Community Robustness in Temporal Networks under Edge Addition

Moyi Tian* and Pablo Moriano†

* Division of Applied Mathematics, Brown University † Computer Science and Mathematics Division, Oak Ridge National Laboratory

Background

- A **community** is a subset of nodes within the graph such that connections between the nodes are denser than connections with the rest of the network. It often represents essential functional or behavioral units in complex networks. Communities in real networks with unknown properties are usually detected by algorithms based on various methods.
- **Community robustness** [2] tells how strong a community structure is against network perturbations such as edge rewiring and node/edge removal. It is important since real-world networks usually evolve over time.
- We study community robustness under edge addition in synthetic and empirical temporal networks.

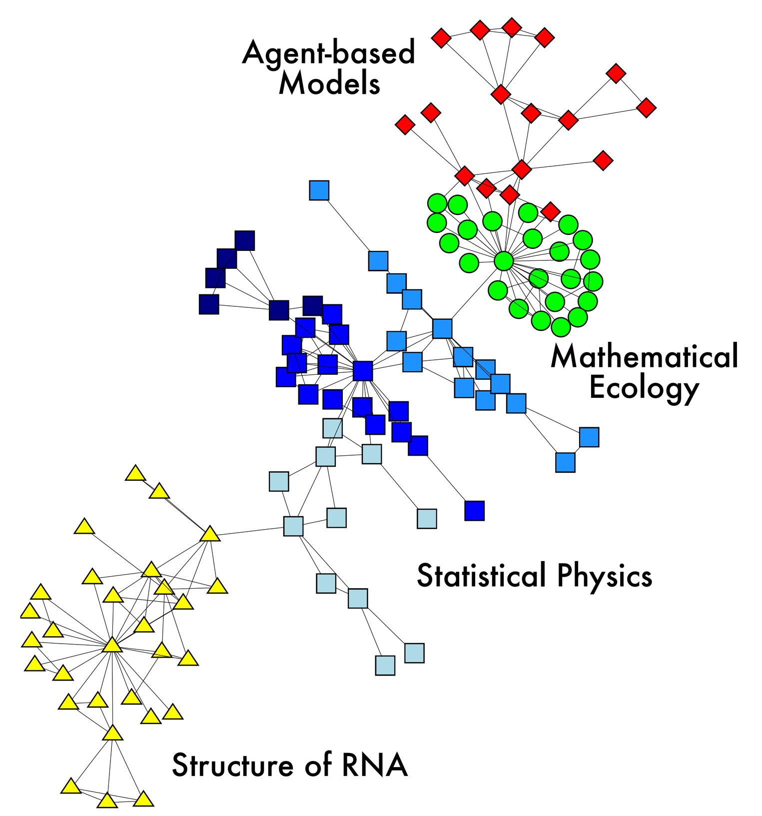


Figure: The Santa Fe Institute collaboration network where vertex shapes indicate primary divisions detected by algorithm [1].

Methods

Networks:

- Synthetic: Lancichinetti-Fortunato-Radicchi (LFR) benchmark graphs [3] with different mixing parameter, μ
- Empirical: Preprocessed sub-networks from temporal email networks

Community Detection:

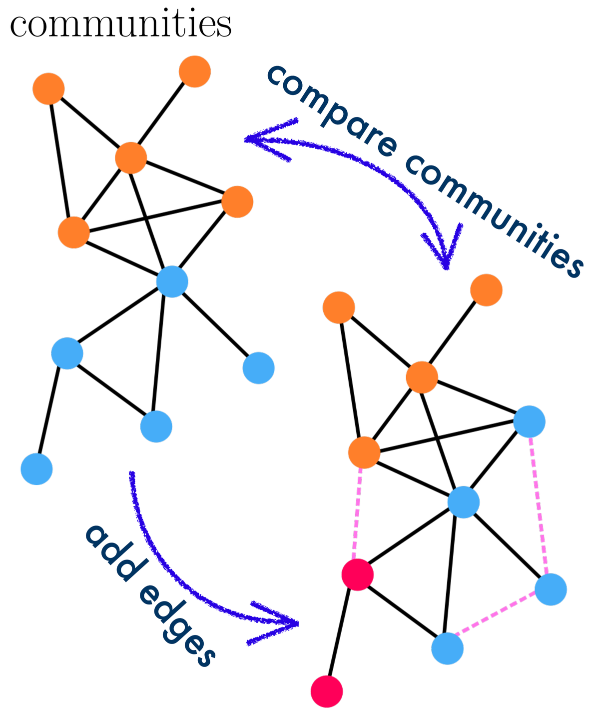
- Information-theoretic-based algorithm **(a) Infomap**,
- Message-passing-based algorithm **(b) Label Propagation**,
- Modularity-based algorithms **(c) Leiden** and **(d) Louvain**

Community Similarity: ranging from 0 (least similar) to 1 (most similar)

- Normalized Mutual Information (NMI) [4]
- Element-Centric Clustering Similarity [5] - better deals with bias in the number of clusters but has larger time complexity compared with NMI

Procedures:

- Synthetic
  - Generate LFR benchmark graphs with ground-truth communities
  - At each step:
    - (i) Add a number of edges selected from all or across-community non-edges uniformly at random (analogous to random failures and deliberate attacks resp.) while avoiding multi-edges
    - (ii) Detect communities in the perturbed network
    - (iii) Compute the community similarity metric score with respect to the original ground-truth communities
    - (iv) Repeat steps (i), (ii) and (iii) for independent realizations and average scores

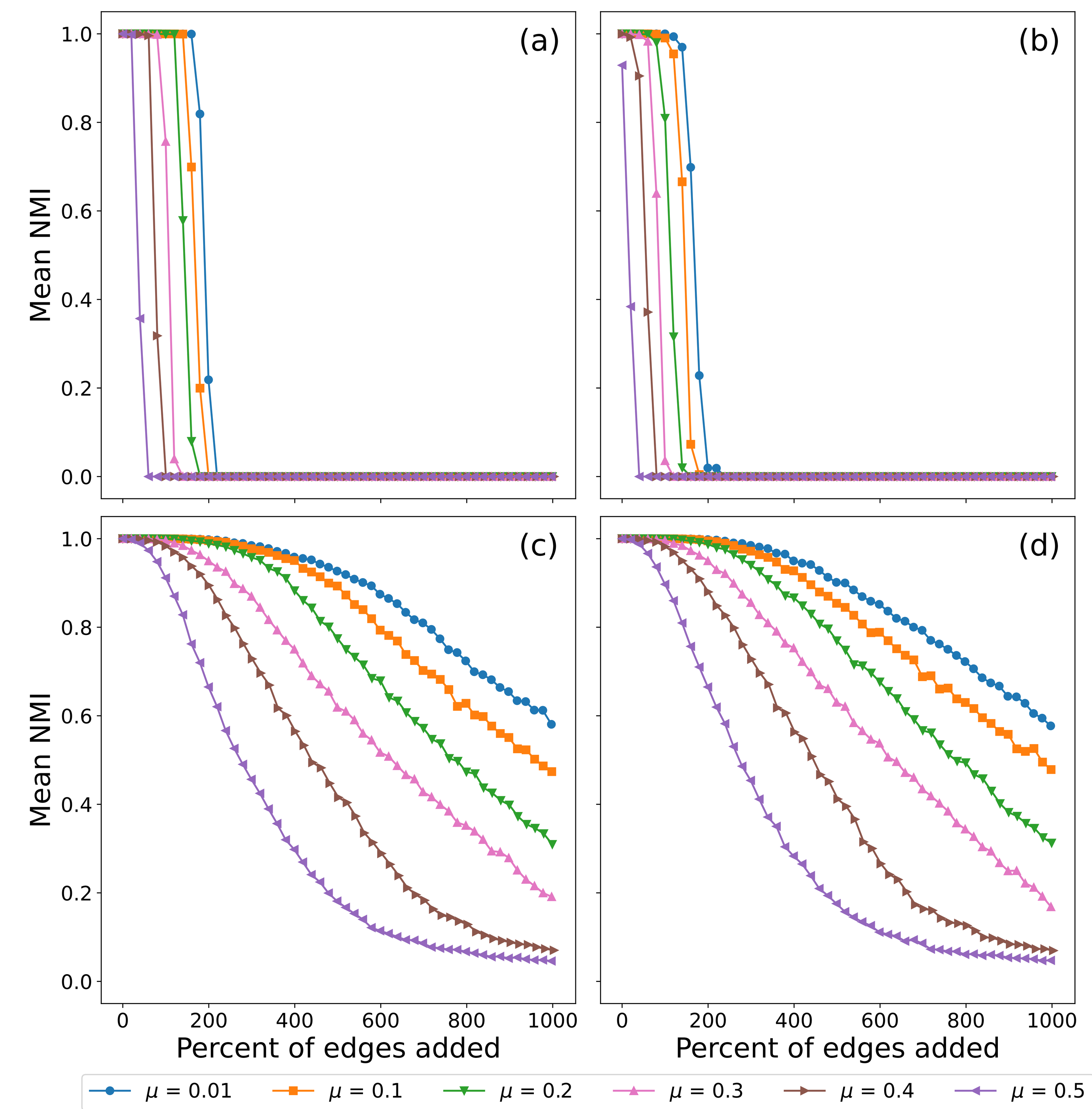


- Empirical
  - At initial time, use fast concensus [6] to find initial communities
  - At each time step:
    - (i) Add a number of temporal edges
    - (ii) Detect communities in the evolved network
    - (iii) Compute the community similarity metric score with respect to the initial communities
    - (iv) Repeat steps (ii) and (iii) for independent detections and average the scores

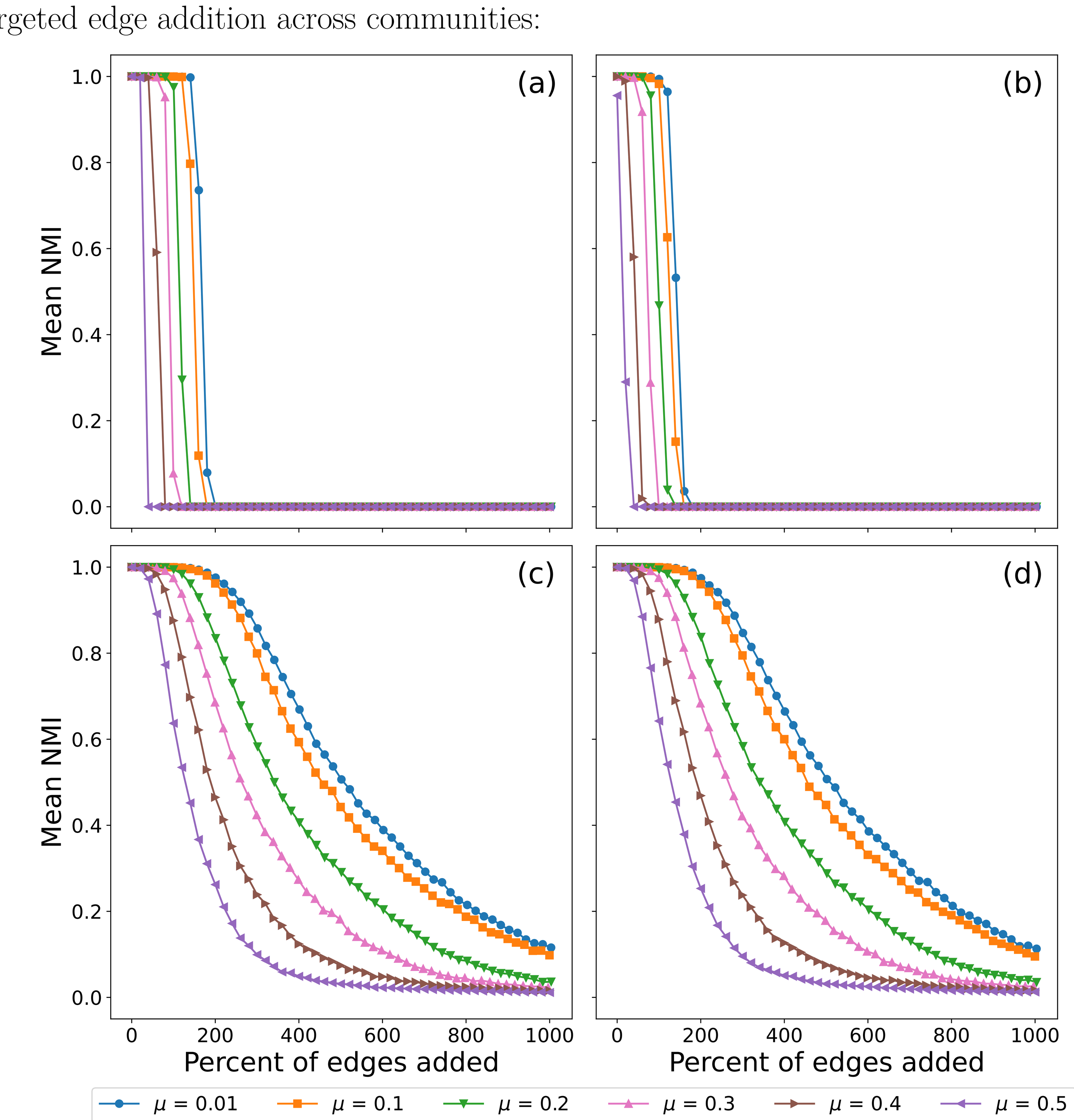
Synthetic Networks

LFR benchmark graphs with 1000 nodes. Results on 10000 nodes and with Element-Centric Similarity metric demonstrate similar qualitative trends.

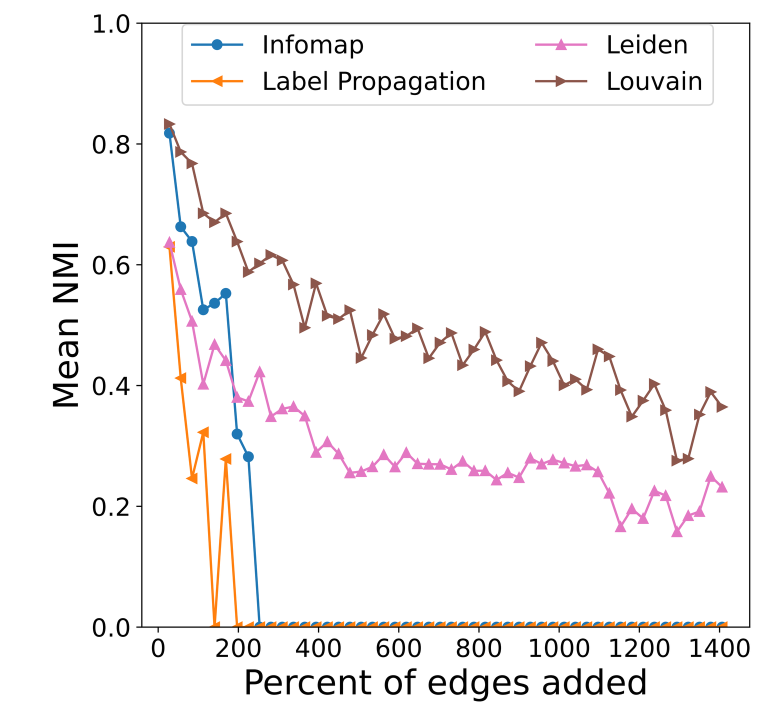
Random edge addition:



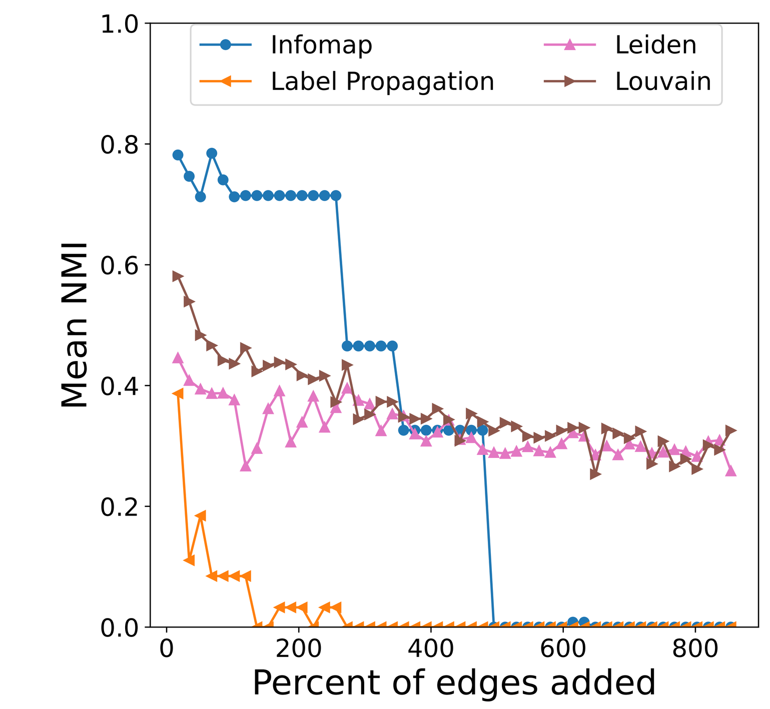
Targeted edge addition across communities:



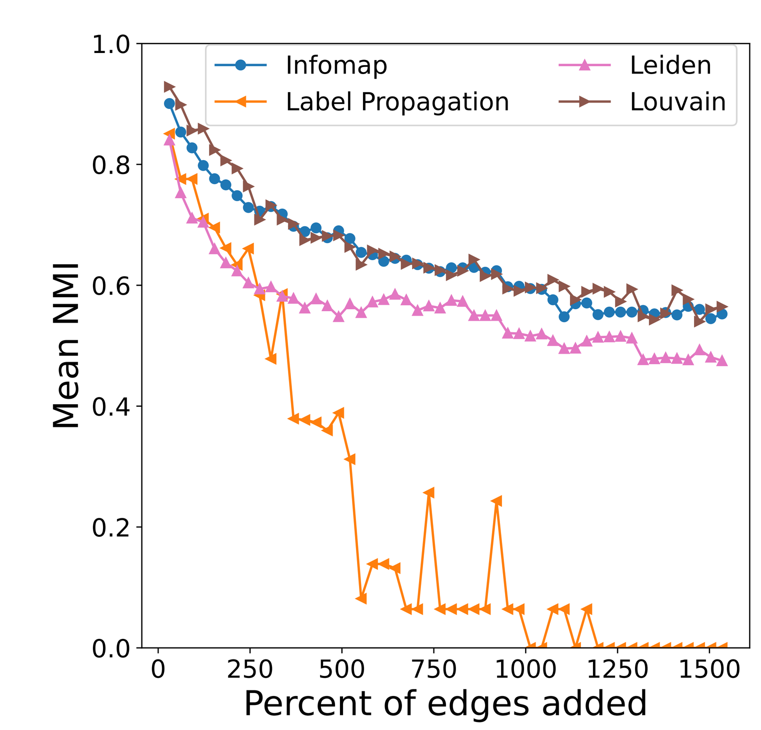
Empirical Networks



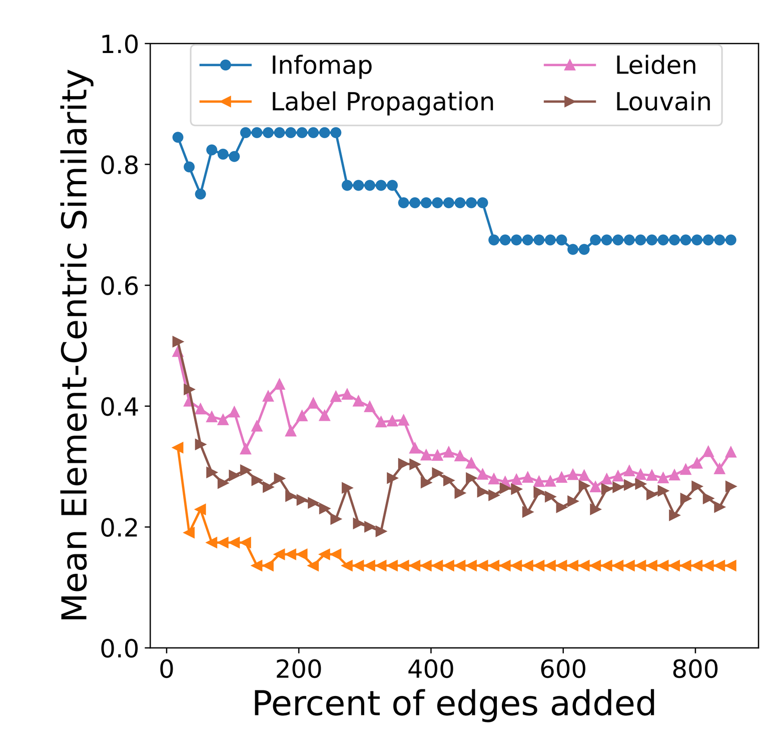
sub-network from ia-radoslaw-email network with 74 nodes



sub-network from Enron-all network with 120 nodes



sub-network from email-Eu-core-temporal network with 282 nodes



sub-network from Enron-all network with 120 nodes

Conclusions

- The robustness of communities under edge addition depends strongly on the community detection algorithms. In particular, we observe:
  - Modularity-based algorithms are able to detect communities in the perturbed networks that are closer to the initial community structure in synthetic experiments.
  - Label Propagation is not good at detecting communities similar to the initial communities after perturbations in both synthetic and empirical experiments.
- In synthetic networks:
  - Communities with lower μ (i.e. clearer divisions into modules) are more robust.
  - Targeted edge addition leads to faster drops in community similarity as expected.
- A future direction - In empirical networks:
  - Intrinsic network properties (degree distributions, edge-addition rules...) and community similarity metric also play roles in the community robustness.

References

[1] M. Girvan and M. E. J. Newman. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences*, **99**, (2002) 7821.

[2] B. Karrer, E. Levina and M. E. J. Newman. Robustness of Community Structure in Networks. *Phys. Rev. E*, **77**, (2008) 046119.

[3] A. Lancichinetti, S. Fortunato and F. Radicchi. Benchmark Graphs for Testing Community Detection Algorithms. *Phys. Rev. E*, **78**, (2008) 046110.

[4] A.L.N. Fred and A.K. Jain. Robust Data Clustering. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, **2**, (2003) II-II.

[5] A.J. Gates, I.B. Wood, W.P. Hetrick et al. Element-Centric Clustering Comparison Unifies Overlaps and Hierarchy. *Sci Rep*, **9**, (2019) 8574.

[6] A. Tandon, A. Albeshri, V. Thayananthan, W. Alhalabi and S. Fortunato. Fast Consensus Clustering in Complex Networks. *Phys. Rev. E*, **99**, (2019) 042301.