Community detection in growing networks with aging

Matúš Medo
UESTC, Chengdu
University of Fribourg, Fribourg
Inselspital, Bern

The PIIK: 2nd Symposium on Network Science
13 March 2018, Zurich

Introduction to community detection

- Many networks have community structure:
  - Some nodes are densely connected with each other (community)
  - Communities in social networks can be due to language, age, race, ...

Introduction to community detection

- Many networks have community structure:
  - Some nodes are densely connected with each other (community)
  - Communities in social networks can be due to language, age, race, …
- Importance:
  - Can help us understand how the system works
  - Communities often have properties that differ a lot from the average network properties

Introduction to community detection

- Many networks have community structure:
  - Some nodes are densely connected with each other (community)
  - Communities in social networks can be due to language, age, race, …
- Importance:
  - Can help us understand how the system works
  - Communities often have properties that differ a lot from the average network properties

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

Introduction to community detection

- We focus on growing networks in particular: information and social networks, for example
- Detecting communities in such systems can help to understand:
    1. Group formation
    2. Opinion polarization
    3. Spreading of misinformation
    4. ...

Network modularity (static)

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

Network modularity (static)

- Popular approach to community detection: maximize the modularity function (Girvan & Newman, 2002)
- Its version for directed networks (Arenas *et al.*, 2007):

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

Network modularity (static)

- Popular approach to community detection: maximize the modularity function (Girvan & Newman, 2002)
- Its version for directed networks (Arenas *et al.*, 2007):

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

Growing networks with community structure

- Modeling growing networks (Medo *et al.*, 2011):

$$P(i|j,t) \sim \underbrace{k_i(t)}_{\text{degree}} \times \underbrace{f_i}_{\text{fitness}} \times \underbrace{D(t)}_{\text{aging}}$$

  - Preferential attachment and node fitness are optional
  - Node aging: timescale Θ

Growing networks with community structure

- Modeling growing networks (Medo *et al.*, 2011):

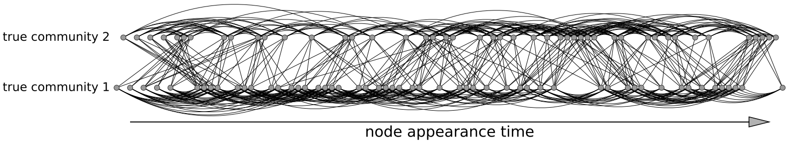
$$P(i|j,t) \sim \underbrace{k_i(t)}_{\text{degree}} \times \underbrace{f_i}_{\text{fitness}} \times \underbrace{D(t)}_{\text{aging}}$$

  - Preferential attachment and node fitness are optional
  - Node aging: timescale Θ
- Community structure can be easily introduced
  - Assign each node to a ground-truth community C
  - Multiply $P(i|j,t)$ with

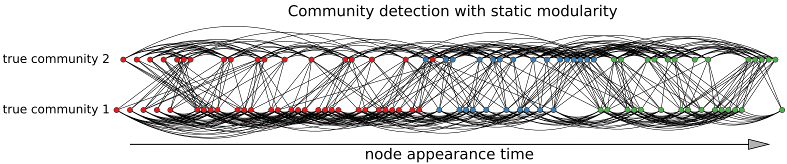
$$\mu[1-\delta(C_i,C_j)]+(1-\mu)\delta(C_i,C_j)$$

  - $\mu = 0$: only nodes from the same community can connect

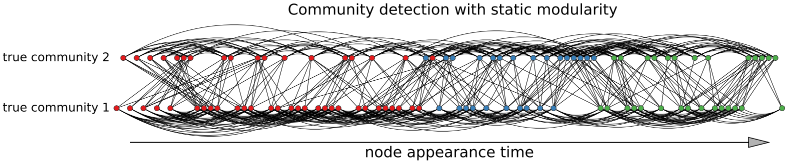
The problem in growing networks



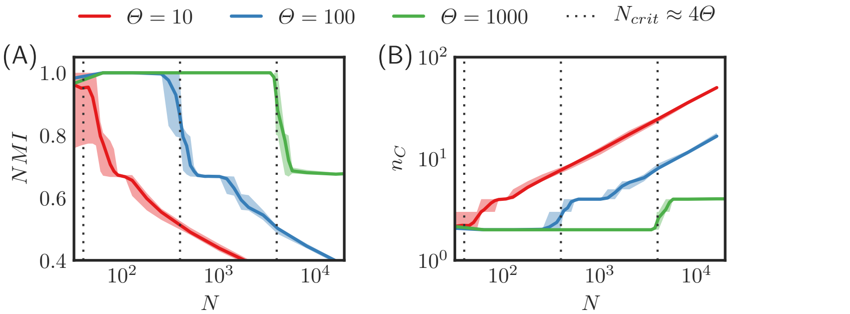
The problem in growing networks



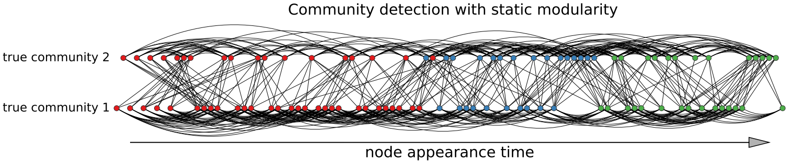
The problem in growing networks



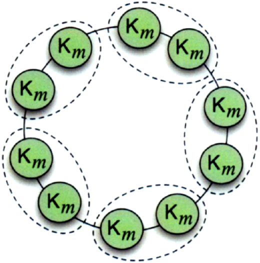
Problem even when $\mu = 0$



The problem in growing networks



Opposite to the well-known “resolution limit” of modularity



Fortunato & Barthélemy, 2007

Modularity for growing networks (to be submitted)

- **The reason of failure:**
  If time matters, the link expectation term is wrong

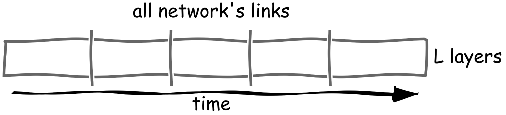
$$Q = \frac{1}{m} \sum_{i,j} \left(A_{ij} - \boxed{\frac{k_i^{out} k_j^{in}}{m}} \right) \delta(c_i, c_j)$$

Modularity for growing networks (to be submitted)

- The reason of failure:
  If time matters, the link expectation term is wrong

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

- Dynamic Configuration Model (Ren *et al.*, 2018?):

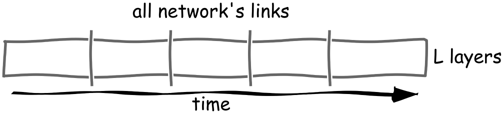


Modularity for growing networks (to be submitted)

- The reason of failure:
  If time matters, the link expectation term is wrong

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

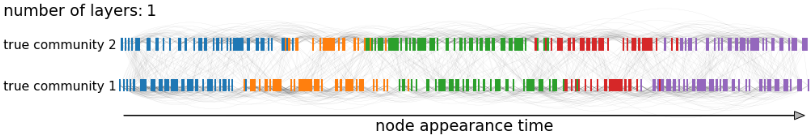
- Dynamic Configuration Model (Ren *et al.*, 2018?):



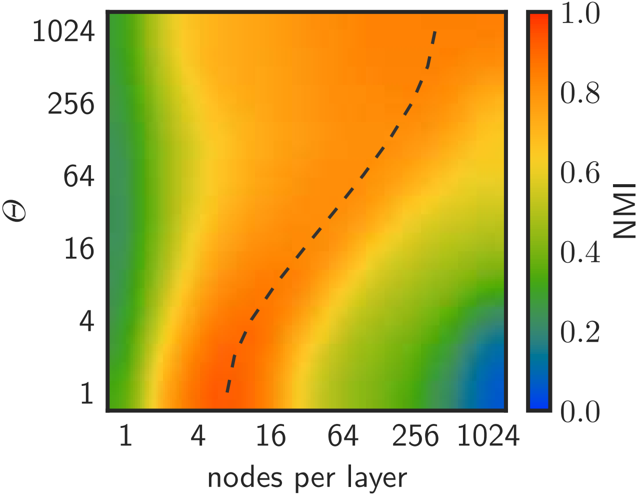
Temporal modularity computes link expectation from L layers

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

Temporal modularity in action



Temporal modularity in action



Dashed line corresponds to choosing L by the median link span

Real data

- Two real datasets
    1. Subsets of the APS citation data from years 1893–2013 corresponding to the second level in the PACS classification (*e.g.*, 89.75.* = “Complex systems”)
    2. Subsets of a news citation dataset (Spitz and Gertz, 2015) corresponding to individual newspapers

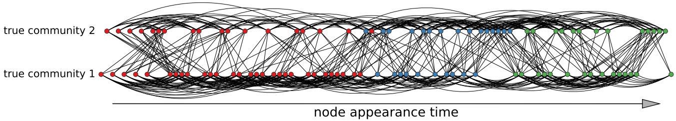
Real data

- Two real datasets
  1. Subsets of the APS citation data from years 1893–2013 corresponding to the second level in the PACS classification (*e.g.*, 89.75.* = "Complex systems")
  2. Subsets of a news citation dataset (Spitz and Gertz, 2015) corresponding to individual newspapers
- But... How to evaluate network partitions in real data without the ground truth?

Custom-made evaluation metric

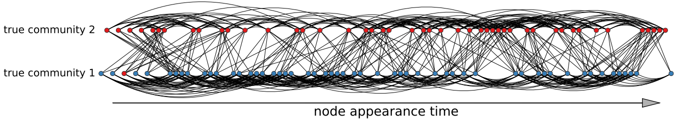
bad

Community detection with static modularity



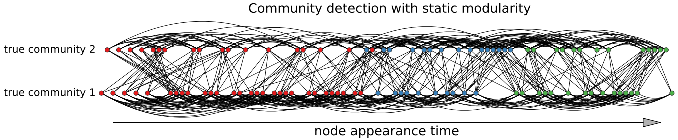
good

Community detection with temporal modularity

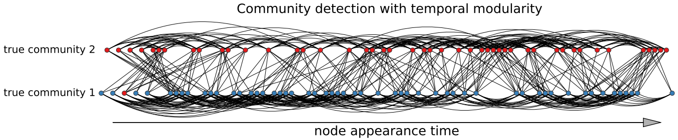


Custom-made evaluation metric

bad



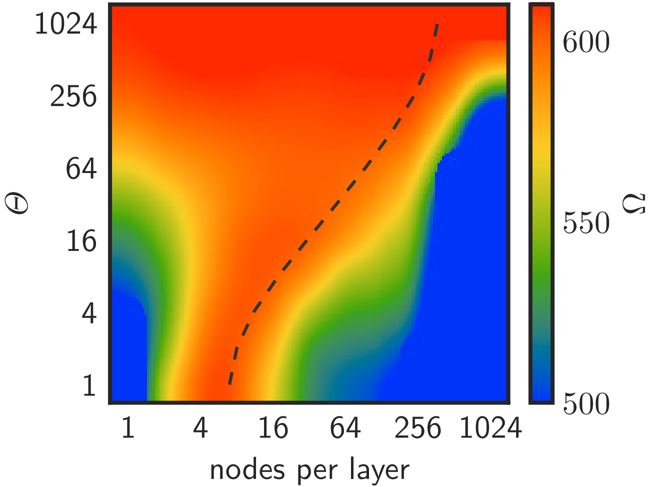
good



New metric: average community span* Ω

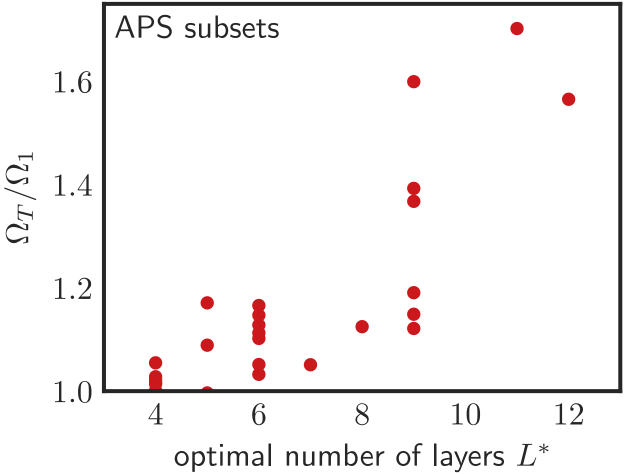
*: Span of a community is the difference between the 20th and 80th percentile of node IDs in the community, and the average is weighted by community size.

Custom-made evaluation metric

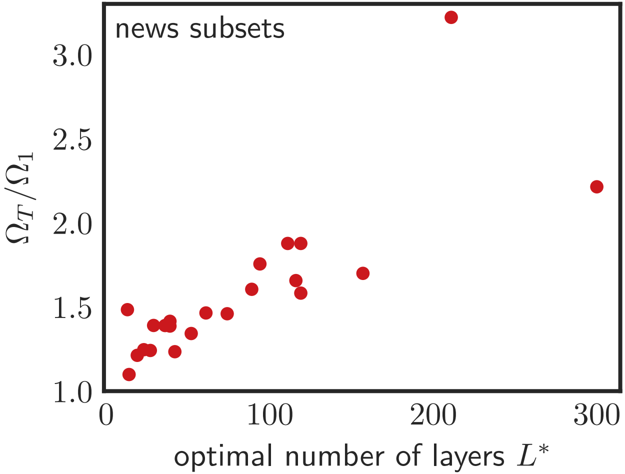


Look back at the model data

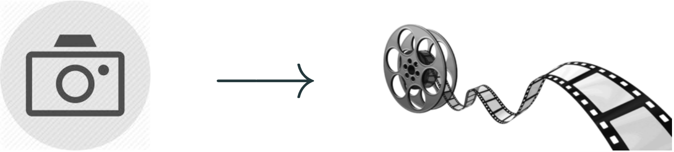
Real data results



Real data results



Take-home message

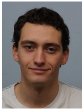


1. Static modularity fails when aging is fast, temporal modularity just works
2. Models help you assess old tools and devise new ones
3. In many systems, taking time into account improves the results

Further related work:

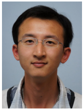
1. H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, M.-Y. Zhou, Ranking in evolving complex networks, Physics Reports 689, 1-54, 2017
2. Z.-M. Ren, M. S. Mariani, Y.-C. Zhang, M. Medo, Randomizing growing networks with a time-respecting null model arXiv:1703.07656
3. G. Vaccario, M. Medo, N. Wider, M. S. Mariani, Quantifying and suppressing ranking bias in a large citation network, Journal of Informetrics 11, 766-782, 2017
4. M. Medo, G. Cimini, Model-based evaluation of scientific impact indicators, Physical Review E 94, 032312, 2016
5. A. Vidmer, M. Medo, The essential role of time in network-based recommendation, EPL 116, 30007, 2016
6. M. Medo, M. S. Mariani, A. Zeng, Y.-C. Zhang, Identification and modeling of discoverers in online social systems, Scientific Reports 6, 34218, 2016

Web site: www.ddp.fmph.uniba.sk/~medo/physics/



Manuel Mariani

Zhuo-Ming Ren



An Zeng

Yi-Cheng Zhang

Thank you for your attention!

Questions?