Comparative analysis of two network anonymization settings using Integer Linear Programming

For the course CSE3000 Research Project - Author: Mike J.J.S. Erkemeij; Supervisor Dr. Anna L.D. Latour

1. NTRODUCTION

Network anonymity is a critical privacy concern, especially in social networks . These kinds of networks often contain sensetive personal information that are not always adequatly protected. Researchers are also effected by this, as they often hesitate to publish or share data, ultimatly hindering open science [6].

2. STATE OF THE ART

In an unpublished work by Latour a complete anonymization method is implemented as an Integer Linear Programming (ILP) encoding.

- (n,m)-k-anonymity
- deleting edges
- Gurobi

3. MOTIVATION

While deleting edges is observed to be computationally more effieient, in social networks, adding edges is often regarded as the better privacy-preserving technique [4].

- Better in preserving graph properties in scale-free networks.
- creating connections is more natural
- Preserves original data

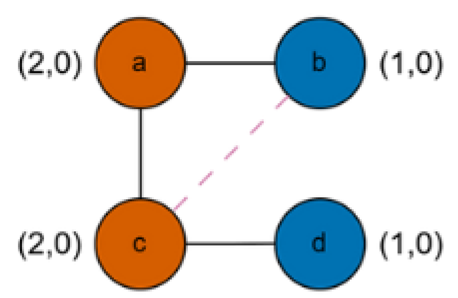
4. RESEARCH QUESTION

How does an ILP implementation of the setting in which we add edges compare to the one in which we delete edges, in terms of solving time, memory consumption and quality of the solution?

5. APPROACH

We build on top of the existing ILP encoding. L is the set of edge variables and T is the set of triangle variables.
Constraints are the same as we still employ (n,m)-k-anonymity

delete



Variables

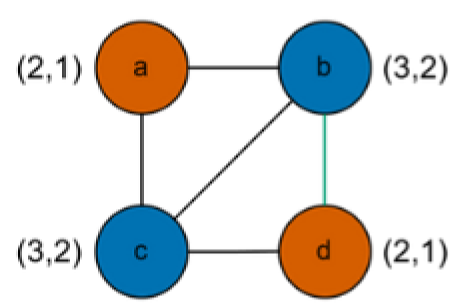
$$L := \{\ell_{(u,v)} \mid \{u,v\} \in E\}$$

$$T := \{t_{(u,v,w)} \mid \{u,v,w\} \in \Theta\}$$

objective function

$$\max \sum_{\ell \in L} \ell$$

add



Variables

$$L := \{\ell_{(u,v)} \mid \{u,v\} \notin E, u \neq v\}$$

$$T := \{t_{(u,v,w)} \mid \{u,v,w\} \notin \Theta, u \neq v \neq w\}$$

objective function

$$\min \sum_{\ell \in L} \ell$$

6. HEURISTIC

As deleting edges has been shown to be computationally more efficient on social networks [3], we introduce a heuristic, intended to reduce running time and memory usage.

- Only adds edges within a node's 2-neighborhood
- Mimics natural connections through mutals

7. DATASETS

Social Datasets:

- Networkrepository [5]

Synthetic Datasets:

- Erdös-Rényi [2]
- Barabási-Albert [1]

Dataset	$\|V\|$	$\|E\|$	$\|\Theta\|$	GT
insecta-ant-trophallaxis-colony1-day1	30	37	12	0.14
insecta-ant-trophallaxis-colony1-day8	37	68	39	0.17
mammalia-baboon-grooming-group07	16	41	51	0.27
mammalia-raccoon-proximity-24	14	41	162	0.66
mammalia-raccoon-proximity-3	23	50	96	0.48

8. METHODOLOGY

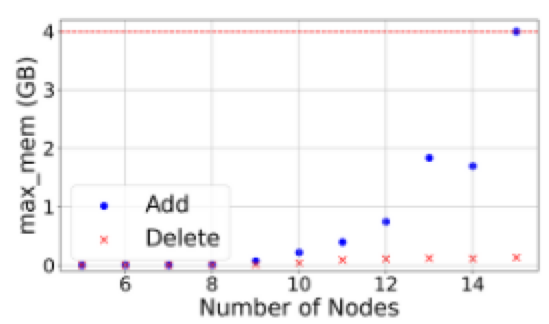
To answer our research question, we run the different approaches on the synthetic networks and social networks. These are compared on running time, memory usage and quality of the solution. We aim to gain insight on the following aspects:

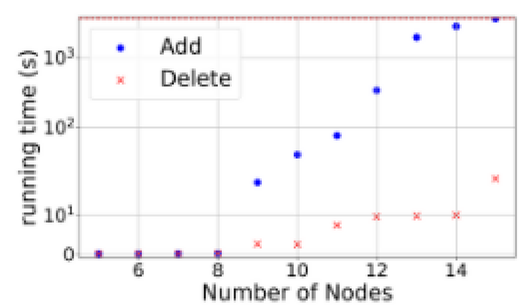
- Scalability on network size
- Behavior on network topology
- Impact on graph properties

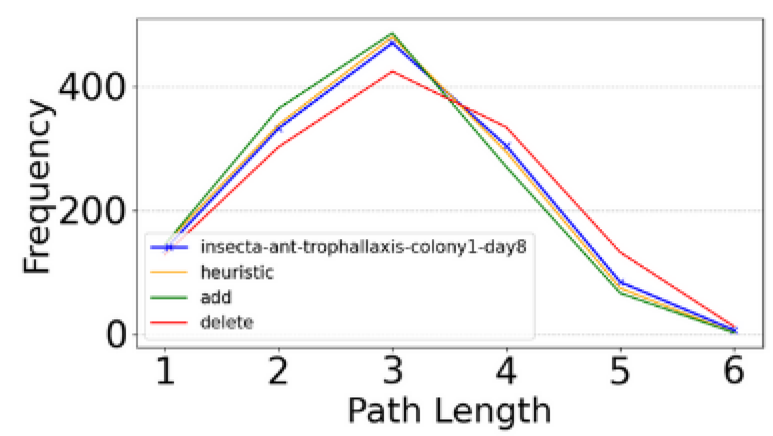
9. RESULTS

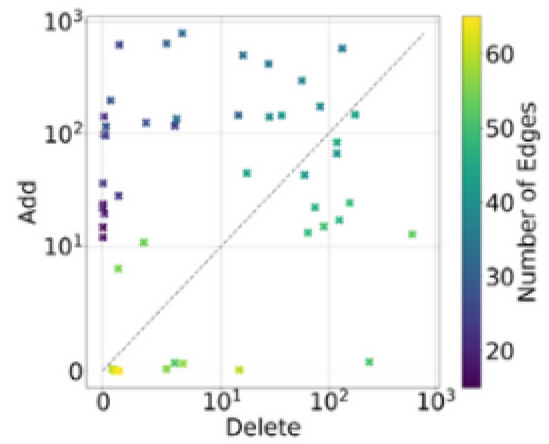
The solution quality is measured on the utility metrics Global transitivity, shortest path and Closeness Centrality. Data on the left was created using synthetic networks. Data on the right was created using social networks.

Dataset	Approach	Running time (s)	Avg Mem (GB)	Max Mem (GB)	$\|E\|$	GT
insecta-ant-trophallaxis-colony1-day1	delete	0.097	0.002	0.008	27	**0.13**
	add	630	1.33	5.27	32	0.22
	heuristic	10.8	0.0170	0.053	32	0.19
	original	-	-	-	30	0.14
insecta-ant-trophallaxis-colony1-day8	delete	2.96	0.004	0.017	33	0.13
	add	1830	5.32	20.8	41	**0.17**
	heuristic	421	0.405	0.807	41	0.21
	original	-	-	-	37	0.17
mammalia-baboon-grooming-group07	delete	1.04	0.004	0.021	13	0.22
	add	1900	1.29	1.85	20	**0.29**
	heuristic	2140	1.10	1.45	22	0.35
	original	-	-	-	16	0.27
mammalia-raccoon-proximity-24	delete	22.1	0.109	0.177	9	**0.64**
	add	14100	1.50	2.47	22	0.70
	heuristic	935	0.587	0.893	22	0.72
	original	-	-	-	14	0.66
mammalia-raccoon-proximity-3	delete	49.6	0.120	0.195	45	0.38
	add	8530	1.64	3.03	55	**0.46**
	heuristic	1530	0.635	0.919	56	0.52
	original	-	-	-	50	0.48

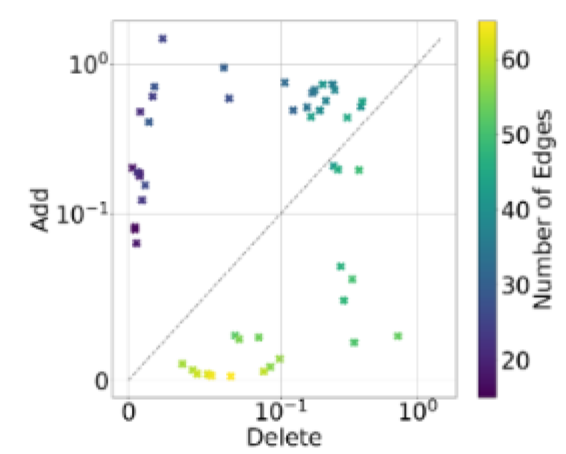




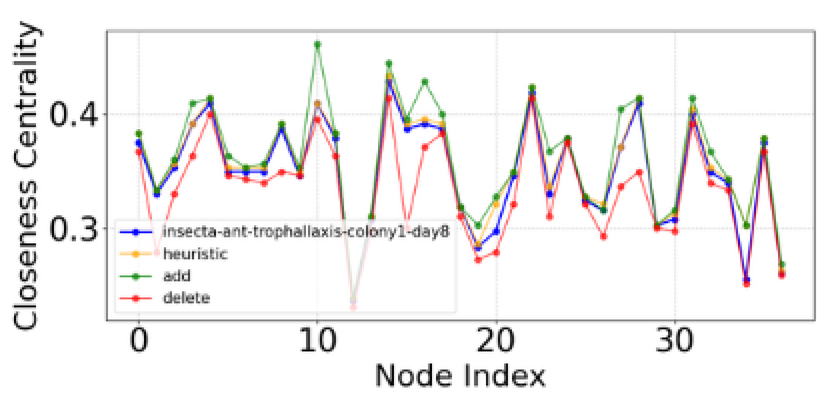




Running time (s).



Max memory (GB).



REFERENCES

[1] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. science, 286(5439):509-512, 1999.
[2] P. Erdös and A. Rényi. On the evolution of random graphs, pages 38-82. Princeton University Press, Princeton, 2006.
[3] Xinyue Xie. Anonymization algorithms for privacy-sensitive networks. Master's thesis, LIACS, Leiden University, 2023.
[4] Jordi Casas-Roma. An evaluation of edge modification techniques for privacy-preserving on graphs. In Modeling Decisions for Artificial Intelligence: 12th International Conference, MDAI 2015, Skövde, Sweden, September 21-23, 2015, Proceedings 12, pages 180-191. Springer, 2015.
[5] Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In AAAI, 2015.
[6] Rachel G. de Jong, Mark P. J. van der Loo, and Frank W. Takes. A systematic comparison of measures for k-anonymity in networks, 2024.