Temporal-Based Ranking in Heterogeneous Networks

Chen Yu, Ruidan Li, Dezhong Yao, Feng Lu, Hai Jin

Temporal-based Ranking in Heterogeneous Networks

Chen Yu, Ruidan Li, Dezhong Yao, Feng Lu, Hai Jin

Services Computing Technology and System Lab
Cluster and Grid Computing Lab
School of Computer Science and Technology
Huazhong University of Science and Technology, Wuhan, 430074, China

Abstract. Ranking is a fundamental task for network analysis, benefiting to filter and find valuable information. Time information impacts results in content that is sensitive to trends and events ranking. The current ranking either assumes that user's interest and concerns remain static and never change over time or focuses on detecting recency information. Meanwhile most prevalent networks like social network are heterogeneous, that composed of multiple types of node and complex reliance structures. In this paper, we propose a general Temporal based Heterogeneous Ranking (TemporalHeteRank) method. We demonstrate that TemporalHeteRank is suitable for heterogeneous networks on the intuition that there is a mutually information balance relationship between different types of nodes that could be reflected on ranking results. We also explore the impact of node temporal feature in ranking, then we use the node life span by carefully investigating the issues of feasibility and generality. The experimental results on sina weibo ranking prove the effectiveness of our proposed approach.

Keywords: Heterogeneous Networks, Heterogeneous Ranking, Diverse Rank, Information Flow Propagation, Hotspot Detection

1 Introduction

In recent years, the rapid development and flexible application of networks have revolutionized the way people discover, share, and these rapid changes simultaneously have a serious effect like massive data generated. Those enormous amount of data lead to find the information of user's interest is extremely difficult, making the network analysis techniques emerged. Ranking as one of these is becoming a burning topic gradually, serving to Internet researchers and academics. The common practice is the graph based ranking [1][2]. However, those approaches either assumed that user's interest and concerns remain static and never change over time [3][4] or focused on detecting recency sensitive information [5][6][7]. Simple aggregation and recency extraction can overshadow the temporal trends that could potentially provide valuable signals for better ordering of information, while lots of demands are not satisfied yet like the temporal based rules.

In this paper, we take sina weibo (or weibo) ranking for the instance. Weibo as one of the most popular on-line short message communication platform, provides tremendous information. In particular, it focuses on recent hot-spots since user can express opinions immediately. Due to the highly temporal nature, incorporating time information into weibo ranking is crucial. The conventional approaches of weibo rank are based on content or the interaction of users, such as forwarding, comments and following. They all have several deficiencies. First, the characteristics used to rank are relatively simple that all nodes in the network were regarded as the same type. Second, for the weibo content tend to be over-entertainment, not all the information is valuable. Third, the temporal factor is an important measurement for ranking results. To solve the problems described above, we introduce temporal based heterogeneous ranking, i.e. TemporalHeteRank, by integrating the information flow propagation in heterogeneous linked nodes and the temporal feature of nodes to enhance the precision and contribute to detect the hotspot. To summarize, the contribution of this paper are described as follows:

- We study the ranking on heterogeneous networks, where the network actually contains multiple types of nodes and complex dependence structures.
- Proposing a method to use the information flow propagation of multiple types node to capture the correlation between different types of node.
- Integrating the temporal feature i.e. life span of nodes to explore the effect on the process of ranking.
- Performing experiments on the most prevalent social network, sina weibo, as the ranking application on hotspot detection to demonstrate the feasibility.

The remaining of the paper is organized as follows. We present the related work in Section 2, and introduces the fundamental concept and necessary preliminaries in Section 3. We describe the specific process of TemporalHeteRank in Section 4. We carry out the performance evaluation and application in Section 5, whereas our conclusions are drawn in Section 6.

2 Related Work

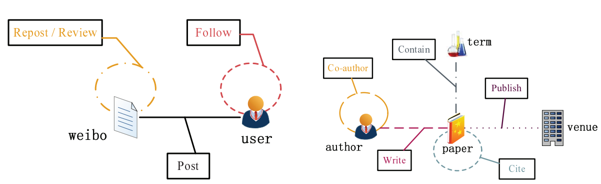
The fundamental goal of ranking is to filter and extract most relevant information from tremendous data. Thus, ranking could save users time and find informative content [8][9] simultaneously. However, few studies concentrate on or relevant to the heterogeneous networks ranking problem from the past [1][10] to the current[3][4]. The conventional methods [1][10] are both classical ranking method playing an important role in homogeneous networks. The research of [3] puts forward the Tri-HITS algorithm on tweet ranking by using the cross-link between tweet and web document to construct the heterogeneous network. After that, combing the reliability feature of the web documents and heterogeneous information iteratively propagation to improve the ranking quality. However, ranking tweet without considering the node temporal feature can lead to meaningless and unvalued information. As the tweet may be out of popularity time [11] and the ranking may not satisfy users demand like tracking the news or capturing the hot topics. [4] mainly rank the venues and authors on DBLP network. They proposed the authority ranking principles based on the rules, that if the node highly ranked then the other linked nodes should be ranked higher reciprocally. While the defection is that none temporal information has considered on the rankings In [12], they present supervised mathematical method of transfer learning called "learn to rank" to solve the complex ranking issues on heterogeneous networks, but the label information of dataset which needed in the supervised learning are extremely expensive and difficult to obtain in the real world. Those aforementioned approaches either assumed that user's interest and concerns remain static and never change over time or simple focused on detecting recency-sensitive information[5][6][7], for instance [6] proposed a temporal query model, using temporal features for query performance predict. Also many studies like [11][13][14], they all demonstrate that the popularity and influence of tweet varies over time.

3 Concepts and Preliminaries

3.1 Heterogeneous Information Network

An information network represents an abstraction of the real world, focusing on the nodes and the interactions between the nodes. Formally, [4] define an information network as the directed graph $G = (V, E)$ on $V = \{V_1 \cup \ldots \cup V_N\}$ and $E = \{E_1 \cup \ldots \cup E_M\}$. When the types of nodes $N > 1$ or the types of relationship $M > 1$, the network namely is the heterogeneous information networks. Here we give some networks for example.

1. Sina weibo network [15]. The sina weibo consist of two different types nodes (i.e. weibo and user) and many relationships between different types of nodes. Relationships can emerge in same type like weibo forwarding, and different types like user post weibo (see in Fig.1a).



(a) (b)

Fig. 1: Temporal trend of weibo and DBLP networks

2. DBLP bibliographic network [16]. DBLP contains four types of node, namely papers, authors, terms and venues (conferences or journals). Links exist between authors and papers by the relation of write (or written by), between papers and terms by mention (see in Fig.1b).

Information imbalance exists in heterogeneous linked nodes. Taking weibo and web document for example, weibo possesses the qualities of real-time and massive, whilst the messy weibo makes it uninformative and unreliable. Web could not provide the real time information, but they always come from organizations that reliable inherently. Thus we use the flow propagation to make the information of heterogeneous linked nodes mutually reinforced. The connection of weibo and web document's built through semantic similarity.

The weibo heterogeneous network, defined as graph G inherited from the information network, composes of weibo, web document, and user. Namely $V = \{V_w \cup V_u \cup V_d\}$ and $E = \{E_w \cup E_{wu} \cup E_u \cup E_{wd} \cup E_d\}$.

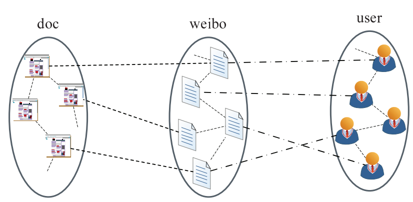


Fig. 2: Weibo heterogeneous network

3.2 Preliminaries

We briefly introduce the work of [2] for the diverse rank in homogeneous networks. Diverse rank or DivRank is a random walk ranking algorithm. In contrast to PageRank, DivRank assumes that the transition probabilities change over time, and the ranking score of nodes varies accordingly. After the z-th iterations, the transition probability matrix M becomes:

$$M(z) = \alpha \cdot M(z-1) \cdot R(z-1) + \frac{(1-\alpha)}{|V|} \cdot E \tag{1}$$

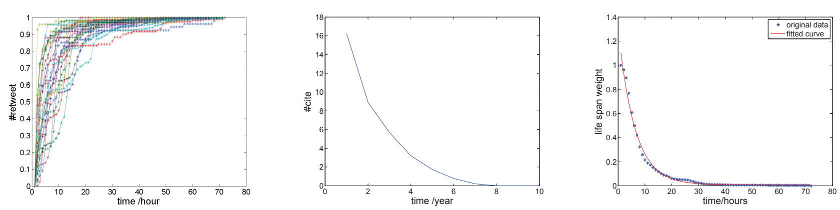
4 TemporalHeteRank Method

To make the ranking draw attention as much as possible, we define the informative as the measurement of weibo rank. Our basic assumption of ranking is the heterogeneous information flowing propagation: 1) Highly ranked weibo may attract many forwarding amount and reviews generated by highly ranked users, verse vice; 2) Highly ranked weibo aligns with many highly ranked web document content. As the web contains abundant information and comes from formal genre, so it can be used to reinforce the weibo content quality; 3) The recently released weibo should be given the corresponded promotion, as minor forwarding amount and comments that can reveal the process of the information propagation explicitly.

After crawling all the weibo within a specified time window, we first use the weibo forwarding pattern to analyze weibo temporal information. Then we define queries based on the top terms in weibo, and use the Google Search API to retrieve the titles of the top m web for those queries ($m = 5$ for our experiments).

4.1 Ranking the Graph

Life Span Analysis The life span is an important measurement to evaluate the ranking qualities. Currently there are several approaches to measure the weibo life span, like the temporal variation of hot topic that weibo related to and the weibo forwarding amount. The forwarding can explicitly reflect the information dissemination process, hence we adopt the forwarding to measure the life span. To prove the generality, the life span could be extended to DBLP network [17][18], and we employ the cite amount to analysis the paper temporal trend in Fig.3b.



(a) (b) (c)

Fig. 3: Temporal trend of weibo and DBLP networks. a) describes thirty weibo's temporal repost; b) shows DBLP's temporal cite amount; c) presents the life span of weibo.

In [11] and [19] they all proved a rule that almost 90% of weibo are rarely forwarded after 72 hours since they are posted. Fig.3a depicts thirty weibo forwarding pattern in 72 hours. It shows that the weibo forwarding amount quickly increases with the time after release but saturates after reaching the time of thirtieth hours after birth and forwarding amount reaches 90% of the total. From the description above we draw a conclusion that weibo forwarding approximates to the sigmoid curve. Then we use the follow function to model the weibo life span:

$$cW_{\cdot t}^{life} = d \cdot \exp^{c-b*t} - a \tag{2}$$

where $W_{\cdot t}^{life}$ is the weibo life span weight at time t. Parameter a and d are used to convert the horizontal position of curve and the b and c factors are defined to control the decay rate. In Fig. 3c, the simulated trend almost totally overlaps with the dynamic of the real world data forwarding process. The sum of residuals square are all fall below 0.6 in 95% confidential interval.

Initialize Weibo and Web Ranking As the weibo ranks on multiple topics, so we would like the output keep diversity. DivRank algorithm could achieve diversity by iteratively selecting the most prestigious or popular nodes and continuously updating the transposition probability matrix. At each step, the algorithm updates the dynamical transition matrix Eq.(1). Hence after z-th iterations, the ranking score of weibo and web document become:

$$R(z) = \alpha \cdot [M(z)]^T \cdot R + \frac{(1-\alpha)}{|V|} \cdot E \tag{3}$$

Accordingly the temporal based weibo ranking R^w calculates as $R^w \cdot W^{life}$. The weight of two weibo(or web) w_i and w_j denotes the cosine similarity of them. Each weibo can be treated as a short document, then we employ the TF-IDF method to weight the terms of the weibo words. Each entry of the adjacent matrix M stands for the text similarity of the weibo or web in the graph, and is defined as follows:

$$M_{ij} = \frac{sim(w_i, w_j)}{\sum_k sim(w_i, w_k)}, sim(w_i, w_j) = \frac{\boldsymbol{w_i} \cdot \boldsymbol{w_j}}{\|\boldsymbol{w_i}\| \cdot \|\boldsymbol{w_j}\|} \tag{4}$$

In Eq.(4), the $sim(\cdot)$ denotes the cosine similarity between two weibo (or web) and the $\boldsymbol{w_i}$ represent the TF-IDF vector of the weibo (or webs) w_i. Also TF represents term frequency, IDF said the reciprocal of documents.

Initialize User Ranking The aforementioned user graph $G_u = (V_u, E_u)$ is a directed and weighted graph. At first we use the following relationship to establish the users graph. When the user u_i follows user u_j, we add a edge (u_i, u_j) to the following adjacent matrix M^{uf}. Thus M^{uf} is defined as follows:

$$M_{ij}^{uf} = \frac{f(u_i, u_j)}{\sum_k f(u_k, u_j)}, f(u_i, u_j) = \begin{cases} 1 & (u_i, u_j) \in E_u \\ 0 & (u_i, u_j) \notin E_u \end{cases} \tag{5}$$

Moreover, we take the credibility of users into consideration, as the prestige is not absolutely coordinate with the credibility. We define the credibility weigh between user u_i and user u_j as M_{ij}^{uc} , according to the number of interactions, for example mentions, reposts and reviews. The creditable weight between two users u_i and u_j is described as follow:

$$M_{ij}^{uc} = \frac{actions\_from\_u_i}{actions\_of\_u_j} \tag{6}$$

Furthermore the users relation matrix M^u becomes $M^{uf} \cdot M^{uc}$. In Eq.(6), $actions \in \{mention, repost, review\}$ represent user interactions with weibo. The $actions\_from\_u_i$ denotes the reciprocal interactions between u_i and u_j. The $action\_of\_u_i$ denotes the alternation of the optional user u_k and u_j. Naturally, we apply DviRank random walk model on user graph using matrix M^u, and compute the ranking score of each user.

4.2 Affinity Matrices

According to the previous description of E_{wd} and E_{wu}, we define two adjacent matrices M^{wd} and M^{wu}. Matrix M^{wd} represents the weight between the weibo and the web documents, and measured by the cosine similarity of document and weibo content.

$$M_{ij}^{wd} = \begin{cases} weight_{ij} & , weight_{ij} > \delta_{wd} \\ 0 & , others \end{cases}, weight_{ij} = \frac{sim(w_i, d_j)}{\sum_k sim(w_i, d_k)} \tag{7}$$

Matrix M^{wu} represents the weight between weibo and user. We use a set of weibo that a user posts such as w_m in a period of time to compute the cosine similarity with the weibo w_i, if the similarity exceeds the threshold we set the user link to the weibo in M^{wu}.

$$M_{ij}^{wu} = \begin{cases} \max_{sim(w_i,w_k)} w_k & , max_{sim(w_i,w_k)} > \delta_{wu} \\ 0 & , others \end{cases} \tag{8}$$

w_k in Eq.(8) indicates the element of weibo set that a user posts in a period of time. $sim(\cdot)$ describes the cosine similarity between pairwise weibo.

4.3 Flow Propagation

The tripartite weibo graph comprises three homogeneous graphs and two heterogeneous graphs. The weibo-document denotes the content align inter-relation between the weibo and the web document, and the weibo-user means the implicit relationship between the user and the weibo. Based on the ranking assumptions described at foremost of this section, we use the following iterative information flowing propagation to formulate the procedure:

Step 1 Starting from web document R^d, the update process considers both the last ranking score and the information flow propagation from connected weibo R^w, which can be expressed as:

$$R^d(z+1) = (1-\lambda_d) \cdot M^d(z) \cdot R^d(z) + \lambda_d \cdot M^{wd} \cdot R^w(z) \tag{9}$$

Step 2 In the same way, we define the information flow propagation from weibo R^w to user R^u as:

$$R^u(z+1) = (1-\lambda_u) M^u(z) \cdot R^u(z) + \lambda_u \cdot M^{wu} \cdot R^w(z) \tag{10}$$

Step 3 Each weibo R^w can be influenced by the information propagation from both web document and user, then compute weibo ranking scores:

$$\begin{aligned} R^w(z+1) =&(1-\lambda_d-\lambda_u)M^w \cdot R^w(z+1) \\ &+\lambda_d \cdot M^{dw} R^d(z) + \lambda_u \cdot M^{uw} R^u(z) \end{aligned} \tag{11}$$

where the parameter λ is to balance the importance of weibo, user, and document. $R^w(z)$, $R^d(z)$ and $R^u(z)$ are the ranking score matrix of weibo, web document and user at z-th iteration. To guarantee the iteration converges, we normalize R^w, R^d and R^u after each iteration using $R(z+1) = R(z+1)/\|R(z+1)\|$. The algorithm typically converges when the difference between the scores computed at two successive iterations for any weibo falls below a threshold ξ (set as 0.001 in our method).

5 Experiment and Application

5.1 Dataset

Sina weibo is the most popular microblogging service in China. The dataset in [20] collected a complete network between 1,700,000 users and all the weibo posted by those users between Jul. 28th, 2012 and Oct. 29th, 2012. We choose the three most popular topics in Aug, 2012 (described in Tab.1) and study how to rank weibo in heterogeneous information network. We also study how life spans of weibo influence the ranking results.

Table 1: Sina Weibo dataset description

Dataset	Users	Follow-relationship	Original-microblogs	Retweets
Sina Weibo	1,776,950	308,489,739	300,000	23,755,810

5.2 Evaluation Metric

For evaluation, we employ two widely used metrics: $MeanAveragePrecision$ (MAP) and $DiscountedCumulativeGain$ (DCG) [21]. In particular, we measure the MAP and DCG on the top-n results, denote as MAP@n and DCG@n respectively. Instead of DCG@n, we adopt $NormalizedDiscountedCumulativeGain$ (NDCG) [22], which is a normalization of DCG in the range [0, 1] calculated as:

$$NDCG@n = \frac{DCG@n}{IDCG@n}, DCG@n = \sum_{i=1}^{n} \frac{2^{rel_i}-1}{\log_2(i+1)} \tag{12}$$

where IDCG@n is the ideal DCG@n, i.e. the maximum possible DCG value up to the ranking position n. We apply MAP on ranking output under the assumption that the top ranked weibo are more relevant to the hot topic and the rest are less:

$$MAP = \frac{1}{|T|}\sum_{j=1}^{|T|} P_i^j \cdot r_i^j \tag{13}$$

where T represent the topic set. In our experiments we set $n = 5, 10, 25, 50$ i.e. MAP@5, NDCG@10.

5.3 Experiment on Ranking

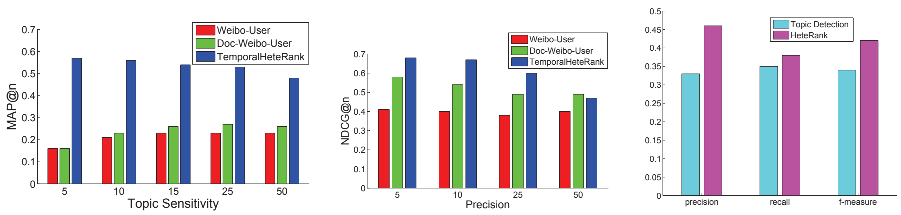
In our experiment, we primarily show three different kinds of ranking methods in Tab.2 to verify the feasibility.

Table 2: Description of three kinds of analysis method

Methods	Description
1.Weibo-User	Using information propagation between weibo and users purely to rank on weibo-user network
2.Doc-Weibo-User	Ranking by combining the web document and information flow propagation on doc-weibo-user network
3.TemporalHeteRank	Weibo temporal constraint life span included beside information propagation

The Sina Weibo official study points out that there were three drastically discussed topics during August 2012, namely 'Liu Xiang', 'Lin Dan', and 'Diaoyu Islands'. According to the studies, the experiments of ranking falls into two parts: the topic sensitivity and the precision compared with the ground truth ranking. We intuitively rank the topics based on the time it happened. The topic sensitivity is to figure up the text similarity between weibo and topic, and the weibo is much more similar to the higher ranked topic indicates the weibo is relevant to this topic. The relevant is 1 and 0 otherwise, then get the topic sensitivity by MAP. By employing the mutually annotated weibo as the ground truth we use the NDCG to evaluate the precision that compared with the ground truth ranking. Our results are summarized in Figure 4a and 4b. Fig 4a shows the models that elicited above performance. In Figure 4b, it provides results when model performance is evaluated against the gold standard ranking obtained from the weibo network.

Figure 4a shows first method that ranks only on weibo and user perform worst, that implies weibo based on the independent user rank is unable to extract significant information like hot topics for weibo tend inclined to the entertainment. The second performs better than the first. The crucial factor is the information flow propagation between web document and weibo. The results also validate our previous assumptions that making use of web document containing abundant information and formal genre can improve the accuracy of weibo ranking. Comparing with the two methods described above, the TemporalHeteRank i.e. the third indicates the temporal feature of node has great impact on ranking and fully satisfy the demand of topic detection.



(a) (b) (c)

Fig. 4: Describing the MAP and DCG of the three rankings. a) shows topic sensitivity; b) represents precision compared with ground truth; c) denotes the comparison between TemporalHeteRank and topic detection model.

We randomly choose weibo from the dataset manually annotated as the ground truth ranking of our reference. Following the annotation guidelines defined by [3], five annotators parallelize each assigned weibo a grade in a 5-star likert scale. When the label difference between annotators is 1, the lower grade is selected. When the label difference is greater than 1, those tweets are re-annotated until the label difference falls below 1. From the Figure 4b, the TemporalHeteRank method constantly performs superior to the other two methods.

5.4 Experiment on Hotspots Detection Application

As pointed out in the introduction of this paper, the weibo ranking can be applied to hot-spots detection. We compare our approach with the state-of-the-art topic detection model [23]. All the models are subject to use the same dataset and the standard results attested by sina weibo. The detection model (Topic Detection) optimizes the feature selection and weight computation method to filter out those topic-unrelated weibo, and uses a new vector distance calculation method to update the center vector. Fig.4c describes the experimental results on hot-spots detection. The TemporalHeteRank based on information flow propagation and weibo life span consistently outperforms the topic detection model, as the topic detection model never takes the node temporal feature into account. It generates the same hot-spots at any point. Our TemporalHeteRank algorithm models life span regarding the weibo and integrate the information flow propagation to rank. Moreover, it attempts to mine the informative weibo by invoking web document. Both the instances are evaluated by precision, recall rate and F-measure. The data in Fig.4c indicates the hot topics are unexpected and sent by many users from multiple groups, ranking can promote user concern and experience.

6 Conclusion

This paper has investigated the temporal based ranking on the heterogeneous network, and takes the most prevalent sina weibo for the experiments. After crawling the weibo dataset, we analyse the temporal information via weibo forwarding pattern and fit the time-vary life span curve of weibo firstly. Secondly, we use the traditional approaches to filter the noisy weibo and mine the valuable information out from the weibo heterogeneous network. According to the characteristic of entertainment, we improve the ranking precision of weibo resorting to the web thirdly. In fourth step, by adopting the information flowing propagation, the model balance the heterogeneous linked information. Finally, the TemporalHeteRank model integrates temporal weighted ranking results to obtain hotspot of weibo. The proposed TemporalHeteRank method is easy to implement, and the followed experiment shows that it is more efficient and more effective than other conventional methods.

Acknowledgements

The work is partly supported by Technology Innovation Fund of Huazhong University of Sci. and Tech. (No.CXY13Q018) and Ministry of Education and China Mobile Communications Corporation (MoE-CMCC) Research Founding (No.MCM20130382)

References

1. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems **30**(1-7) (1998) 107–117
2. Mei, Q., Guo, J., Radev, D.: Divrank: The interplay of prestige and diversity in information networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '10 (2010) 1009–1018
3. Huang, H., Zubiaga, A., Ji, H., Deng, H., Wang, D., Le, H.K., Abdelzaher, T.F., Han, J., Leung, A., Hancock, J., Voss, C.R.: Tweet ranking based on heterogeneous networks. In: Proceedings of the 24th International Conference on Computational Linguistics. COLING '12 (2012) 1239–1256
4. Sun, Y., Han, J.: Mining heterogeneous information networks: a structural analysis approach. SIGKDD Explorations **14**(2) (2012) 20–28
5. Li, X., Croft, W.B.: Time-based language models. In: Proceedings of the 12th ACM International Conference on Information and Knowledge Management. CIKM '03 (2003) 469–475
6. Keikha, M., Gerani, S., Crestani, F.: Time-based relevance models. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '11 (2011) 1087–1088
7. Choi, J., Croft, W.B.: Temporal models for microblogs. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. CIKM '12 (2012) 2491–2494
8. Zhukovskiy, M., Khropov, A., Gusev, G., Serdyukov, P.: Fresh browserank. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '13 (2013) 1029–1032

9. Yan, R., Lapata, M., Li, X.: Tweet recommendation with graph co-ranking. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. ACL '12 (2012) 516–525
10. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5) (1999) 604–632
11. Kong, S., Feng, L., Sun, G., Luo, K.: Predicting lifespans of popular tweets in microblog. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '12 (2012) 1129–1130
12. Yang, Z., Tang, J., Li, J.: Learning to rank in heterogeneous network. Sciencepaper Online **4** (2011) 273–279
13. Cheng, S., Arvanitis, A., Hristidis, V.: How fresh do you want your search results? In: Proceedings of the 22nd ACM International Conference on Conference on Information; Knowledge Management. CIKM '13 (2013) 1271–1280
14. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web. WWW '10 (2010) 591–600
15. Sina. http://weibo.com
16. Sun, Y., Han, J., Aggarwal, C.C., Chawla, N.V.: When will it happen?: relationship prediction in heterogeneous information networks. In: Proceedings of the 15th ACM International Conference on Web Search and Data Mining. WSDM '12 (2012) 663–672
17. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: Densification laws, shrinking diameters and possible explanations. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. KDD '05 (2005) 177–187
18. Gehrke, J., Ginsparg, P., Kleinberg, J.: Overview of the 2003 kdd cup. ACM SIGKDD Explorations Newsletter **5**(2) (2003) 149–151
19. Ma, H., Qian, W., Xia, F., He, X., Xu, J., Zhou, A.: Towards modeling popularity of microblogs. Frontiers of Computer Science **7**(2) (2013) 171–184
20. Zhang, J., Liu, B., Tang, J., Chen, T., Li, J.: Social influence locality for modeling retweeting behaviors. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence. IJCAI '13 (2013) 2761–2767
21. Järvelin, K., Kekäläinen, J.: IR evaluation methods for retrieving highly relevant documents. In: Proceedings of the 23rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '00 (2000) 41–48
22. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of ir techniques. ACM Trans. Inf. Syst. **20**(4) (2002) 422–446
23. Zhao, X., Zhu, F., Qian, W., Zhou, A.: Impact of multimedia in sina weibo: Popularity and life span. In: Semantic Web and Web Science. Springer Proceedings in Complexity. (2013) 55–65