

Measuring Credibility of Users in an E-learning Environment

Wei Wei
The Royal Institute of
Technology, Sweden
weiwe@kth.se

Jimmy Lee
The Chinese University of
Hong Kong
jlee@cse.cuhk.edu.hk

Irwin King
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. In this paper, we propose a novel method to rank credit authors in the LV system. We first propose a k -EACM graph to describe the article citation structure in the LV system. 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. Furthermore, we design a graph-based ranking algorithm, the Credit Author Ranking (CAR) algorithm, which can be applied to rank nodes in a graph with negative edges. Finally, we perform experimental evaluations by simulations. The results of evaluations illustrate that the proposed method works pretty well on ranking the credibility of users in the LV system.

Categories and Subject Descriptors

J.1 [ADMINISTRATIVE DATA PROCESSING]: Education; J.4 [SOCIAL AND BEHAVIORAL SCIENCES]: Sociology

General Terms

Algorithms, Measurement

Keywords

Education Platform, Author Ranking, HITS Algorithm

1. INTRODUCTION

Learning Villages (LV) is an E-learning platform in which we use a novel method to evaluate the credibility of users in the LV communities. The citation structure of articles in the community can be rich source of information to rank authors performance on the condition that we can build an effective model to describe the relationship among articles and authors. Citation analysis has been widely used for bibliometrics ranking, but the author ranking in the bibliometrics domain is paid relatively little attention. Much research related to citation analysis focuses on publication-based ranking algorithms [4, 1, 2]. The method proposed in this paper focuses on authors ranking in an E-learning system.

Copyright is held by the author/owner(s).
WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.
ACM 978-1-59593-654-7/07/0005.

2. CITATION MODELS DEFINITION

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

DEFINITION 1 (BACM GRAPH). *BACM 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: $E = \{(p, q) | p, q \in V \text{ and the corresponding article } p \text{ citing the article } q\}$. W is a set of weighted values corresponding to edges belong to E . The weight value of an edge $E(p, q)$ is w_{pq} .*¹

The k -Extended Article Citation Model. To take into account indirect impact from nodes, we extend the BACM graph to an k -Extended Article Citation Model (k -EACM).

DEFINITION 2 (k -DISTANCE LINK). *In a BACM graph, if there is a shortest path without considering 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 .*

DEFINITION 3 (k -EACM). *A k -EACM is a weighted directed graph $G'(V, E', W')$ based on BACM 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: $E' = \{(p, q) | p, q \in V \text{ and the corresponding articles } p, q \text{ share a } l\text{-distance link } (l \leq k)\}$. 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:*²

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

User Citation Model. To describe the implicit citations among authors, we map k -EACM graph to the corresponding User Citation Model namely 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.*

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: $E'' = \{(u, v) | u, v \in V' \text{ and the corresponding}$*

¹ $w_{pq} = 1$, if article p supports article q ; $w_{pq} = -1$, if article p is against article q .

² R_{pq} represents one shortest route from node p to node q ; $\{R_{pq}\}$ represents a set of all the shortest routes existing between p and q ; $\text{edges}(r)$ 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 $\{R_{pq}\}$.

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:³

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

3. CREDIT AUTHOR RANKING ALGORITHM

Credit Author Ranking (CAR) algorithm is designed for ranking authors in k -UCM graph. In the CAR algorithm, we associate each user u in the k -UCM graph with a credible 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 $\{y^{<u>}\}$. Two operations denoted by α and β are defined as:⁴

- The α operation updates the x -weights:

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

- The β operation updates the y -weights:

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

To compute the reinforcing equilibrium values for credible attribute weights and trouble attribute weights of each users in k -UCM graph, CAR algorithm does the same iterative process as HITS [3] in Algorithm 1. Considering the $x^{<u>}$ representing a user u 's credible attribute and $y^{<u>}$ representing a user u 's trouble attribute, we introduce $z^{<u>}$ to represent the user u 's overall performance and define $z^{<u>}$ as: $z^{<u>} = x^{<u>} - y^{<u>}$.

Algorithm 1 Iterative Process in CAR

```

Iterate( $V, t$ )
1:  $V$ : a set of nodes in  $k$ -UCM graph
2:  $t$ : a natural number
3: Let  $w$  denote the vector  $(1, 1, 1, \dots, 1) \in R^n$ .
4: Set  $x_0 := w$ .
5: Set  $y_0 := w$ .
6: for  $i = 1, 2, \dots, t$  do
    Apply  $\alpha$  operation to  $(x_{i-1}, y_{i-1})$ , obtaining new  $x$ -weights  $x'_i$ .
    Apply  $\beta$  operation to  $(x_i, y_{i-1})$ , obtaining new  $y$ -weights  $y'_i$ .
    Normalize  $x'_i$ , obtaining  $x_i$ .
    Normalize  $y'_i$ , obtaining  $y_i$ .
7: end for
8: END
9: return  $(x_t, y_t)$ .

```

4. EXPERIMENT

We conduct experimental evaluations of the proposed method on the data from simulations. Three types of users are defined in the simulation including: G type (good users), B type (bad users) and A type (average users). G type of users have three parameters θ_{GG} , θ_{GB} and θ_{GA} which respectively

³ $\text{Atk}(u)$ represents a set of articles in k -EACM of an author u ; $\text{Cby}(p)$ represents a set of nodes q and (p, q) is an edge in k -EACM graph.

⁴ 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 .

indicate support probabilities to users of G type, B type and A type. The other two type of users have the same support probability parameters. Therefore a 3×3 matrix GBA is defined to control the parameters of support probabilities among users. The evaluation process is to examine the percentage of G type users will be ranked in top 50 by the proposed method.

We perform three simulations with different parameters setting. In each simulation, 50 G type users, 50 B type users and 200 A type users are defined and the simulation are run for 100 cycles. The GBA matrix $\begin{vmatrix} \theta_{GG} & \theta_{GB} & \theta_{GA} \\ \theta_{BG} & \theta_{BB} & \theta_{BA} \\ \theta_{AG} & \theta_{AB} & \theta_{AA} \end{vmatrix}$ used in the three experiments are respectively as follows:

Table 1: GBA matrix for 3 simulations

Simulation 1			Simulation 2			Simulation 3		
0.9	0.1	0.5	0.9	0.1	0.9	0.9	0.1	0.8
0.1	0.7	0.5	0.1	0.7	0.1	0.1	0.7	0.1
0.8	0.2	0.5	0.5	0.5	0.5	0.9	0.1	0.8

We utilize the proposed method to analyze the data from three simulations. The percentages of G type users in top 50 are also calculated for each simulation. The results are shown in Table 2. In Table 2, we can find that the proposed

Table 2: Percentage of G type users in top 50

Top	Simulation 1	Simulation 2	Simulation 3
10	100%	100%	100%
20	100%	100%	95.0%
30	100%	96.7%	90.0%
40	97.5%	90.0%	87.5%
50	86.0%	78.0%	82.0%

method rank good users from G type as top credible users in high percentage, which is consistent with our expectation.

5. CONCLUSION

Our research in this paper focus on author ranking in the LV system. We propose k -EACM graph and k -UCM graph to describe the citation relations among articles and authors and furthermore design the CAR algorithm which can be applied to rank nodes in a graph with negative edges. By analyzing the data from three simulations, the experimental results demonstrate that the proposed method works pretty well on ranking the credibility of users in the LV system.

6. REFERENCES

- [1] Y. An, J. Janssen, and E. E. Milios. Characterizing and mining the citation graph of the computer science literature. *Knowledge and Information Systems*, 6(6):664–678, 2004.
- [2] E. Garfield. Journal impact factor: a brief review. *Canadian Medical Association Journal*, 161(8):979–980, October 1999.
- [3] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.
- [4] A. Sidiropoulos and Y. Manolopoulos. A citation-based system to assist prize awarding. *SIGMOD Record*, 34(4):54–60, 2005.