

A Threat to Friendship Privacy in Facebook

Francesco Buccafurri^(✉), Gianluca Lax, Serena Nicolazzo,
and Antonino Nocera

DIIES, University Mediterranea of Reggio Calabria, Via Graziella,
Località Feo di Vito, 89122 Reggio Calabria, Italy
{bucca,lax,s.nicolazzo,a.nocera}@unirc.it

Abstract. The rapid growth of social networks, primarily Facebook, has coincided with an increasing concern over personal privacy. This explains why more and more users personalize their Facebook privacy settings. As a matter of fact, the list of friends is often one of the profile sections kept private, meaning that this information is perceived as sensible.

In this paper, we study the robustness of this privacy protection feature, showing that it can be broken even in the less advantageous conditions for the adversary. To do this, we exploit both the potential information extracted from user alter accounts in Twitter and a social network property, recently demonstrated for Twitter, called interest assortativity. The preliminary experimental results reported in this paper, give a first evidence of the effectiveness of our attack, which succeeds even in the most difficult case that is when the information about the victim are minimum.

Keywords: Facebook privacy · Assortativity · Identity management

1 Introduction

Facebook's privacy setting gives a user the possibility to choose who is allowed to see their profile information. Hence, a user who wants not to reveal his friend list information to everyone can specify to hide such an information in the privacy setting page. By default, everyone can see the friends of a user. As a matter of fact, there are many Facebook users having a private list of friends, meaning that this information is often (reasonably) perceived as sensible. In this paper, we study the robustness of this privacy protection feature, showing that it can be broken even in the less advantageous conditions for the adversary. Obviously, if the adversary knows the victim in the real life or has information about the contexts in which the victim lives, he can easily guess some Facebook profiles owned by a real-life friend of the victim whose friend list is public. It is rather intuitive that only few seeds are enough to discover incrementally large portions of private friends, as usually friends form highly connected clusters. Therefore, this case is trivial. But we want to consider the most difficult case. The adversary has only the name of the victim and the link to his Facebook profile, he can

guess only some general information about him (nationality, for example), but has no information about his real life, his job, his interests, etc. This case, for example, may occur in Web investigation. Again, the privacy of the list of friends can be broken once only a few friends (even one) is found with public profile. But, how to find them? Since guessable general information selects a very large portion of Facebook users, it would seem that the only way for the adversary is to try an infeasible guess-and-check attack. In this paper, we show that, by exploiting a social network property recently demonstrated for Twitter [9], a much more efficient attack is possible, allowing the adversary to break the privacy of the victim in the most cases. This short paper includes the first experimental evidence of this result, thus encouraging us to more deeply analyze this issue in the next future.

The plan of this paper is as follows. In Sect. 2, we present our approach to discover private friendships in Facebook. Section 3 describes the preliminary experimentation carried out to study the effectiveness of our technique. Section 4 deals with literature related to our work. Finally, in Sect. 5, we draw our conclusions.

2 Approach Formalization

In this section, we describe the technique we propose to discover (at least a part of) friends of a social network user who decided to make private his friend list. First, we observe that our approach works for social networks, such as Facebook, in which friendship relations are symmetric (that is, if the user u_1 is friend of u_2 , then also u_2 is friend of u_1).

The intuition underlying our approach is that privacy setting of an account is indicated by the account owner, meaning that u_1 can choose to make private his friendship with u_2 , whereas u_2 can choose to make public his friendship with u_1 . Consequently, by looking at u_2 's account, the friendship between u_1 and u_2 can be inferred even though u_1 tries to hide it. It is worth noting that the mere execution of the strategy sketched above has a strong limitation that makes this trivial search unfeasible. Indeed, due to the huge number of social network accounts, the search space of possible friends is limitless. Moreover, this strategy returns at most one friend for each possible friend analyzed.

To overcome these drawbacks, we designed a technique more sophisticated than the above one to reach two important advantages. The first one is to provide a relatively limited number of accounts to analyze (say *candidates*), thus reducing the search space and making this solution feasible. The second advantage is that, thanks to a suitable selection of each candidate, the processing of each candidate account is able to return more friends of the initial account u_1 (to obtain this, we exploit the mechanism of friend community present in social networks).

The technique used to discover private friends basically relies on three procedures, *find alter accounts*, *select candidates*, and *find common friends*. For the sake of presentation, we describe at high level how our proposal works, whereas the detailed implementations of the above three procedures are provided in Sects. 2.1, 2.2 and 2.3, respectively.

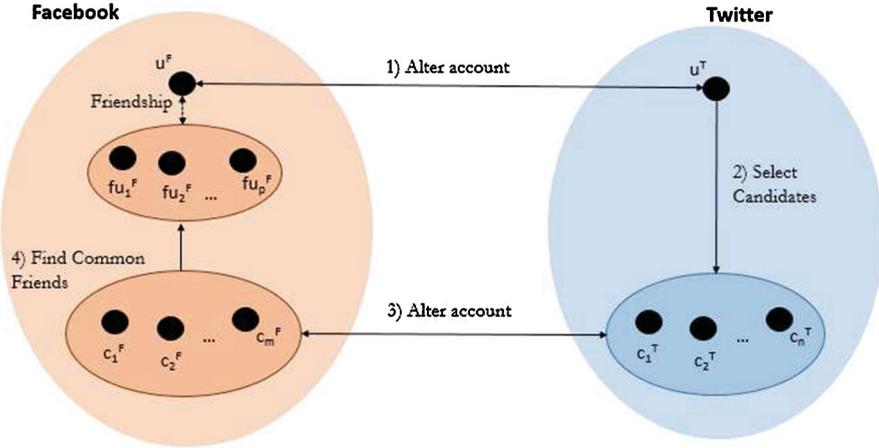


Fig. 1. Graphical representation of the approach.

The input of our technique is u^F , which is the account of the user u in a social network F that supports symmetric friendship relations. In the follow, we instantiate F with Facebook, the most popular social network. Clearly, the friend list of u^F is private. The output of our technique is a set of accounts $f_i u^F$ that are friends of u^F in Facebook. Our approach is schematized in Fig. 1 and consists of 4 steps.

1. In order to discover private friends of u^F , the first step we run is *finding alter accounts* of u^F . This step aims to identify a secondary account of u in another social network (how to perform this task is described in Sect. 2.1). It is well known that users register account on different social networks and use them for different purposes [8, 26]. Among all the social networks in which u has registered an account, we are interested in his secondary account in Twitter: indeed, Twitter is a very famous and common social networks, which is used to exchange very short messages. At the end of this step, we obtain u^T , which is the account of the user u in Twitter.
2. Now, we run the second step of our technique, that is *select candidates*. This step aims to identify Twitter accounts that are in the same community of u and that can lead to discover the private friends of u^F . Among the three steps, this is the core one and is deeply explained in Sect. 2.2. Let c_1^T, \dots, c_n^T be the set of candidates outputted at this step.
3. In this step, we run the procedure *finding alter accounts* for each candidate c_j^T , in order to discover his/her account on Facebook, say c_j^F . At the end of this step, we have found some accounts on Facebook that hopefully are in the same Facebook community of u^F (because they are alter accounts of users in the same community of u in Twitter).
4. In the last step, for each c_j^F , we run the procedure *finding common friends* in order to find the list of friends in common between c_j^F and u^F . This procedure,

which can appear *magic*, is instead provided by many social networks in order to give members the opportunity to find new friends. The detail on the implementation of this step is given in Sect. 2.3. At the end of this step, we obtain the set $f_i u^F$ containing some (hopefully all) private friends of u^F .

From the high-level description of how our technique works, it is clear we cannot guarantee that this approach is always able to break the privateness of friend list: indeed, it is necessary that alter accounts are found and friend community exists. However, as we will show in Sect. 3, we experimented that for many real-life accounts, the execution of this technique is able to discover at least a portion of private friend list.

In the next sections, we will describe the implementation of the three procedures used in our technique.

2.1 Finding Alter Accounts

Many social networks provide their users with the possibility to add in their own profiles a link toward one of their accounts in another social site or external website. This feature is typically enabled during the creation of the user profile. This information is extremely useful in our approach because it allows the identification of the accounts belonging to the same person in a multi-social network scenario.

Technically speaking, users who explicitly declare their alter accounts via social network tools, physically create special links among social networks. These special links are referred as *me* edges [10, 22]. Information about these inter-social network links can be extracted in several ways. The basic strategy leverages the use of social network APIs, a set of methods and services, typically available for social network developers, allowing the interaction with social-network data and functionalities to create new software on top of them.

However, not all social networks provides APIs to extract this information. Therefore, another possibility to extract alter accounts relies on the XFN (*XHTML Friends Network*) standard, an HTML microformat to represent relationship among user accounts. This is obtained by empowering the set of values that the `rel` attribute of the HTML tag `<a>` (which represents a link) can assume. In particular, the value “me” (i.e., `rel='me'`) is used to indicate that the corresponding `<a>` link represents a *me* edge.

Finally, in cases in which users do not declare explicitly their alter accounts, several approaches proposed by the scientific community can be applied to detect missing *me* edges (see, for example, [12, 19, 27, 33]).

In summary, alter accounts can be retrieved by using social network APIs, XFN, or techniques such as those defined in [12, 19, 27, 33]. Observe that, in the case of Facebook and Twitter, XFN is adopted for declaring alter accounts.

2.2 Selecting Candidates

The approach we follow to obtain the set of candidates leverages the concept of *assortativity* in social networks. Assortativity is an empirical measure describing

a positive correlation in personal attributes of people socially connected with each other [29]. Hence, if a network is assortative with respect to a given attribute, it means that the majority of its users tend to act as their friends when it comes of the aspect expressed by that particular attribute.

It is proved that Twitter shows assortative behavior with respect to user interests [11], where interest assortativity is defined as the preference for friends to share the same interest (e.g., sport, music). Indeed, in Twitter there exist accounts belonging to *public figures*, which, due to their influence w.r.t. a specific topic, act as a sort of representative for that topic [15]. This way, the abstract concept of interest (or topic) can be mapped to the concrete entity of a public figure.

Assortativity w.r.t. an interest I , say IA_I , is given by the difference between the fraction of the users interested in I having at least one friend interested in I measure in the real network, and that computed in the *random* graph corresponding to the real network, said *null model* [29]. Indeed, the random graph models the case in which no assortativity occurs and is obtained by preserving the nodes of the social network and replacing the deterministic occurrence of edges by a random variable in such a way that node degree distribution is unaltered. More in details, the fraction referred above has as numerator the number of nodes that: (1) follow a public figure associated with the interest I and (2) have at least a friend who is follower of any other public figure representative of I ; the denominator is the total number of nodes following any public figures associated with I .

Thanks to the assortative behavior of Twitter users, we can find people belonging to his clique by searching on the neighbors of a public figure of interest for a given user, obtaining a set of suitable candidates for our technique.

2.3 Finding Common Friends

By following the reasoning described in Sect. 2.2, we can extract a set of Twitter accounts, which are potentially “close” to the user, say u , for whom we want to reconstruct the friend list. Now, for each of the candidates obtained we can adopt the strategy described in Sect. 2.1 to find their alter-accounts in Facebook, thus obtaining a set of Facebook-candidate accounts.

At this point, if these candidates have a public friend list, we can verify whether u is present in any of these lists. Each time we find an account in any of these lists, due to the bidirectional nature of Facebook friendships, we have discovered an element of the friend list of u . Because the set of candidates may be very big, it is very likely that we can reconstruct the entire friend list of u . However, this strategy would require a very large number of checks to verify the membership of u in any of the friend list of the candidates.

Fortunately, Facebook provides a very powerful set of APIs, namely Graph API [1], that can give us some advantages in our objective. Graph APIs are a low-level HTTP-based APIs useful to retrieve data from Facebook.

Specifically, it is possible to discover the mutual friends between two Facebook users by performing an HTTPS request at the following link:

`https://www.facebook.com/friendship/<screen-name1>/<screen-name2>`

where the parameters `<screen-name1>` and `<screen-name2>` represent the string ids used to univocally identify the two users in Facebook. Interestingly, this API method works also if one of the two accounts (that of u in our case) has the friend list private. Due to the tendency of Facebook users to form cliques [31] the above API method allows us to discover a very large set of items of the friend list of u with a single call. This reduces the number of operations required to reconstruct the entire friend list.

3 Preliminary Evaluation

Our experiments were performed on a machine equipped with a 2 Quad-Core E5440 processor and 16 GB of RAM. The operating system was Linux Ubuntu Server 14.04.4 LTS, with kernel version 4.2.0-35, Java Virtual Machine version 1.8.0 45 (64-Bit) and Twitter4J [2] as external library for Twitter API support. We wrote our implementations in Java.

To obtain the initial set of Facebook profiles to test our approach performance, we could not rely on existing datasets as they do not provide information about `me` edges. To extract necessary data, we exploited the SNAKE system [14] which allows the extraction of profile contact information from a very large set of social networks. One of the main issues in this activity was the detection of the profiles showing the right features for our investigation (i.e., an alter account in Twitter and a private friend list). However, using one of the classical crawling technique [30] is not suitable for two main reasons:

1. First, the percentage of accounts with a `me` edge is very low [13], so it is extremely difficult to find these particular users. This implies that an almost complete visit should be performed to obtain necessary information. However, full structural information of the network is not needed, because we are interested only in Facebook users with an alter account in Twitter.
2. Secondly, a crawling technique may privilege the visit of some nodes with particular structural properties (i.e., very high degree) introducing some biases in our results.

As a consequence, we decided to perform uniform sampling. Although, uniform sampling is not a trivial task in general, for Facebook and Twitter, it is facilitated by how user identifiers are organized. Indeed, both social networks adopt 64-bit identifiers for user accounts. However, because we are looking for private Facebook accounts, we cannot start our sampling from Facebook. We started by uniformly sampling Twitter to collect accounts having a `me` edge towards Facebook¹. In particular, the URL address of the profile page of Twitter has the following structure: `http://twitter.com/account/redirect_by_id?id=xxx`, where `xxx` is a 64-bit positive integer. Hence, to obtain a uniform sampling, we generated numbers uniformly at random in a suitable interval and then we checked if it

¹ Observe that, the direction of the `me` edge is not relevant for our objective, because we only need to know that the two accounts belong to the same person.

corresponds to a real account with a me edge towards Facebook and whose alter account in Facebook has a private friend list. From this sampling step we obtained a set of 355 accounts.

We proceeded by analyzing the set of Twitter friends for each of the accounts above, and we selected public figures among them. In our case, because the set of candidates (see Sect. 2) is a subset of followers of these public figures, we neglected those public figures who have a too high in-degree and considered only accounts with an in-degree ranging from 500 to 1500. Clearly, this choice was made only to guarantee computation feasibility, and did not affect the performance of our technique as will be shown in the following.

After this, we considered the Facebook alter accounts of the candidates, extracted in the previous step, and continued by calling the Graph API method described in Sect. 2.3 to verify whether we were able to reconstruct the private friend list of the original set of 355 users. As a result, we succeeded in 259 cases, thus obtaining a probability of success of 0.73.

4 Related Work

The more social networks take a central role in people everyday life, the more users privacy becomes a critical issue for researchers. As stated in [4], although a considerable part of Facebook users are not aware of privacy options or do not use them, there is an always increasing number of active users having more online privacy literacy.

Indeed, a number of studies [5, 16, 18, 23, 24, 28] analyse users behaviour when it comes of privacy in social networks. In particular, the authors of [28] measure the disparity between what users desire and their actual privacy settings, and perform an analysis of problems emerging from a not proper management of privacy. They found that privacy settings match users' expectations only in 37% of the time, exposing content to more users than expected. Hargittai et. al [23] examine how users privacy practices have changed over time according to modifications on Facebook privacy settings. However, concerning the problem faced in this paper, no change of privacy setting can affect neither the issue nor the solution. The study presented in [16] aims at exploring relationships between concern about mediated lurking and strategic ambiguity on Facebook, and Facebook privacy management. Moreover, the authors of [5, 24] focus on privacy settings of adolescents on Facebook. As explained in [17] unintentional disclosure to friends and acquaintances on Facebook can led to bullying/meanness and unwanted contacts, especially among adolescents. Other work surveys users' awareness, attitudes, and privacy concerns towards profile visibility and show that only a minority of users change the default privacy preferences on Facebook [3, 21].

From Facebook birth, privacy flaws have continued to keep appearing. For instance, in 2006 Facebook introduced a new functionality, called "News Feed", which tracks and displays the online activities of a user's friends on start pages of the user. Although none of the individual actions were private, users felt deprived of their sense of control over their information and began to form protest groups on Facebook [6]. Subsequently, Facebook introduced privacy controls that allowed users to determine what was shown on the news feed and to whom.

The information gained can be exploited for a number of reasons [21, 25]. An attacker could, for instance, deduce social security numbers (which are often derived from name, gender, and date of birth) from the information posted on the user profiles

[7,21]. Moreover, data on relationships or common interests in groups can be exploited for phishing [25].

In order to mitigate privacy threats, authors of [20] propose a recommender system that suggests privacy settings automatically learned for a given profile (cluster) of users. Whereas, in [32] the authors investigate what strategies undergraduate students have developed in addition to the use of the default privacy settings, to protect their privacy on Facebook. These are excluding contact information, using the limited profile option, untagging and removing photographs, and limiting friendship requests from strangers.

Despite these attempts and in addition to the low level of users privacy awareness, there exist more complex ways to bypass privacy setting on Facebook. Indeed, our work shows how, leveraging Twitter data, a user can discover the private Facebook friend list of a victim. This combined use of multi-social network resources adds further privacy concerns to the above literature.

5 Conclusion

Although a considerable part of Facebook users is not aware of privacy options or do not use them, there is an always increasing number of active users who decide to protect their personal information by restricting the access to their profile. One of the most sensitive profile information is the friend list. In this paper, we described a possible attack on the privateness of the list of friends.

For this purpose, we started by the observation that if the adversary has information about the contexts in which the victim lives, he can easily guess some Facebook profiles owned by the victim real-life friends, and from them reconstruct some portions of the victim friend list (as usually friends form highly connected clusters).

Therefore, our attention moved toward the most difficult case: The adversary knows the minimum information that is only the name of the victim and his Facebook profile. In this scenario, it would seem that the only way for the adversary is to try an infeasible guess-and-check attack. In this paper, we showed that, a much more efficient strategy is possible, allowing the adversary to bypass Facebook privacy settings and break the privacy of the victim in most cases. Our attack exploits the concept of alter accounts combined with a recently studied property, named interest assortativity.

Starting from the victim Facebook profile, we first identify his alter account on Twitter (if any), and then, thanks to interest assortativity, we are able to select some suitable candidates that can lead to some public friends in common with the victim, thus breaking his privacy. The attack incrementally proceeds, by discovering the most of private friends.

This short paper includes the first experimental evidence of this result, thus encouraging us to more deeply analyze this issue in the next future.

Acknowledgment. This work has been partially supported by the Program “Programma Operativo Nazionale Ricerca e Competitività” 2007–2013, Distretto Tecnologico CyberSecurity funded by the Italian Ministry of Education, University and Research.

References

1. Facebook Graph API Documentation (2016). <https://developers.facebook.com/docs/graph-api>
2. Twitter4J (2016). <http://twitter4j.org/>
3. Acquisti, A., Gross, R.: Imagined communities: awareness, information sharing, and privacy on the facebook. In: Danezis, G., Golle, P. (eds.) PET 2006. LNCS, vol. 4258, pp. 36–58. Springer, Heidelberg (2006)
4. Bartsch, M., Dienlin, T.: Control your facebook: an analysis of online privacy literacy. *Comput. Hum. Behav.* **56**, 147–154 (2016)
5. Beldad, A.: Sealing one’s online wall off from outsiders: determinants of the use of facebook’s privacy settings among young dutch users. *Int. J. Technol. Hum. Interact. (IJTHI)* **12**(1), 21–34 (2016)
6. Boyd, D.: Facebook’s privacy trainwreck. *Convergence: Int. J. Res. New Media Technol.* **14**(1), 13–20 (2008)
7. Buccafurri, F., Lax, G.: Implementing disposable credit card numbers by mobile phones. *Electron. Commer. Res.* **11**(3), 271–296 (2011)
8. Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A.: Comparing twitter and facebook user behavior: privacy and other aspects. *Comput. Hum. Behav.* **52**, 87–95 (2015)
9. Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A.: Interest assortativity in twitter. In *Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST)*, WEBIST (2016)
10. Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A.: A model to support design and development of multiple-social-network applications. *Inf. Sci.* **331**, 99–119 (2016)
11. Buccafurri, F., Lax, G., Nocera, A.: A new form of assortativity in online social networks. *Int. J. Hum. Comput. Stud.* **80**, 56–65 (2015)
12. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: Discovering links among social networks. In: Flach, P.A., De Bie, T., Cristianini, N. (eds.) ECML PKDD 2012, Part II. LNCS, vol. 7524, pp. 467–482. Springer, Heidelberg (2012)
13. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: Moving from social networks to social internetworking scenarios: the crawling perspective. *Inform. Sci.* **256**, 126–137 (2014)
14. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: A system for extracting structural information from social network accounts. *Softw. Pract. Experience* **45**(9), 1251–1275 (2015)
15. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in twitter: the million follower fallacy. *ICWSM* **10**(10–17), 30 (2010)
16. Child, J.T., Starcher, S.C.: Fuzzy facebook privacy boundaries: exploring mediated lurking, vague-booking, and facebook privacy management. *Comput. Hum. Behav.* **54**, 483–490 (2016)
17. Christofides, E., Muise, A., Desmarais, S.: Risky disclosures on facebook the effect of having a bad experience on online behavior. *J. Adolesc. Res.* **27**(6), 714–731 (2012)
18. Debatin, B., Lovejoy, J.P., Horn, A.-K., Hughes, B.N.: Facebook and online privacy: attitudes, behaviors, and unintended consequences. *J. Comput. Mediated Commun.* **15**(1), 83–108 (2009)
19. Gani, K., Hacid, H., Skraba, R.: Towards multiple identity detection in social networks. In: *Proceedings of the International Conference Companion on World Wide Web (WWW 2012 Companion)*, Lyon, France, pp. 503–504. ACM (2012)

20. Ghazinour, K., Matwin, S., Sokolova, M.: Yourprivacyprotector, a recommender system for privacy settings in social networks (2016). arXiv preprint [arXiv:1602.01937](https://arxiv.org/abs/1602.01937)
21. Gross, R., Acquisti, A.: Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, pp. 71–80. ACM (2005)
22. Gyongyi, Z.I.: Social affinity on the web, 15 October 2013. US Patent 8,560,605
23. Hargittai, E., et al.: Facebook privacy settings: who cares? *First Monday* **15**(8) (2010)
24. Hofstra, B., Corten, R., van Tubergen, F.: Understanding the privacy behavior of adolescents on facebook: the role of peers, popularity and trust. *Comput. Hum. Behav.* **60**, 611–621 (2016)
25. Jagatic, T.N., Johnson, N.A., Jakobsson, M., Menczer, F.: Social phishing. *Commun. ACM* **50**(10), 94–100 (2007)
26. Jain, P., Kumaraguru, P., Joshi, A.: @ i seek‘fb. me’: Identifying users across multiple online social networks. In: Proceedings of the 22nd International Conference on World Wide Web Companion, pp. 1259–1268. International World Wide Web Conferences Steering Committee (2013)
27. Korula, N., Lattanzi, S.: An efficient reconciliation algorithm for social networks. In: Proceedings of the International Conference on Very Large Data Bases (VLDB 2014), pp. 377–388, Hangzhou, Cina. VLDB Endowment (2014)
28. Liu, Y., Gummadi, K.P., Krishnamurthy, B., Mislove, A.: Analyzing facebook privacy settings: user expectations vs. reality. In: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, pp. 61–70. ACM (2011)
29. Newman, M.E.: Assortative mixing in networks. *Phys. Rev. Lett.* **89**(20), 208701 (2002)
30. Rao, N., Iyengar, S., DeSaussure, G.: The visit problem: visibility graph-based solution. In: Proceedings of the 1988 IEEE International Conference on Robotics and Automation, pp. 1650–1655. IEEE (1988)
31. Van Cleemput, K.: Friendship type, clique formation and the everyday use of communication technologies in a peer group: a social network analysis. *Inf. Commun. Soc.* **15**(8), 1258–1277 (2012)
32. Young, A.L., Quan-Haase, A.: Privacy protection strategies on facebook: the internet privacy paradox revisited. *Inf. Commun. Soc.* **16**(4), 479–500 (2013)
33. Zafarani, R., Liu, H.: Connecting users across social media sites: a behavioral-modeling approach. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, pp. 41–49 (2013)