How to Publish Friends Privately

Hyoungshick Kim
hk331@cl.cam.ac.uk

Joseph Bonneau
jcb82@cl.cam.ac.uk

Supervisor: Ross Anderson
Ross.Anderson@cl.cam.ac.uk

Public Search Listings in Facebook

Facebook (www.facebook.com) is one of the popular social networking websites. Facebook permits anybody to access friends information of each user through "public search listing" which includes a simple profile and a list of 8 friends simple profiles for promotional purposes. An example is shown in Figure 1.

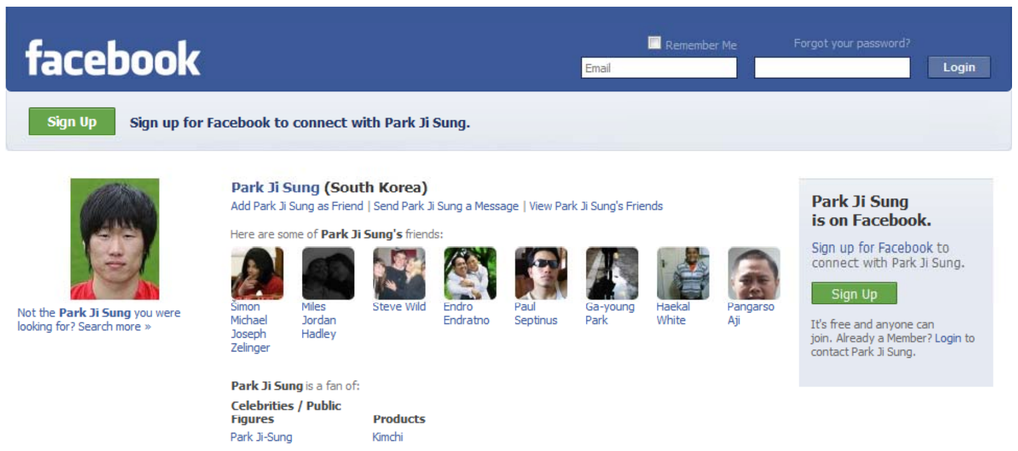

Figure 1: An example of public search listings

I Know How Many Friends You Have

An adversary can collect data about the network using "public search listing" and build up a public graph G_P. Figure 2 shows the construction process of G_P.

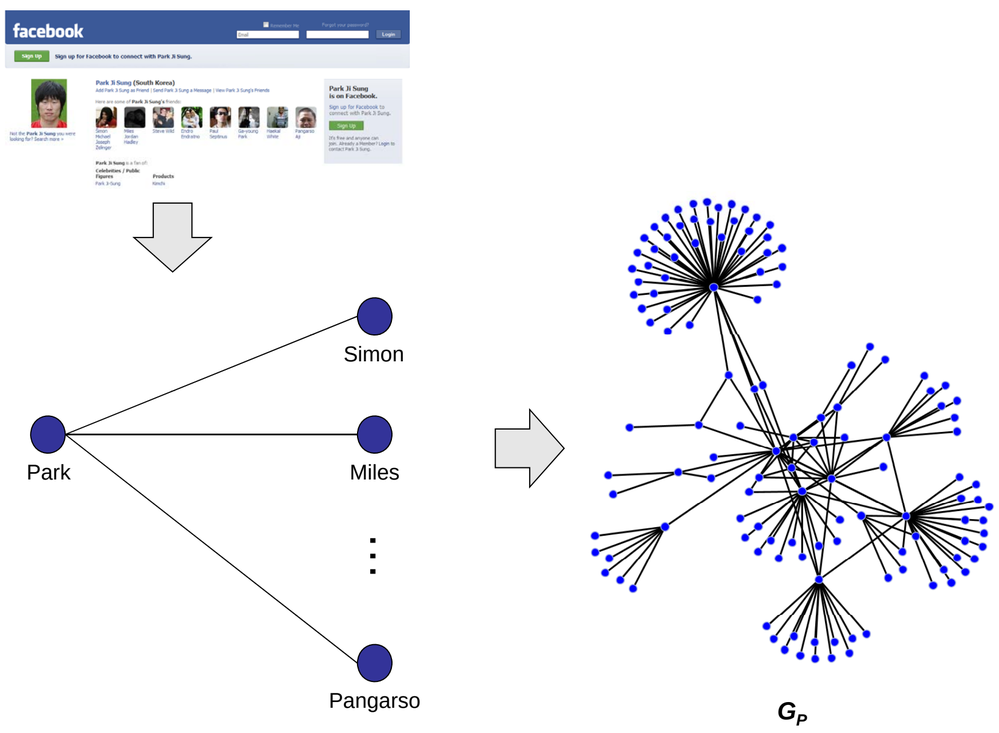

Figure 2: Construction of a public graph G_P

At first glance, "public search listing" seems to be a well designed compromise solution between privacy and utility, revealing an equal number of edges for each user. However, Bonneau at al. showed that the constructed public graph G_P leaks critical information about the original network structure [1]. In particular, the adversary can estimate degree information of each node with high accuracy. This is because links associated with each user are revealed by not only her own public list but all of her friends' public lists. Facebook failed to control the incoming edges of each node.

Threat Model

Our adversary's goal is simply to choose a few "influential members" of the network. For example, a marketer may wish to advertise their product to a few nodes within a network and maximise the reach of subsequent word-of-mouth promotion. To evaluate this goal, we define the edge coverage metric which is the ability to maximise the fraction of edges in the network covered by a target set of users *S*. An adversary's effectiveness can be defined as the amount of edges covered by the target set.

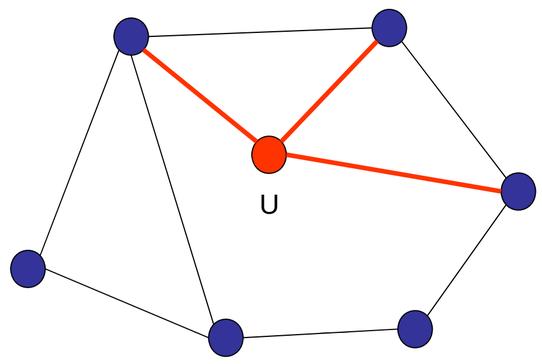

Figure 3: The covered edges by node *U*

Finding the optimal target set *S* is *NP-hard* even if the underlying network is completely given [2]. However, heuristic algorithms based on nodes' degree have been shown to perform well in practice. In addition, Figure 4 shows the attack is still effective with the public graph G_P instead of the original network.

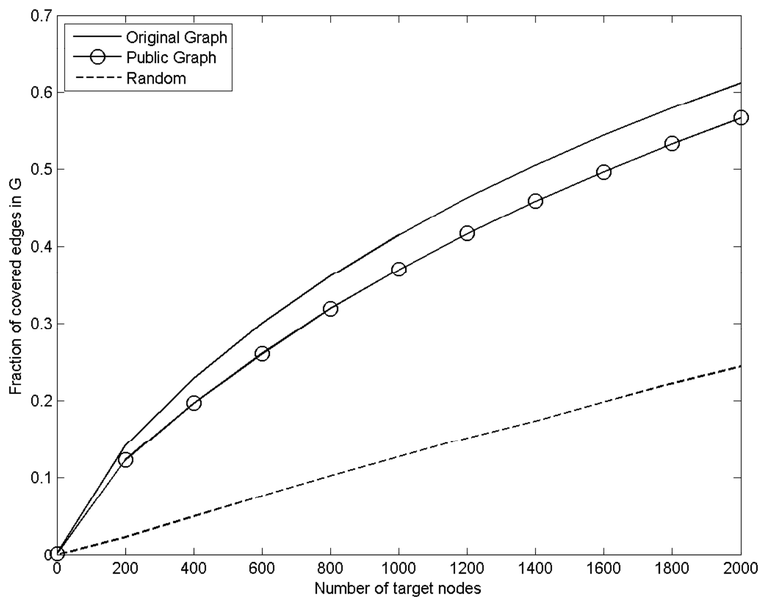

Figure 4: The covered edges by node *U*

How to Design New Public Search Listings

We observe that a high-degree user's degree in the original network is still maintained in the public graph because the user is exposed in her friends' public search listings in proportion to the user degree. In order to offset the user degree, we propose the public search listings via the weighted sampling. The main idea is to select a low-degree user with higher probability compared to other high-degree friends.

The weighted sampling performed much better leaking less information against the attack while still being useful for promotional purposes.

References

[1] Joseph Bonneau, Jonathan Anderson, Ross Anderson, and Frank Stajano. "Eight Friends are Enough: Social Graph Approximation via Public Listings", In Proceedings of Second ACM Workshop on Social Network Systems, 2009.

[2] R. M. Karp. "Reducibility Among Combinatorial Problems", In Complexity of Computer Computations, 1972.

UNIVERSITY OF CAMBRIDGE
Computer Laboratory
Security Group

Microsoft Research
Summer School 2009