Practical Privacy-Preserving Friend Recommendations on Social Networks

William Brendel †, Fangqiu Han †, Luís Marujo †, Luo Jie †, Aleksandra Korolova ‡
Snap Research † University of Southern California ‡

ACM Reference Format:
William Brendel †, Fangqiu Han †, Luís Marujo †, Luo Jie †, Aleksandra Korolova ‡. 2018. Practical Privacy-Preserving Friend Recommendations on Social Networks. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3184558.3186954

1 INTRODUCTION

Friend recommendations, whose goal is to expand the connections between users and increase their engagement on the network, is an essential problem for social networks. A particular challenge in friend recommendations is in making recommendations in a *cold-start* situation. This situation occurs when a new user has just registered and, as result, the model does not yet have sufficient information to directly provide recommendations.

Friend recommendations also raise privacy concerns, as they may leak friendship relationships between people on the social network. Knowledge of such relationships may reveal sensitive information about a user, namely their political or sexual preferences [14], medical issues [8], or even de-anonymize their anonymous identities [6, 9]. The easiest and most common way to learn people's relationships is through a brute-force attack that creates fake identities on the graph, connects them to the target user, and then observes friend recommendations that are based on the target user's friends, and therefore, leak their social graph [2, 6].

As more users access social networks through their mobile phones, their phone contact books represent a valuable source of information for bootstrapping the recommendations in the cold-start situation. Our main contribution is to propose that the phone contact book can also be used to better protect the privacy of the users' friend graphs when making friend recommendations, describe a straw-man approach for doing so, and measure its impact on recommendation quality through experiments.

2 PREVIOUS WORK: THEORY VS. PRACTICE

Approaches to privacy in friend recommendations in current social media i) don't promise privacy, ii) hope for privacy by obscurity (e.g., by using many signals in recommendations and not giving explanations as to why recommendations are made [7]) and iii) use ad-hoc thresholds to deter the simplest attacks (typically, thresholds are applied to the number of common friends two users have before one can be recommended to the other). The only attempt to rigorously address the question in practice we are aware of is a very recent one by Signal [13], but it solves a much more restricted problem than ours – privately learning whether contacts in an address book are Signal users.

Theoretical work suggests taking the approaches of structured graph perturbation [15] and randomization of probabilities of recommendation [1, 11] in order to achieve the rigorous privacy guarantee of differential privacy [4]. Although differential privacy has gained traction in practice in other data-mining applications [5], none of the proposed theoretical approaches for friend recommendations have been deployed, possibly due to their prohibitive trade-offs between privacy and utility [12] (i.e., they often recommend people one has no chance of knowing, which has a significant negative impact on the quality of the user experience).

3 OUR APPROACH

Rather than trying to protect privacy against all possible attackers, we observe that most practical attacks i) create a finite number of fake accounts, interconnect them between themselves and with some of the existing nodes in the graph whose privacy is to be violated, and observe the recommendations [2, 6] ii) utilize an auxiliary graph to de-anonymize an anonymized graph [10, 16]. Since we are not releasing an anonymized graph, our proposed solution focuses on making type i) attacks difficult.

Hence, rather than modifying an existing friend prediction algorithm to make it privacy-preserving, we modify the graph used by the algorithm: we create a candidate graph that is more robust than the original graph against the practical brute-force attacks. Our modification takes into account the contact book graph when deciding which subgraph of the social graph can be used. Although there is extensive work on using auxiliary graph information for de-anonymization (e.g., [10]), to the best of our knowledge, we are the first to use an auxiliary graph to improve privacy.

Friend recommendation algorithms typically operate on a friend graph, which may be directed and weighted according to user interaction metrics. From this friend graph, for each user A for whom we'd like to recommend friends, we create a candidate graph (Fig. 1c) as follows:

(a) we consider the set of users, denoted by V_p^C in Figure 1a), who have A's phone number in their contact books;
(b) we add to the candidate graph of A all users in V_p^C and all friends of users in V_p^C who have at least k friends in V_p^C (see Fig. 1b).⋆

We then run the algorithm we would have run on the friend graph on the created candidate graph.

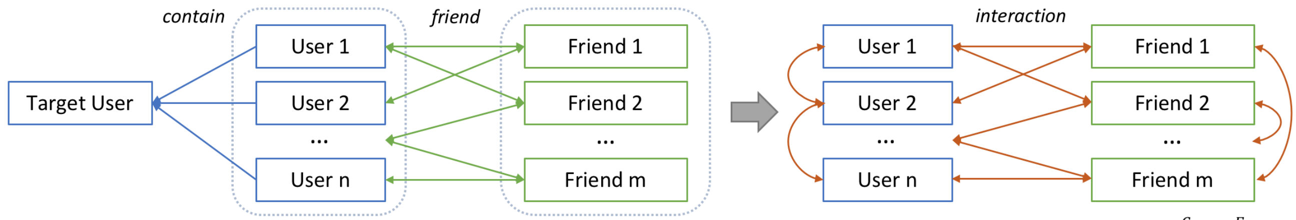
The rationale for (a) is inspired by PageRank: it may be easy for an attacker to create "fake accounts" and connect to the user they are trying to target, but it is difficult to get the user to add the fake accounts to his phone contact book. The rationale for (b) is that

† firstname.lastname@snap.com | Snap Research, 63 Market St, Venice, CA 90291
‡ lastname@usc.edu | USC, 941 Bloom Walk, Los Angeles, CA 90089

WWW '18 Companion, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3186954

⋆ V_p^C is for $V_{privacy}^{Contacts}$, and V_k^F is for $V_{with\ at\ least\ \ k\ connections}^{Friends}$.



(a) Node set V_p^C (b) Node set V_k^F (c) Candidate graph $G(V, E)$ with $V = V_p^C \cup V_k^F$

Figure 1: Building the friend candidate graph for a target user. (a) V_p^C = Users in the friend graph that contain the target user's phone number in their contact book. (b) V_K^F = friends of V_p^C that have at least k friendship connections with them. (c) The friend candidate graph has the node set = $V_p^C \cup V_K^F$. Edge weights encode candidate interactions, e.g., the number of chat messages exchanged, amount of content created, etc.

	V_p^C	$V_p^C \cup V_0^F$	$V_p^C \cup V_k^F$	V_d^C	$V_d^C \cup V_0^F$	$V_d^C \cup V_k^F$
$\|V\|$	36.8	666.8	396.9	47.1	649.4	398.5
$\|E\|$	37.3	4913.4	2524.9	49.4	4491	2372.6
UES	42.6	**829.9**	629.6	40.52	704.8	537

Table 1: Avg. # of graph nodes $|V|$, of edges $|E|$, and avg. top5 UES for multiple candidate graph variations. In *blue*: our final node set.

tricking k users into doing that is even more difficult, as an attacker need not only create a clique of fake friends in the network, but also to connect the fake clique to real users, while creating a pattern.

The choice of k represents a trade-off between privacy and utility. In practice, it makes sense to choose k in a way that depends on the friendship graph sparsity but with an element of randomness; we use $k \leftarrow \text{rand}\big[\max\{k_{med} - \delta,\ k_{min},\ 2\}, \min\{k_{med} + \delta, k_{max}\}\big]$, where $\delta \geq 2$, and k_{min}, k_{max}, k_{med} are the min, max and median number of friends per user among the original pool of candidates.

We note that Snapchat uses a user's contact book only if the user has opted-in to share its contents. Furthermore, we utilize Bloom filters [3] to encode and check contact book relationships, rather than explicitly storing the contact books.

4 EXPERIMENTS

We now experimentally study the impact that using the friend candidate graph instead of the actual friend graph may have on utility. We measure utility in terms of the characteristics of the graph and the user engagement score (UES) calculated internally by Snapchat (based on chat messages and snaps sent in a time window) for such recommendations.

The node set V of our proposed candidate graph $G(V, E)$ is the union of the phone number containing contacts V_p^C (Figure 1a) and the pruned 1-hop friends V_k^F (Figure 1b). We also consider the following variations of graphs that could potentially be used: $V = V_p^C$, $V = V_d^C$, $V = V_p^C \cup V_0^F$, $V = V_d^C \cup V_k^F$ and $V = V_d^C \cup V_0^F$. We define V_d^C as the set of all contacts found in a target user's contact book that are also on the network; V_0^F as the whole set of 1-hop friends in the network's friend graph of the previously selected contacts V_p^C or V_d^C, with 0 instead of k in the subscript meaning we don't prune the selected 1-hop friends based on their number of connections with V_p^C or V_d^C in the network friend graph. We test these candidate graph variations on a dataset of 1.4 million of Snap users using a leave-one-out strategy to simulate new users.

Our utility results are encouraging and can be seen in Table 1. In particular, we find that using the friend candidate graph V_p^C instead of V_d^C slightly improves user engagement performances for a similar structure of the friend candidate graph. Further, including 1-hop friend nodes (V_0^F or V_k^F) dramatically boosts the number of potential choices for recommended friends and the user engagement, by a factor of 13.25 to 19.5. Although using V_k^F instead of V_0^F gives a much smaller graph, i.e., 40.5% fewer nodes and 48.6% fewer edges, the potential user engagement gain decreases by less than 24.2%, which may be an acceptable cost to pay for raising the privacy bar.

Finally, we A/B tested our system at full scale on the entire Snap network. We found that compared to random candidate selection using $V_p^C \cup V_0^F$, our framework produced a gain of 12.7% new friendships and improved the user engagement up to 6.5%.

5 CONCLUSION

We proposed a framework to recommend friends on social networks while raising the bar on privacy of our users' friends lists against a brute-force attacker. We found that by utilizing the phone contact book, one can raise the privacy bar and improve the utility of recommendations, as compared with other privacy-preserving alternatives.

REFERENCES

[1] Rediet Abebe and Vasileios Nakos. 2014. Private Link Prediction in Social Networks. Technical Report, Harvard University. (2014).
[2] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. 2007. Wherefore Art Thou R3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography. In *Proc. of the WWW*.
[3] Peter C Dillinger and Panagiotis Manolios. 2004. Bloom filters in probabilistic verification. In *Inter. Conf. on Formal Methods in Computer-Aided Design*.
[4] Cynthia Dwork. 2011. A Firm Foundation for Private Data Analysis. In *Communications of the ACM*.
[5] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proc. of the CCS*.
[6] Ashley Feinberg. Mar 30, 2017. This Is Almost Certainly James Comey's Twitter Account. In *Gizmodo*.
[7] Kashmir Hill. Aug 25, 2017. Facebook Figured Out My Family Secrets, And It Won't Tell Me How. In *Gizmodo*.
[8] Kashmir Hill. Aug 29, 2016. Facebook recommended that this psychiatrist's patients friend each other. In *Splinter News*.
[9] Kashmir Hill. Oct 11, 2017. How Facebook Outs Sex Workers. In *Gizmodo*.
[10] S. Ji, W. Li, N. Gong, P. Mittal, and R. Beyah. 2015. On Your Social Network De-anonymizablity: Quantification and Large Scale Evaluation with Seed Knowledge. In *NDSS*.
[11] Vishesh Karwa, Sofya Raskhodnikova, Adam Smith, and Grigory Yaroslavtsev. 2011. Private analysis of graph structure. *PVLDB* 4, 11 (2011).
[12] Ashwin Machanavajjhala, Aleksandra Korolova, and Atish Das Sarma. 2011. Personalized Social Recommendations - Accurate or Private? *PVLDB* 4, 7 (2011).
[13] M. Marlinspike. Sep 26, 2017. Private contact discovery for Signal. (Sep 26, 2017).
[14] A. Mislove, B. Viswanath, K. Gummadi, and P. Druschel. 2010. You Are Who You Know: Inferring User Profiles in Online Social Networks. In *WSDM*.
[15] Prateek Mittal, Charalampos Papamanthou, and Dawn Song. 2013. Preserving link privacy in social network based systems. In *NDSS*.
[16] Arvind Narayanan and Vitaly Shmatikov. 2009. De-anonymizing social networks. In *30th IEEE Symposium on Security and Privacy*.