Privacy Preservation in Social Networks with Sensitive Edge Weights

Lian Liu, Jie Wang, Jinze Liu, and Jun Zhang

Department of Computer Science, University of Kentucky, Lexington, KY 40506

ABSTRACT

- With the development of emerging social networks, such as Facebook and MySpace, security and privacy threats arising from social network analysis bring a risk of disclosure of confidential knowledge when the social network data is shared to the public. We study a situation in which weights are attached to network edges that are considered to be confidential.
- We consider perturbing the weights of some edges to preserve data privacy when the network is published, while retaining the shortest path and the approximate cost of the path between some pairs of nodes in the original network. We develop two privacy-preserving strategies for this application. The first one is based on a Gaussian randomization multiplication, the second one is a greedy perturbation algorithm based on graph theory. In particular, the second strategy not only yields an approximate length of the shortest path while maintaining the shortest path between selected pairs of nodes, but also maximizes privacy preservation of the original weights.

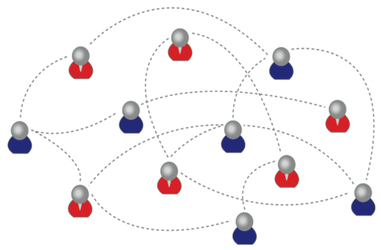
Social Network Analysis: Promising but Challenging

Source: Facebook.com

Social networks become more and more **popular**! 200 million active users in Facebook as of Feb 2009.

Public Awareness: **Privacy** Protection.

Weight Privacy



In social networks, how often you connect with someone is weight. High weight probably stands for a close personal relationship which somebody want to hide/perturb. "I DON'T WANT MY MYSPACE GIRLFRIEND KNOWS MY FACEBOOK GIRLFRIEND."

Our Goals

- **Data Privacy** --- all edge's weights
- **Data Utility** --- shortest paths and

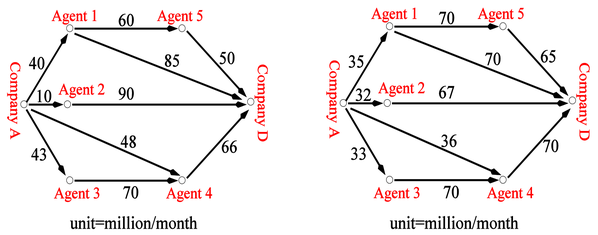
Challenges:
How do we carefully change individual weights without unacceptable impacts on shortest paths and lengths?

Our goals:
- Perturb weights as much as possible,
- keep shortest paths and corresponding lengths the same as original ones as much as possible.

Gaussian Perturbation

$$w^*_{i,j} = w_{i,j}\,(1-x_{i,j}),$$

Here x_{ij} is a randomly generated number from the Gaussian distribution $N(0,\sigma^2)$.

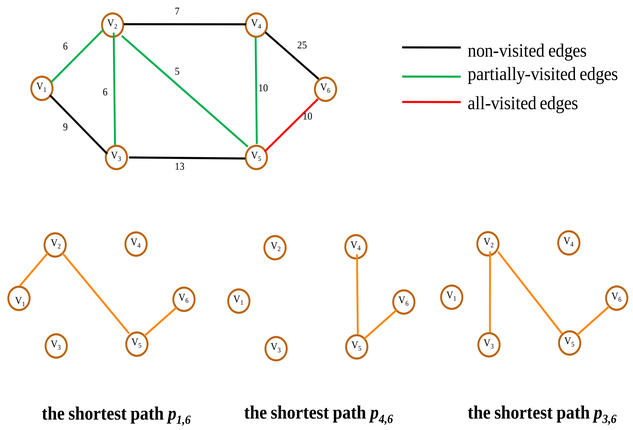


- Privacy: almost all weights are changed.
- Utility: Same shortest path between Company A and Company D and length is 99.

It is high possible for analysts to keep the shortest paths if we carefully choose σ such that σ is smaller than the ratio of the length of a shortest path to the length of the second shortest path.

Reference:
Lian Liu, Jie Wang, Jinze Liu, and Jun Zhang. *Privacy Preservation in Social Networks with Sensitive Edge Weights*. 2009 SIAM International Conference on Data Mining (SDM09), Sparks, Nevada, April 30--May 2, 2009.

Edge Categorization



Greedy Perturbation

Claim 1: For a non-visited edge, increasing its weight will NOT change all shortest paths (and lengths) in *H*.

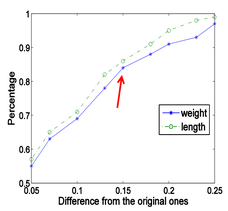
Claim 2: For an all-visited edge, decreasing its weight will NOT change all shortest paths in *H*, but decrease the length of corresponding shortest paths.

Claim 3: For a partial-visited edge, if we want to increase its weight, we should guarantee the shortest paths, which go through it, will still go by this edge after perturbation.

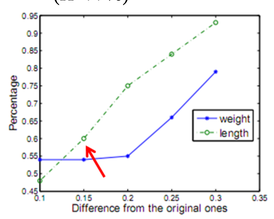
Claim 4: For a partial-visited edge, if we want to decrease its weight, we should guarantee the shortest paths, which do not go through it, will not change after perturbation.

Experiment

Gaussian Perturbation (σ=0.1)

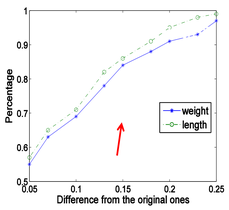


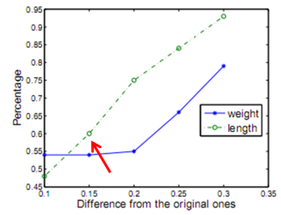
Greedy Perturbation (H=77%)



Simply, the blue line stands for data privacy (bigger value → worse privacy), while the dashed line is for data utility (bigger value → better utility).

Discussion on Experiments





	Data Utility	Data Privacy
Gaussian	Lengths of shortest paths are better preserved, cannot guarantee maintain the exact shortest path.	Low
Greedy	Length is not well preserved compared to Gaussian. But the shortest paths are exactly maintained.	High

Conclusions

(What do we want to do?)
Keep weight privacy and the shortest path utility.

(Why do we want to do?)
Weights in some social cases are sensitive and confidential.

(How do we do?)
Gaussian perturbation and greedy perturbation arep proposed to achieve the balance between data utility and data privacy in different conditions.

(What we do is applicable?)
It seems that the two strategies do meet the expectation of our purpose.

UNIVERSITY OF KENTUCKY College of Engineering