

Influential Analysis in Micro Scholar Social Networks

Li Weigang^{1,2}, Icaro Araújo Dantas¹, Ahmed Abdelfattah Saleh², Daniel L. Li³

¹TransLab, Department of Computer Science, University of Brasilia, Brasilia, Brazil

²PPMEC, Department of Mechanical Engineering, University of Brasilia, Brasilia, Brazil

³Coleman Research, Raleigh, North Carolina-NC, USA

weigang@unb.br, icaro.a.dantas@gmail.com, ahmdsalh@yahoo.com, daniel.lezhi.br@gmail.com

Abstract

Scholar citation is a basic activity in scientific community. Some academic search engines have been developed in Web such as Google Scholar and Microsoft Academic Search. Efficient flexible querying method is essential for researchers to effectively follow trends within related topics of their research field. In this paper, we propose a procedure to construct *Micro Scholar Social Networks (MSSN)* from *Google Scholar* and then develop a querying and ranking method to find the influential researchers or articles in MSSN. An extension to the *Follow Model (Extended Follow Model)* is proposed in this paper and applied to describe the paper-citation and author-follow relationships. It is also coupled with different ranking algorithms, namely, *PageRank*, *AuthorRank* and *InventorRank* to study a MSSN in Air Traffic Management. The case study shows that *Extended Follow Model* is robust and efficient for ranking and mining a heterogeneous academic network. In spite the fact that study was done on *Google Scholar*, but the proposed data mining method is applicable for other academic search engines.

1 Introduction

With the development of Internet technology and applications, there are at least 114 million English-language scholarly documents accessible on the Web [Khabsa and Giles, 2014]. The term “*scholarly documents*” here refers to journal and conference papers, books, dissertations and theses, technical reports and working papers. The size of scholarly documents accessible through the web differs from one academic search engine to the other; *Google Scholar*¹, for example, comprises nearly 100 million scholarly documents and also available advanced search for general consulting.

¹ <http://scholar.google.com/>

Copyright © 2015 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

In Google Scholar, the most cited paper is “*A short history of SHELX*” [Sheldrick, 2007] with 49,792 citations. The authors who cited this paper have composed a special society or a network. Understanding the relations in this society is valuable to the researchers.

In this massive academic network, efficient querying models of academic search engines or databases is crucial for a researcher to conduct his research while following up the development trends in a specific research topic of particular scientific field.

There are two problems that should be deeply studied: 1) developing efficient method and system (in Web tool level) to construct *Micro Scholar Social Network (MSSN)* for an especial topic or field from large scholar social networks, such as *Google Scholar*, *Microsoft Academic Search*, *Web of Science* or others; 2) developing efficient mining algorithms to analyze this MSSN for scholar’s diversity objectives.

In literature, some research developed the mining methods of heterogeneous information networks [Sun et al. 2012]. Ahmedi et al. focused on the study of the property of the Co-authorship Networks [Ahmedi et al., 2011].

In recent years, many researches proposed solutions to these problems. Liu et al. [2005] demonstrated *AuthorRank*. *AMiner* has been developed by [Tang et al., 2008] as a scholar platform with the database and search interface. *W-entropy* was proposed to measure the influence of the members from social networks [Weigang et al., 2011]. Sandes et al. [2012] introduced the concept of *Follow Model* for the development of advanced queries on social networks. Du et al. [2015] demonstrated the way of analyzing importance of nodes in heterogeneous networks.

In this paper, *Extended Follow Model (EMF)*, an extension to the *Follow Model* presented by [Sandes et al, 2012], is proposed. EMF is applied to describe the paper-citation and author-follow relationships. It is also coupled with different ranking algorithms, namely, *PageRank*, *AuthorRank* and *InventorRank* to study a MSSN in Air Traffic Management. The case study shows that *Extended Follow Model* is robust and efficient for ranking and mining a heterogeneous academic network. In spite the fact that the MSSN used in this study was constructed using *Google Scholar*, but the study is applicable, as well, for other academic search engines.

2 Micro Scholar Social Networks (MSSN)

This section introduces the concept of *Micro Scholar Social Networks (MSSN)* and explains its basic elements and relations using Google Scholar as an example. In a MSSN, authors and publications are considered the main objects. However, in addition to authors, there are also editor of the journal, editor of the book, conference chairs. As for publications, there are journal papers, conference papers, books, book chapters, reports and others. In this study, we use paper to refer to all different types of publications, and author for all contributors. The specification of other objects will be considered in our future data mining studies.

The “*Micro Scholar*” term refers to a specific research field, while “*Social Network*” term refers to the network constructed from the relations between papers and authors of that field. Figure 1 shows the MSSN constructed from 165 papers and 249 authors of the “*Air Traffic Management (ATM)*” research field. As seen in figure 1, MSSN-ATM is constructed of directed graphs, whose vertices are papers (sub-graph a) or authors (sub-graph b) while the edges represent the relations among those elements.

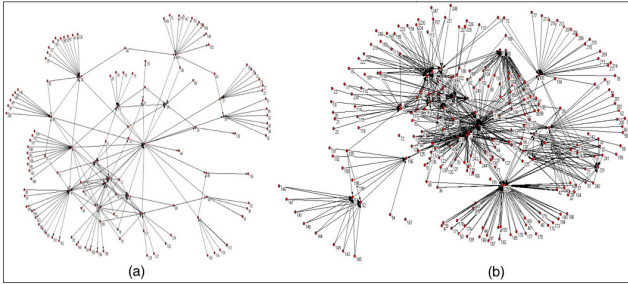


Figure 1: ATM-MSSN of Google Scholar; a) 165 papers and their citations; b) Follow relations of 249 authors.

2.1 Elements and Relations in MSSN

The basic elements in MSSN are papers, authors, venue and publishers. This paper focuses on the information related to paper and author. *Citation* and *co-authoring* are the core relations between papers and authors. As such, a MSSN is constructed using the citation relations among papers and authors, in addition to the co-author relations among different authors. These relations among MSSN elements can be explained as follows:

Relations between papers

- *Citation relation*: Citations of a paper p are those papers that were cited by p .
- *Cited In relation*: Cited Ins of a paper p are those papers that cited p .
- *Both-cited relation*: Two papers a and b are considered Both-Cited, in the case that paper a cites b and paper b cites a .

Follow relation between authors

Citation relation can be extended to describe the relation among authors. Where, when a paper cites another paper, in actuality, authors are simply citing authors with prior studies

relevant to their paper. As such, the follow relation, as per the *Follow Model* introduced by [Sandes et al, 2012], can be used to describe the citation relation between authors.

- *Followee (Citation relation)*: Citations of a paper p are those papers that were cited by p . The authors of these papers are *followee* of the authors of p .
- *Follower (Cited In relation)*: Cited Ins of a paper p are those papers that cited p . The authors of these papers are *followers* of the author of paper p .
- *R-Friends (Both-cited relation)*: Two papers a and b are considered Both-Cited, in the case that paper a cites b and paper b cites a . The authors of those two papers are *R-Friends*.
- *Self-Following (Self-Citation)*: If an author has cited one of his own papers.

Co-author relation between authors

Another important relation among authors is that of co-authorship. Where, for any particular paper, there are one or more authors. The relationship among those authors can be referred to as *co-author*. An author may be a co-author for several authors in one or more papers. This paper presents a weighing formula to assign a representative weight for each author depending on the *order of authorship* of different papers.

2.2 Types of MSSN

Scientific papers are characterized by multiple attributes (e.g. authors, venue, publish time, editor of the journal, editor of the book, conference chairs, etc.) in addition various relations among these attributes. As such, MSSN is considered a kind of heterogeneous information network that contains multiple types of elements and links [Sun et al., 2012], [Kim and Leskovec, 2012].

According to the nature of the elements used to construct a Micro Scholar Social Networks, MSSN’s can be divided into two types; *i) Homogenous MSSN*, with vertices (nodes) created using the same elements (i.e. papers or authors); and *ii) Heterogeneous MSSN*, where the networks vertices include different classes of elements (i.e. papers and authors) with their subsequent relations.

2.2.1 Constructing Homogenous MSSN

Regarding the relations in MSSN, there are citation relations (*citation*, *cited in* and *both cited*) between papers, as well as co-author relations and Follow relations (*followee*, *follower*, *r-friends*) between authors. These relations form three sets of homogenous MSSN’s.

A. ATM-MSSN-Papers

Figure 1(a) shows a MSSN of Air Traffic Management research topic, which is represented by a graph whose vertices are papers and the citation relation is its edges. The data reflected in this graph was collected from Google Scholar in January 26, 2015.

To create the graph, a citation relation matrix Pc is introduced. Pc is a square matrix of size $(N \times N)$, where N is the total number of papers in the ATM-MSSN. pc_{ij} is an element

of the Pc matrix, with $i, j= 1, 2, \dots, N$. As such, if a paper i cited a paper j then $pc_{ij} = 1$.

B. ATM-MSSN-Authors

Figure 1(b) shows a graph with follow relations among 249 authors. A close relation can be observed between the two subfigures, where, subfigure 1(b) is based on 1(a), the only difference is that one paper can be written by more than one author.

The author relation matrix A is introduced to create this graph. A is a square matrix of size $(M \times M)$, where M is the total number of authors in the ATM-MSSN. a_{ij} is an element of the A matrix, with $i, j= 1, 2, \dots, M$. such, a_{ij} represents the number of times an author i follows an author j .

C. ATM-MSSN-CoAuthors

Figure 2(a) shows the co-author relations among 249 authors. An author can be a co-author with more than one author.

The co-author relation matrix Ac is introduced to create this graph. Ac is a square matrix of size $(M \times M)$, where M is the total number of authors in the ATM-MSSN. ac_{ij} is an element of the matrix Ac , with $i, j= 1, 2, \dots, M$. such, ac_{ij} represents the number of papers in which authors i and j are co-authors.

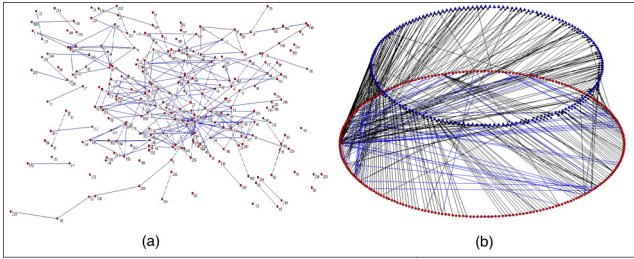


Figure 2: a) Homogenous ATM-MSSN-CoAuthors; b) Heterogenous ATM-MSSN.

2.2.2. Constructing Heterogeneous MSSN

Heterogeneous MSSN is formed of multiple elements and/or relations. As such, a heterogeneous MSSN can be constructed using multiple elements (e.g. papers and authors) and connected using a particular relationship and/or multiple relations for the same element class (e.g. co-author and follow relations for authors).

A. ATM-MSSN-Author-Paper graph

Figure 2(b) shows ATM-MSSN, which is a heterogeneous MSSN constructed using two classes of elements, papers and authors. Where, authors' nodes are represented by red circles while the papers' nodes are represented by blue triangles. The citation relation connects the paper nodes, while follow relation connects the authors. The two classes of nodes are then connected together using weighted co-authorship relation, where every author-paper edge has a weight that reflects the level of involvement (order) of this author in the authorship of that paper as seen in equation 1. According to figure 2(b), a matrix PA can be used to present the relations between authors and papers in the ATM-MSSN model. PA is

a matrix of size $(M \times N)$, with M is the total number of authors in the model, while N is the total number of papers. The matrix element pa_{ij} represents the weight of author i in writing the paper j . In other words, if a paper j has only one author I then $pa_{ij} = 1$. While for a paper i written by more than one author, then the value of pa_{ij} depends on the order of authors who wrote that paper [Du et al., 2015], the equation is modified as

$$pa_{ij} = \begin{cases} 1 - \sum_{i=2}^M pa_{ij} & \text{for } Order_j(i) = 1 \\ \frac{1}{Order_j(i) + (M - 1)} & \text{for } Order_j(i) > 1 \end{cases} \quad (1)$$

B. ATM-MSSN-Author-CoAuthor graph

Other type of heterogenous graphs is that constructed using the same class element (e.g. authors) and then connected by different types of relations (e.g. follow and co-author relations).

Therefore, constructing MSSN's graphs is the core step in building an efficient data-mining model that is used to perform complex analytical queries. As such, suitable heterogeneous and/or homogenous graph representation is selected to achieve the intended goal of the mining study. These graphs could be extended to include various attributes (conferences, journals, publishers, etc.) in order to develop a data-mining model that is capable of analyzing the relations among these attributes.

3 Extended Follow Model and Querying

In this section we extend the *Follow Model*, introduced by [Sandes et al. 2012], as the best way to model MSSN's and perform effective queries. In addition, *PageRank* and other ranking methods can be coupled with *Extended Follow Model (EFM)* to perform advanced queries.

3.1 Extended Follow Model (EFM)

MSSN can be best described in the form of directed graph $G = (V, E)$ where the vertices set V represents the papers and/or authors, while the directed edges $E: V \times V$ represents the relations between them. For heterogeneous MSSN, there are more types of relations; E can be noted as E_a, E_p or $E_a \cup E_b$.

The author follow relation $(v, u) \in E_a$ means that author v follows author u , and the graph $G_a = (V_a, E_a)$ presents the author relationship. While, $(a, b) \in E_p$ means that paper a cites paper b , and the graph $G_p = (V_p, E_p)$ presents the paper relationship.

As such, the *Extended Follow Model (EFM)* can efficiently describe the relations between the MSSN classes as mentioned in section 2.1. Where two authors can be related as either; followee, follower or r-friends while two papers are related as cited (followee), cited in (follower) or both-cited (r-friends).

Using these relations, one can construct data subsets for

big data querying. These data subsets can be extracted using the following functions:

$$f_{out}(u)=\{v|(u,v)\in E_a\}, \quad (2)$$

where, $f_{out}(u)$ is the followee function to present the subset, V_{out} , of all followees, v , of author u , $V_a \rightarrow V_{out}$, $V_a \subset V$; $|f_{out}(u)|$ is the number of the elements (authors) in the followee subset of author u ; $f_{out}^p(u)=\{p(v)|(u,v)\in E_a\}$, $p(v)$ is a value of the author v , such as the order in the subset or h-index of the author etc. $f_{out}^w(u)=\{w(v)|(u,v)\in E_a\}$, $w(v)$ is a weight value of the link between the authors u and v , such as the number of citations etc.

$$f_{in}(u)=\{v|(v,u)\in E_a\}, \quad (3)$$

where, $f_{in}(u)$ is the follower function to present the subset, V_{in} , of all followers, v , of author u , $V_a \rightarrow V_{in}$, $V_a \subset V$; $|f_{in}(u)|$ is the number of the elements (authors) in the follower subset of author; $f_{in}^p(u)=\{p(v)|(v,u)\in E_a\}$, $p(v)$ is a value of the author v , such as the order in the subset or h-index of the author etc. $f_{in}^w(u)=\{w(v)|(v,u)\in E_a\}$, $w(v)$ is a weight value of the link between the authors u and v , such as the number of citations.

$$f_r(u)=f_{out}(u) \cap f_{in}(u), \quad (4)$$

where, $f_r(u)$ is the r-friend function to present the subset, V_a , of all r-friends of author u , $V_a \rightarrow V_r$, $V_a \subset V$; $|f_r(u)|$ is the number of the elements (authors) in the r-friend subset of author; $f_r^p(u)=\{p(\cdot)\}$, $p(\cdot)$ is a value of the r-friends of author u , such as the order in the subset or h-index of the author etc. $f_r^w(u)=\{w(\cdot)\}$, $w(\cdot)$ is a weight value of the link between the author u and his f-friends, such as the number of co-author, etc.

With these basic definitions, *EFM* has both numeric $|f(\cdot)|$ and symbolic $f(\cdot)$ representations for more sophisticated relationships between users.

The *Follow Model* is also characterized by three properties: *reverse relationship*, *compositionality*, and *extensibility* [Sandes et al. 2012, Weigang et al., 2014]. Joining functions allow us to create many other relationship functions. For example: $f_{in}f_{out}(u)$ represents the followers of followees of u ; $f_{in}^2(u)$ represents the followers of followers of u ; $f_{out}^2(u)$ represents the followees of followees of u ; $f_r^2(u)$ represents the r-friends of r-friends of u .

In this research, beside *EFM* is applied as a querying method for MSSN-AUTHOR, it is also used in MSSN-PAPER by three functions: 1) $f_{out}(p)$ is a function to present all papers which are cited by paper p ; 2) $f_{in}(p)$ is a function to present all the papers which cited the paper p ; and 3) $f_r(p)$ is a function to present the paper p 's both-cited, which are papers that cited p and were cited by p .

3.2 Querying Google Scholar using Follow Model

EFM can be applied for querying in MSSN from Google Scholar or Microsoft Academic Search to satisfy the needs of users of these academic search engines. For example:

- An author of a paper p may be interested in the papers that cited his paper. In this case Follow model can be used, $P(p) = f_{in}(p)$, where $P(p)$ is the set of all papers that cited the paper p (Cited Ins).
- The same author may be interested in listing authors who follow him, therefore according to follow model, $A(p) = f_{in}(p)$, where $A(p)$ is the set of authors that cites paper p (i.e. followers of the author).
- One of many other interesting queries in the same context is obtaining the list of papers cited by those papers that cited p , or in other words, the list of followees of the author's followers $A(p)$. Follow model can simplify this query using $A(p) = f_{out}(f_{in}(p))$
- In addition an author may be interested in the set of papers that cites a particular paper p_i (cited by his paper as well) in addition to his paper p . As such, $f(p_i) = f_{in}(p) \cap f_{in}(p_i)$, where $P(p_i)$ is the set of all papers that cited his paper p in addition to p_i .
- Other users may be interested in the citations of paper p . Using Follow model, $P(p) = f_{out}(p)$, where $P(p)$ is the set of the papers that were cited by p .
- Other queries include: finding out the set of top- x (x may be 5 or more) papers, in terms of number of *Cited Ins*, for the papers cited by p . Follow model can present this query as $P(p) = f_{in}(f_{out}(p))_5$, where $f_{out}(p)$ is a function containing the set of papers (\mathbf{Pc}) that were cited by paper p ; $f_{in}(\mathbf{Pc})_5$ is a function generating the top 5 papers, cited papers of the set \mathbf{Pc} , that have the highest number of *Cited-Ins*.
- Also, finding out the set of top- x papers, in terms of the influence of the paper or the authors, for the papers that were cited by the papers that cited p . Follow model can present this query as $P(p) = f_{out}(f_{in}(p))_{10}$, where $f_{in}(p)$ is a function containing the set of papers (\mathbf{Pc}) that cited paper p ; $f_{out}(\mathbf{Pc})_{10}$ is a function generating the top 10 papers, that were cited by papers of the set \mathbf{Pc} , that have the highest influence. Different ranking algorithms, explained in the following section, can be used to determine the influence of papers and authors.

4 Influential Scholar Ranking Models

Ranking algorithms can be used to find the influential scholars in a MSSN. This section presents three ranking methods: *PageRank*, *AuthorRank* and *InventorRank*. All these models are presented in the form of *Extended Follow Model*.

4.1 PageRank and AuthorRank

PageRank [Brin and Page, 1998] can be presented in the form of EFM as follows:

$$PR(i) = (1 - d) + d \left[\frac{s(f_{in}^p(i))}{|f_{out}(f_{in}(i))|} \right] \quad (5)$$

where, i is an author, $f_{in}^p(i)$ is the set of the values of all authors who linked (followed) to i , and $s(f_{in}^p(i))$ is the sum of the values in this set. $|f_{out}(f_{in}(i))|$ is the number of followers of the followee of i .

On the other hand, *AuthorRank* is an indicator of the impact of an individual author in the network [Liu et al. 2005]. This algorithm is considered as an improvement of *PageRank* algorithm. Where, weights of nodes represent the number of times by which an author was co-author with another. Using EFM, *AuthorRank* can be represented as follows:

$$AR(i) = (1 - d) + d \left\{ s \left[\left(\frac{f_{in}^w(i)}{s(f_{out}^w(f_{in}(i)))} \right) \cdot f_{in}^p(i) \right] \right\} \quad (6)$$

Where, the $f_{in}^w(i)$ represents the weights of the followee of i and $f_{out}^w(f_{in}(i))$ represents the weights of the followers of the followee of i .

4.2 InventorRank

In the study of the data model of the inventor-ranking framework, [Du et al. 2015] demonstrates how to perform analysis of important nodes in heterogeneous networks. EFM is used to present one of the three rules described by [Du et al. 2015] for determining influential authors based on co-authorship. Where, highly ranked authors tend to co-author with other highly ranked authors. The first rule of *InventorRank* is determined using the following equation:

$$R_i(k) = \alpha_{ii} \{ [s(f_r^w(i) \cdot f_r^p(i) \cdot w_{ii}) + R_i(k)(1 - w_{ii}) \cdot |f_r(k)|] \} \quad (7)$$

where $R_i(k)$ is the rank of author k , $f_r(k)$ is the set of the all co-authors of k . See other rules in [Du et al. 2015].

4.3 Adjusting PageRank and Inventor Rank with SJR

Based on the fact that articles are usually published in some events or journals, González-Pereira [2010] proposes a way to classify the influence of a journal, based on the weight of citations and eigenvector centrality, in heterogeneous networks, this model is called *SCImago Journal Rank (SJR)*. Siebelt et al. [2010] e Macedo et al. [2010] are examples of some researchers that suggest using *SJR* to calculate the importance of an article. It is suggested that, like Du et al. [2015] used the classification of journals for classifying authors, it is possible to use periodicals classification as a mean for calculating journals or authors importance. As such this paper proposes adding journals weights in the previously defined classification algorithm. Thus, modified equations can be as follows:

PageRank:

$$SR(i) = PR(i) * SCImago_Rank \quad (8)$$

InventorRank:

$$R_p(j) = R_p(j) * SCImago_Rank \quad (9)$$

Note that *AuthorRank* was not modified to use this technique due to the fact that this method is restricted to classification of a network that contains only authors.

5 ATM-MSSN Ranking Case Study

In this section, the ranking case study in ATM-MSSN is described in details, to demonstrate how EFM can be coupled with three classification algorithms; PageRank, AuthorRank and InventorRank; to achieve effective querying. It also demonstrates the application of SJR to obtain influential rankings.

In figure 2(b), Heterogenous ATM-MSSN includes a total of 249 authors and 165 papers. Citation relationships between the papers (where an article cites others) and between the authors (an author cites / follows others) are illustrated. There is also another kind of relationship between authors and articles which is defined as the co-authoring relationship.

For all these tests, the parameters of the models are defined by the following pattern: For PageRank and AuthorRank, parameter d was set at 0.5. For the parameters, α_{ij} , α_{ip} e α_{pi} , the values were 0.4, 0.4 and 0.2 respectively. The w_{ii} and w_{ip} were both set as 0.5. The classification values were all obtained from SJR's site: <http://www.scimagojr.com/>. If there is no journal, it is considered as null classification and does not influence the ranking calculation.

5.1 Ranking authors and papers without SJR

The first result obtained is related to the author's ranking in ATM-MSSN. The ranking results are diferent due to the diferent characteristics of each model.

From Table 1, it is possible to observe that each profile at the top of the ranking reflects the characteristics that most affect the model.

Table 1. Author's Ranking for ATM-MSSN

Ranking	AuthorRank	InventorRank	PageRank
1	J F Butler	E Ferons	A R Odoni
2	H N Psaraftis	L Kang	D Trivizas
3	G Roger	J P Clarke	H N Psaraftis
4	Dear	B Delcairet	E P Gilbo
5	B G Sökkappa	W D Hall	Dear
6	L Gippo	H Idris	G Roger
7	M Cini	R Bhuva	D J Bertsimas
8	S S Patterson	AR Odoni	A Hormann
9	C F Dayl	R Hoffman	S S Patterson
10	G Andreatta	D J Bertsimas	B G Sökkappa

Author A R Odoni appears first according to PageRank because his papers have the most citations in ATM-MSSN, at

a total of 13 citations (followers). E Ferons, best ranked by InventorRank model, co-authored with A R Odoni, B Delcairet, H Idris, J P Clarke, W D Hall e B Delcairet, all well ranked authors in ATM-MSSN. This shows that although A R Odoni received many citations, a total of 89 (followers), it was not significant enough to affect his ranking because those that work with him are not the best ranked in the network. In case of AuthorRank, J F Butler received top ranking because his papers received most of the citations from A R Odoni, G Andreatta and B G Sokkapa.

Table 2, which lists the Top 5 papers by PageRank and InventorRank. For InventorRank, the paper's ranking is affected mostly by the authors' influence, and vice-versa. This correlation is not observed in PageRank.

Table 2. Top 5 papers by PageRank and InventorRank

Ranking	InventorRank	PageRank
1	Queuing model for taxi-out time estimation.	A dynamic programming approach for sequencing groups of identical jobs.
2	Collaborative decision making in air traffic flow management	Traffic Control and Transport Planning: A Fuzzy Sets and Neural Networks Approach.
3	Observations of departure processes at Logan airport to support the development of departure planning tools.	The flow management problem in air traffic control
4	Input-output modeling and control of the departure process of congested airports	The traffic flow management-rerouting problem in air traffic control: A dynamic network flow approach.
5	A comparison of formulations for the single-airport ground-holding problem with banking constraints	The dynamic scheduling of aircraft in the near terminal area.

5.2 Ranking authors and papers considering SJR

When analyzing the best papers ranked by InventorRank, it is observed that the position of the authors influence dictates the paper's ranking. PageRank does not yield similar results and relations, because it is not well structured to evaluate heterogeneous networks. The tests below analyze how these models operate when adding new characteristics to the network.

From table 3, it is possible to observe that the InventorRank differed in the ranking of a few others, which indicates that its results obtained in table 1 may not be representative. Table 3 shows that the weight of SJR for the journals where E Ferons' articles were published did not contribute to his rankings. Instead, A R Odoni, M O Ball and others are well ranked.

Comparing Table 2 to Table 4, we observe that PageRank is altered significantly when considering SJR. This was a result of either the quality of the journals where the articles were published, or in some cases where the articles were not

even published as they were simply graduate dissertations or internal reports within the institution.

Table 3 Top 10 Authors Ranking Considering SJR

Ranking	Name
1	A R Odoni
2	M O Ball
3	W D Hall
4	D J Bertsimas
5	E Ferons
6	R Hoffman
7	G Lulli
8	H Idris
9	J P Clarke
10	G L Nemhauser

On the other hand, the InventorRank demonstrates its robustness even with the addition of different features to the network. Its ranking of papers remained consistent and very similar to the first classification in table 2, where SJR was not taken into consideration.

Table 4. Top 5 Papers Considering SJR in Ranking

Ranking	InventorRank	PageRank
1	Collaborative decision making in air traffic flow management	A dynamic programming approach for sequencing groups of identical jobs.
2	Observations of departure processes at Logan airport to support the development of departure planning tools.	Applications of operations research in the air transport industry
3	Queuing model for taxi-out time estimation.	Stochastic and dynamic networks and routing.
4	The multi-airport ground-holding problem in air traffic control	The traffic flow management-rerouting problem in air traffic control: A dynamic network flow approach.
5	Collaborative Decision-Making in Air Transportation	The flow management problem in air traffic control

6 Conclusion

This paper presents the construction of *Micro Scholar Social Networks* (MSSN) for specific research topic using *Google Scholar* academic search engine. *Extended Follow Model* (EFM) was proposed as a comprehensive way of creating efficient data-mining model for querying homogenous and heterogeneous MSSNs. By means of the advantages, EFM was coupled with ranking algorithms to achieve a full querying and ranking models for scholarly documents of *Google Scholar*.

Comparing the results of such algorithms shows that *InventorRank* is a much more robust and accurate model, especially when considering the amount of information used for classification and ranking. Where, *InventorRank* can be easily adapted to adding new features to the network, in addition to its flexibility resulted from the ability to set the

degree of importance of each term of the algorithm. On the other hand, changes in the network directly affect classification ability of *Pagerank* algorithm.

It is worth mentioning that *Extended Follow Model* provides a simple and efficient means for representing several existing ranking models. Such representation facilitates the coding routine for developers, as complex equations are represented as an easy to understand and code algorithms.

This study also showed that the ranking system can be further modified to consider the level of influence of the journal where the paper is published. However, such modification requires caution as models such as PageRank is very sensitive to alterations, and may incorrectly classify articles of high quality.

Thanks to the mentioned authors and their papers in ATM-MSSN from Google Scholar.

References

- [Brin and Page, 1998] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, 30(1): 107-117, 1998.
- [Ahmedi et al., 2011] Lule Ahmedi, Lejla Abazi-Bexheti, Arbana Kadriu. A Uniform Semantic Web Framework for Co-authorship Networks. In *Proceedings of IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pages 958-965. Sydney, Australia, 2011. IEEE.
- [Du et al., 20015] Yong-ping Du, Chang-qing Yao, Nan Li. Using Heterogeneous Patent Network Features to Rank and Discover Influential Inventors. To appear in *Frontiers of Information Technology & Electronic Engineering*, 2015. doi:10.1631/FITEE.1400394
- [González-Pereira et al., 2010] Borja González-Pereira, Vicente P. Guerrero-Bote, and Félix Moya-Anegón. A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of informetrics*, 4(3): 379-391, 2010.
- [Hirsch, 2005] Jorge E. Hirsch. An index to quantify an individual's scientific research output. In *Proceedings of the National academy of Sciences of the United States of America*, 102(46), 16569-16572, 2005.
- [Khabsa and Giles, 2014] Madian Khabsa and C. Lee Giles. The Number of Scholarly Documents on the Public Web. *PloS one*, 9(5): e93949, 2014.
- [Kim and Leskovec, 2012] Myunghwan Kim and Jure Leskovec. Multiplicative Attribute Graph Model of Real-World Networks. *Internet Mathematics*, 8(1-2): 113-160, 2012.
- [Liu et al., 2005] Xiaoming Liu, Johan Bollen, Michael L. Nelson, Herbert Van de Sompe. Co-authorship networks in the digital library research community. *Information processing & management*, 41(6): 1462-1480, 2005.
- [Macedoa et al., 2010] Luciana G. Macedoa, Mark R. Elkinsb, Christopher Mahera, Anne M. Moseleya, Robert D. Herberta, Catherine Sherrington. There was evidence of convergent and construct validity of Physiotherapy Evidence Database quality scale for physiotherapy trials. *Journal of Clinical Epidemiology*, 63(8): 920-925.
- [Sandes et al., 2012] Edans F. Sandes, Li Weigang and Alba C. de Melo. Logical Model of Relationship for Online Social Networks and Performance Optimizing of Queries. In *Proceedings of Web Information Systems Engineering*, pages 726-736, Paphos, Cyprus, November 2012. Springer Berlin Heidelberg.
- [Sheldrick, 2007] George M. Sheldrick. A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography*, 64(1): 112-122, 2007.
- [Siebel et al., 2010] Michiel Siebel, Teun Siebelt, Peter Pilot, Rolf Bloem, M. Bhandari and Rudolf Poolman. Citation analysis of orthopaedic literature; 18 major orthopaedic journals compared for Impact Factor and SCImago. *BMC Musculoskeletal Disorders*, 11(1), 4, 2010.
- [Sun et al., 2012] Yizhou Sun, Jiawei Han, Charu C. Aggarwal, Nitesh V. Chawla. When will it happen?: relationship prediction in heterogeneous information networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*. Pages: 663-672, New York, NY, USA, 2012. ACM.
- [Tang et al., 2008] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990-998, August 2008. ACM.
- [Weigang et al., 2014] Li Weigang, Edans F. Sandes, Jianya Zheng, Alba C. de Melo, and Lorna Uden. Querying dynamic communities in online social networks. *Journal of Zhejiang University - Science C*, 15(2):81-90, 2014.
- [Weigang et al., 2014] Li Weigang, Jianya Zheng, and Daniel Li. W-entropy index: the impact of members on social networks. In *Proceedings of the Web Information Systems and Mining Conference*, LNCS 6987 (I), pages 226-233, Taiyuan, China, 2011. Springer Berlin Heidelberg.
- [Yu and Van de Sompel, 1965] Pingkang Yu and H. Van de Sompel. Networks of scientific papers. *Science*, 169: 510-515, 1965.