Improving (n,m)-greedy edge deletion anonymisation using global heuristic

When does greedy not stay ahead?

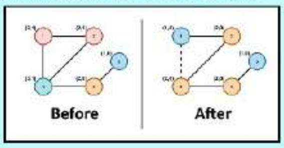
Student: Jakub Matyja
Supervisor: Dr. Anna L.D. Latour

TU Delft

1. Background

Network anonymisation is a crucial process in ensuring that constituent nodes **cannot be uniquely identified**. In particular, this quality is sought-after in **publicly available** networks. This problem is proven to be **NP-hard**, however, various methods and heuristics were introduced in the literature.

Greedy edge deletion method provides **outstanding** reduction in graph uniqueness for limited number of edge deletions but the solution quality tends to **stagnate** w.r.t the number of iterations when more edges are removed from the network. This method relies on (n,m)**-anonymity**, which determines whether nodes in the network can be identified based on the **number of neighbours n and incident triangles m.**



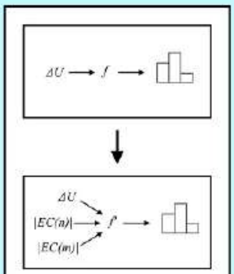
2. Research Question

How does including the size of equivalence classes in edge evaluation influence the performance of (n,m)-greedy edge deletion anonymisation?

- How does the final uniqueness of the network change?
- How does the number of available deletions impact the change in network uniqueness?
- How does the time required to complete the algorithm change?

3. Approach

- The original algorithm finds **locally** optimal edge deletions, which lead to a promising decrease in network uniqueness for a **small number of iterations**. It ranks and removes the edges iteratively according to the predicted change in the uniqueness caused by removing them.
- Literature shows that greedy anonymisation is **outperformed** in the context of solution quality by other strategies for many edge deletions. In particular, prioritising **low-degree** nodes is suggested to contribute to their success. We speculate that this preference is influenced by equivalence classes and their sizes, based on which the uniqueness of the network is determined.
- The aim is to devise a **globally-oriented** heuristic capturing the greedy algorithm's performance for few deletions and improving the solution quality provided by it as more edges are removed from the network.
- Hence, we propose modifications to the algorithm that rank and remove the edges based on the existing estimation of the change in uniqueness combined with the **sizes of equivalence classes** of nodes connected by the considered edge — $|EC(n)|$ and $|EC(m)|$.



4. Results

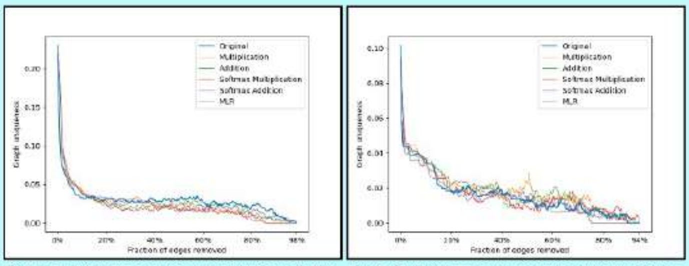


Figure 1a: Performance on the social network email-univ. **Figure 1b**: Performance on the sparse network mahindas.

Figure 1: Performance in terms of graph uniqueness of the proposed algorithms against the original algorithm for selected networks.

5. Conclusions

- The chosen heuristic **improves solution quality** for large number of deletions while maintaining **comparable results** for fewer removals on all inspected types of networks other than dense graphs, i.e. social and sparse networks. Proposed algorithms using this heuristic also **require considerably fewer iterations** to fully anonymise these networks.
- Introduced method leads to **constant computational and memory overhead** for each iteration of the algorithm. It performs computations solely on **already existing information and data**. Indirectly, using fewer edge deletions to anonymise the network also contributes to lower total running time.
- Depending on the network, there exist **intervals or thresholds** of the number of deletions, for which at least some of the proposed algorithms using the chosen heuristic **outperform** greedy in terms of graph uniqueness.
- Deletion-based anonymisation leads to **higher prevalence of low-degree equivalence classes** and their **larger sizes.** This phenomenon likely contributes to the improved performance of the devised modifications, which **favour larger ECs.**

6. References

[1] X. Xie, *Anonymization Algorithms for Privacy-Sensitive Networks*, LIACS, Leiden University, 2023.
[2] Chester et al., *Complexity of social network anonymization*, Social Network Analysis and Mining, 2012.
[3] R. G. de Jong, M. P. J. van der Loo, and F. W. Takes, *The anonymization problem in social networks*, 2025.
[4] R. G. de Jong, M. P. J. van der Loo, and F. W. Takes, *A systematic comparison of measures for k-anonymity in networks*, 2024.