Sampling Networks from Modular Compression of Network Flows

NetSci 2023, Vienna, Austria

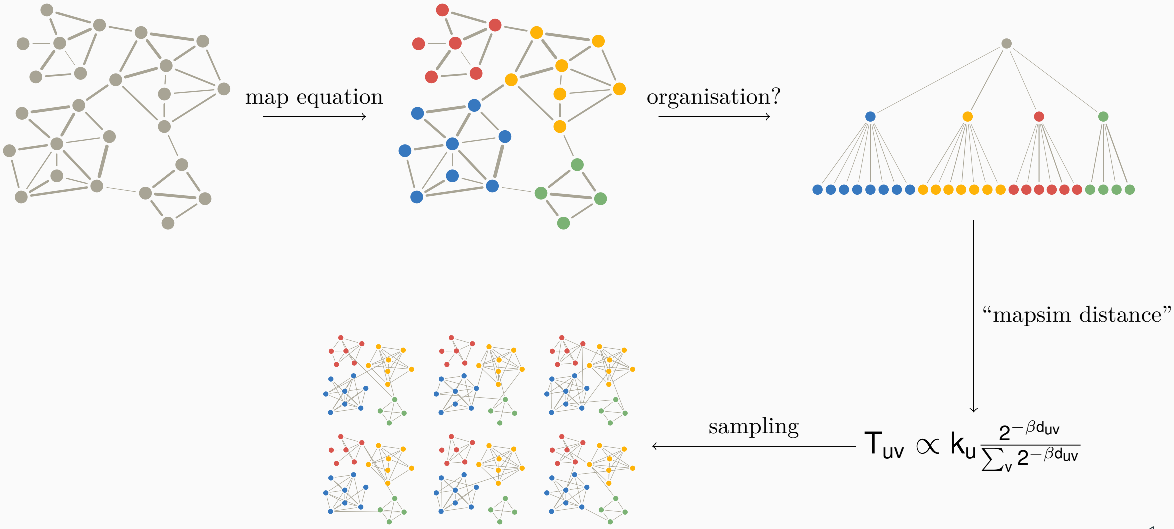
Christopher Blöcker[1,2], Jelena Smiljanić[2], Ingo Scholtes[1], Martin Rosvall[2]

`christopher.bloecker@uni-wuerzburg.de`

[1]Chair of Machine Learning for Complex Networks
Center for Artificial Intelligence and Data Science
Julius-Maximilians-Universität Würzburg
Würzburg, Germany

[2]Integrated Science Lab
Department of Physics
Umeå University
Umeå, Sweden

TL;DR



map equation
organisation?
“mapsim distance”
$T_{uv} \propto k_u \frac{2^{-\beta d_{uv}}}{\sum_v 2^{-\beta d_{uv}}}$
sampling

Motivation

- Benchmark models generate networks with given structural characteristics
  - degree distribution
  - degree correlations
  - community structure
- Typically not considering (higher-order) dynamical processes
- But: structure and dynamics are inter-dependent

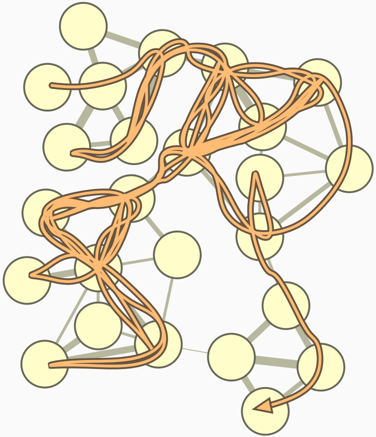
Idea

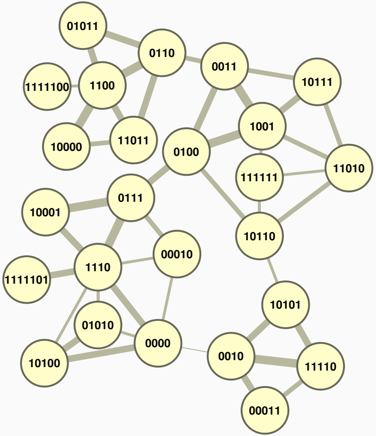
Generate networks from the modular compression of network flows that captures characteristics of dynamical processes

The Map Equation

→ Rosvall and Bergstrom; PNAS 2008

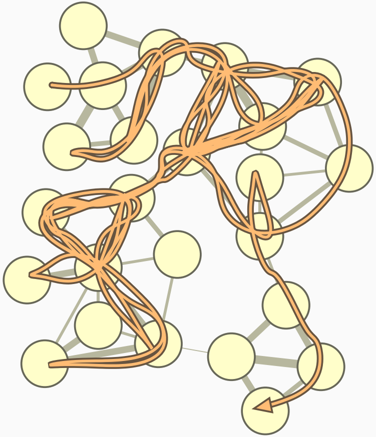
The Map Equation

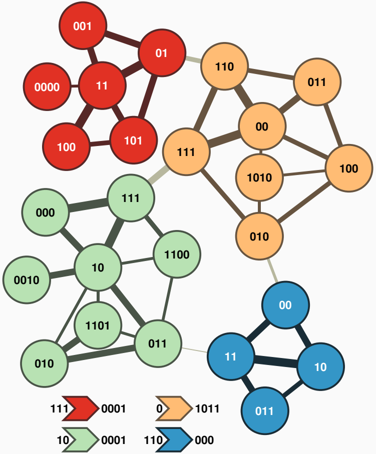




$$L(M) = H(P) = -\sum_{p} p \log_2 p \text{ [bits]}$$

The Map Equation





$$L(M) = qH(Q) + \sum_{m \in M} p_m H(P_m) \ \text{[bits]}$$

Note: we don’t actually simulate random walks or assign codewords!

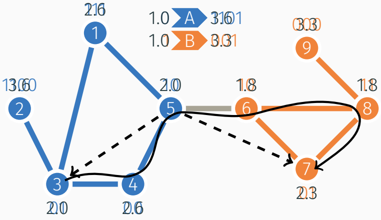
Map Equation Similarity

→ Blöcker, Smiljanić, Scholtes, and Rosvall; PMLR 2022

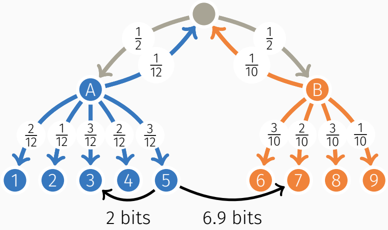
Map Equation Similarity

Key Idea

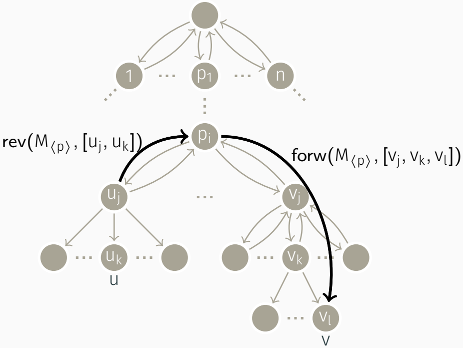
We can use the coding scheme to describe **any link or non-link** in the network.



description: 0 01 00 10 1101 1 10 11 01



Map Equation Similarity



random walker transition rate

$$\begin{aligned} \mathrm{mapsim}\,(M, u, v) &= \mathrm{rev}(M_{\langle p \rangle}, \mathrm{addr}(M_{\langle p \rangle}, u)) \\ &\quad \cdot \mathrm{forw}(M_{\langle p \rangle}, \mathrm{addr}(M_{\langle p \rangle}, v)) \end{aligned}$$

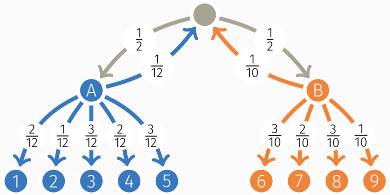
→ Blöcker, Smiljanić, Scholtes, and Rosvall; PMLR 2022

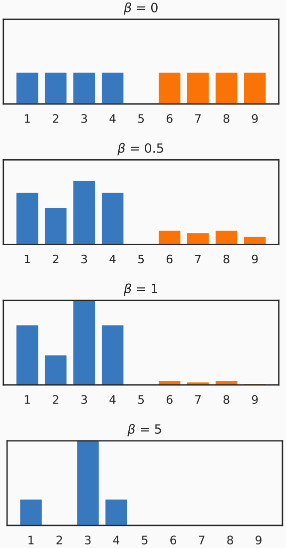
cost in bits (“distance”)

$$d_{uv} = -\log_2 \mathrm{mapsim}\,(M, u, v)$$

Sampling Networks from Modular Compression of Network Flows

Sampling



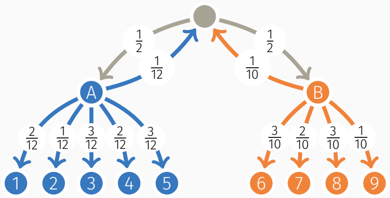


Link probability

$$T_{uv} \propto k_u \cdot \frac{2^{-\beta d_{uv}}}{\sum_v 2^{-\beta d_{uv}}}$$

$$\beta \overset{!}{=} 1 \Rightarrow T_{uv} \propto k_u \cdot \text{mapsim}\,(M, u, v)$$

Sampling



Link probability

$$T_{uv} \propto k_u \cdot \frac{2^{-\beta d_{uv}}}{\sum_v 2^{-\beta d_{uv}}}$$

$$\beta \stackrel{!}{=} 1 \Rightarrow T_{uv} \propto k_u \cdot \text{mapsim}\left(M, u, v\right)$$

Expected out-degree

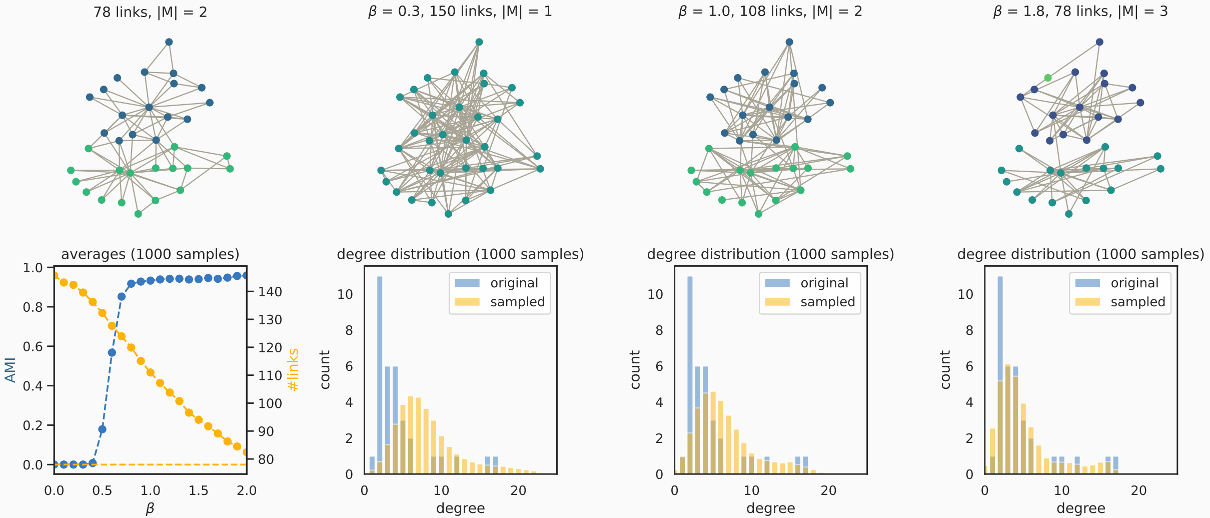
$$E\left[k_u\right] = \sum_v k_u \cdot \frac{2^{-\beta d_{uv}}}{\sum_v 2^{-\beta d_{uv}}} = k_u$$

Expected in-degree

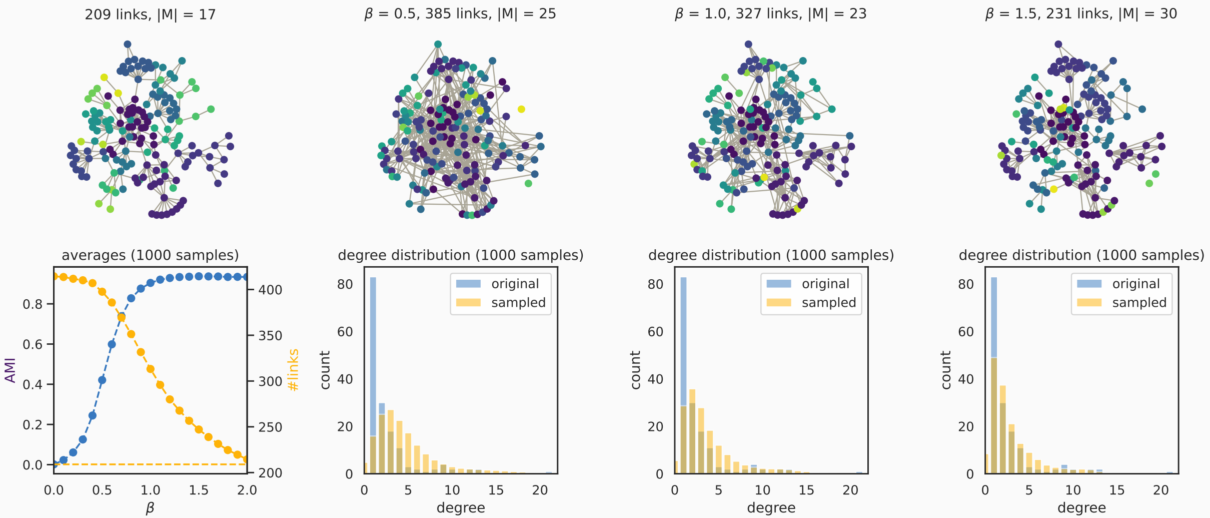
$$E\left[k_v\right] = \sum_u k_u \cdot \frac{2^{-\beta d_{uv}}}{\sum_v 2^{-\beta d_{uv}}} \neq k_v$$

Experiments

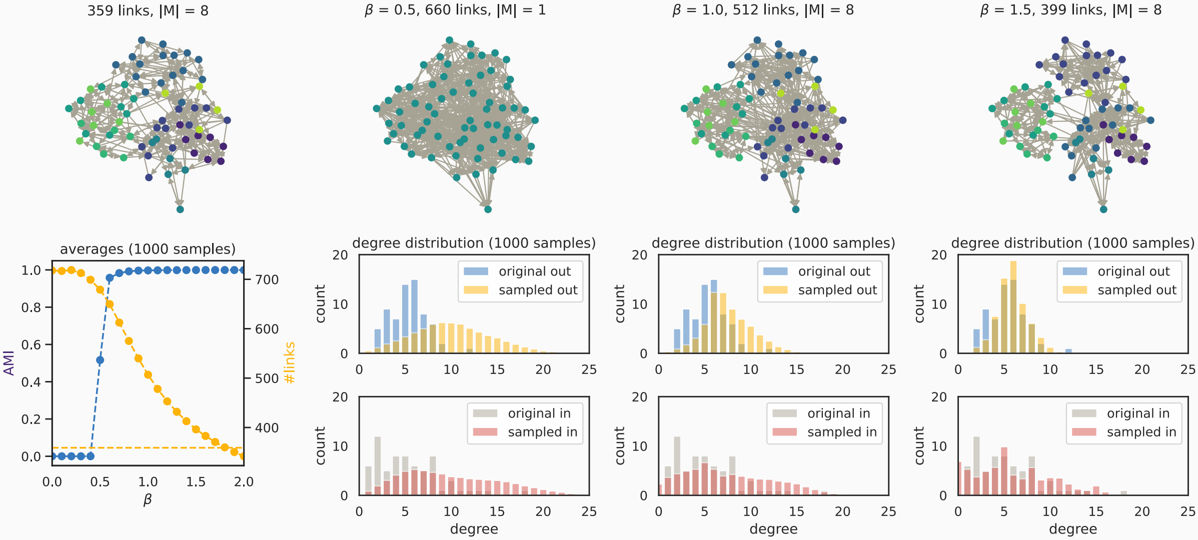
Karate Club


78 links, |M| = 2
β = 0.3, 150 links, |M| = 1
β = 1.0, 108 links, |M| = 2
β = 1.8, 78 links, |M| = 3
averages (1000 samples)
AMI
#links
β
degree distribution (1000 samples)
original
sampled
count
degree

Interactome



Highschool


359 links, |M| = 8
β = 0.5, 660 links, |M| = 1
β = 1.0, 512 links, |M| = 8
β = 1.5, 399 links, |M| = 8
averages (1000 samples)
AMI
#links
β
degree distribution (1000 samples)
count
original out
sampled out
original in
sampled in
degree

Conclusion

Conclusion

Motivation

Generate networks whose structure is guided by a dynamical process

Our Approach

We sample links based on map equation similarity, an information-theoretic similarity measure based on the compressions of random walks provided by the map equation.

Results

- we recover the community structure
- out degrees are preserved (if β not too high)
- in degrees are randomised

Naturally generalises to hierarchical organisation and higher order

Thank you for your attention!

Questions?