Community detection in growing networks

Matúš Medo
IFFS, UESTC, Chengdu
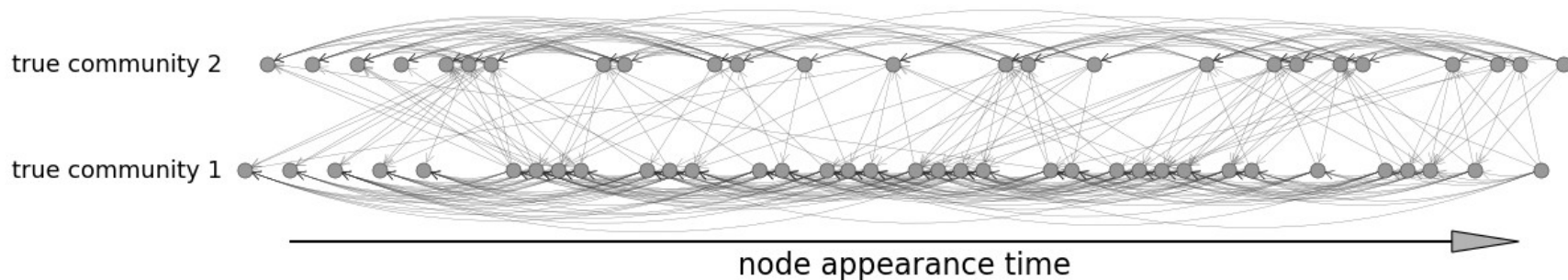
Motivation

"As long as there will be networks, there will be people looking for communities in them."
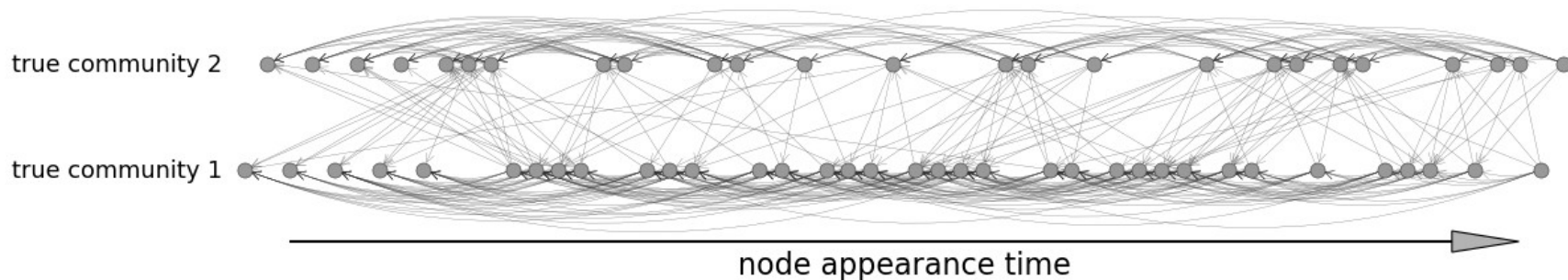
— Fortunato and Hric, 2016

A model growing network

true community 2

true community 1

node appearance time

- Typical information network:
  - News relating to each other
  - Scientific citations
  - …

A model growing network

true community 2

true community 1

node appearance time

Goal: Find the communities

Network modularity (static)

- Popular approach to community detection: maximize the modularity function (Girvan & Newman, 2002)

in the same community

↓

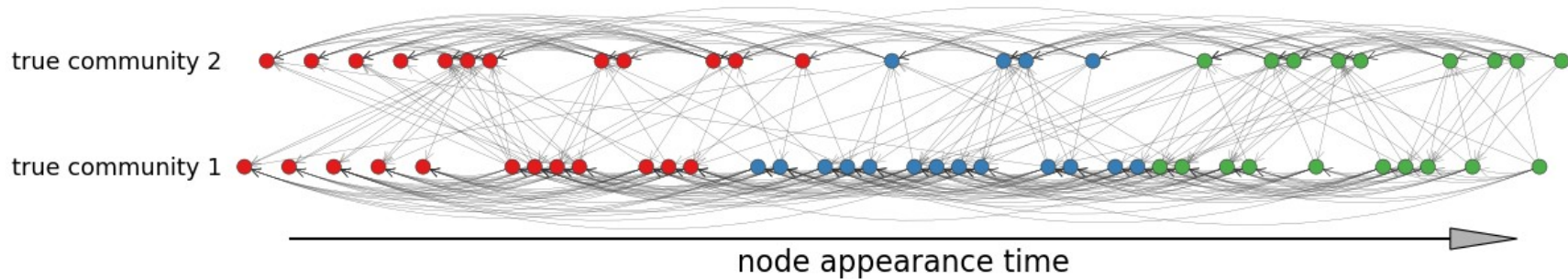$$Q = \frac{1}{m} \sum_{i,j} \left(A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right) \delta(c_i, c_j)$$
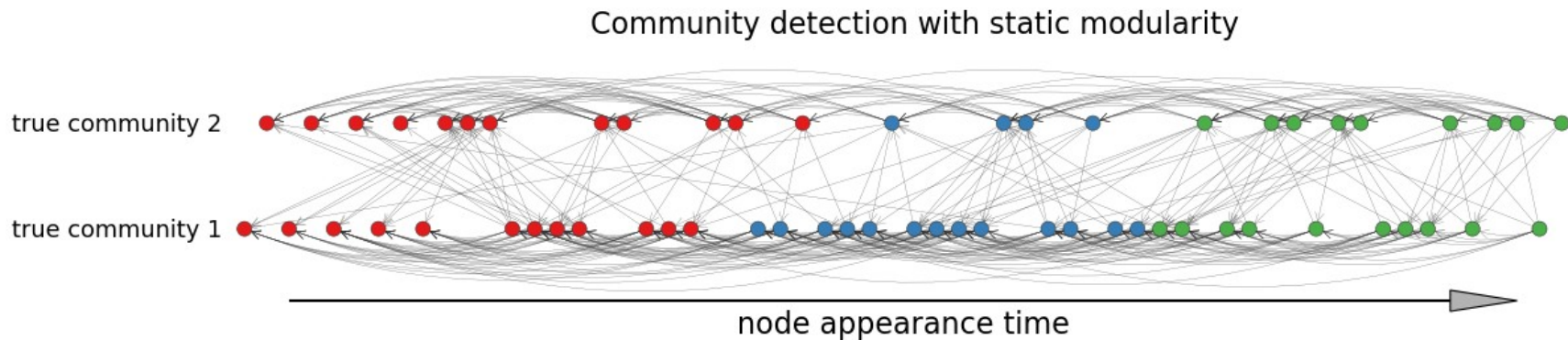
number of links   connected or not   link expectation

# The failure

Community detection with static modularity

true community 2

true community 1

node appearance time

The failure

Community detection with static modularity

true community 2

true community 1

node appearance time

…that makes you smile :-)

What went wrong

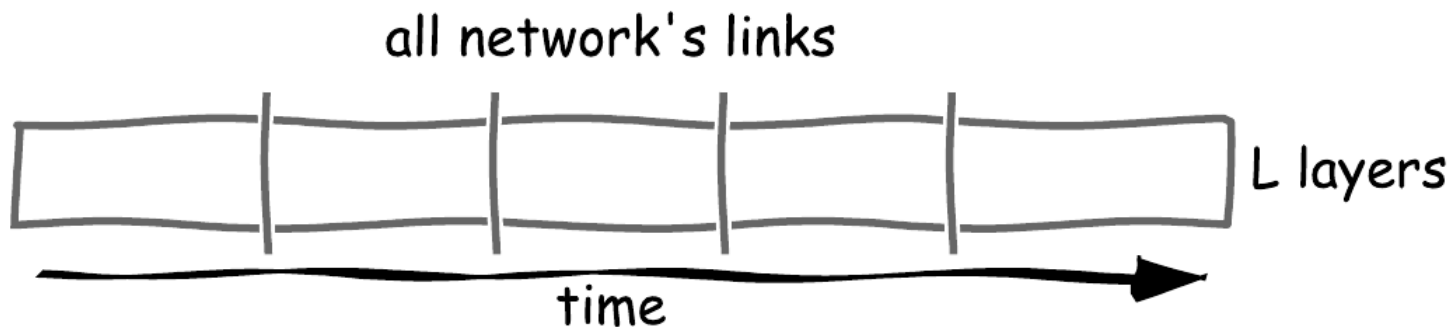$$Q = \frac{1}{m} \sum_{i,j} \left(A_{ij} - \boxed{\frac{k_i^{out} k_j^{in}}{m}} \right) \delta(c_i, c_j)$$

If time matters, the link expectation term is wrong

Modularity for growing networks

all network's links
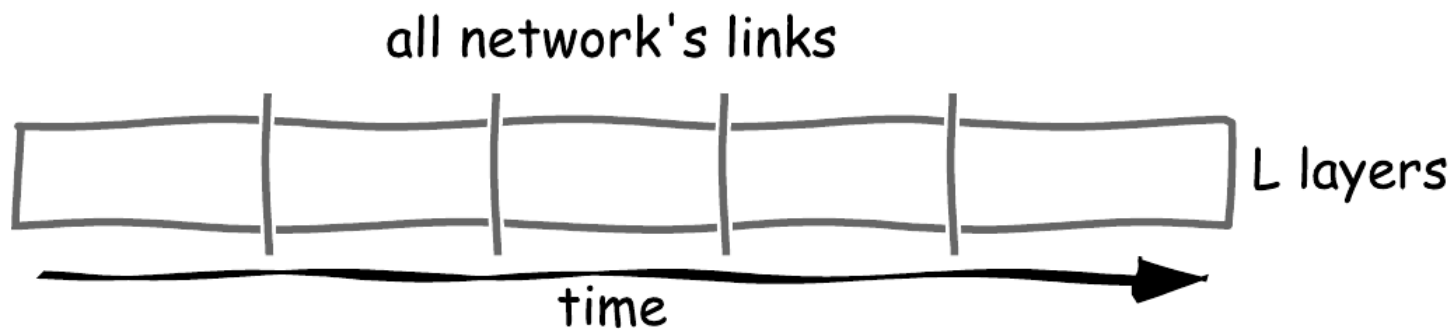
L layers

time

Ren et al., 2018 (PRE, arXiv:1703.07656)
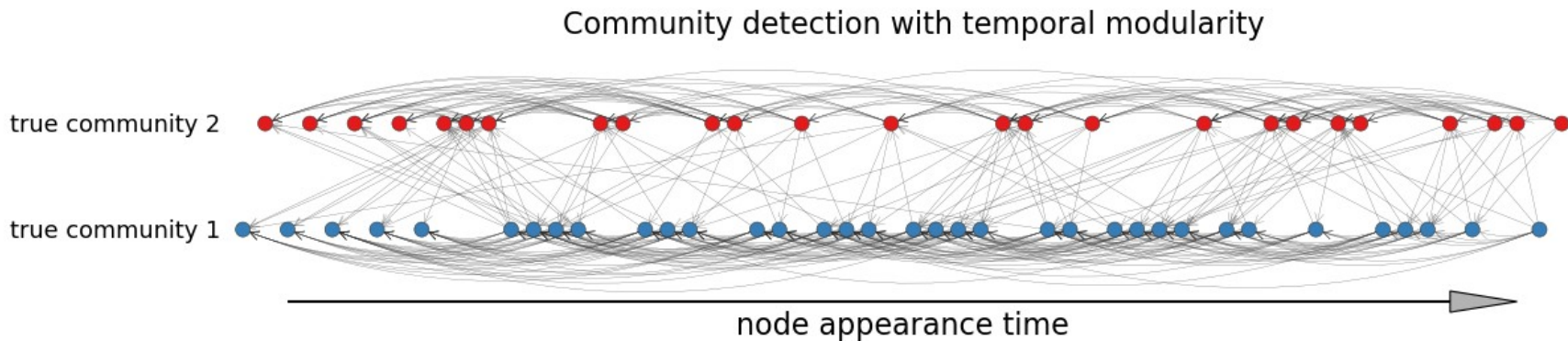
Modularity for growing networks

all network's links

L layers

time

Temporal modularity

$$Q_T(L) = \frac{1}{m} \sum_{i,j} \left(A_{ij} - \sum_{l=1}^{L} \frac{\Delta k_{i,l}^{out} \Delta k_{j,l}^{in}}{m_l} \right) \delta(c_i, c_j)$$

The success

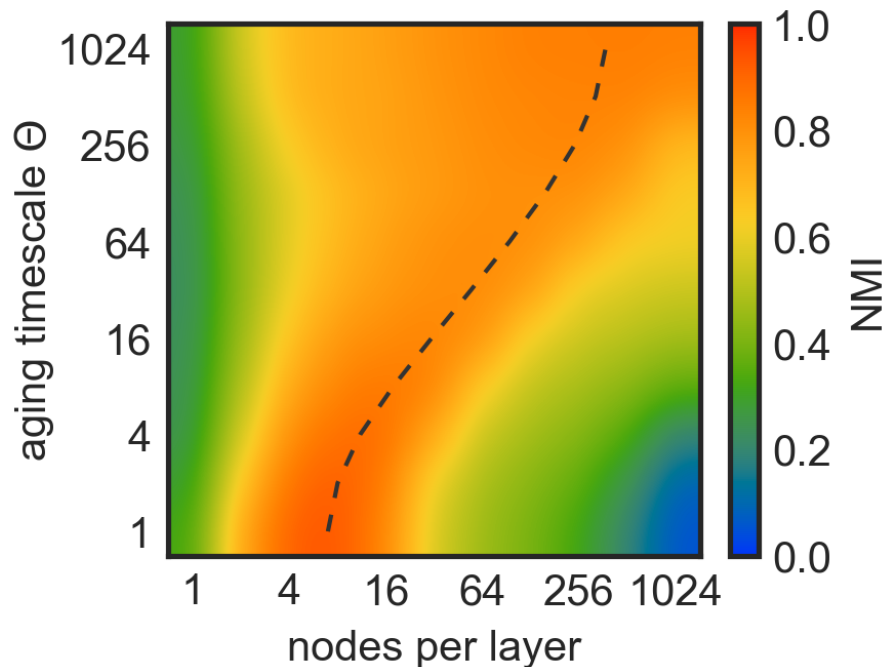Community detection with temporal modularity

Temporal modularity with 8 layers (L = 8)

How many layers?

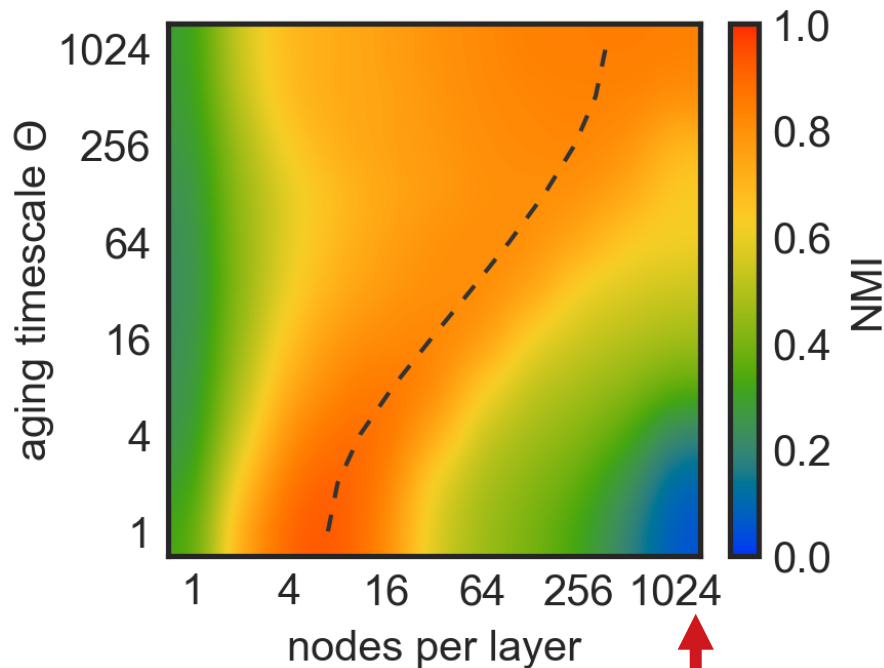1) Compute median link timespan λ

2) Use $L* = N / \lambda$ layers

Fast aging ⇨ Short links ⇨ Many layers

Temporal modularity in action

Dashed line: *L* chosen by the median link timespan

# Temporal modularity in action
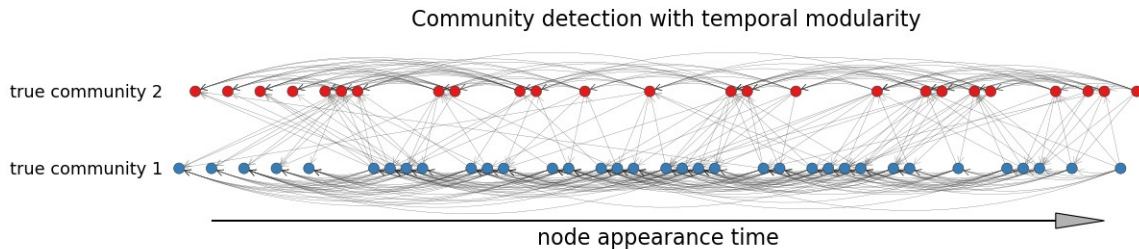

Usual static modularity

Real data

1) APS citation data:
   Subsets defined by PACS codes

2) News citation data (Spitz and Gertz, 2015):
   Subsets represent different news outlets

Evaluation on real data

- Observation on model data:

  - Wrong "static" communities are short

Community detection with static modularity

true community 2

true community 1

node appearance time
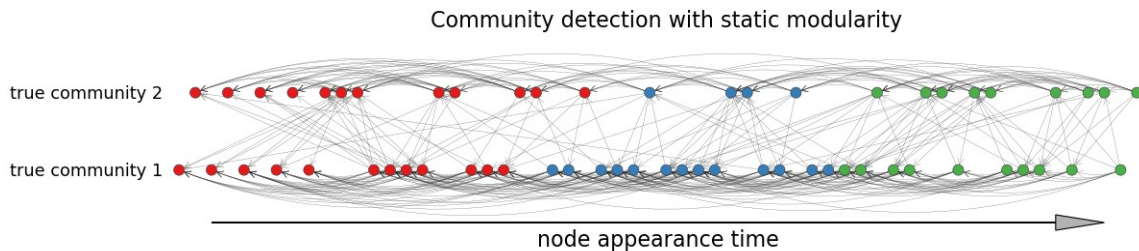
  - Correct "temporal" communities are long

Community detection with temporal modularity

true community 2

true community 1
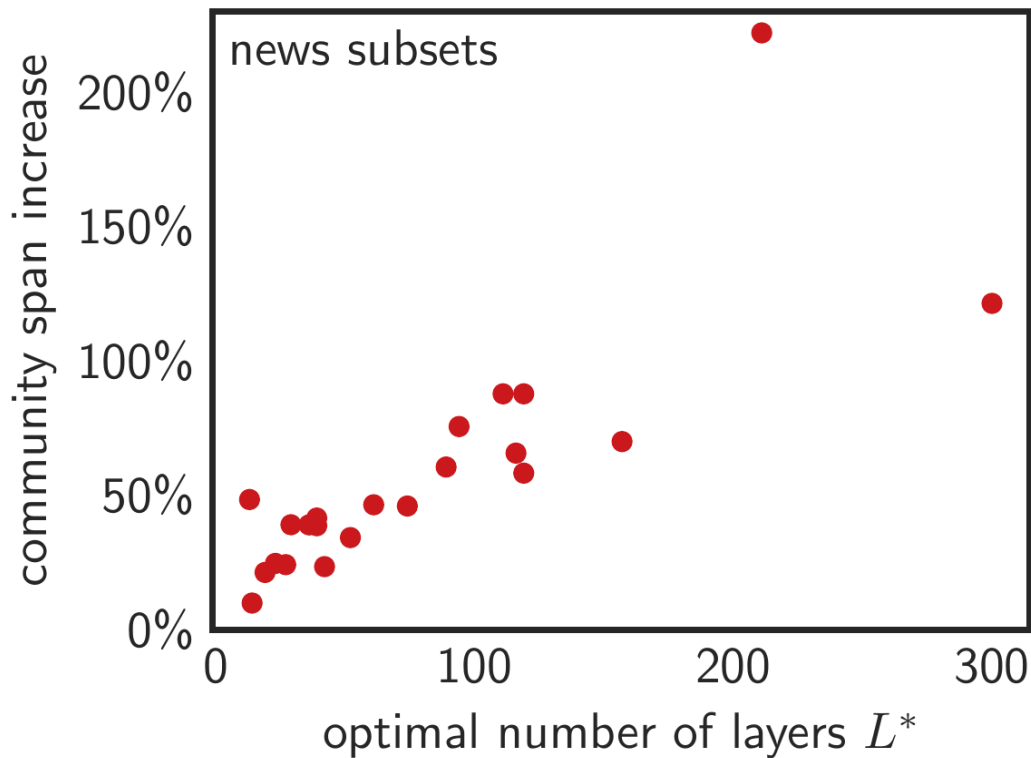
node appearance time

Evaluation on real data

Communities determined by node age are short

Increasing community length is a good sign

Results on real data

Summary

1) Static modularity fails on growing networks when aging is fast

2) Temporal modularity simply works

Many thanks to colleagues:
Manuel Mariani, Zhuo-Ming Ren, An Zeng, and Yi-Cheng Zhang