDC and PRC yield the same Kendall correlations for Academia as IDEAS and SSRN.

The centrality top-k lists received from follower-following networks of RG match up to 94%, 86%, and 92%, respectively, with the lists of IDEAS, SSRN, and Clarivate in general. Similarly, the centrality top-k lists received from follower-following networks of Academia match up to 94%, 88%, and 90%, respectively, with the lists of IDEAS, SSRN, and Clarivate in general. The highest correlation is achieved in both cases compared to IDEAS global ranks.

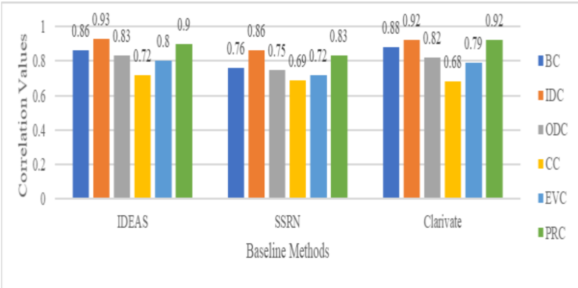


Figure 7: Spearman Correlation for Global Ranking in RG

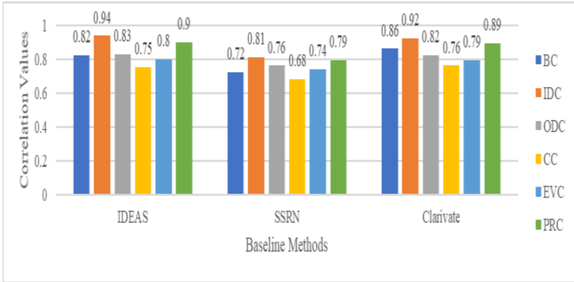


Figure 8: Kendall Correlation for Global Ranking in RG

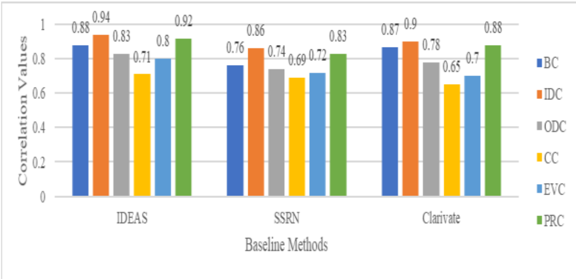


Figure 9: Spearman Correlation for Global Ranking in Academia

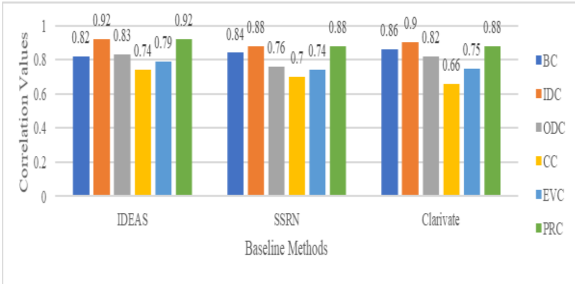


Figure 10: Kendall Correlation for Global Ranking in Academia

The efficient results of IDC and PRC suggest that the majority of matching top-k researchers have a huge number of followers, and the majority of those followers are directly or indirectly influential in a network. Moderate correlations in the case of BC suggest many researchers depend on one another to make connections with others. A wider set of researchers in the constructed networks has many links crucial to influence propagation. EvC and PRC also achieve moderate but overall low correlations than BC due to fewer researchers in each researcher's network. The lowest correlation of CC suggests that few researchers have ties to several influential researchers in the constructed network.

Discussion

The existing approaches in scientific influence identification majorly work with already established co-authorship and co-citation networks. Other networks like followers and followings of researchers are not explored in the literature. Social collaborations regarding followers and followings on digital scholarly platforms provide a novel paradigm in influence identification. In this age of high-tech progressions and advancements, it is essential to consider these social collaborations while gauging the influence of researchers. The proposed methodology provides a novel approach to treat scientific information as a connected network of researchers regarding followers and followings. The aim is to anatomize the impact of follower and following relations among researchers on influence identification.

For our analysis, we render scientific information from well-known scholarly platforms: ResearchGate (RG) and Academia. From the collected linked information, the follower-following networks of researchers are constructed. Standard centrality algorithms are implemented on constructed networks to identify the list of top-k researchers. The results are compared with the researcher's influence indicator provided by RG and Academia and recognized global ranking lists. The highest correlation achieved for RG is 89% and 94%

compared to RGScore and the global landscape, respectively. The highest correlation achieved for Academia is 92% and 94% when compared to AcadScore and the global landscape, respectively.

From the results, it is observed that for RG and Academia, the lists of identified influential researchers using followers-followings are highly correlated with the influence indicators of RG and Academia, respectively. Also, the generated lists have wide coverage compared to global ranking benchmarks. From the obtained results, it is perceived that like SNA, social collaborations in terms of followers and followings among researchers significantly impact influence identification in the scientific domain.

Conclusion

This research aims to identify the significance of followers and followings on researchers' influence identification. Two well-known digital scholarly platforms – RG and Academia, are analyzed for our aim. The results indicate the significant impact of social collaborations i.e., followers and following, on researchers' influence identification. This research and the obtained efficient outcome outset an epitome to consider SNA-inspired followers and followings in scientific influence identification.

In the future, the expertise of identified influential researchers can be analyzed to perform various scientific activities such as inviting these influential researchers for expert talks, to be reviewers in domain-specific conferences/journals, for research collaborations, and more. Follower-following-based researchers' networks can be explored for SNA-inspired applications such as recommendation systems, community detection, and clustering. In the future, the researcher's scholarly demographics on scholarly platforms can also be assembled with centrality algorithms to identify influence more accurately.

References

Alizade Zowj, H., Ghane, M.R. & Ehsanifar, F. (2019). Identifying information retrieval research trends using author co-citation network. *International Journal of Information Science and Management*, 17(2), 99-117. Retrieved from https://ijism.ricest.ac.ir/article_698301_76b4ccbfeac12f9015005979e782f9a5.pdf

Amjad, T., Daud, A. & Aljohani, N. R. (2018). Ranking authors in academic social networks: A survey. *Library Hi Tech*, 36(1), 97-128. https://doi.org/10.1108/LHT-05-2017-0090

Amjad, T., Daud, A., Akram, A. & Muhammed, F. (2016). Impact of mutual influence while ranking authors in a co-authorship network. *Kuwait Journal of Science*, 43(3), 100-109. Retrieved from https://journalskuwait.org/kjs/index.php/KJS/article/view/941/141

Amjad, T., Daud, A., Che, D. & Akram, A. (2016). MuICE: mutual influence and citation exclusivity author rank. *Information Processing & Management*, 52(3), 374-386. https://doi.org/10.1016/j.ipm.2015.12.001

Ansari, J. A. N. & Khan, N. A. (2020). Exploring the role of social media in collaborative learning the new domain of learning. Smart Learning Environments, 7(9). https://doi.org/10.1186/s40561-020-00118-7

Bai, X., Zhang, F., Hou, J., Lee, I., Kong, X., Tolba, A. & Xia, F. (2018). Quantifying the impact of scholarly papers based on higher-order weighted citations. *PLoS ONE*, 13(3), e0193192. https://doi.org/10.1371/journal.pone.0193192

Bhattacharya, S., & Sarkar, D. (2021). Study on information diffusion in online social network. In *Proceedings of International Conference on Frontiers in Computing and Systems* (pp. 279-288). Springer, Singapore.

Bibi, F., Khan, H. U., Iqbal, T., Farooq, M., Mehmood, I. & Nam, Y. (2018). Ranking authors in an academic network using social network measures. *Applied Sciences*, 8(10), 1824. https://doi.org/10.3390/app8101824

Desai, M., Mehta, R. G. & Rana, D. P. (2019). An empirical analysis to identify the effect of indexing on influence detection using graph databases. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(9S), 414-421. Retrieved from file:///C:/Users/drghane/Downloads/An_Empirical_Analysis_to_Identify_the_Ef.pdf

Desai, M., Mehta, R. G. & Rana, D. P. (2021). RGNet: The novel framework to model linked ResearchGate information into network using hierarchical data rendering. In *Advances in Machine Learning and Computational Intelligence* (pp. 37-45). Springer, Singapore.

Desai, M., Mehta, R. G. & Rana, D. P. (2022a). Anatomising the impact of ResearchGate followers and followings on influence identification. *Journal of Information Science*. https://doi.org/10.1177/01655515221100716

Desai, M., Mehta, R. G. & Rana, D. P. (2022b). ScholarRec: A scholars' recommender system that combines scholastic influence and social collaborations in academic social networks. *International Journal of Data Science and Analytics*, 16, 203–216. https://doi.org/10.1007/s41060-022-00345-w

Desai, M., Mehta, R. G., & Rana, D. P. (2023). Contextual analysis of scholarly communications to identify the source of disinformation on digital scholarly platforms. *Kybernetes*. https://doi.org/10.1108/K-07-2022-0998 (in print)

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*, 5(1), 187-203. https://doi.org/10.1016/j.joi.2010.10.008

Espinoza Vasquez, F. K., & Caicedo Bastidas, C. E. (2015). Academic social networking sites: A comparative analysis of their services and tools. IConference 2015 proceedings.

Gasparyan, A. Y., Nurmashev, B., Yessirkepov, M., Endovitskiy, D. A., Voronov, A. A. & Kitas, G. D. (2017). Researcher and author profiles: Opportunities, advantages, and limitations. *Journal of Korean Medical Science*, 32(11), 1749-1756. https://doi.org/10.3346/jkms.2017.32.11.1749

Hashmi, H. B. U. H., Kayani, H. U. R., Toor, S. K., Mansoor, A., & Raheem, A. (2020). The Impact of social media: A Survey. *International Journal of Scientific & Technology Research*, 9(12), 341-348. Retrieved from http://www.ijstr.org/final-print/dec2020/The-Impact-Of-Social-Media-A-Survey.pdf

Jordan, K. (2019). From social networks to publishing platforms: A review of the history and scholarship of academic social network sites. *Frontiers in Digital Humanities*, 6, 5. https://doi.org/10.3389/fdigh.2019.00005

Keller, R. C. A. (2019). Information and tips related to search engines like Google Scholar and other ways your work is and can be more visible on the net. A personal perspective. Retrieved from file:///C:/Users/drghane/Downloads/Search%20enginescomparison.pdf

Kesht-Karan, S., Ghane, M. R., & Danesh, F. (2021). Dimensions of the Scientific Collaborations of the Researchers Affiliated with Shiraz University of Medical Sciences. *International Journal of Information Science and Management (IJISM)*, 19(2), 31-48. Retrieved from https://ijism.ricest.ac.ir/article_698348_dcb524c97aae58ebd64bef5f3323f4e7.pdf

Khvatova, T., Dushina, S. & Nikolaenko, G. (2017, December). Do the online activities of scientists in social professional networks influence their academic achievements? *In European Conference on Management, Leadership & Governance* (pp. 217-226). Academic Conferences International Limited.

Liang, Z., Mao, J., Lu, K., Ba, Z. & Li, G. (2021). Combining deep neural network and bibliometric indicator for emerging research topic prediction. *Information Processing & Management*, 58(5), 102611. http://dx.doi.org/10.1016/j.ipm.2020.102475

Maia, L. F. M. P., Lenzi, M., Rabello, E. T. & Oliveira, J. (2019). Scientific collaboration in Zika: identification of the leading research groups and researchers via social network analysis. Colaborações científicas em Zika: identificação dos principais grupos e pesquisadores através da análise de redes sociais. *Cadernos de saude publica*, 35(3), e00220217. https://doi.org/10.1590/0102-311X00220217

Makkizadeh, F., Dehghan, A. & Mostafavi, E. (2020). Investigating association between social influence, productivity, and performance in co-author network of researchers in medical ethics. *Iranian Journal of Medical Ethics and History of Medicine*, 13(1), 240-252. Retrieved from https://ijme.tums.ac.ir/article-1-6177-en.pdf [in Persian]

Malinen, S. & Koivula, A. (2020). Influencers and targets on social media: Investigating the impact of network homogeneity and group identification on online influence. First Monday, 25(4). https://doi.org/10.5210/fm.v25i4.10453

Ortega, J. L. (2017). Toward a homogenization of academic social sites: A longitudinal study of profiles in Academia. edu, Google Scholar Citations and ResearchGate. *Online Information Review*, 41(6), 812-825. https://doi.org/10.1108/OIR-01-2016-0012

Samie, M. E., Biranvand, A., Rahmaniyan, S. & Varnamkhasti, E. M. (2022). The impact of the activity of industrial engineering researchers in various scientific-citation networks on improving their scientific authority status. *International Journal of Information Science and Management (IJISM)*, 20(1), 257-271. https://dorl.net/dor/%2020.1001.1.20088302.2022.20.1.14.7

Savov, P., Jatowt, A. & Nielek, R. (2020). Identifying breakthrough scientific papers. *Information Processing & Management*, 57(2), 102168. https://doi.org/10.1016/j.ipm.2019.102168

Valsesia, F., Proserpio, D. & Nunes, J. C. (2020). The positive effect of not following others on social media. *Journal of Marketing Research*, 57(6), 1152-1168. https://doi.org/10.1177/0022243720915467

Wan, Z., Mahajan, Y., Kang, B. W., Moore, T. J. & Cho, J. H. (2020). A Survey on Centrality Metrics and Their Network Resilience Analysis. *IEEE Access*, 9, 104773-104819.

Wang, Y., Ding, Z., Wei, X. X. & Long, J. (2021). Scholars influence evaluation based on time series heterogeneous network. In *2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (pp. 868-871). IEEE.

Wu, D., Fan, S. & Yuan, F. (2021). Research on pathways of expert finding on academic social networking sites. *Information Processing & Management*, 58(2), 102475. https://doi.org/10.1016/j.ipm.2020.102475

Wu, X. & Zhang, C. (2019). Finding high-impact interdisciplinary users based on friend discipline distribution in academic social networking sites. *Scientometrics*, 119(2), 1017-1035. https://doi.org/10.1007/s11192-019-03067-2

Yaghtin, M., Sotudeh, H., Mohammadi, M., Mirzabeigi, M. & Fakhrahmad, S. M. (2019). A correlation study of co-opinion and co-citation similarity measures. *International Journal of Information Science and Management (IJISM)*, 17(2),

Zhang, S., Zhao, D., Cheng, R., Cheng, J. & Wang, H. (2016). Finding influential papers in citation networks. In *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)* (pp. 658-662). IEEE.