User Identity Linkage with Accumulated Information from Neighbouring Anchor Links

Xiang Li[1,2,3(✉)], Yijun Su[1,2,3], Wei Tang[1,2,3], Neng Gao[2,3], and Ji Xiang[2,3]

[1] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[2] State Key Laboratory of Information Security, Chinese Academy of Sciences, Beijing, China
[3] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{lixiang9015,suyijun,tangwei,gaoneng,xiangji}@iie.ac.cn

Abstract. User identity linkage is to identify all the users belonging to the same individual in different networks and has been widely studied along with the increasing popularity of diverse social media sites. Generally, a pair of probable corresponding users on different networks may form a true "Anchor Link". Most existing methods identify a user based on unique features (username, interests, friends, etc.) and neglect the importance of users local network structure. Therefore, one challenging problem is how to address the user identity linkage problem if only structural information is available. In this paper, we explore techniques for dealing with the fundamental and accumulated information from neighbouring anchor links. Furthermore, we design a Trustworthy Predicting Approach (TPA) for computing the *authority* of an anchor link, inferring the *trustworthiness* of a candidate anchor link being true and predicting whether an anchor link is able to be veritably formed. Experiments illustrate the effectiveness of our proposed algorithm.

1 Introduction

With the vigorous development of Internet in the world, online social networks have revolutionized our daily life and brought us in a "second life". Social networks, such as Facebook, Twitter, Flickr and LinkedIn, make people easy to share their information with other familiar or unfamiliar people. According to the statistical data about Facebook, Twitter and Youtube from 2017 Pew Research Center report[1], more than half of the users tend to acquire information from multiple social media sites as shown in Table 1.

The problem of identifying users across online social networks (also known as *User Identity Linkage*) is valuable and particularly challenging. Mapping users from diverse social platforms can bring many benefits. Discrepant information

[1] http://www.pewresearch.org/fact-tank/2017/11/02/more-americans-are-turning-to-multiple-social-media-sites-for-news/.

© Springer Nature Switzerland AG 2018
H. Hacid et al. (Eds.): WISE 2018, LNCS 11234, pp. 335–344, 2018.
https://doi.org/10.1007/978-3-030-02925-8_24

Table 1. % of each site's news users who get news from...

	Facebook	Twitter	Youtube
Only that site	50%	18%	22%
2 sites	30%	37%	39%
3 or more	20%	45%	39%

of the same user on different social platforms helps to construct a better portrait for corresponding natural person and provide precise and personalized recommendations or advertisements [1,5].

Recent work in user identity linkage often leverages user profiles and user-generated content. However, there are several difficulties for subsequent exploitation. For example, due to the personal preferences and privacy demand, researchers have to face the dilemma that user profiles and user-generated content often behave truthless, incomplete and inconsistent. Therefore, a more challenging and interesting scenario emerges when only social circle is available.

A basic intuition to use social cycle is that when most of a person's friend say "account u on one social platform and account $u^{'}$ on another social platform belong to that person", it seems believable that these two accounts indeed belong to that person. Based on above intuition, a concept "shared identified friends" has been presented and widely used. In this paper, a detailed analysis on the basic concept Shared Identified Friends has been conducted. The main idea behind our method is to differentiate each identified friend according to his *authority*. As a result, we propose a key component called Authority-Trustworthiness Analysis Model to iteratively compute the *trustworthiness* and *authority* of each probable anchor link. By combining Authority-Trustworthiness Analysis Model with the process of anchor link inference, a Trustworthy Predicting Approach (TPA) is presented for solving the problem of user identity linkage purely based on structure information.

The remainder of this paper is organized as follows: Sect. 2 reviews some existing work on user identity linkage. Section 3 describes our analysis model and approach in detail. Experimental evaluation and comparison to other methods are shown in Sect. 4. Finally, Sect. 5 concludes the paper with a brief discussion.

2 Related Work

Existing approaches mainly use user's unique attributes (e.g., name, age, hometown, interests) and content (e.g., post, comment). Great efforts have been made on feature engineering [7,8,11,13,14]. For example, [11] considers distance-based profile features and neighborhood-based network features and iteratively identify unknown user identity pairs. More comprehensively, [8] models heterogeneous behaviors including distance-based profile features, style-based content features, trajectory-based content features and neighborhood-based network features in a semi-supervised manner. Other techniques such as embedding [7,13]

also have been utilized to learn better features. Besides feature extraction, recent works make progress in designing better models like Energy-based model [16] and Latent User Space model [9].

Nevertheless, due to the aforementioned drawbacks (e.g., truthless, incomplete and inconsistent), a fundamental problem is how to solve user identity linkage problem by making full use of structure information. [10] proposes a propagation algorithm to find new links by computing the match score of all probable links based on degree and feedback from previously constructed anchor links. [12] firstly uses the distance vector to initial seed anchors. Then, the authors compare the local network structure by randomized spanning trees and recursive sub-graph matching. Similarly, [2] designs an Unified Similariy (US) measurement by combining the degree centrality, closeness centrality, betweenness centrality of nodes and the relative distance to the initial seed anchors. Besides, [15] presents a local degree-based method and a global embedding-based method for identifying users by only utilizing structure information. [4] gives a better search strategy for generating candidate links to be computed. In the ith phase of the alogrithm, it only allows nodes of degree roughly $D/2^i$ and above to be matched, where D is a parameter related to the largest node degree.

3 Proposed Method

A social platform can be viewed as an undirected network and each node in the network can respresent an account on the social platform. Let $\mathbf{S} = \{S_1, S_2, \ldots, S_{n_s}\}$ denotes a set of different networks and for each $S_i \in \mathbf{S}$, $S_i = (V_i, E_i)$ where $V_i = \{v_1^i, v_2^i, \ldots, v_{n_i}^i\}$ denotes the set of nodes on S_i and $E_i \subseteq (V_i \times V_i)$ denotes undirect links on S_i. Without loss of generality, we focus on two networks S_1, S_2 in this study. This is reasonable because solving the problem of two sites can be easily generalized to the problem of n_s networks in a pairwise manner.

All true anchor links between S_1 and S_2 is denoted as $T = \{(v_i^1, v_j^2) | v_i^1 \in S_1, v_j^2 \in S_2\}$ and $T^p \subseteq T$ represents prior true anchor links known in advance. For each node v_i^m, $N(v_i^m)$ represents the set of nodes linked to node v_i^m on network S_m and $F(v_i^m)$ denotes the set of matched friends among its neighbouring nodes.

3.1 Authority-Trustworthiness Analysis Model

Before presenting the ananlysis model, it is necessary to introduce the basic concept "Shared Identified Friend (SIF)". As shown in Fig. 1, we already know account pair (v_1^1, v_1^2) belong to user A, (v_2^1, v_2^2) belong to user B and so on. In this case, (v_1^1, v_1^2) (or (v_2^1, v_2^2)) is called a shared identified friend for (v_3^1, v_3^2). The set of shared identified friends for (v_3^1, v_3^2) can be represented as $SIF(v_3^1, v_3^2) = F(v_3^1) \cap F(v_3^2)$.

In this paper, only structure information can be taken into consideration. To better solve the problem, we differentiate the function of different shared identified friend. This kind of difference originates from the *authority* of distrinct

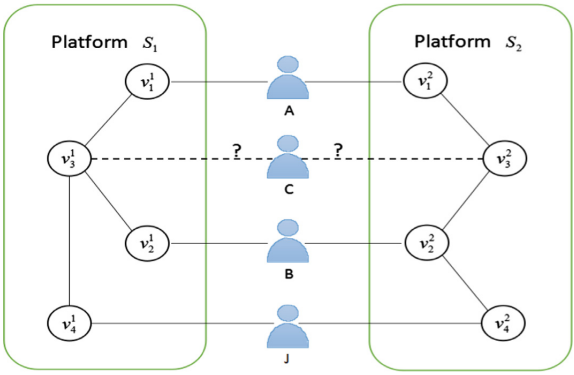

Fig. 1. An example for some notations. A solid line with an arrow on both sides denotes an true anchor link and a dash line is a probable anchor link.

people. By common sense, we have summarized two basic intuitions: (1) The authority of a person can be evaluted by the trustworthiness of judgements he has made. (2) A person that provides mostly true judgements for many objects will likely provide true judgements for another objects. Furthermore, two conclusions, which are interactively promoted, have been drawn from these two intuitions: (1) Conclusion 1: A judgement is more trustworthy if people returning this judegment are more authoritative; (2) Conclusion 2: A person is more authoritative if the judgements returned by this person are more probable to be correct. Based on above intuitions and conclusions, the definitons of two new concepts *Trustworthiness* and *Authority* naturally arises as follows:

Definition 1 ***(Trustworthiness of a judgement):*** *The trustworthiness of a judgement is the probability of this judgement being correct, according to the best of our knowledge.*

Definition 2 ***(Authority of an anchor link):*** *The authority of an anchor link* (v_i^1, v_j^2) *is the expected trustworthiness of the judgements provided by* (v_i^1, v_j^2).

To formulate the Authority-Trustworthiness model expediently, some notations and notions should be provided in advance. We denote $trust(v_i^1, v_j^2) \in [0, 1]$ as the trustworthiness score of a judgement for probable anchor link (v_i^1, v_j^2) and $auth(v_i^1, v_j^2) \in [0, 1]$ as the authority score of an anchor link (v_i^1, v_j^2). For a fixed order of all probable account pairs between two networks, $trust \in R^{n_1 n_2 \times 1}$ and $auth \in R^{n_1 n_2 \times 1}$ separately represent the trustworthiness and authority of all probable account pairs. In addition, a "transition" matrix $M \in R^{n_1 n_2 \times n_1 n_2}$ is defined with the same order of account pairs in *trust* and *auth*. When $(v_p^1, v_q^2) \in SIF(v_i^1, v_j^2)$, the value of $M[v_i^1, v_j^2][v_p^1, v_q^2]$ is set to 1. Otherwise, the value is equal to 0. Naturally, a diagonal matrix D can be defined. Each of element in D is the sum of corresponding row in M.

According to above two conclusions and definitions, our iterative computation model for authority and trustworthiness can formulated as $trust =$

$D^{-1}M \cdot auth$ and $auth = (D^{-1}M)^T \cdot trust$. By viewing this iterative procedure as a HITS algorithm [3], the authroity-trustworthiness analysis model necessarily converges as the number of iteration increases arbitrarily due to the convergence proof of HITS. For each pair (v_i^1, v_j^2), its authority and trustworthiness can be computed as:

$$\begin{aligned} trust(v_i^1, v_j^2) &= \frac{\sum_{(v_p^1, v_q^2) \in SIF(v_i^1, v_j^2)} auth(v_p^1, v_q^2)}{|SIF(v_i^1, v_j^2)|} \\ auth(v_i^1, v_j^2) &= \frac{\sum_{(v_p^1, v_q^2) \in SIF(v_i^1, v_j^2)} trust(v_p^1, v_q^2)}{|SIF(v_i^1, v_j^2)|} \end{aligned} \tag{1}$$

From the view of weighted majority voting, above model only allows shared identified friends to vote. However, everyone identified has the right to vote. For example, for a candidate pair (v_3^1, v_3^2), we know $SIF(v_3^1, v_3^2) = \{A, B\}$ in Fig. 1. If we compute the trustworthiness of this pair as above, we ignore the matched person J, which is unreasonable. Noting that if a matched person such as J is not a shared identified friend of a certain candidate pair, it means this person disagree the corresponding account pair belongs to a real person. In this paper, we think this kind of matched person has no contribution to the trustworthiness and authority of corresponding account pairs, which means the value is zero. Therefore, our analysis model can be modified as:

$$trust(v_i^1, v_j^2) = \frac{\sum_{(v_p^1, v_q^2) \in SIF(v_i^1, v_j^2)} auth(v_p^1, v_q^2)}{|F(v_i^1) \cup F(v_j^2)|} \tag{2}$$

$$auth(v_i^1, v_j^2) = \frac{\sum_{(v_p^1, v_q^2) \in SIF(v_i^1, v_j^2)} trust(v_p^1, v_q^2)}{|F(v_i^1) \cup F(v_j^2)|} \tag{3}$$

3.2 Trustworthy Predicting Approach

In this paper, the analysis model described above is not able to predict. Therefore, a semi-supervised algorithm called Trustworthy Predicting Approach is designed for integrating the authority-trustworthiness anaylsis model and anchor link inference process. As shown in Algorithm 1, in each iteration, we generate the candidate set for each people in identified set. For example, assuming (v_i^1, v_j^2) has been identified, the candidate set of this person is $C(v_i^1, v_j^2) = \{(v_l^1, v_h^2) | (v_l^1, v_h^2) \in (N(v_i^1) \times N(v_j^2) - SIF(v_i^1, v_j^2)), |N(v_i^1) - N(v_j^2)| < window\}$). Then, the trustworthiness of probable anchor links in the candidate set can be computed. After computing the trustworthiness of candidate set, the authority-trustworthiness analysis model is applied to acquire the stable authority and trustworthiness for each link. During the process of analysis process, we repeat the iteration by only considering anchor links whose trustworthiness score is above a low bound.

4 Experiment Study

In this section, we compare the proposed approach with existing baseline methods. The main comparsion methods used in experiments include:

Algorithm 1. Trustworthy Predicting Approach

Input: $S_1 = (V_1, E_1), S_2 = (V_2, E_2)$, prior anchor links $T^p \neq \emptyset$, the threshold *right_low_bound*, degree window size *window*, number of iterations *iter_num*
Output: all identified anchor links T^*

$T^* = T^p$
for k=1,2,...,iter_num **do**
  $temporal\_cand = \emptyset$
  for $(v_i^1, v_j^2) \in T^*$ **do**
    $temporal\_cand = temporal\_cand \cup C(v_i^1, v_j^2)$
  $T^* = T^* \cup temporal\_cand$
  while not reach stable state **do**
    for $(v_i^1, v_j^2) \in (T^* - T^p)$ **do**
      compute trustworthiness score by (2)
      if trust$(v_i^1, v_j^2) < right_low_bound$ **then**
        remove (v_i^1, v_j^2) from T^*
    for $(v_i^1, v_j^2) \in T^*$ **do**
      compute authority score by (3)

- Local Method (LM) [15]: Nodes who has maximum number of identified friends in its network in each iteration are considered as an anchor.
- Global Method (GM) [15]: By constructing the normalized laplacian matrix for each network, the algorithm gives spectral embedding for each node and learns a linear transformation between initial seed nodes in two networks.
- UserMatch [4]: The algorithm uses the concept shared identified friends to calculate the similarity of a candidate node pair and designs a better propagation strategy for reducing the size of candidate set.
- Trustworthy Predicting Approach (TPA): Our proposed method utilize the structure information based on the Authority-Trustworthiness Analysis Model.

Experiment Setups. To evaluate the performance of comparison methods, *Recall* and *F-measure*($F1 = \frac{2*Precision*Recall}{Recall+Precision}$) are considered in this paper. Considering the labeled data required by semi-supervised model, the ratio of prior anchor links shown as (4) needs to be investigated during experiments.

$$prior\_ratio = \frac{|T^p|}{|T^a|} \tag{4}$$

4.1 Experiments on Social Networks

Datasets. Facebook is one of most popular social platforms currently. We use facebook data in Standford Large Network Dataset Collection [6] to assess the performance of different measures. Specific information about facebook dataset is shown in Table 2. Then, we generate two networks from facebook dataset by edge sampling. For each edge, a random value p with the uniform distribution

Table 2. Information about networks

Name	Nodes	Edges	Average degree
Facebook	4029	88234	43.691
Data Mining	20680	71130	6.879
Artificial Intellegence	25674	76141	5.931

Table 3. Experimental results under different prior ratio, $\alpha_o = 0.5, \alpha_s = 0.5$

Metric	Method	0.05	0.1	0.2	0.3	0.4	0.5
Recall	LM	0.0	0.00087	0.00327	0.00225	0.00351	0.00314
	GM	0.00275	0.11283	0.15389	0.17848	0.19315	0.20923
	UserMatch	0.03714	0.07114	0.13425	0.21785	0.27963	0.34399
	TPA	**0.10839**	**0.18047**	**0.28978**	**0.36145**	**0.46752**	**0.53225**
F1	LM	0.0	0.00087	0.00328	0.00225	0.00351	0.00315
	GM	0.00275	0.11283	0.15389	0.17848	0.19315	0.20923
	UserMatch	0.04963	0.08961	0.16273	0.25868	0.32558	0.40110
	TPA	**0.10928**	**0.18219**	**0.29145**	**0.36316**	**0.46875**	**0.53435**

in [0, 1] is generated. Then, if $p \leq 1 - 2\alpha_s + \alpha_o\alpha_s$, the edge will be removed; if $1 - 2\alpha_s + \alpha_o\alpha_s < p \leq 1 - \alpha_s$, the edge will only be kept in first sub-network; if $1-\alpha_s < p \leq 1-\alpha_s\alpha_o$, the edge will only be kept in second sub-network; if $p > 1-\alpha_s\alpha_o$, this edge will be kept in both two sub-networks. Based on above strategy, we know the overlap level is $overlap\_level = \frac{2*(1-(1-\alpha_s\alpha_o))}{1-(1-2\alpha_s+\alpha_o\alpha_s)+1-(1-\alpha_o\alpha_s)} = \alpha_o$. As a result, we call α_o "edge overlap level" and α_s "sparsity level".

Results and Comparison. To evaluate the performace of our TPA method in its entirely, we firstly compare it with baseline methods in different settings. Specifically, different prior ratios in $[0.05, 0.1, 0.2, 0.3, 0.4, 0.5]$ with same overlap level $\alpha_o = 0.5$ and same sparsity level $\alpha_s = 0.5$ are tested. From Table 3, when prior ratio is small enough such as 0.05, TPA behaves better than other methods when prior ratio increases from 0.05 to 0.5. Moreover, LM is always the poorest method under all prior ratios and UserMatch is always the best method between LM, GM and UserMatch. The deviation of Recall and F1 between TPA and UserMatch ranges from 6% to 20%.

In addition, different overlap levels $\alpha_o = [0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$ with same prior ratio $prior\_ratio = 0.3$ and same sparsity level $\alpha_s = 0.5$ are tested. From Table 4, similar to Table 3, LM is always the poorest method under all overlap levels and UserMatch is always the best method between LM, GM and UserMatch. However, our TPA approach still behaves better and exceeds about 10% in average than other methods.

Similarly, different sparsity levels $\alpha_s = [0.3, 0.4, 0.5, 0.6, 0.7, 0.8]$ with same prior ratio $prior\_ratio = 0.3$ and same overlap level $\alpha_o = 0.5$ are also tested.

Table 4. Experimental results under different overlap level, $\alpha_s = 0.5, prior\_ratio = 0.3$

Metric	Method	0.4	0.5	0.6	0.7	0.8	0.9
Recall	LM	0.00188	0.00224	0.00186	0.00297	0.00371	0.00258
	GM	0.1509	0.17848	0.23312	0.31501	0.43091	0.62182
	UserMatch	0.09572	0.21785	0.39910	0.55423	0.65565	0.75636
	TPA	**0.19707**	**0.36145**	**0.58859**	**0.76486**	**0.85587**	**0.93107**
F1	LM	0.00188	0.00225	0.00187	0.00297	0.00372	0.00258
	GM	0.15090	0.17848	0.23312	0.31501	0.43091	0.62182
	UserMatch	0.11371	0.25868	0.46210	0.63747	0.73927	0.83111
	TPA	**0.19890**	**0.36316**	**0.59068**	**0.77001**	**0.85874**	**0.93417**

Table 5. Experimental results under different α_s, $\alpha_o = 0.5, prior\_ratio = 0.3$

Metric	Method	0.3	0.4	0.5	0.6	0.7	0.8
Recall	LM	0.00160	0.00307	0.00224	0.00222	0.00328	0.00292
	GM	0.11658	0.14731	0.17847	0.21686	0.22608	0.26931
	UserMatch	0.25801	0.25615	0.21785	0.20015	0.18955	0.17966
	TPA	**0.45592**	**0.41500**	**0.36145**	**0.33791**	**0.32359**	**0.29155**
F1	LM	0.00161	0.00309	0.00225	0.00224	0.00329	0.00292
	GM	0.11658	0.14731	0.17847	0.21686	0.22608	0.26931
	UserMatch	0.32975	0.31246	0.25868	0.23019	0.21620	0.21467
	TPA	**0.45850**	**0.41862**	**0.36316**	**0.34185**	**0.32658**	**0.29406**

From Table 5, the performance of all methods except GM decreases when the sparsity level increases. TPA always behaves best than other methods in Table 5. In fact, when the sparsity level is greater than 0.8, GM can achieve nearly the same performance as our proposed TPA. After observing this phenomenon, we find that this phenomenon often emerges when α_o or $prior\_ratio$ is small enough and sparsity level is large enough by conducting extensive experiments.

4.2 Experiments on Co-author Networks

Datasets. Co-author networks have been widely adopted in user identity linkage problem. Firstly, 10 representative conferences on Data Mining (DM)[2] and 9 representative conferences on Artificial Intellegence (AI)[3]. Then, we crawl data from DBLP and build a co-author network by the authors of papers from January

[2] The conferences selected from the DM field are KDD, SIGMOD, SIGIR, ICDM, ICDE, VLDB, WWW, SDM, CIKM, and WSDM.

[3] The conferences selected from the AI field are AAAI, IJCAI, CVPR, ICML, NIPS, UAI, ACL, EMNLP and ECAI.

Table 6. Experimental results under different prior ratio

Metric	Method	0.05	0.1	0.2	0.3	0.4	0.5
Recall	LM	0.00027	0.00031	0.00056	0.00062	0.00084	0.00091
	GM	0.00027	0.00031	0.00059	0.00036	0.00042	0.00063
	UserMatch	0.00424	0.00814	0.02224	0.03321	0.04017	0.04989
	TPA	**0.02121**	**0.03314**	**0.06013**	**0.08036**	**0.11247**	**0.12627**
F1	LM	0.00028	0.00033	0.00059	0.00069	0.00088	0.00094
	GM	0.00027	0.00031	0.00059	0.00036	0.00042	0.00063
	UserMatch	0.00833	0.01553	0.04008	0.05646	0.06632	0.08029
	TPA	**0.02194**	**0.03415**	**0.06124**	**0.08221**	**0.11470**	**0.12873**

2010 to September 2017 shown as Table 2. Finally, the shared number of same users $|T^a|$ between DM and AI dataset is 4941.

Results and Comparsion. Because the overlap level of DM-AI dataset is fixed, only prior ratio needs to be considered in experiments. As shown in Table 6, TPA exhibits the best performance on predicting anchor links between the two co-author networks on AI and DM. By experiments on social networks, we know the deviation between TPA and other methods is not largely when the overlap level of datasets or the prior ratio is too small. It is shown that TPA exhibits the best performance on predicting anchor links between co-author networks between DM and AI. The deviation between TPA and other methods ranges from 1% to 4.8%. The recall and F1 of TPA raises rapidly with varying prior ratio.

5 Conclusions

In this paper, we addressed the problem of user identity linkage. Unlike user unique attributes, we explore the power of user's social circle. The heart of our idea is that if most your best friends judge the different accounts on different networks is yours, these accounts are believed to belong to you. To acquire the authority of each friend and the trustworthiness of each final judgement, an Authority-Trustworthiness Analysis Model has been presented. Finally, we design a Trustworthy Predicting Approach to resolve the problem of user identity linkage.

Acknowledgments. This work was partially supported by National Natural Science Foundation of China No. U163620068 and Strategy Cooperation Project AQ-1703 and AQ-17014.

References

1. Carmagnola, F., Cena, F.: User identification for cross-system personalisation. Inf. Sci. **179**(12), 16–32 (2009)
2. Ji, S., Li, W., Srivatsa, M., He, J.S., Beyah, R.: Structure based data de-anonymization of social networks and mobility traces. In: Chow, S.S.M., Camenisch, J., Hui, L.C.K., Yiu, S.M. (eds.) ISC 2014. LNCS, vol. 8783, pp. 237–254. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13257-0_14
3. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. In: ACM-SIAM Symposium on Discrete Algorithms, pp. 668–677 (1998)
4. Korula, N., Lattanzi, S.: An efficient reconciliation algorithm for social networks. Proc. VLDB Endow. **7**(5), 377–388 (2014)
5. Kumar, S., Zafarani, R., Liu, H.: Understanding user migration patterns in social media. In: AAAI Conference on Artificial Intelligence, pp. 1204–1209 (2011)
6. Leskovec, J., Krevl, A.: SNAP datasets: stanford large network dataset collection, June 2014. http://snap.stanford.edu/data
7. Liu, L., Cheung, W.K., Li, X., Liao, L.: Aligning users across social networks using network embedding. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 1774–1780 (2016)
8. Liu, S., Wang, S., Zhu, F., Zhang, J., Krishnan, R.: Hydra: large-scale social identity linkage via heterogeneous behavior modeling. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD 2014, pp. 51–62. ACM, New York (2014)
9. Mu, X., Zhu, F., Wang, J., Wang, J., Wang, J., Zhou, Z.H.: User identity linkage by latent user space modelling. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1775–1784 (2016)
10. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: 2009 30th IEEE Symposium on Security and Privacy, pp. 173–187, May 2009
11. Shen, Y., Jin, H.: Controllable information sharing for user accounts linkage across multiple online social networks. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, pp. 381–390. ACM, New York (2014)
12. Srivatsa, M., Hicks, M.: Deanonymizing mobility traces: using social network as a side-channel. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS 2012, pp. 628–637. ACM, New York (2012)
13. Tan, S., Guan, Z., Cai, D., Qin, X., Bu, J., Chen, C.: Mapping users across networks by manifold alignment on hypergraph. In: AAAI Conference on Artificial Intelligence (2014)
14. Zafarani, R., Liu, H.: Connecting users across social media sites: a behavioral-modeling approach. In: 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, vol. Part F128815, pp. 41–49. Association for Computing Machinery, August 2013
15. Zafarani, R., Tang, L., Liu, H.: User identification across social media. ACM Trans. Knowl. Discov. Data **10**(2), 16:1–16:30 (2015)
16. Zhang, Y., Tang, J., Yang, Z., Pei, J., Yu, P.S.: COSNET: connecting heterogeneous social networks with local and global consistency. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1485–1494 (2015)