



# Identifying Influential Scholars in Academic Social Media Platforms

Na Li, Denis Gillet

► **To cite this version:**

Na Li, Denis Gillet. Identifying Influential Scholars in Academic Social Media Platforms. Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Aug 2013, Niagara Falls, Canada. hal-00990287

**HAL Id: hal-00990287**

**<https://telearn.archives-ouvertes.fr/hal-00990287>**

Submitted on 13 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Identifying Influential Scholars in Academic Social Media Platforms

Na Li, Denis Gillet

École Polytechnique Fédérale de Lausanne (EPFL)  
1015 Lausanne, Switzerland  
{na.li, denis.gillet}@epfl.ch

**Abstract**—The emergence of social media has created new ways to publish scientific work, foster collaboration, and build professional connections in the research community. The rich data collected in social media platforms has provided new opportunities for assessing scholars’ impact other than the traditional citation-based approach. In this paper, we investigate the measures of scholars’ influence in academic social media platforms, taking both academic and social impact into account. A real-life dataset collected from Mendeley is used to apply different influence metrics. We first assess the academic influence of scholars based on the scientific impact of their publications using three different measures. Then we investigate their social influence using network centrality metrics. The experiments show that top influencers with high academic impact tend to be senior scholars with many coauthors. Furthermore, academic influence and social influence measures do not strongly correlate with each other, and thus scholars with high academic impact are not necessarily influential from a social point of view. Adding the social dimension could enhance the traditional impact metrics that only take academic influence into account.

**Keywords**—social media; ranking; influence; network centrality

## I. INTRODUCTION

The advent of social media has transformed Web users from passive consumers of information into active co-producers of social content. Next to its wide usage for social interactions among young people, social media has also been increasingly used in the academic community to support research activities. A growing number of academic social media platforms such as Mendeley and Academia.edu have enabled worldwide scholars to build professional connections, share research resources, and foster scientific collaboration.

As academic social media platforms bring researchers from diverse disciplines together, one of the major challenges is to identify influential scholars in specific research fields. Finding them can help young researchers to discover research trends, assist conference organizers to identify competent experts as potential committee members, and provide job recruiters with suitable job candidates. To detect influential scholars in a particular research field, both their *academic* and *social* influence should be taken into account. Academic influence is a measure of their scientific impact in a certain research field, and social influence is an indicator of their social impact in a specific research community. We believe that both measures are useful in different application scenarios. In the case of finding knowledgeable experts for someone who wants to learn a research topic, the expertise of the researchers in that field is

a key concern. However, to seek for influential scholars for the purpose of spreading a “call for papers” announcement, their social impact should be considered as vital criteria because the more influential they are from a social point of view, the wider their voices can spread in the community.

In this paper, we introduce three metrics for measuring scholars’ academic influence based on the popularity of their publications. In addition, we assess scholars’ social influence using the aggregation of network centrality metrics. Our experiments on a real-life dataset reveal interesting insights. First, top influencers with high academic impact tend to be senior scholars with many coauthors. Second, three academic influence metrics appear to identify different types of influencers. Last but not least, academic influence and social influence measures capture different dimensions, and thus scholars with high academic impact are not necessarily influential from a social point of view.

## II. RELATED WORK

A number of research efforts have been made to address the problem of assessing expertise. Some researchers [1], [2], [3], [4] adopted different variants of generative probabilistic models to obtain an expert’s knowledge on a given topic based on the documents that she is associated with. Most of the work attempted to investigate the association of author, document, and topic. Tang et al. [5] introduced an extended model by adding the dimension of conference where the document has been published. Priem et al. [6], [7], [8], [9] proposed a promising approach, altmetrics, to quantify researchers’ scientific impact using the metadata retrieved from Web 2.0 applications. There are also traditional approaches to assess scholars’ impact, such as h-index and g-index.

To measure the social influence of users, considerable work has been conducted. Influence of Twitter users has been measured in [10] using an extension of PageRank algorithm [11] where both topical similarity and link structure between users are taken into account. Other researchers [12], [13] also investigated the measures of influence in Twitter by tracking the diffusion of posts across topics and time. Moreover, Tang [14] described a UserRank algorithm that combines link analysis and content analysis techniques to identify influential users in an online healthcare social network. Additionally, there are trendy Web services that measure users’ influence using their social data, such as klout and peerindex.

In the context of academic social media platforms, we take advantage of the expertise measures and social influence mea-

asures in previous research, and investigate scholars’ domain-specific influence from both academic and social point of view. The proposed approach is discussed hereafter.

### III. PROBLEM DEFINITION AND CONTEXT

#### A. Problem Definition

A typical academic social media platform consists of a set of papers, and a set of users associated with the papers. Each paper is described by a list of metadata including title, abstract, authors, and conference or journal in which it has been published. Each user is also associated with a list of corresponding metadata such as research interests and biographic information. The association between a paper and a user is mostly derived from an authorship relation, i.e., the user has written the paper. However, there could be other types of associations such as sharing, reading, and commenting. The association represents how knowledgeable the user is about the paper’s content. In addition to the association between papers and users, there exist heterogeneous types of relations between users themselves. For instance, two users could directly connect to each other, join the same group, or coedit a paper, which lead to the relations of contact, membership, and co-authorship respectively.

In the context of an academic social media platform, our research aims at identifying influential scholars in a specific research field based on a search query. To determine who are the influential scholars in the field of the search query, two problems should be tackled: to detect from the entire network a group of scholars working in a specific field, and to rank them based on their influence in the corresponding field. There have been extensive research on finding people with knowledge on a given topic [1], [2], [3], [4], [5]. In this paper, we focus on studying the problem of ranking scholars according to their influence in a research field.

#### B. Dataset

The dataset used in this paper is collected from a real-life academic social media platform, namely Mendeley [15]. In Mendeley, users are allowed to create their profiles, upload their papers, add each other as contacts, and join research groups. Data were retrieved via the Mendeley API during the time period from 08/2012 to 10/2012. The raw dataset approximately consists of 1 million user profiles, 0.1 million papers, and different relations connecting the users, such as direct contact, membership, and co-authorship. Each entry in the dataset has associated metadata, including research interests, academic status, and biographic information of user profiles, and title, abstract, conference, and number of readers of papers. After filtering out the non-English entries using the Compact Language Detector library [16], the numbers of entries kept for the experiment are listed in Table I. The coauthorship relation represents that two scholars have written a paper together, the membership relation represents that two scholars have joined the same research group, and the contact relation indicates that two scholars have added each other into their contact lists.

To measure the influence of scholars, we need some minimum amount of information about them. Therefore, we only consider 226,240 scholars who have uploaded at least

TABLE I: Numbers of Entries in the Dataset

Entries	Numbers
Papers	74,627
Scholars	226,240
Coauthorships	13,873,740
Memberships	10,161,464
Contacts	148,582

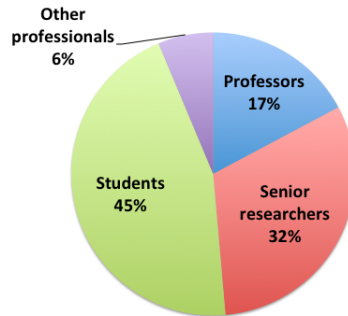


Fig. 1: Percentage distribution of scholars with different academic statuses.

one paper in Mendeley. The percentage distribution of these scholars with different academic statuses is shown in Fig. 1. Professors and senior researchers account for 49% of the population, and students (including PhD, master, and bachelor students) take up 45% of the total.

### IV. METHODOLOGY

The notion of influence is defined by the Oxford Dictionary as the capacity to have an effect on the character, development, or behavior of someone or something. In the fields of sociology, there have been various studies on influence patterns. An important theory, called minority influence [17], states that a minority of people can influence the majority by passing information and advice in the form of word-of-mouth communication. Gladwell [18] has described these most influential people as information specialists, who are central in the network and control the spread of information. In the context of academic social platforms, we focus on measuring scholars’ ability to affect others from both *academic* and *social* perspectives. From an academic point of view, an influential scholar should have popular publications that have been well received by peers. While from a social point of view, an influencer should be well connected, and be able to control the information flow in the network. The assumption is that the two measures capture different dimensions, and thus could be used to identify different types of influential scholars.

In this section, we address the metrics used to quantify academic and social influence. It is worth mentioning that the influence in this paper is inferred based on scholars’ activities in Mendeley and thus could be different from that in their real life. An influential scientist in real life might not be active in Mendeley. The objective of this paper is to give some insights on how influence can be measured by analyzing the Mendeley dataset. More precise influence measure that reflects

the real life could be conducted by aggregating data from multiple platforms such as CiteULike, Zotero, Academia.edu, and ResearchGate. However, that is not the focus of this paper.

### A. Measuring Academic Influence

Traditional metrics for academic impact, such as h-index and g-index, are mainly based on citation of scientific work. In the Web 2.0 era, social media have provided new cues to evaluate academic impact of scholars. In particular, academic social media platforms allow users to tag, annotate, bookmark, and rate scientific literature. Therefore, they collect a large amount of evaluative metadata on scientific content from individual users. Taraborelli [7] argued that such collaborative metadata at the individual level are hardly of interest, but on a large scale can provide information capable of outperforming traditional citation-based impact measures, in terms of coverage, speed, and efficiency.

In the case of Mendeley, the usage of papers is measured through readership, instead of citation. Readership indicates the real usage of papers, as people often read papers but do not necessarily cite them. A paper with a large number of readers is highly visible and thus should have a great scientific impact in a certain field. Additionally, a scholar's academic influence is mostly associated with the impact of her publications. To this end, we examine scholars' academic influence by aggregating the number of readers of their papers. We introduce three metrics to assess the academic influence of a given scholar S. Each of the three metrics can be used to capture a different type of influential scholars in terms of academic impact.

- **Total number of readers:** the total number of readers of all S's papers. It indicates the overall influence of S's work.
- **Maximum number of readers per paper:** the maximum number of readers among all S's papers. It is useful to identify scholars with few but highly visible papers.
- **R-Index:** an analogy for h-index [19]. A scholar with an R-Index of  $n$  represents that she has published  $n$  papers each of which has at least  $n$  readers. R-Index is an indicator of both the productivity and the impact of a scholar's published work.

### B. Measuring Social Influence

Scholars in Mendeley are connected to each other through different types of relations, such as direct contact, coauthorship, and membership. To measure the social influence of scholars, we construct a graph by considering the scholars as nodes and the relations as edges linking these nodes. Two nodes, representing two scholars, are linked by an edge if the two scholars are each other's contact, have written a paper together, or have joined the same research group.

As scholars in Mendeley work in diverse disciplines, those from different disciplines are often not connected to each other. As a consequence, the graph consists of disconnected subgraphs, each of which can be seen as a separate research community. Fig. 2 displays a fraction of the global graph derived from the Mendeley dataset. This fraction consists of four disconnected subgraphs. This reflects the community

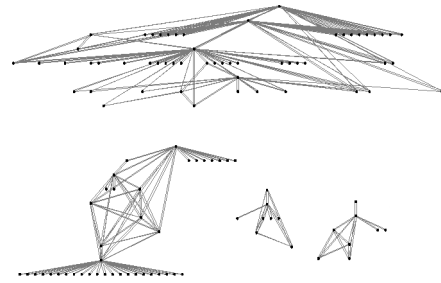


Fig. 2: Part of the global graph that consists of disconnected subgraphs.

clusters of researchers in real life. It is meaningless to measure the social influence of a scholar globally using the entire graph, because scholars mostly have a certain level of social impact in a specific research community, but not in all the communities. For instance, it does not make sense to evaluate the social impact of a physicist in the community of psychology. Therefore, we focus on measuring the *local social influence* of scholars in their corresponding research communities.

For a given search query, in order to find the corresponding research community, the following steps are conducted: we first concatenate the title, abstract, and conference or journal name of all the papers, and then filter out the noise by applying the Part-of-Speech Tagging technique [20]. After that, we use Latent Dirichlet Allocation topic model [21] to analyze the text and generate topic vectors for the papers. Furthermore, we select a set of most relevant papers to the given query by comparing the topic vector of the query and that of the papers. Cosine similarity measure [22] is used to compute the similarity between topic vectors. Finally, we produce a *local graph* by including the authors of the most relevant papers, and other scholars who are connected with those authors via contact, coauthorship, and membership relations. Those authors and the scholars connected with them are considered as nodes, while the relations linking the nodes are considered as edges in the local graph. This local graph is seen as the corresponding research community that matches a given query. It is used for inferring scholars' local social influence hereafter.

The influence a node has on other nodes in the network has been measured by different node centrality metrics [23], [24], [25]. The most widely used ones are the three measures formalized by Freeman [26]:

- **Degree:** the number of nodes that a target node is connected to, and it indicates the involvement of the node in the network.
- **Closeness:** the inverse sum of shortest distances to all other nodes from a target node. It measures how fast a node can reach other nodes in the rest of the network.
- **Betweenness:** the degree to which a node lies on the shortest path between two other nodes. A node with high betweenness is able to control the flow in the network.

In our work, to assess a scholar's social influence in a local graph, we aggregate these three measures by computing

the Euclidean norm of the three scores. Before aggregating the three scores, we first normalize them. Let  $D_i$ ,  $C_i$ , and  $B_i$  be the degree score, closeness score, and betweenness score of node  $i$  respectively. Let  $n$  be the number of nodes in the local graph. As in (1), (2), and (3), the normalized degree, closeness, betweenness scores,  $\overline{D}_i$ ,  $\overline{C}_i$ , and  $\overline{B}_i$ , are calculated using the maximum and minimum scores of all the nodes in the local graph. Finally, the social influence score of the node  $i$ ,  $SI_i$ , is obtained by calculating the Euclidean norm of the vector  $(\overline{D}_i, \overline{C}_i, \overline{B}_i)$ , as shown in (4).

$$\overline{D}_i = \frac{D_i - \min(\{D_1, \dots, D_n\})}{\max(\{D_1, \dots, D_n\}) - \min(\{D_1, \dots, D_n\})} \quad (1)$$

$$\overline{C}_i = \frac{C_i - \min(\{C_1, \dots, C_n\})}{\max(\{C_1, \dots, C_n\}) - \min(\{C_1, \dots, C_n\})} \quad (2)$$

$$\overline{B}_i = \frac{B_i - \min(\{B_1, \dots, B_n\})}{\max(\{B_1, \dots, B_n\}) - \min(\{B_1, \dots, B_n\})} \quad (3)$$

$$SI_i = \sqrt{(\overline{D}_i)^2 + (\overline{C}_i)^2 + (\overline{B}_i)^2} \quad (4)$$

## V. EXPERIMENT AND DISCUSSION

In this section, we apply the metrics of academic and social influence introduced in section IV on the Mendeley dataset, and elaborate our findings about the following questions.

- 1) Are top influencers senior scholars?
- 2) What types of influential scholars can we detect using the academic influence measures?
- 3) What are the correlations between the three academic influence measures: total number of readers, maximum number of readers per paper, and R-Index?
- 4) Do top influencers tend to have many co-authors?
- 5) Are top influencers in terms of academic impact also influential from a social point of view?

### A. Overall Characteristics of Academic Influence

Academic influence scores are computed using the three measures introduced in section IV. To understand the overall characteristics of scholars' academic influence, we plot the distributions of total number of readers, maximum number of readers per paper, and R-Index among scholars as illustrated in Fig. 3, 4, and 5. The figures reveal that three distributions roughly follow a power law pattern. This indicates that the majority of scholars have very low scores, while only a few top influencers have disproportionately high scores.

To investigate how top influencers distribute among different academic statuses, we select the top 1% scholars according to their R-Index scores, and plot the percentage distribution of different academic statuses in Fig. 6. The figure shows that the majority of top influencers (81%) are professors and senior researchers. Although students account for a large proportion (45%) of the overall population in the Mendeley dataset (as

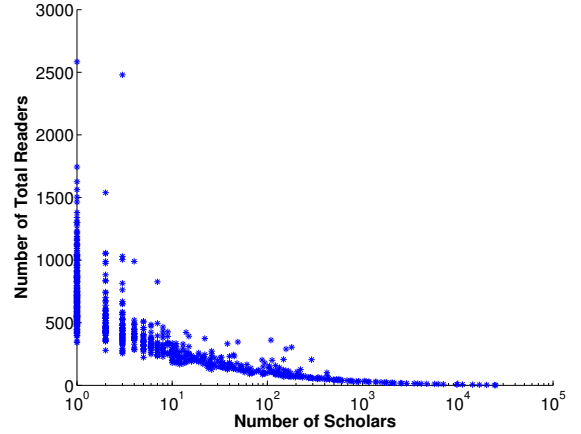


Fig. 3: Distribution of number of total readers among scholars.

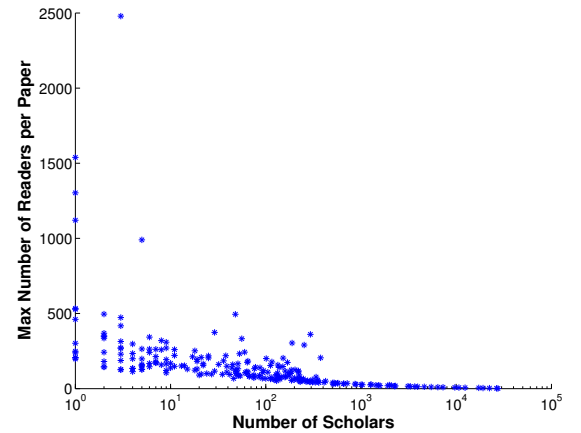


Fig. 4: Distribution of maximum number of readers per paper among scholars.

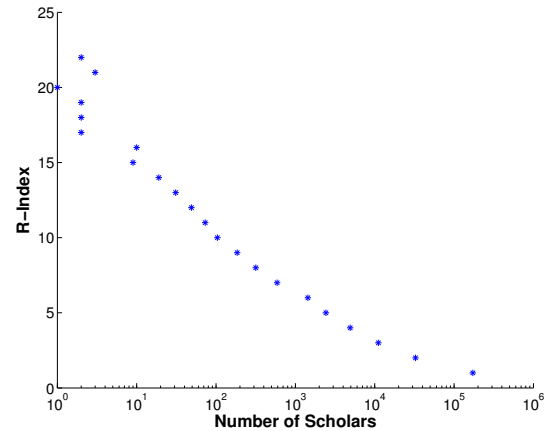


Fig. 5: Distribution of R-Index among scholars.

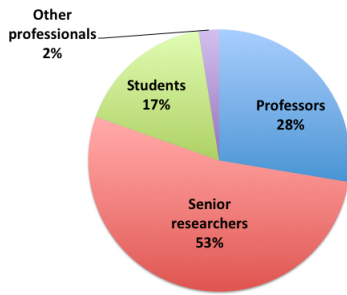


Fig. 6: Percentage distribution of top influencers with different academic statuses, selected by academic influence measure.

shown in Fig. 1), top influencers found by R-Index tend to be senior scholars. The percentage distributions according to total number of readers and maximum number of readers per paper provide similar results. The result is consistent with the fact that in real life senior scholars usually have higher academic impact than young scholars.

### B. Compare Three Academic Influence Measures

To figure out what types of influential scholars are detected by three different academic influence measures, we take a closer look at the top 100 scholars ranked according to three different measures respectively. Total number of readers is an indicator of a scholars’ overall academic impact as it represents the size of audience of one’s work. However, it tends to have a bias towards scholars with a big number of papers. For instance, scholar A has 15 papers, each of which has only 1 reader, and scholar B has one paper with 10 readers. Using the measure of total number of readers, scholar A gets a higher score than scholar B. However, scholar B should have a stronger scientific impact than scholar A as the visibility of A’s papers is generally low. Furthermore, the top influencers selected according to maximum number of readers per paper are scholars who have at least one great publication that gets lots of attention from others. Those scholars are mostly authors of fundamental works such as “Building Theories from Case Study Research” [27], “Error Bars in Experimental Biology” [28], and “Pattern Recognition and Machine Learning” [29]. Finally, R-Index is used to identify scholars whose work has a strong impact in terms of both productivity and quality. By observing the profiles of scholars with top R-Index scores, we discover that they are mostly active researchers who explore research trends and provide promising insights. Therefore, R-Index is useful for discovering trendy research topics. Three academic influence measures can be used to capture different types of influential scholars.

To get a deeper insight on how three measures differ from each other, we select the top 10% (22624 scholars) and top 1% (2262 scholars) influential scholars according to each measure, and compute the fraction of scholars who are common. Table II illustrates the overlap fraction of top influencers selected by three academic influence measures. One observation is that there is a high overlap between the set of scholars found by maximum number of readers per paper and total number of readers, while the overlap between other pairs of measures is very low. In addition, as the rank goes up, the overlap appears

TABLE II: Overlap of Top Influencers Selected by Three Academic Influence Measures

Overlap	Top 10%	Top 1%
R-Index vs Max reader count	24.44%	7.87%
R-Index vs Total reader count	45.52%	29.75%
Max reader count vs Total reader count	77.71%	62.51%

to decrease for all the pairs. This implies that three measures rank very different scholars in the top.

To investigate how three measures correlate with each other, we employ Spearman’s Rank Correlation Coefficient to calculate the correlation of each pair of measures. Spearman’s Rank Correlation Coefficient is a nonparametric measure of statistical dependence between two variables. It can vary between -1 and 1. -1 indicates a negative correlation, 0 indicates no linear correlation, and 1 indicates a positive correlation. We first rank scholars according to three academic influence measures respectively. Scholars with identical scores are assigned a rank equal to the average rank among them. For each pair of ranked list, we calculate Spearman’s Rank Correlation Coefficient  $\rho$  as defined in (5), where  $x_i$  and  $y_i$  denote the ranks of scholars according to two different academic influence measures, and  $n$  denotes the number of scholars in total.

$$\rho = 1 - \frac{6 \sum (x_i - y_i)^2}{n(n^2 - 1)} \quad (5)$$

Table III illustrates the correlation between each pair of academic influence measures. The maximum number of readers per paper strongly correlates (0.9641) with total number of readers. It suggests that scholars with at least one great publication tend to get lots of attention in general as well. However, R-Index does not seem to strongly associate (0.3455) with the maximum number of readers per paper. Given that scholars with very few readers have low ranks across all three measures, it could create a bias in the overall correlation. To this end, we also examine the correlation of three measures using only top 10% and top 1% scholars selected according to their R-Index scores. The results show that there is still a strong association between the maximum number of readers per paper and total number of readers (0.8976 and 0.8589). The correlation between R-Index and total number of readers increases from 0.5296 to 0.6736 as the sample size reduces. It reveals that scholars with high R-Index scores tend to attract a large number of readers in total. However, R-Index appears to be less associated with the maximum number of readers per paper even for top influencers. This suggests that scholars who produce lots of papers with relatively high quality do not necessarily have one particular publication with a high number of readers. For all the observations, the p-value is smaller than 0.001, which indicates that the correlations observed are statistically significant.

### C. Top Influencers and Number of Coauthors

To examine if top influencers have many coauthors, we calculate the average number of coauthors of all scholars, top 50%, 25%, 10%, 5%, and 1% scholars based on three different academic influence measures. As shown in Fig. 7, the average

TABLE III: Correlation between Three Academic Influence Measures

Correlation	ALL	Top 10%	Top 1%
R-Index vs Max reader count	0.3455	0.3839	0.4099
R-Index vs Total reader count	0.5296	0.6441	0.6736
Max reader count vs Total reader count	0.9641	0.8976	0.8589

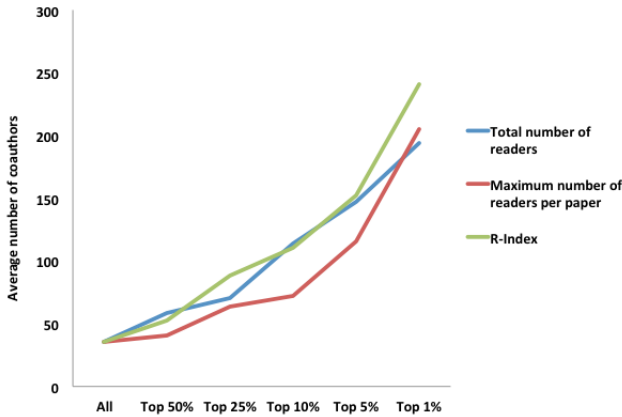


Fig. 7: Average number of coauthors of all scholars, top 50%, top 25%, top 10%, top 5%, and top 1% scholars.

number of coauthors is 35 when taking all scholars into account. As the rank increases, the average number of coauthors increases dramatically. The results are consistent across all influence measures. Therefore, we draw the conclusion that top influencers tend to have many coauthors.

#### D. Overall Characteristics of Social Influence

After analyzing the academic influence of scholars, we also compute the social influence scores of scholars within local graphs. Using the approach introduced in section IV, we first need to define the search queries based on which the local graphs are created. We find 5 experts specialized in 5 different research fields: human-computer interaction, control engineering, machine learning, computer network, and social media. Each expert is asked to specify 5 search queries associated with her research field. It is required that 5 queries vary from general to specific terms. After that, for each of the 25 queries, we detect the corresponding research community represented by a local graph. The social influence score (Eq. 4) of each scholar is obtained by calculating the Euclidean norm of the three normalized centrality scores (Eq. 1, 2, 3).

We select the top 1% influencers in each local graph based on scholars' social influence scores, and check their academic statuses. Compared to academic influence measures that tend to identify senior scholars, top scholars with high score in terms of social influence appear to be young researchers. The percentage distribution of top 1% influencers with different academic statuses is shown in Fig. 8. Fig. 8a, Fig. 8b, and Fig. 8c represent the communities derived from three search queries: Human-Computer Interaction, Recommender Systems, and Data Mining. The communities created based on the other 22 queries also give similar patterns.

TABLE IV: Correlation between Social and Academic Influence in the Community of Human-Computer Interaction

Influence Measures	Correlation
Centrality Norm vs Total reader count	0.0457
Centrality Norm vs Max reader count	0.0261
Centrality Norm vs R-Index	0.0599

TABLE V: Correlation between Social and Academic Influence in the Community of Recommender Systems

Influence Measures	Correlation
Centrality Norm vs Total reader count	-0.0115
Centrality Norm vs Max reader count	0.0154
Centrality Norm vs R-Index	-0.0212

TABLE VI: Correlation between Social and Academic Influence in the Community of Data Mining

Influence Measures	Correlation
Centrality Norm vs Total reader count	-0.0661
Centrality Norm vs Max reader count	-0.0768
Centrality Norm vs R-Index	-0.0492

#### E. Compare Academic Influence and Social Influence

In order to investigate if the top influencers in terms of academic impact are also influential from a social point of view, we again employ Spearman's Rank Correlation Coefficient to calculate the correlation between social influence and academic influence measures. The results derived from three search queries (Human-Computer Interaction, Recommender Systems, and Data Mining) are displayed in Table IV, V, and VI. There are 7818, 6428, and 3285 nodes (scholars) in three local graphs respectively. The results show that the correlation values stay close to zero for every pair of social influence and academic influence measures among all communities. The correlation values from 22 other search queries give similar results as well. Additionally, the p-value for all the correlation observations stays smaller than 0.001, representing that the correlations observed are statistically significant. This implies that there is no linear correlation between academic influence and social influence measures. In other words, scholars with high academic impact are not necessarily influential from a social point of view.

In summary, academic influence and social influence measures capture different dimensions, and thus can be used to identify different types of influencers that fit in different use cases. For the purpose of finding knowledgeable experts, academic influence measure should be considered as important criteria. But in the case of identifying well-connected scholars, who are able to control the information flow in a specific network, social influence measure should be used. Moreover, by aggregating the two measures, one can find influential scholars in terms of both academic and social impact.

## VI. CONCLUSION AND FUTURE WORK

Thanks to the wide usage of social media, Web users have created a large amount of data collaboratively. This provides

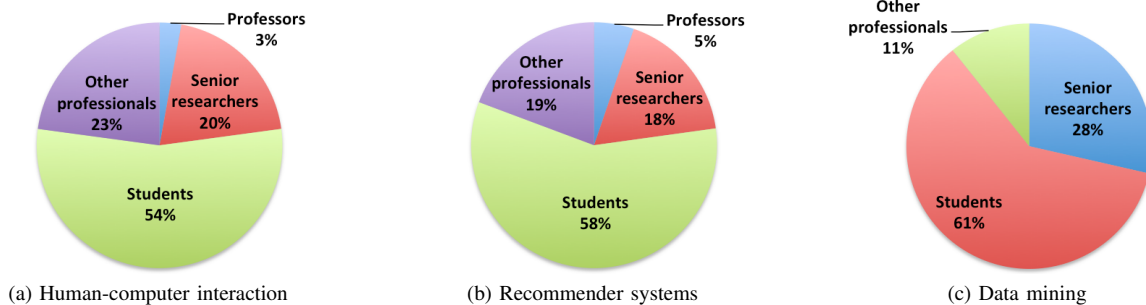


Fig. 8: Percentage distribution of top influencers with different academic statuses in the community of human-computer interaction, recommender systems, and data mining, selected by social influence measure.

new opportunities to assess scholars' impact other than the traditional citation-based approach. In the context of academic social media platforms, we investigate the influence measures of scholars from both academic and social point of view. Three readership-based measures of academic influence are introduced: total number of readers, maximum number of readers per paper, and R-Index. To measure social influence, we employ the norm of three network centrality metrics. We apply the influence measures on the Mendeley dataset and provide interesting findings. First, academic influence measures tend to discover senior scholars with many coauthors. Second, three measures of academic influence appear to identify different types of influencers. Third, scholars with high academic impact are not necessarily influential from a social point of view. Adding the social dimension could enhance the traditional impact metrics that mainly focus on academic influence.

As we conduct our experiments only on the Mendeley dataset, the results might not completely reflect scholars' influence in real life. Aggregating data from multiple platforms for more precise influence measures is planned in our research agenda. Moreover, other than readership, more metadata such as bookmarks, annotations, and ratings should be taken into account in the future. Also, regarding the local graph for computing social influence, we do not consider the weight of different relations among scholars in current work. Appropriate weighting schemes should also be further investigated.

## REFERENCES

- [1] K. Balog, L. Azzopardi, and M. De Rijke, "Formal models for expert finding in enterprise corpora," in *Proc. SIGIR*, 2006, pp. 43–50.
- [2] H. Fang and C. Zhai, "Probabilistic models for expert finding," in *Proc. European conference on IR research*, 2007, pp. 418–430.
- [3] M. Karimzadehgan and C. Zhai, "Constrained multi-aspect expertise matching for committee review assignment," in *Proc. ACM CIKM*, 2009, pp. 1697–1700.
- [4] D. Mimno and A. McCallum, "Expertise modeling for matching papers with reviewers," in *Proc. ACM SIGKDD*, 2007, pp. 500–509.
- [5] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Armetminer: extraction and mining of academic social networks," in *Proc. ACM SIGKDD*, 2008, pp. 990–998.
- [6] J. Priem and B. H. Hemminger, "Scientometrics 2.0: New metrics of scholarly impact on the social web," *First Monday*, vol. 15, no. 7, 2010.
- [7] D. Taraborelli, "Soft peer review: Social software and distributed scientific evaluation," in *Proc. COOP*, 2008, pp. 99–110.
- [8] C. Neylon and S. Wu, "Article-level metrics and the evolution of scientific impact," *PLoS biology*, vol. 7, no. 11, 2009.
- [9] P. Groth and T. Gurney, "Studying scientific discourse on the web using bibliometrics: A chemistry blogging case study," *WebSci10*, 2010.
- [10] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proc. ACM WSDM*, 2010, pp. 261–270.
- [11] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.
- [12] E. Bakshy, J. Hofman, W. Mason, and D. Watts, "Everyone's an influencer: quantifying influence on twitter," in *Proc. ACM WSDM*, 2011, pp. 65–74.
- [13] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proc. ICWSM*, 2010, pp. 10–17.
- [14] X. Tang and C. Yang, "Identifying influential users in an online healthcare social network," in *Proc. IEEE ISI*, 2010, pp. 43–48.
- [15] K. Jack, J. Hammerton, D. Harvey, J. J. Hoyt, J. Reichelt, and V. Henning, "Mendeleys reply to the datatel challenge," *Procedia Computer Science*, vol. 1, no. 2, pp. 1–3, 2010.
- [16] Chromium compact language detector. [Online]. Available: <http://code.google.com/p/chromium-compact-language-detector/>
- [17] W. Wood, S. Lundgren, J. A. Ouellette, S. Busceme, and T. Blackstone, "Minority influence: a meta-analytic review of social influence processes." *Psychological bulletin*, vol. 115, no. 3, pp. 323–345, 1994.
- [18] M. Gladwell, *The tipping point: How little things can make a big difference*. Little, Brown, 2000.
- [19] J. Hirsch, "An index to quantify an individual's scientific research output," *PNAS*, vol. 102, no. 46, pp. 16 569–16 572, 2005.
- [20] K. Toutanova, D. Klein, C. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proc. HLT-NAACL*, 2003, pp. 252–259.
- [21] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *JMLR*, vol. 3, pp. 993–1022, 2003.
- [22] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. Pearson Education India, 2007.
- [23] S. Borgatti, "Centrality and network flow," *Social networks*, vol. 27, no. 1, pp. 55–71, 2005.
- [24] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, vol. 32, no. 3, pp. 245–251, 2010.
- [25] M. Newman, "Analysis of weighted networks," *Physical Review*, vol. 70, no. 5, 2004.
- [26] L. Freeman, *The development of social network analysis: a study in the sociology of science*. Empirical Press Vancouver, 2004.
- [27] K. Eisenhardt, "Building theories from case study research," *Academy of management review*, vol. 14, no. 4, pp. 532–550, 1989.
- [28] G. Cumming, F. Fidler, and D. Vaux, "Error bars in experimental biology," *J. Cell Biology*, vol. 177, no. 1, pp. 7–11, 2007.
- [29] C. Bishop, *Pattern recognition and machine learning*. Springer, 2007.