PhD proposal on

Anonymization in Social Networks

Cristina Bazgan, Université Paris-Dauphine, bazgan@lamsade.dauphine.fr
and Janka Chlebikova, University of Portsmouth, janka.chlebikova@port.ac.uk

With the enormous extension of various social networks, the protection of privacy has become an important and active field of research [9, 10]. Social networks may contain sensitive information about individuals. The usual representation of social networks is based on graphs, where nodes correspond to individuals or other social entities and edges correspond to relationships between entities. In many situations, the goal is to preserve some general properties of the networks while ensuring anonymity of its entities. Anonymity of an entity is usualy achieved by a slight modification of its features to make several entities similar and thus undistinguishable.

A classical way to anonymize data is encryption and hashing. However, these methods are insufficient for a real protection since the malicious users can relatively easily analyze the data and produce inferences based on prior knowledge of their target.

Several more sophisticated anonymization techniques have been proposed in the literature. Generalization gathers records according to certain criteria in order to hide individual records. Data perturbation consists of modifying the data without disrupting the statistical significance of these data. k-anonymization is also a common method where each sample of data is considered anonymous if each record is indistinguishable from at least $(k-1)$ others. We refer to Wu et al. [9] for a survey of anonymization models.

Several such models were studied from an algorithmic point of view. The problem of checking if a graph contains an anonymous family was considered in [4]. Several algorithms for anonymization by clustering are already known, see [1, 2] for a detailed study of various clustering methods. The computational complexity of making a given graph k-anonymous were studied and heuristics were proposed either through vertex/edge deletions [3, 8] or through vertex/edge insertions [5, 6, 7]. A main task is to define and study optimization problems based on existing or new models of anonymization.

Generally, most of these optimization problems are NP-hard. The goal of this PhD thesis is to identify polynomial-time solvable cases and establish exact and approximation algorithms for solving NP-hard variants. Other approaches, as exact and approximate parameterized algorithms, may also be considered.

The PhD candidate must have a good background in combinatorial optimization, exact and parameterized complexity, approximation algorithms, and graph theory. This fully funded position will take place in Paris-Dauphine University but also allow visiting stays at the University of Portsmouth.

References

[1] F. Abu-Khzam, C. Bazgan, K. Casel, H. Fernau, Clustering with Lower-Bounded Sizes, *Algorithmica*, to appear.

[2] G. Aggarwal, R. Panigrahy, T. Feder, D. Thomas, K. Kenthapadi, S. Khuller, A. Zhu, Achieving anonymity via clustering, *ACM Trans. Algorithms*, 6(3), 49, 2010.

[3] C. Bazgan, R. Bredereck, S. Hartung, A. Nichterlein et G. Woeginger, Finding large degree-anonymous subgraphs is hard, *Theoretical Computer Science*, 622, pp. 90–110, 2016.

[4] A. Bettinelli, L. Liberti, F.Raimondi, D. Savourey, The Anonymous Subgraph Problem, *Computers & OR*, 40(4), pp. 973–979, 2013.

[5] S. Chester, B. M. Kapron, G. Ramesh, G. Srivastava, A. Thomo, S. Venkatesh, Why Waldo befriended the dummy? k-Anonymization of social networks with pseudo-nodes, *Social Network Analysis and Mining*, 3(3), pp. 381–399, 2013.

[6] S. Hartung, A. Nichterlein, R. Niedermeier, O. Suchý, A Refined Complexity Analysis of Degree Anonymization on Graphs, *Information and Computation*, 243, pp. 249–262, 2015.

[7] K. Liu, E. Terzi, Towards identity anonymization on graphs, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 93–106, 2008.

[8] H. Moser, D. M. Thilikos, Parameterized Complexity of Finding Regular Induced Subgraphs, *Journal of Discrete Algorithms*, 7(2), pp. 181–190, 2009.

[9] X. Wu, X. Ying, K. Liu, L. Chen, A survey of privacy-preservation of graphs and social networks, in: *Managing and Mining Graph Data*, Springer, pp. 421–453, 2010.

[10] E. Zheleva, E. Terzi, L. Getoor, Privacy in Social Networks, *Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers*, 3(1), pp. 1–85, 2012.