

Article

# Tagging Users' Social Circles via Multiple Linear Regression

Hailong Qin <sup>1</sup>, Jing Liu <sup>2</sup>, Chin-Yew Lin <sup>2</sup> and Ting Liu <sup>1,\*</sup>

<sup>1</sup> Harbin Institute of Technology, Harbin 150001, China; liuting@ir.hit.edu.cn

<sup>2</sup> Microsoft Research Asia, Beijing 100086, China; liudiani@microsoft.com (J.L.); cyl@microsoft.com (C.-Y.L.)

\* Correspondence: liuting@ir.hit.edu.cn; Tel.: +86-451-8641-3683

Academic Editor: Remo Pareschi

Received: 4 March 2016; Accepted: 24 May 2016; Published: 24 June 2016

**Abstract:** A social circle is a category of strong social relationships, such as families, classmates and good friends and so on. The information diffusion among members of online social circles is frequent and credible. The research of users' online social circles has become popular in recent years. Many scholars propose methods for detecting users' online social circles. On the other hand, the social meanings and the tags of a social circle are also important for the analysis of a social circle. However, little work involves the tags discovery of social circles. This paper proposes an algorithm for social circle tag detection by multiple linear regression. The model solves the data sparse problem of tags in social circles and successfully combines different categories of features in social circles. We also redmap the concept of the social circle into "reference circles" of an academic paper. We evaluate our method in datasets of both Facebook and Microsoft Academic Search, and prove that it is more effective than other relevant methods.

**Keywords:** tag detection; social circle; multiple linear regression

---

## 1. Introduction

Social media is a popular communication platform. Compared with other information networks, user relationships promote more effective dissemination of information on social networks. There are different functional social medias, such as online academic networks, recommendation networks and social network services (SNS) and so on. In these networks, users communicate with their friends and share information about their similar interests. Users usually have strong relationships with and similar backgrounds to their friends.

A user has many categories of strong relationships in social media. These strong relationships are constituted of users' online social circles. A user has several social circles on Twitter and Facebook, such as classmates in a school, and colleagues in a company and so on (Figure 1). A social circle reflects an individual's social environment that can often be leveraged to infer important information about that individual's attitudes, behaviors, and decisions [1–7]. However, this type of social circle is different from the traditional community.

Perceived in a graph view, the distribution of edges is not only globally, but also locally inhomogeneous, with high concentrations of edges within special groups of vertices, and low concentrations between these groups. This feature of a real network is called community structure [8]. A node of a community may refer to people, and it also could be compared to a computer in Internet or a gene in a gene network. Firstly, a traditional community is larger than a general social circle which may just have three members (such as three family members). On the other hand, a residential area can be a community in a mobile communication network. The area may have several thousand people, but it is not anyone's social circle. Secondly, although there are some common tags or profiles of each member in the same community, these members may not know each other. In contrast,

there are strong relationships between each member in the same social circle. In social science, weak ties are human relationships (acquaintance, loose friendship, etc.), that are less binding than family and close friendship but might, according to Granovetter [9], yield better access to information and opportunities [10]. The tie-strength can measure the quality of a community [11,12]. A community may be a weak tie, and sometimes members of a community even have no explicit social relationships. However, social circles are typically strong ties. Namely, a community of mobile network depends on the location, not explicit social relationships. As suggested above, community detection is focused on finding arbitrary highly interconnected subgraphs within larger networks, and social circle detection will instead discover several groups of strong social relationships including one or more specific individuals [13].

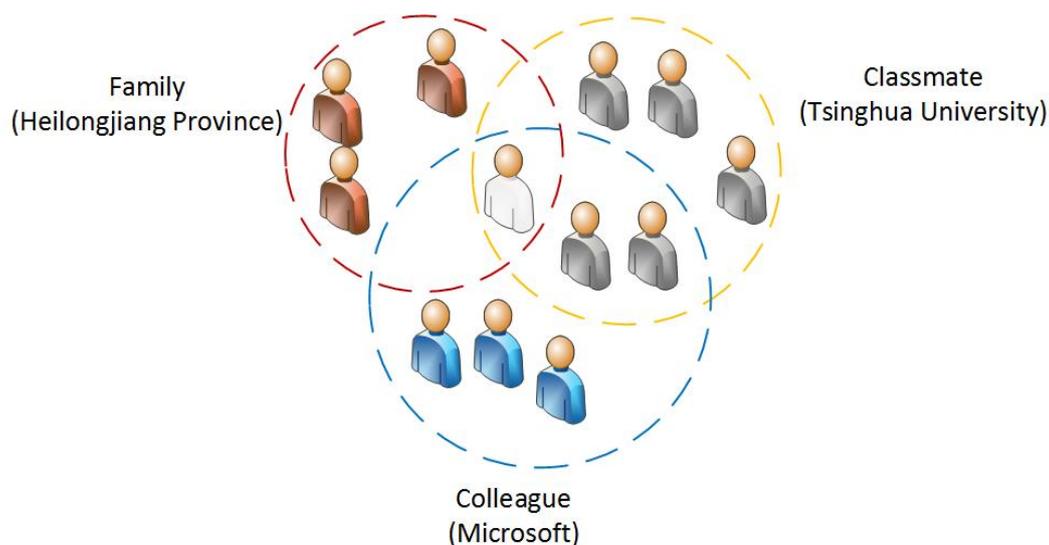


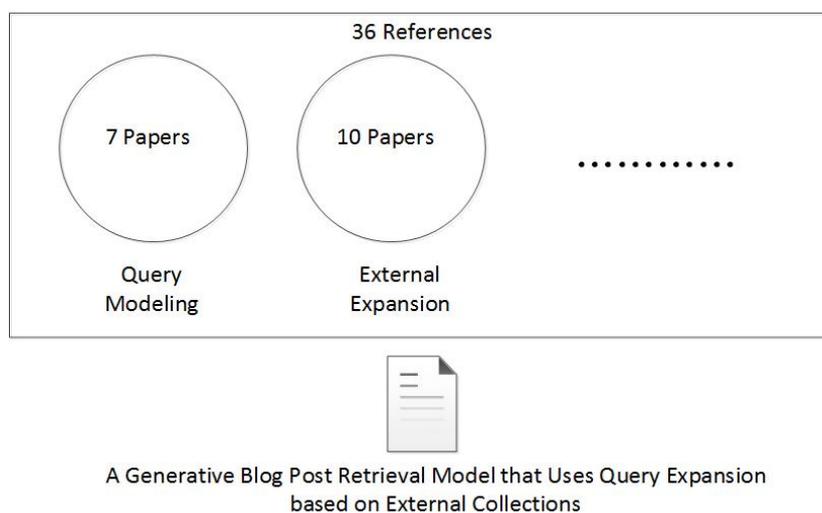
Figure 1. A Social Circle.

According to these characteristics, detecting and analyzing user's social circles is valuable for research on social network and user behavior. Although there is a lot of work about community detection, many scholars propose algorithms for detecting and analyzing social circles specially [13–17]. Furthermore, the concept of a circle in academic networks means a group of papers which have strong relationships. A paper may cite other papers with several intentions. Some of the references may be relevant according to the research problems and other references may be relevant according to the methodologies. If a paper cites another paper with similar topics or themes, we can regard these two papers are friends, and they have strong relationships with respect to content (Figure 2).

Every social circle should have its tags which can represent its social meanings. For example, a social circle of families may have some tags about a city or a town (such as Beijing), and a social circle of classmates may have some tags about a university and a major (such as Tsinghua University, Computer Science). So far, there are no public social circle datasets with annotated tags. The ground-truth of a social circle is also difficult to collect for privacy issues. Everyone has her/his own social circles, and can only annotate the members and the social meanings of her/his own social circles. So the scale of social circle datasets is usually small. Therefore, it is not easy to evaluate the performance of a social circle detection algorithm.

Research of online social circles has already been strongly concerned by scholars in recent years. In particular, a lot of relevant work in terms of social circle detection has appeared, and this work distinguishes this area from the traditional community detection. However, detecting and mining of social circles is just a foundation of online social circle research, and little work involves tag detection for ground-truths of social circles. One reason for the lack of research is that the purpose of traditional

social tag recommendation is to recommend tags for social resources, not users and social groups generally. On the other hand, the amount of members in social circles is small and most users only have few tag items. Some users even have no tags. So, general topic models are not adapted to tags detection of social circles. Therefore, the existing algorithms are difficult to apply to tag detection for social circles. The above factors make social meanings of social circles difficult to identify.



**Figure 2.** The main references of this paper "A Generative Blog Post Retrieval Model that Uses Query Expansion Based on External Collections" are relevant to two themes: Query Modeling and External Expansion.

In this paper, we propose an algorithm for detection of social circle tags via multiple linear regression. The algorithm detects profiles which can represent a social circle according to members' information of the social circle. By many features of members' relationships and tags, the model trains parameters according to characteristics of social circle data. Then, it gives a weight to every tag for a social circle. Top ranked tags can be representative tags of the social circle. Considering the network topologies, the important members' tags will have higher weights. This can make up the lack of members' tags in social circles. In the dataset of Facebook, the model can detect social circles' tags precisely. In the dataset of the Microsoft Academic Search, it detects keywords for reference circles and filters other redundant words. In the two datasets, it improves precision by 11% and 23% respectively compared to the relevant methods which are based on text information.

The rest of this paper is organized as follows: in the next section, we describe related work. Then we introduce our methodology of tags detection of social circles in Section 3. We describe datasets and the experiment in Section 4. Finally, we conclude our work and point out avenues for future research.

## 2. Related Work

### 2.1. Social Circle

The social circle is closely relevant to the term community which has two interpretations: one is the geographical notion of community and another one is relational. The second one is mainly concerned with people's relationships, without reference to location [18]. In this paper, we mainly consider the online social circles that carry the second meaning and mainly concern people's relationships. The detection of social circles is a new research area which is emerging with the popularity of social media. It is a clustering problem within the ego network. Members within a social circle do not only have dense relationships, but also have some common tags. Generally, a social circle is a group of strong social relationships with a specified social meaning.

A growing number of scholars study their subjects in view of social circles. The recommendation algorithm based on users' circles performs no worse than those based on the full network [19]. Question recommendation, question popularity analysis and prediction based on social circles get a better performance [20,21]. It is also a main feature of linking users across online social networks [22].

Huberman, B.A. et al. proposed that it is necessary to mine users' real friends [23]. The evolution of online social groups is analyzed and predicted by [24]. Qu and Liu propose a semi-supervised method to detect social circles in Twitter [25]. However, some members in user groups are not users' real friends. A lot of groups just classify different types of followees, many of these followees are not users' bilateral friends and strong relationships. Based on Google+, [26] explores motivations of users creating social circles and sharing information in social circles. The paper observes users' behavior to identify the strategies for improving sharing precision through selective sharing. And it also analyzes the names of social circle, such as family or friends. It proves that members of most social circles usually have strong ties, but it does not explore the specified profiles and tags of social circles. The authors of [27] propose a visualization tool for social interaction, and it also can be used to visualize users' social circles. Since 2012, some specified social circle detection algorithms have been proposed [13–17]. These algorithms can detect users' strong relationships in social media. These works of social circle detection are foundations of social circle analysis.

## 2.2. Tag Detection

Reference [28] studies the behavior of tagging in Twitter. Users give tags for their tweets for filtering the topics of tweets. The paper proposed that these tags can indicate the lifetime of topics in Twitter. However, it does not involve user's or social circles' tags. The most straightforward unsupervised method for tag detection is using TF-IDF (the detail is in Section 3.2) [29] to rank candidate tags and selecting the top-M as tags. TF-IDF ranks candidate keywords only according to their tags. This will fail to consider the topological structure among social circle members. To our best knowledge, there is little work about tag detection for social circle. In 2012, Liu proposes frequency-based keyword extraction (FKE) for detect users' tags based on text on social network [30]. However, social tags are sparser than text and it is usually difficult to collect users' text. The authors of [16] detect tags of social circles with members' following accounts. The method cannot detect tags of a social circle by members' own characteristics. Moreover, many scholars explore the relationships of scientific topics since a lot of academic data will be released. Topic model is a common technology for the evolution of research themes [31,32] and discovery of high quality papers [33]. The work of citation prediction combines features of links and topics [34], this can also improve the precision of topic detection [35,36]. Bolelli clusters papers in different topics within sparse citations [37]. However, the work is still absent in terms of clustering an academic paper's references. Furthermore, in the situations of data sparsity, these algorithms are difficult to apply to tag detection of ego social circles.

## 3. Tag Detection of Social Circles

### 3.1. Overview

In social networks, every user usually has many social circles, and every social circle has its social meanings, such as families and colleagues. Predicting social meanings of a social circle is significant for the analysis of social circles in social media. However, it is difficult to identify social meanings of social circles via members' tags immediately.

There are many common tags in most social circles, such as degree type. Some tags only belong to a specified user, such as user ID. These kinds of tags cannot be representative tags and attributes of one social circle. On the other hand, representative tags of a social circle should be owned by most members of the circle. However, the lack of individuals' tags means that this does not always work.

For solving these problems of discovery of meaningful tags, we propose a model of multiple linear regression for detecting tags of social circles by combining features about the topological

structure and the members' tags of social circle. We regard tag detection as a problem of tag ranking in a social circle. The model gives every tag a score for a social circle, tags with higher score are more likely tags of social circles.

### 3.2. Features

User relationship is an effective feature in a lot of work for social circle detection [14,15]. User relationships can reflect users' importance in a social circle. More authoritative users can contribute more important tags. So we choose relevant features of both users' relationships and users' tags. We use all users' tags and every user's tag can describe her/his profile, such as school and location, and so on (Table 1). A user may have one or several items for every type of tag, and different items usually have different values.

**Table 1.** Types of tags.

---

last_name, first_name, birthday, name, gender
locale, hometown-name, hometown-id, education-school-name, education-school-id
education-type, education-year-name, education-year-id, education-concentration-name
education-concentration-id, id, location-name, location-id, education-classes-from-name
education-classes-from-id, education-classes-with-name, education-classes-with-id
education-classes-name education-classes-id, work-position-name work-position-id
work-start_date, work-end_date work-employer-name, work-employer-id
work-location-name, work-location-id, languages-name, languages-id
middle_name, work-projects-name, work-projects-id, education-with-name
education-with-id, work-projects-with-name, work-projects-with-id, work-description
education-degree-name, education-degree-id, work-projects-start_date, work-with-name
work-with-id, work-projects-from-name, work-projects-from-id
education-classes-description, work-from-name, work-from-id, political, religion
work-projects-end_date, work-projects-description, location

---

(1) **Percentage of members who own the tag**

$t$  is a tag and  $|MEM_t|$  is the amount of members who own this tag.

$$Fea_1 = \frac{|MEM_t|}{Total\ of\ Circle\ Members} \quad (1)$$

(2) **The members' average centrality who own the tag**

$u$  is a circle member who owns the tag,  $InnerDegree(u)$  is the number of this user's friends in the circle.

$$Fea_2 = \frac{\sum InnerDegree(u)}{Total\ of\ Circle\ Members - 1} \quad u \in Circle \quad (2)$$

(3) **The tag's TF (Term Frequency) value in a circle**

In our work, we regard the set of all members' tags in a social circle as a document, and every tag in the set as a word of this document. For example, a user has two tags user:id:27 and school:id:10. The two items are words and all members' words constitute the tag document of this social circle.  $Count(tag)$  is the amount of a tag item in the circle.

$$Fea_3 = \frac{Count(Tag)}{Total\ of\ Profile\ Items\ in\ Circle} \quad (3)$$

- (4) **The tag's IDF (Inverse Document Frequency) value in a circle**

$$Fea_4 = \log \frac{\text{Total of Circles}}{\text{Count of Circles Having the Tag}} \quad (4)$$

- (5) **The tag's TF-IDF value**

$$Fea_5 = Fea_3 \times Fea_4 \quad (5)$$

- (6) **If only one user owns the tag**

If only one user owns this tag,  $Fea_6$  is 1, otherwise, this  $Fea_6$  is 0.

- (7) **If only one social circle owns the tag**

If only one social circle owns this tag,  $Fea_7$  is 1, otherwise,  $Fea_7$  is 0.

- (8) **Prefix of the tag**

Some tags cannot be tags of social circles since they can only belong to a single user, such as user:id. We filter types of all tags and if a tag might be a social circle's tag,  $Fea_8$  is 1, otherwise,  $Fea_8$  is 0.

### 3.3. Multiple Linear Regression

The algorithm computes a score for every tag in a social circle by multiple linear regression. The model uses all features which are mentioned in the previous section. We set  $C$  is a circle, and  $t$  is a tag.  $F(C, t)$  is computed by Equation (6). In the training set, we give tags of every circle high weights, and scores of negative tags as 0. The loss function is Equation (7). We use QR decomposition to find all  $\theta$ s for more numerically stable results. QR decomposition is a method of matrix factorization, it is often used to solve the linear least squares problem.

$$F(C, t) = \sum \theta_n Fea_n \quad (6)$$

$$Loss = \frac{1}{2n} \sum_{k=1}^n (Score(C, t)^k - F(C, t)^k)^2 \quad (7)$$

## 4. Experiment

### 4.1. Dataset

The algorithm is evaluated in both Facebook and academic network. The dataset of Facebook is released in Kaggle [38]. It includes 60 users and their social circles. The 60 users annotate their social circles and its members in Facebook. There are 17,115 friends in all social circles, while every user has 19.73 social circles and every social circle has 28.91 friends averagely. The task is detecting tags for these social circles. The dataset includes all users' networks and tags, it also has the ground-truth of every user's memberships of social circles. We annotate ground-truths of social circle tags by members' tag. There are 227 social circles in the train set and 315 social circles in the test set. We take characteristics of users' relationships and tags as features of multiple linear regression. The results of tags can represent social attributes of circles.

The dataset of this academic paper is extracted from Microsoft Academic Search [39]. There are 50 papers in dataset. The related work is divided into many sections in these papers. All references of every section are relevant to a research problem or a research methodology. So these sections of references can be regarded as ground-truth of reference-circles. The citations among references can be regarded as social relationships. The task is detecting topics and keywords for these circles. We annotate technology terminology in titles and abstracts of these papers, regarding these terminologies

as candidate keywords of every single paper. There are 46 reference circles in the train set and 62 reference circles in the test set.

#### 4.2. Baseline

There is no specified algorithm for tag detection of social circles, and topic analysis methods are also difficult to apply to such sparse data of social circle tags. In the respect of tags mining, we choose popular tags, TF-IDF and frequency-based keyword extraction (FKE) as the baselines. FKE measures weights of all members' tags in the social circle. We rank candidate tags by term-frequency and member-frequency (TF-MF). Given  $T$  is the set of all members' tags of the circle, the tag  $t \in T$ .  $TF_t$  of a tag  $t$  represents occurrence times of  $t$  in  $T$ , and  $|t|$  is the length of tag  $t$ . We define member-frequency as Equation (8) and TF-MF as Equation (9).

$$MF_t = \frac{|\{m : t \in m\}|}{|Members|} \quad (8)$$

$$TF - MF_t = TF_t \times \log_2(MF_t + 1) \times |t| \quad (9)$$

#### 4.3. Result Analysis

A social circle may have several tags that can represent the circles attributes. Therefore, we select the top 10 tags in algorithm results as tags of a social circle. The evaluation metric is precision@10 (The correct tags in top 10). in every social circle (Equations (10) and (11)).

$$P\_Circle = \frac{Correct\ Profiles\ in\ Top\ 10}{10} \quad (10)$$

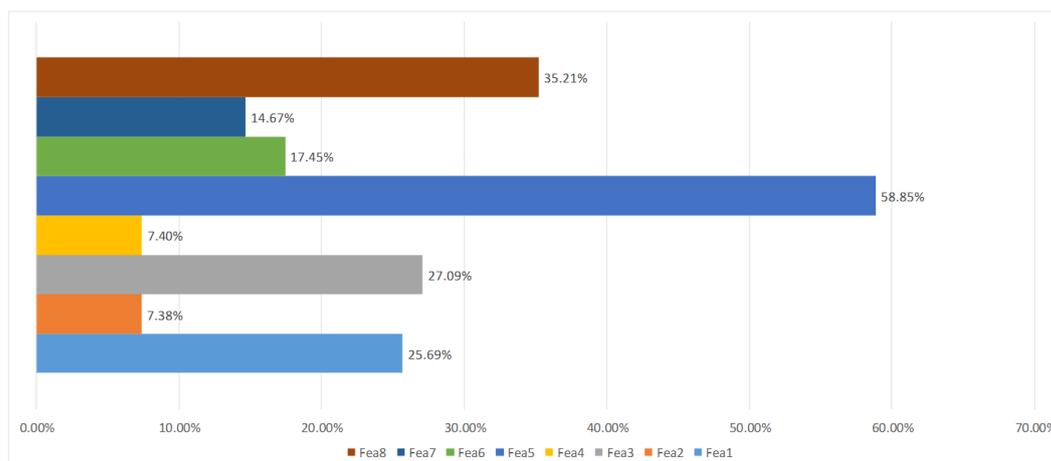
$$Precision = \frac{\sum P\_Circle_i}{Total\ of\ Circles} \quad (11)$$

In both datasets of Facebook and Microsoft Academic Search, our method is better than baselines (Table 2). Our method can extract key tags and keywords of different social circles and different parts of academic related works. The tag detection method from text (FKE) will achieve good results in a large number of posts. However, tag data are sparse, and cannot also consider users' topological structure. The improved precisions are 11% and 23% in two datasets, respectively. That proves that the performances of our work are good in both problems of tag detection for social circles and keyword detection for reference circles.

We run multiple linear regression by every single feature in the Facebook dataset (Figure 3). The results show that TF-IDF is the strongest feature in the tag detection of the social circle. When the model uses only one feature, the trained  $\theta$  can get the best performance by TF-IDF. On the basis of TF-IDF, multiple linear regression combines it with other features and improves detection precision effectively. At the same time, the model can easily transfer to other similar problems about tag detection for social circles.

**Table 2.** Result of Circle Tag Detection.

		Facebook	Microsoft Academic Search
Popular	P@10	28.29%	N/A
FKE	P@10	12.01%	15.08%
TF-IDF	P@10	60.02%	17.10%
<b>Multiple Linear Regression</b>	P@10	<b>71.54%</b>	<b>40.63%</b>



**Figure 3.** Precision of Multiple Linear Regression with Every Single Feature.

## 5. Conclusion

In this paper, we propose a tags detection algorithm for social circles by multiple linear regression. The model infers social meanings of social circles by all members' memberships and their tags. Following the detection of social circles, this work can deeply analyze the attributes of users' social circles. At the same time, this paper transfers the concept of the social circle into the network of academic papers. The model can detect keywords of papers' reference circles. It is beneficial for understanding the topics of paper's references more precisely and in a focused way. In the future, we will try to complement users' tags with their friends in the same social circles. We will also analyze author circles in an academic network according to their research area and co-author relationships.

**Acknowledgments:** This work was supported by the National Basic Research Program (973 Program) of China (No. 2014CB340503), National Natural Science Foundation of China (No. 61133012 and No. 61472107).

**Author Contributions:** Hailong Qin and Jing Liu has developed the model and wrote the manuscript under the guidance of Chin-Yew Lin and Ting Liu.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ajzen, I.; Fishbein, M. *Understanding Attitudes and Predicting Social Behaviour*; Pearson: London, UK, 1980.
2. Roserberg, M. *Society and the Adolescent Self-Image*, revised edition; Wesleyan University Press: Middletown, CT, USA, 1989.
3. Yen, I.H.; Syme, S.L. The social environment and health: A discussion of the epidemiologic literature. *Ann. Rev. Public Health* **1999**, *20*, 287–308.
4. Ståhl, T.; Rütten, A.; Nutbeam, D.; Bauman, A.; Kannas, L.; Abel, T.; Lüschen, G.; Rodriguez, D.J.; Vinck, J.; van der Zee, J. The importance of the social environment for physically active lifestyle—Results from an international study. *Soc. Sci. Med.* **2001**, *52*, 1–10.
5. Wei, C. *Formation of Norms in a Blog Community*; University of Minnesota: Saint Paul, MN, USA, 2004.
6. Goldberg, M.; Kelley, S.; Magdon-Ismael, M.; Mertsalov, K.; Wallace, A. Finding overlapping communities in social networks. In Proceedings of the 2010 IEEE Second International Conference on Social Computing (SocialCom), Minneapolis, MN, USA, 20–22 August 2010; pp. 104–113.
7. De Klepper, M.; Sleebos, E.; van de Bunt, G.; Agneessens, F. Similarity in friendship networks: Selection or influence? The effect of constraining contexts and non-visible individual attributes. *Soc. Netw.* **2010**, *32*, 82–90.
8. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826.
9. Granovetter, M.S. The strength of weak ties. *Am. J. Soc.* **1973**, *78*, 1360–1380.

10. Ferrara, E.; de Meo, P.; Fiumara, G.; Proveti, A. The role of strong and weak ties in Facebook: A community structure perspective. *Commun. ACM* **2012**, doi:10.1145/2629438.
11. Petróczi, A.; Nepusz, T.; Bacsó, F. Measuring tie-strength in virtual social networks. *Connections* **2007**, *27*, 39–52.
12. Gilbert, E.; Karahalios, K. Predicting tie strength with social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Boston, MA, USA, 4–9 April 2009; pp. 211–220.
13. Burton, S.H.; Giraud-Carrier, C.G. Discovering social circles in directed graphs. *ACM Trans. Knowl. Discov. Data* **2014**, *8*, 21.
14. Leskovec, J.; McAuley, J.J. Learning to discover social circles in ego networks. In Proceedings of the Twenty-Sixth Annual Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, CA, USA, 3–8 December 2012; pp. 539–547.
15. Qin, H.; Liu, T.; Ma, Y. Mining User's Real Social Circle in Microblog. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), Istanbul, Turkey, 26–29 August 2012; pp. 348–352.
16. Liu, T.; Qin, H. Detecting and tagging users' social circles in social media. *Multimed. Sys.* **2014**, *22*, doi:10.1007/s00530-014-0435-4.
17. Wang, M.; Morrison, D.; Hayes, C. Information fusion methods for the automatic creation of Twitter lists. *Int. J. Soc. Netw. Min.* **2015**, *2*, 19–43.
18. Gusfield, J.R. *Community: A Critical Response*; Harper & Row: New York, NY, USA, 1975.
19. Sharma, A.; Gemici, M.; Cosley, D. Friends, strangers, and the value of ego networks for recommendation. In Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, Boston, MA, USA, 8–10 July 2013.
20. Liu, T.; Zhang, W.N.; Zhang, Y. SocialRobot: A big data-driven humanoid intelligent system in social media services. *Multimed. Syst.* **2014**, *22*, doi:10.1007/s00530-014-0374-0.
21. Liu, T.; Zhang, W.N.; Cao, L.; Zhang, Y. Question Popularity Analysis and Prediction in Community Question Answering Services. *PloS ONE* **2014**, *9*, e85236.
22. Liu, J.; Zhang, F.; Song, X.; Song, Y.I.; Lin, C.Y.; Hon, H.W. What's in a name? An unsupervised approach to link users across communities. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, Rome, Italy, 4–8 February 2013; pp. 495–504.
23. Huberman, B.A.; Romero, D.M.; Wu, F. Social networks that matter: Twitter under the microscope. *First Monday* **2009**, *14*, 1–5.
24. Kairam, S.R.; Wang, D.J.; Leskovec, J. The life and death of online groups: Predicting group growth and longevity. In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, Washington, DC, USA, 8–12 February 2012; pp. 673–682.
25. Qu, Z.; Liu, Y. Interactive group suggesting for Twitter. In *HLT-Short '08 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; Volume 2, pp. 519–523.
26. Kairam, S.; Brzozowski, M.; Huffaker, D.; Chi, E. Talking in circles: Selective sharing in Google+. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; ACM: New York, NY, USA, 2012; pp. 1065–1074.
27. Savage, S.; Forbes, A.; Toxtli, C.; McKenzie, G.; Desai, S.; Höllerer, T. Visualizing targeted audiences. In Proceedings of the 11th International Conference on the Design of Cooperative Systems (COOP 2014), Nice, France, 27–30 May 2014; pp. 17–34.
28. Huang, J.; Thornton, K.M.; Efthimiadis, E.N. Conversational tagging in Twitter. In Proceedings of the 21st ACM Conference on Hypertext and Hypermedia; ACM: New York, NY, USA, 2010; pp. 173–178.
29. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523.
30. Liu, Z.; Chen, X.; Sun, M. Mining the interests of Chinese microbloggers via keyword extraction. *Front. Comput. Sci.* **2012**, *6*, 76–87.
31. Wang, X.; Zhai, C.; Roth, D. Understanding Evolution of Research Themes: A Probabilistic Generative Model for Citations. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: New York, NY, USA, 2013; pp. 1115–1123.

32. He, Q.; Chen, B.; Pei, J.; Qiu, B.; Mitra, P.; Giles, L. Detecting topic evolution in scientific literature: How can citations help? In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*; ACM: New York, NY, USA, 2009; pp. 957–966.
33. Lu, Z.; Mamouliis, N.; Cheung, D.W. A Collective Topic Model for Milestone Paper Discovery. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*; ACM: New York, NY, USA, 2014; pp. 1019–1022.
34. Chang, J.; Blei, D.M. Relational topic models for document networks. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, Clearwater, FL, USA, 16–19 April 2009; pp. 81–88.
35. Nallapati, R.M.; Ahmed, A.; Xing, E.P.; Cohen, W.W. Joint latent topic models for text and citations. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2008; pp. 542–550.
36. Guo, Z.; Zhang, Z.; Zhu, S.; Chi, Y.; Gong, Y. A Two-Level Topic Model Towards Knowledge Discovery from Citation Networks. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 780–794.
37. Bolelli, L.; Ertekin, S.; Giles, C.L. Clustering scientific literature using sparse citation graph analysis. In *Knowledge Discovery in Databases: PKDD 2006*; Springer: Berlin Heidelberg, Germany, 2006; pp. 30–41.
38. Learning Social Circles in Networks. Available online: <https://www.kaggle.com/c/learning-social-circles/data> (accessed on 15 December 2015).
39. Microsoft Academic Search. Available online: <http://libra.msra.cn/> (accessed on 15 December 2015)



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).