Compressing regularised dynamics improves link prediction in sparse networks

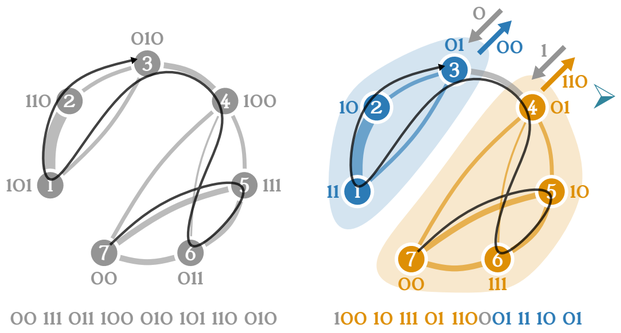
Maja Lindström, Umeå University
Department of Computing Science, Integrated Science Lab (IceLab) and Siftlab AB

Why link prediction?

Predicting gene-gene interactions in biological systems paves the way for breakthroughs in **genetic research**, identifying friends in online **social networks** enhances user engagement, and suggesting items in **retail recommendation systems** boosts revenue and customer satisfaction. In these link prediction applications, both **performance** and **interpretability** are crucial.

The map equation

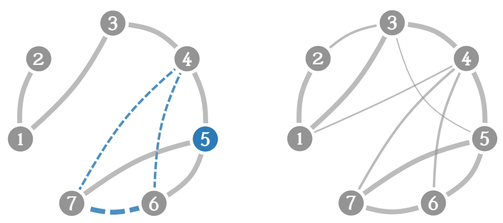
- The map equation is a **flow-based community detection method** that combines principles from information theory and coding theory with the concept of random walks [1].



- The map equation measures the quality of a network partition as the expected per-step description length — the **codelength** — for a random walk on the network.
- Detecting the optimal communities is a search problem that involves **minimising the map equation** by identifying sets of nodes where the random walker stays for a relatively long time.

Local regularisation

- In some cases, global regularisation can obscure local regularities in networks. In response, we evaluate three local regularisation techniques:

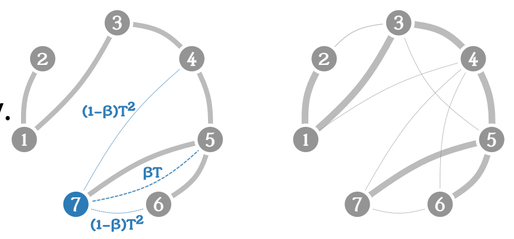


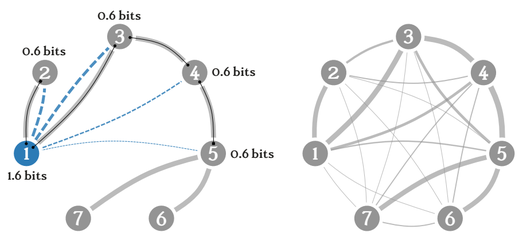
Common Neighbours

- If two nodes have a common neighbour — connect them!

Mixed Markov Time (MMT)

- Connect nodes that are two steps away.



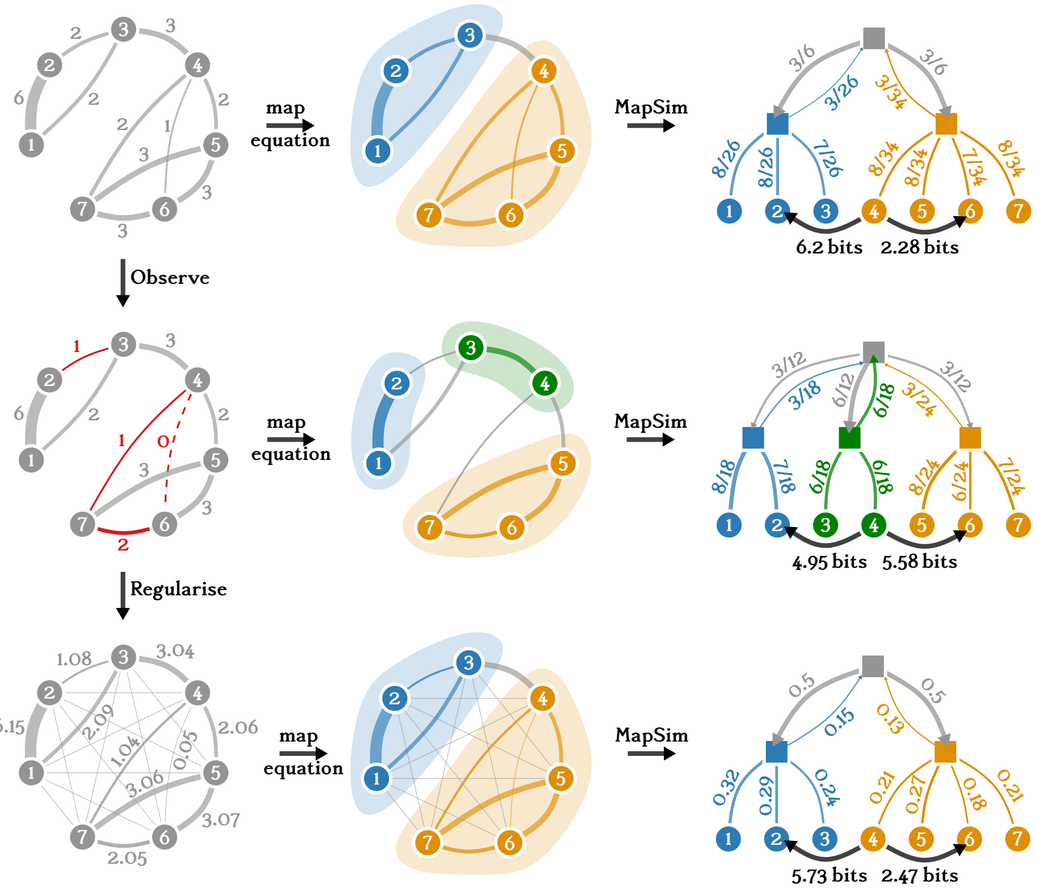


Variable Markov Time (VMT)

- Add links between nearby low-degree nodes.

MapSim and global regularisation

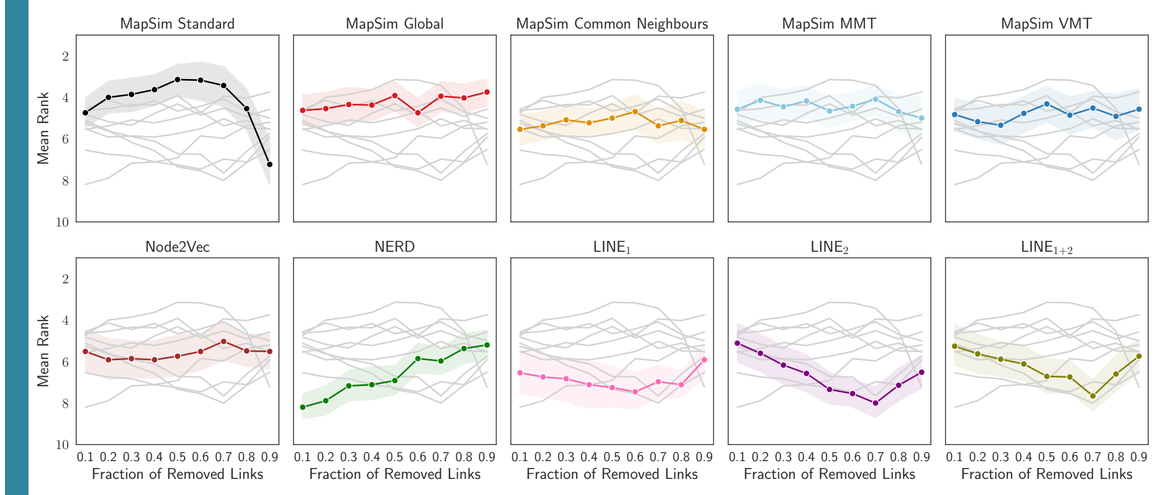
- **MapSim** is a node-similarity measure based on the map equation introduced for similarity-based link prediction [2].
- Nodes in the same module are generally considered more similar since a transition between them can be described more efficiently with **shorter codewords**, corresponding to **more probable links**.



- MapSim works well on dense and complete networks [2]. In sparse networks, the map equation can over-partition the network, sometimes degrading the link prediction performance. To overcome this issue, we incorporate a **global regularisation** method based on a Bayesian estimate of the transition rates [3].

Results

- Experiments on 35 real-world networks.
- Mean rank based on **AUC-score**.



- **Global regularisation** consistently outperforms standard MapSim and other state-of-the-art embedding methods in **highly sparse networks**.
- Although standard MapSim excels in denser networks, global regularisation maintains **stable performance** also in sparse networks, making it a good choice when the **network density is unknown**.

References

1. M. Rosvall and C. T. Bergstrom. ***Maps of random walks on complex networks reveal community structure.*** PNAS, 105(4):1118–1123, 2008.
2. C. Blöcker, J. Smiljanić, I. Scholtes and M. Rosvall. ***Similarity-based link prediction from modular compression of network flows.*** PMLR 198:52:1–52:18, 2022.
3. J. Smiljanić, C. Blöcker, D. Edler, and M. Rosvall. ***Mapping flows on weighted and directed networks with incomplete observations.*** J. Complex Netw., 9(6), 12 2021.