The 5th International Conference on
Computer Science & Education
Hefei, China. August 24–27, 2010

Measuring Credibility of Users in an E-learning Social Network

Wei Wei
Department of Computer and Information Science
Norwegian University of Science and Technology
wwei@idi.ntnu.no

Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong
king@cse.cuhk.edu.hk

***Abstract*—Learning Villages (LV) is an E-learning platform for people's online discussions and frequently citing postings of one another. It will greatly improve learning efficiency if credible users can be accurately identified in the E-learnning community. In this paper, we propose a novel method to rank credibility of users in the LV system. We first propose a k-EACM graph to describe the article citation structure in the LV system. The k-EACM graph not only describes the citation attitude between articles but also takes into account indirect citation links. We further build a weighted graph model k-UCM graph to reveal the implicit relationships between users hidden behind the citations among their articles. Finally, we design a graph based ranking algorithm, called Credible Author Ranking (CAR) algorithm, which can be applied to rank nodes in a graph with negative edges. We perform experiments on three simulated data sets. The experimental results show that our proposed method works well to rank credibility of users in the LV system. The comparison of results on average credibility of users in each set and composition of users in top N of three cases demonstrates that the CAR algorithm ranks users credibility in consistent with the predefined ground truth.**

I. Introduction

The focus of our study is on an E-learning platform in which users have to participate in online discussions and frequently cite postings of one another. If you were a new comer in an E-learning community, would you want to know that who are the credible users in the community? Do you believe that the user with a good score really do great contributions to the community? Most people will answer "yes" to the first question as people are more likely to be impacted by credibility. However, to the second question the answer will be a little bit tricky depending on the way the system scores its users.

As the internet becomes widespread, E-learning communities become more and more popular. The score of each user in the E-learning system not only stands for the reputation of the user but also is the rank of his/her impact in the community. To objectively score the users in an E-learning system is not a trivial task. Some E-learning systems simply count the number of articles of each user as the user's score in the community. This is not a good measure since user A who produces 20 expert articles has higher impact than user B who produces 40 trivial articles. Therefore when evaluating a user's performance in the community we have to take into account the quality of articles of the user.

Learning Villages (LV)[1] is deployed to some 3,000 junior secondary school students in Hong Kong, which is an E-learning platform for users' online discussions and frequently citing postings of one another. An important objective of LV is to train students the skills of structured discussions, which takes place in the form of article posting. During the discussions, users might write articles citing one another with explicitly selecting attitudes (support or against) towards the cited articles. In this paper, we propose a novel method to model relationships among users based on the citation structures of their articles and further design an algorithm to rank the credibility of users modeled in the graph. Citation structure analysis has been widely studied for academic publications (e.g., [25], [24], [1], [11], [8], [9], [10], [2], [24]). However, most of the existing works focus on analyzing citation structures of articles. Our work differs from existing works in that: (1) our work is to analyze citation structures of users, i.e., articles' authors, that is constructed based on citation structures among their articles; (2) unlike citation structure of articles in academics that only contains positive edges in the structure graph, the citation structure analyzed in our work contains both positive edges and negative edges, on which we propose an algorithm to rank credibility of users modeled in the graph. The contribution of this paper is as follows:

- We propose a k-EACM graph to describe the article citation structure in the LV system. The k-EACM graph not only describes the citation attitude between articles but also takes into account indirect citations.
- We propose a weighted graph model named k-UCM graph to reveal the implicit relationship between authors hidden behind the citations among their articles.
- We design a graph based ranking algorithm, named the Credible Author Ranking algorithm, which is applied to rank nodes in a graph with negative edges.

The rest of this paper is organized as follows. In Section II, we review some related works with citation analysis and explain the similarity and difference between bibliometric citation analysis and citation analysis in the LV system. In section III, we first introduce the idea of building ACM graph, k-EACM graph and map the k-EACM graph to generate k-UCM graph.

[1] http://www.learningvillages.com/en/index.php

Further we propose a graph-based ranking algorithm to rank nodes in k-UCM graph. In section IV, we give some effective examples as experiment to show how k-UCM and CAR algorithm work, followed by the Section V that concludes the paper.

II. Related work

Citation analysis in bibliometrics. Citation analysis has been widely used for bibliometrics ranking (e.g., [23], [17], [26], [21], [19], [6]). The ranking algorithms in bibliometrics can be separated into two classes depending on what the nodes represent in the graph model. In the first class, the nodes in the citation graph represent publications. The edge from node x to node y represents a citation from paper x to paper y. We refer to this class of ranking algorithms as publication-based ranking algorithms. The citation analysis in this class is very similar to web graph analysis [12], [22], [14]. The PageRank algorithm [20] and the HITS algorithm [13] are notable in this class. In the second class, the nodes in the citation graph represent collections. Collections may be journals (e.g., [16], [4], [28], [15], [27], [18], [3]), conferences (e.g., [5]), or even authors. The weighted edges in the citation graph represent the total number of citations made from one collection to another. We refer to this class of ranking algorithms as collection based ranking algorithms. The ISI impact factor [7] belongs to this class.

Authors ranking in the bibliometrics. The author ranking algorithms in bibliometrics should belong to the second class above. However, author ranking in the bibliometrics domain are seldom focused by research while there exist a large number of research works related to citation analysis focused on publication based ranking algorithms. In [24], an aggregate methods of ranking authors was proposed, but the method is still based on publication ranking. The difficulty in directly ranking authors in bibliometrics is that there is not an effective method to model authors in a graph.

Users ranking in the LV system. The user ranking process in the LV system is similar to but not the same as author ranking in bibliometrics. Both ranking are based on citation analysis. In the LV system, however when an article cites another article it has one of two attitude: support or against. In other words, if an article A is against an article B, even if there is a link from A to B, it should not improve B's authors's ranking in the community. In our work, we propose a method to generate a user citation model (k-UCM) graph derived from an extended article citation model (k-EACM) graph. Then we intorduce a new graph based ranking algorithm named Credible Author Ranking (CAR) algorithm to rank the users in the k-UCM graph.

III. Users Credibility Ranking in LV

In this section, we formulate the credibility ranking method in the Learning Villages system. In subsection III-A, we introduce a k-EACM graph, namely k-Extended Article Citation Model graph, derived from an ACM graph, namely Article Citation Model graph, to describe the relationships among all articles in the system. In subsection III-B, we map the k-EACM to the corresponding UCM graph, namely User Citation Model graph. In subsection III-C, we introduce our CAR algorithm, namely Credible Author Ranking algorithm, to rank the users in the constructed k-UCM graph.

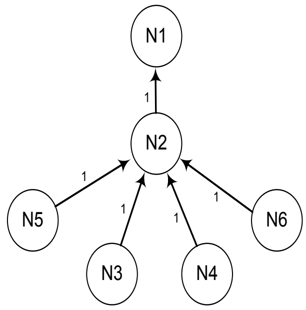


Fig. 1: An example of ACM graph model

A. Article Citation Graph

In this section, we first give a definition on an ACM graph and then define a k-EACM graph.

1) Article Citation Model: Articles in the LV system form a citation relation set. We use a graph to model the citation relationship. We define the ACM graph as:

Definition 1 (ACM Graph)**.** *ACM graph is a weighted directed graph $G(V, E, W)$. V is a set of vertices where each vertex represents an article in the system. E is a set of edges between the vertices: $E = \{(p, q)|p, q \in V$ and the corresponding article p citing the article $q\}$. W is a set of weighted values corresponding to edges belong to E. The weight value w_{pq} of an edge $E(p, q)$ is defined as:*

- *$w_{pq} = 1$ if the attitude of article p was specified as "support" to an article q.*
- *$w_{pq} = -1$ if the attitude of articles p was specified as "against" to an article q.*

An example of an ACM graph is shown in Fig. 1. The article $N2$ cites and supports article $N1$ and the article $N2$ is also cited and supported by articles $N3$, $N4$, $N5$, $N6$.

2) The k-Extended Article Citation Model: In an ACM graph, an edge between node $N2$ to node $N1$ stands for the explicitly citing from $N2$ to $N1$. However, the LV system does take into account even indirect citations. Let us look at another example of ACM graph in Fig. 2. Node $N7$ has one indegree similar to node $N1$ in Fig. 1. Does node $N7$ actually hold the same impact as node $N1$? It is obvious that node $N1$ indirectly impact more peers than node $N7$ does. It is unfair if we treat $N1$'s impact as same as $N7$'s without noting $N1$'s indirect impact. To take into account this kind of indirect impact from nodes, we extend the ACM graph and propose the k-EACM graph. Before we present the k-EACM model here, we will give the definition of k-distance link and the meaning of symbols we use in Definition 3.

Definition 2 (k-distance link)**.** *In an ACM graph of articles in the system, if there is a shortest path without considering*

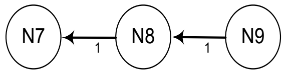


Fig. 2: Another example of ACM graph model

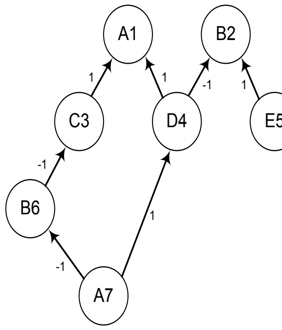


Fig. 3: A third example of ACM graph model

weight from node x to node y in k hops, we say that there is a k-distance link between the corresponding article x to the corresponding article y.

We also have the following notations defined as:

- R_{pq}represents one shortest route from node p to node q in the graph.
- $\{R_{pq}\}$ represents a set of all shortest routes existing between p and q in the graph.
- $edges(r)$ represents a set of all edges in a route r ($r \in R_{pq}$).
- w_{eg} represents the weight value of an edge eg in ACM graph.
- $|\{R_{pq}\}|$ represents the number of elements in the set $\{R_{pq}\}$.

According to Definition 2, the node $C3$ in Fig. 3 has a 1-distance link to $A1$. And the node $A7$ has a 2-distance link to $A1$. Now we present the definition of a k-Extended Article Citation Graph as below.

Definition 3 (k-EACM)**.** *A k-EACM is a weighted directed graph $G^{'}(V, E^{'}, W^{'})$ based on ACM graph (V, E, W). V is a set of vertices where each vertex represents an article in the system. $E^{'}$ is a set of edges between the vertices: $E^{'} = \{(p, q)|p, q \in V$and the corresponding articles p, q share a l-distance link $(l < k)\}$. W is a set of weighted values corresponding to edges belong to E. The weight value kw_{pq} of an l-distance link edge (p,q) in graph k-EACM is defined as:*

$$kw_{pq} = \begin{cases} w_{pq} & l = 1 \\ \frac{1}{|\{R_{pq}\}|} \sum\limits_{\forall r \in R_{pq}} (1 - \frac{l-1}{k}) \prod\limits_{\forall edge \in edges(r)} w_{edge} & l > 1 \end{cases} \quad (1)$$

Actually, the ACM graph model in Definition 2 is a 1-EACM graph model. Here we give another example to explain the k-EACM graph model. In Fig. 3, we let a new node B9 come and join to be against A7. In Fig. 4a, we get an updated ACM graph of Fig. 3 which is also a 1-EACM graph model. The Fig. 4b shows a 2-EACM graph model extended from Fig. 4a by explicitating all the 2-distance links between nodes. The Fig. 4c shows a 3-EACM graph model extended from Fig. 4a by expliciting all the 2-distance links and 3 distance links.

B. User Citation Model

The citation links between two authors is not explicit as their articles. However, the citation structure among articles can reflect the implicit citation links among their authors. In this part, we formulate how we map k-EACM graph to the corresponding k-UCM graph.

Definition 4 (Author Citation Link)**.** *All the articles in our system form a k-EACM graph, we say the author A has an author citation link to author B if and only if at least one article of author A has an edge to one article of author B in the k-EACM.*

The notation will be used in Defenition 5 are defined as:

- $Atk(u)$ represents a set of all nodes p in k-EACM and ps author is u.
- $Cby(p)$ represents a set of all nodes q and (p, q) is an edge in k-EACM articles graph.

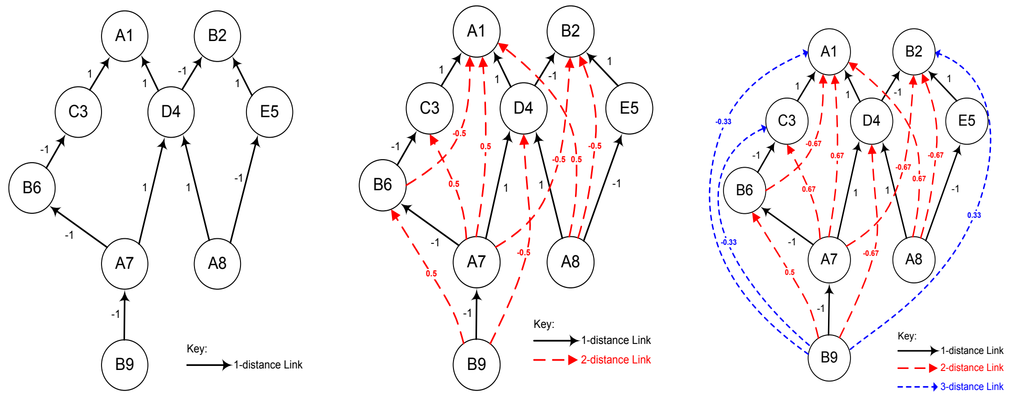
Definition 5 (k-UCM Graph)**.** *An k-UCM graph is a weighted directed graph $G^{''}(V^{'}, E^{''}, W^{''})$ derived from k-EACM graph $G^{'}(V, E^{'}, W^{'})$ where $V^{'}$ is a set of all the users in the system. $E^{''}$ is a set of edges between the vertices: $E^{''} = \{(u, v)|u, v \in V^{'}$ and the corresponding user u, v have an Author Citation Link$\}$. $W^{''}$ is a set of weighted values corresponding to edges in $E^{''}$. The weight value aw_{uv} of an edge (u, v) in graph k-UCM is defined as:*

$$aw_{uv} = \sum_{\forall p \in Atk(u)} \sum_{\forall q \in \{Cby(p) \bigcap Atk(v)\}} aw_{pq}$$

In order to explain how we map an k-EACM graph to the corresponding k-UCM users graph, in Fig.4a we suppose that, the nodes $A1$, $A7$, $A8$ is produced by a user A,the nodes $B2$, $B6$, $B9$ is produced by a user B, the node $C3$ is produced by a user C, the node $D4$ is produced by a user D, and the node $E5$ is produced by a user E. We give the corresponding 1-UCM, 2-UCM, 3-UCM in Fig. 5.

C. Credible Author Ranking Algorithm

After we have constructed a k-UCM graph to describe the users citation relationship, we still lack an effective graph based ranking algorithm to rank the users in the graph. In this subsection, we propose a graph-based ranking algorithm, named Credible Author Ranking (CAR) algorithm. Unlike the Pagerank algorithm [20] and the HITS algorithm [13], the main contribution of the CAR algorithm is that it allows to operate on graph such as k-UCM with negative value weight edges. A positive weight edge between two nodes in k-UCM indicates that the citing node support to the cited node. A negative weight edge between nodes in k-EACM indicates that the citing node is against to the cited node.

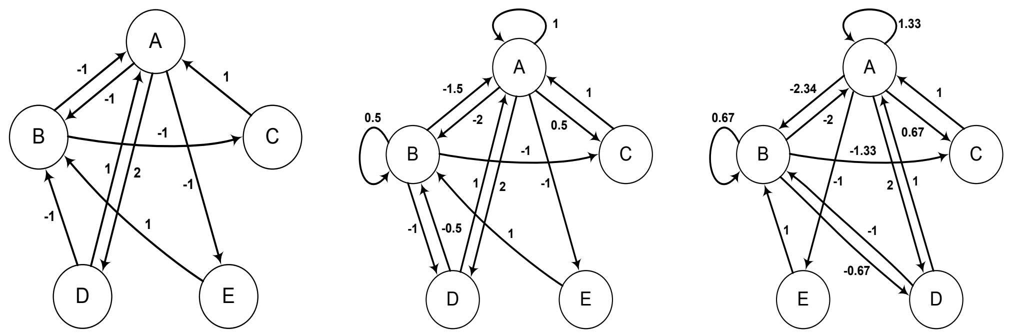


(a) An example of 1-EACM graph

(b) 2-EACM graph of extended from (a)

(c) 3-EACM graph of extended from (a)

Fig. 4: EACM Graphs



(a) 1-UCM corresponding to Fig. 4a

(b) 2-UCM corresponding to Fig. 4b

(c) 3-UCM corresponding to Fig. 4c

Fig. 5: UCM Graphs.

We think of two kinds of attributes of a user in the LV system: canonical attribute and trouble attribute. Therefore, we associate two values $x^{<u>}$ and $y^{<u>}$ with a user u respectively to represent u's canonical attribute weight and trouble attribute weight. If a user's canonical attribute weight is high, it means that most of this user's behavior in LV system is canonical or reasonable to the community. If a user's trouble attribute weight is high, it means that this user always gives trouble to the community. We also believe that canonical attribute weight and trouble attribute weight exhibit a mutually reinforcing relationship: a user A with high canonical attribute weight is a result of A supporting other users with high canonical attribute weights and also going against other users with high trouble attribute weights. Similarly, a user B with low trouble attribute weight is a result of B being supported by users with high canonical attribute weights and being gone against by users with high trouble attribute weights.

In the CAR algorithm, we associate each user u in the k-UCM graph with a canonical attribute weight $x^{<u>}$ and a trouble attribute weight $y^{<u>}$. The $x^{<u>}$ and $y^{<u>}$ for all every users in k-UCM graph respectively form a vector $\{x^{<u>}\}$ and a vector $\{y^{<u>}\}$. Before we process with $\{x^{<u>}\}$ and $\{y^{<u>}\}$, we define two operations denoted by α and β as:

- The α operation updates the x-weights:

$$x^{<u>} \leftarrow \sum_{v:(u,v)^{[1]}\in E} |aw_{uv}| \times x^{<v>} + \sum_{v:(u,v)^{[-1]}\in E} |aw_{uv}| \times y^{<v>}$$

- The β operation updates the y-weights:

$$y^{<u>} \leftarrow \sum_{v:(v,u)^{[1]}\in E} -|aw_{vu}| \times x^{<v>} + \sum_{v:(v,u)^{[-1]}\in E} -|aw_{vu}| \times y^{<v>}$$

The notations used above are explained here:

- E represents a set of all edges in k-UCM graph.
- $(u,v)^{[1]} \in E$ means that edge (u,v) belongs to E and node u has a support citation link to node v.
- $(u,v)^{[-1]} \in E$ means that edge (u,v) belongs to E and node u has an against citation link to node v.
- aw_{vu} represents that weight of edge (u,v) in E.

To compute the reinforcing equilibrium values for canonical attribute weights and trouble attribute weights of each users in k-UCM graph, the CAR algorithm does the same iterative process as HITS in Algorithm 1.

After this iterative process is done, the $x^{<u>}$ and $y^{<u>}$ weight values with each user will converge to their equilibrium values. Considering the $x^{<u>}$ representing a user u's canonical attribute and $y^{<u>}$ representing a user u's trouble attribute, we introduce $z^{<u>}$ to represent the user u's performance and simply define $z^{<u>}$ as: $z^{<u>} = x^{<u>} - y^{<u>}$.

In Table I, we show the result of ranking nodes in a 3-UCM graph of Fig. 5c by the CAR algorithm. From the table, we

Algorithm 1 Iterative Process in CAR

Iterate(V,t)
V: a set of nodes in k-UCM graph
t: a naturall number
Let w denote the vector $(1, 1, 1, ..., 1) \in R^n$.
Set $x_0 := w$.
Set $y_0 := w$.
for $i = 1, 2, ..., t$ **do**
  Apply α operation to (x_{i-1}, y_{i-1}), obtaining new x-weights $x_i^{'}$.
  Apply β operation to $(x_i^{'}, y_{i-1})$, obtaining new y-weights $y_i^{'}$.
  Normalize $x_i^{'}$, obtaining x_i.
  Normalize $y_i^{'}$, obtaining y_i.
end for
END
return (x_t, y_t).

TABLE I: Result of ranking nodes in a 3-UCM by CAR

Order	Node	Perform	Canonical	Trouble
1	A	0.39	0.43	-0.36
2	D	0.18	0.10	-0.28
3	C	0.06	0.02	-0.1
4	E	-0.02	-0.02	0.02
5	B	-0.33	-0.41	0.23

can observe that the node A is with highest canonical attribute value and lowest trouble attribute value. The performance value of the node A is also the best. It is reasonable with our intuition from Fig. 5c.

IV. Experiments

The evaluations of credibility ranking algorithms are always challenging tasks, as we do not have enough ground truth of credibility of each users and furthermore the performance among users are not easy to be compared numerically. In order to clearly show the effectiveness of our proposed method, we design a simulation program to simulate users' behaviors in the LV. In this way, we can easily and clearly predefine the ground truth of each user's credibility to test the efficiency of CAR algorithm. In this section, we will present the experiments of CAR algorithm on data generated from a simulation grogram.

A. Simulation Program Design

In the designed simulation program, we simulated 300 users behavior for 100 cycles. Among 300 users, there are predefined 50 bad users with less credibility in set B, 200 average users with average credibility in set A and 50 good users with high credibility in set G.

In every cycle, each user randomly publishes an article. The new published article may cite to existing articles with probability of θ using "support citing" and probability of $(1-\theta)$ using "against citing". The value of θ depends on the types of sets in which the citing article's author and the cited article's author respectively are. For example, if an article is produced by an author in set G, it may be supported with high probability, but if an article is produced by an author in set B, it may be against with high probability. We define a 3×3 matrix named BAG as follows to describe the θs between authors in two types of sets.

$$\begin{vmatrix} \theta_{BB} & \theta_{BA} & \theta_{BG} \\ \theta_{AB} & \theta_{AA} & \theta_{AG} \\ \theta_{GB} & \theta_{GA} & \theta_{GG} \end{vmatrix}$$

In the above matrix, the users in G set have three parameters θ_{GB}, θ_{GA} and θ_{GG} which respectively indicate support probabilities to users in B set, A set and G set. The users in other two types of set have the same support probability parameters. In this way, different settings of the BAG matrix will produce different citation relationship data for experiments on the CAR algorithm.

B. Evaluation Metrics

The **Average Credibility of Users in each set** implies the credibility level of users in the set. In order to present the average value in a normalized manner, we normalize the credibility value of all users into the range of $[0, 1]$. Intuitively, we expect that, after ranking by **CAR** algorithm the value of average credibility of users in G set will be much higher than the value of average credibility of users in B set and the value of average credibility of users in A set will be between the two.

The **Composition of Users in Top** N examines the proportion of users in each set in constituting the top N users which implies the quality of users in top N. The quality of the top N users is not only decided by the ranking algorithm but also by the whole users' quality in the system. For example, if the average users are not discriminative to tell what is truth knowledge to support the articles from the real credible users and consequently the real credible users will not be recognized, no ranking algorithms can rank the real credible users at the top.

The **Precision of the algorithm** is the proportion of users in set G out of the top N, i.e.,

$$Precision = \frac{N_G}{N},$$

where N_G is the number of users in set G which are ranked in the top N.

The **Recall of the algorithm** is the fraction of users in set G which have been ranked in in the top N users, i.e.,

$$Recall = \frac{N_G}{|G|},$$

where $|G|$ is the number of users in set G.

C. Experimental Results

We perform experiments on the data from simulation programmes with three different settings of BAG matrix which are designed as follows:

From Table II, we can see the three BAG matrix define three different situations in the simulation process. The diagonal

TABLE II: BAG matrix for 3 simulations

BAG-I	BAG-II	BAG-III
$\begin{vmatrix} 1 & 0.2 & 0 \\ 0.2 & 0.5 & 0.8 \\ 0 & 0.8 & 1 \end{vmatrix}$	$\begin{vmatrix} 0.7 & 0.8 & 0.1 \\ 0.8 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.9 \end{vmatrix}$	$\begin{vmatrix} 1 & 0.5 & 0 \\ 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 1 \end{vmatrix}$

elements in the matrix indicate the users support probability to users in the same set. The elements in the second row indicate that average users's support probability to users in B, A and G sets.

- **A rational case.** BAG-I implies a rational case of the online discussion. The predefined ground truth in the simulation indicates that the users in set B and G are 100% consistent with users from the same set, and the users in set A are discriminative to support to users in set G with much more probability than users in set B. The simulated case set by BAG-I is rational and consistent with the expected function of the online discussion platform.
- **A confusing case.** BAG-II implies a confusing case of the online discussion. The predefined ground truth in the simulation indicates that the users in set B and G are high consistent with users from the same set, but the users in set A are confused to support to users in set B with much more probability than users in set G. The simulated case set by BAG-II is confusing and opposite to the expected function of the online discussion platform.
- **A controversial case.** BAG-III implies a controversial case of the online discussion. The predefined ground truth in the simulation indicates that the users in set B and G are 100% consistent with users from the same set, but the users in set A are not clever enough to tell credible users which makes the online discussion in controversy.

We perform experiments on data from simulated cases with the above three BAG settings. We model the users citation relationship with a 3-UCM graph and rank users' credibility in the 3-UCM graph with CAR algorithm. For the rational case (BAG-I) and the confusing case (BAG-II) the iterative process of the CAR algorithm converged very fast and the credibility of each users are computed. For the controversial case (BAG-III), the iterative process do not converge and the algorithm terminated with assigning every users with equal credibility.

*1) **Average Credibility of Users in each Set**:* The average credibility of users in the sets G, A and B are computed and shown in Fig. 6.

From Fig. 6, we can see that in the rational case (BAG-I) the average credibility of users in set G are ranked much higher and the average credibility of users in set B are ranked much lower. However, in the confusing case (BAG-II), because of the average users confound right and wrong, the average credibility of users in set A are ranked highest and the average credibility of users in set B are higher than users in set G. In the controversial case (BAG-III), the average credibility of users in each set is equal and the online discussion need more discriminative users to join to help recognize credible users.

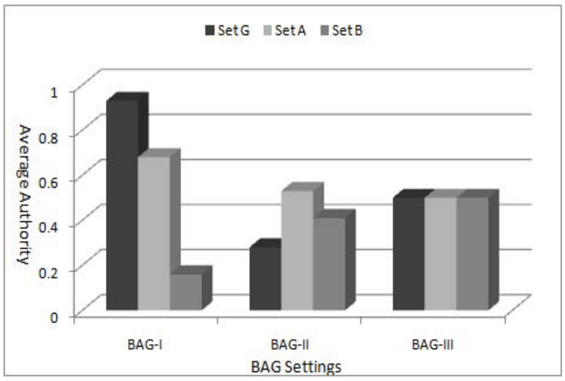


Fig. 6: Average Credibility of Users in Sets for Different BAG Settings

TABLE III: Composition of Users in Top N

	BAG-I			BAG-II		
Top	B	A	G	B	A	G
10	0.000	0.000	100%	30%	70%	0.000
20	0.000	0.000	100%	25%	75%	0.000
30	0.000	3.33%	96.67%	20%	80%	0.000
40	0.000	7.5%	92.5%	25%	75%	0.000
50	0.000	16%	84%	28%	72%	0.000
60	0.000	25%	75%	28.3%	71.7%	0.000
70	0.000	30%	70%	27.1%	72.9%	0.000
80	0.000	37.5%	62.5%	25%	75%	0.000
90	0.000	44.4%	55.56%	24.4%	75.6%	0.000
100	0.000	50%	50%	28%	72%	0.000

2) Composition of Users in Top N: The constitutional proportion of users in each set in top N are shown in Table III.

From Table III, we can see that in the rational case (BAG-I), most of users are in set G within the top N users. However, in the confusing case (BAG-II), most of users are from set A and none of them are from set G. In the controversial case (BAG-III), as all the users have the same credibility, the constitutional proportion of users in each set is equal and not accounted here.

3) Precision and Recall of the Algorithm: In the simulations with settings of BAG-II and the BAG-III, the credible users can not be recognized in the discussion process for the less discriminative ability of average users, so no algorithm can rank credible users based on their citation structures. Therefore, it is meaningless to calculate precision and recall for the algorithm on the data from the confusing case and the controversial case. We only calculate the precision and the recall of the CAR algorithm on the data of the rational case. The results are shown in Fig. 7.

From Fig. 7, we can see that in top 20 ranked users, 100 percent of them are users in G set. As the growth of the value of N, the precision of the algorithm decrease slowly, but the recall of the algorithm increase rapidly. Until the precision reaches to 0.625, the recall of the algorithm reaches to 100% within top 80 users.

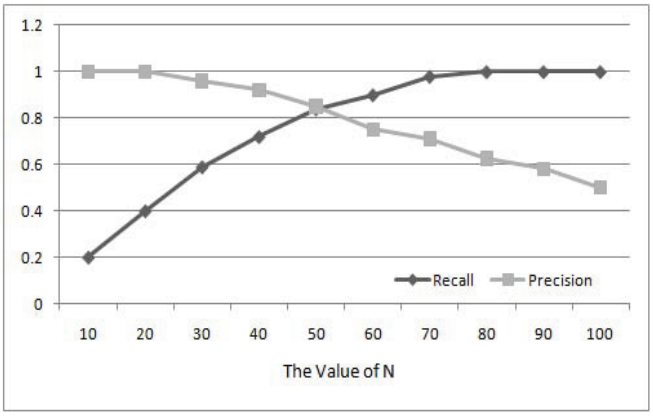


Fig. 7: The Precision and Recall of the CAR Algorithm

V. Conclusion

Learning Villages (LV) is an E-learning platform for people's online discussions and frequently citing postings of one another. It will greatly improve learning efficiency if credible users can be accurately distinguished out in the E-learnning community. In this paper, we use a citation analysis method to evaluate the credibility of users in the LV system. We first propose a k-EACM graph to describe the article citation structure in the LV system. The k-EACM graph not only describes the citation attitude between articles but also takes into account indirect citations. And then we build a weighted graph model k-UCM graph to reveal the implicit relationship between authors hidden behind the citations among their articles. Finally, we design a graph based ranking algorithm, the Credible Author Ranking (CAR) algorithm, which can be applied to rank nodes in a graph with negative edges. In experimental process, we perform experiments on the data generated from the simulated program for three different cases: the **rational case**, the **confusing case** and the **controversial case**. The comparison of results on **average credibility of users in each set** and **composition of users in top** N of three cases demonstrates that the CAR algorithm ranks users's credibility in consistent with the predefined ground truth. The experimental results on testing the variation of the precision and the recall as growth of the value of N for users credibility ranking in the **rational case** demonstrates that the CAR algorithm is an efficient credibility ranking algorithm. In a nutshell, the results of the experiments show that our proposed method works pretty well to rank the credibility of users in the LV system.

Acknowledgement

The authors would like to thank the anonymous reviewers for providing helpful comments. The work described in this paper was supported by the Research Council of Norway under the VERDIKT programme, grants from Microsoft (Project No. CUHK 6902498) and the Direct Grant of the Faculty of Engineering, CUHK (Project Codes: 2050379).

References

[1] Yuan An, Jeannette Janssen, and Evangelos E. Milios. Characterizing and mining the citation graph of the computer science literature. *Knowledge and Information Systems*, 6(6):664–678, 2004.

[2] Rodrigo A. Botafogo, Ehud Rivlin, and Ben Shneiderman. Structural analysis of hypertexts: identifying hierarchies and useful metrics. *ACM Transactions on Information Systems (TOIS)*, 10(2):142–180, 1992.

[3] Leo Egghe. A noninformetric analysis of the relationship between citation age and journal productivity. *JASIST*, 52(5):371–377, 2001.

[4] Xiu fang Wu, Qiang Fu, and Ronald Rousseau. On indexing in the web of science and predicting journal impact factor. *Journal of Zhejiang University Sci. B*, 2008.

[5] Joan M. Francioni. A conference's impact on undergraduate female students. *SIGCSE Bulletin*, 34(2):66–69, 2002.

[6] Lawrence D. Fu and Constantin Aliferis. Models for predicting and explaining citation count of biomedical articles. In *AMIA Annual Symposium Proceedings*, 2008.

[7] Eugene Garfield. Citation analysis as a tool in journal evaluation. *Essays of an Information Scientist*, 1:527–544, 1972.

[8] Eugene Garfield. The impact factor. *Current Contents*, 34(25):3–7, 1994.

[9] Eugene Garfield. Journal impact factor: a brief review. *Canadian Medical Association Journal*, 161(8):979–980, October 1999.

[10] A. Goodrum, K. McCain, S. Lawrence, and C. L. Giles. Scolarly publication in the internet age: a citation analysis of computer science literature. *Information Processing and Management*, 57(2):661–675, 2001.

[11] Baumgartner H. and Pieters R. The influence of marketing journals : a citation analysis of the discipline and its sub-areas. Discussion Paper 123, Tilburg University, Center for Economic Research, 2000.

[12] Peter Ingwersen. The calculation of web impact factors. *Journal of Documentation*, 54(2):236–243, March 1998.

[13] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[14] Zhenjiang Lin, Irwin King, and Michael R. Lyu. Pagesim: A novel link-based similarity measure for the world wide web. In *Web Intelligence*, pages 687–693. IEEE Computer Society, 2006.

[15] Gunther Maier. Impact factors andpeer judgment: The case of regional science journals. 2006.

[16] Henk F. Moed and Thed N. van Leeuwen. Improving the accuracy of institute for scientific informations's journal impact factors. *JASIS*, 46(6):461–467, 1995.

[17] Jeppe Nicolaisen. Bibliometrics and citation analysis: From the science citation index to cybermetrics. *JASIST*, 61(1):205–207, 2010.

[18] Alireza Noruzi. Citation-linking analysis to open access journals. *Webology*, 5(2), 2008.

[19] E. C. M. Noyons, Henk F. Moed, and Marc Luwel. Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *JASIS*, 50(2):115–132, 1999.

[20] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *Technical report, Stanford Digital Libraries*, 1998.

[21] Michael D. Robinson. Applied bibliometrics: Using citation analysis in the journal submission process. *JASIS*, 42(4):308–310, 1991.

[22] A. Schenker, M. Last, H. Bunke, and A. Kandel. *Clustering of web documents using a graph model*, pages 3–18. World Scientific, 2003.

[23] Fred R. Shapiro. Origins of bibliometrics, citation indexing, and citation analysis: The neglected legal literature. *JASIS*, 43(5):337–339, 1992.

[24] Antonis Sidiropoulos and Yannis Manolopoulos. A citation-based system to assist prize awarding. *SIGMOD Record*, 34(4):54–60, 2005.

[25] Antonis Sidiropoulos and Yannis Manolopoulos. A new perspective to automatically rank scientific conferences using digital libraries. *Information Processing and Management: an International Journal*, 41(2):289–312, March 2005.

[26] Dang Yaru. Structural modeling of network systems in citation analysis. *JASIS*, 48(10):946–952, 1997.

[27] Guang Yu, Rui Guo, and Daren Yu. The influence of the publication delay on journal rankings according to the impact factor. *Scientometrics*, 67(2):201–211, 2006.

[28] Guang Yu and Liang Wang. The self-cited rate of scientific journals and the manipulation of their impact factors. *Scientometrics*, 73(3):321–330, 2007.