Chapter 7

When Topology meets Semantics

RECAP: Modeling Choices

Only nodes and edges?

- Directed Networks;
- Weighted Networks;
- Signed Networks;
- Multilayer/Multiplex Networks;
- Temporal Networks;
- Bipartite/Heterogeneous Networks;
- Higher-order Networks.

Goal

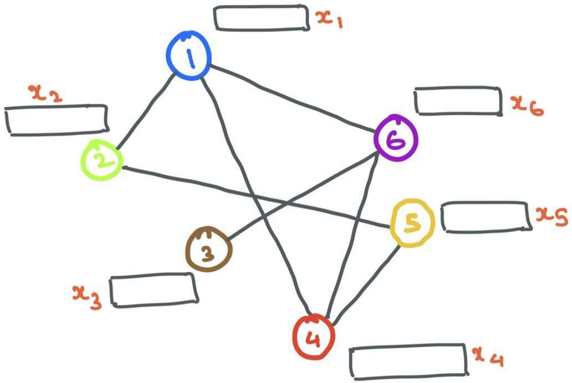
- Matching real-world problems with appropriate models;
- Research Objectives are shaped by the choice of the **network model**;

…Is there still more?

Feature-rich Networks

Feature-rich Networks

- ➔ Feature-rich Networks can integrate non-topological information to any complex topology (pairwise or higher-order networks);
- ➔ Feature-rich Networks as augmented implementations that add more information to a network structure;
- ➔ In the literature, they are often "confused"/mixed only with **node-attributed networks;**
- ➔ Actually, feature-rich networks involve more modeling choices, based on the augmentation on nodes or on links.



a feature-rich network as a network with vector/attributes on the nodes

Interdonato, R., Atzmueller, M., Gaito, S., Kanawati, R., Largeron, C., & Sala, A. (2019). Feature-rich networks: going beyond complex network topologies. *Applied Network Science*

Do I need Feature-rich Networks?

Improving Graph Representations:

A graph is composed of nodes (entities) and links (structure)

Improving the Entities:

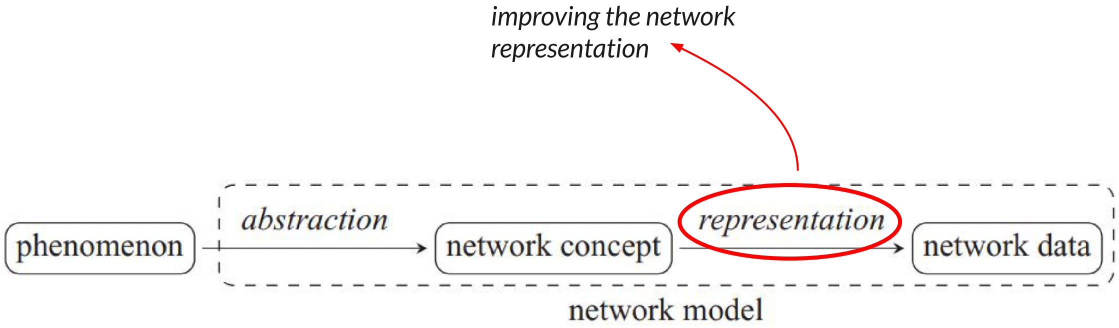
Node-attributed Networks;

Heterogeneous Networks;

Improving the Structure:

Multiplex/Multilayer Networks;

Temporal Networks.



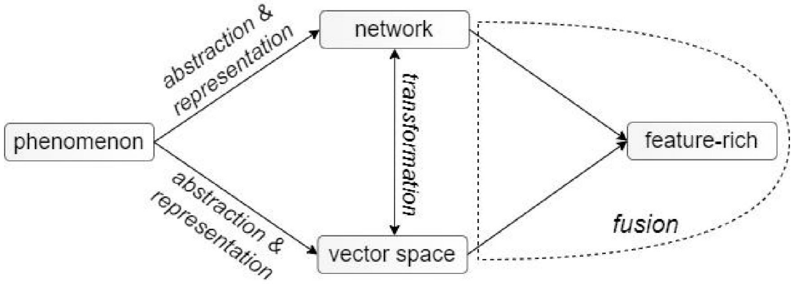
Interdonato, R., Atzmueller, M., Gaito, S., Kanawati, R., Largeron, C., & Sala, A. (2019).
Feature-rich networks: going beyond complex network topologies. *Applied Network Science*

Feature-rich Networks are not...

...Transformations

Feature-rich modeling vs. transformations:

- **Transformation:** “a proper representation of a complex system can be derived from the **F**eatures, **S**imilarity, and **C**onnectivity of the elements contained in the system”;
- **Feature-rich:** “independence” of such elements.



More about transformations:

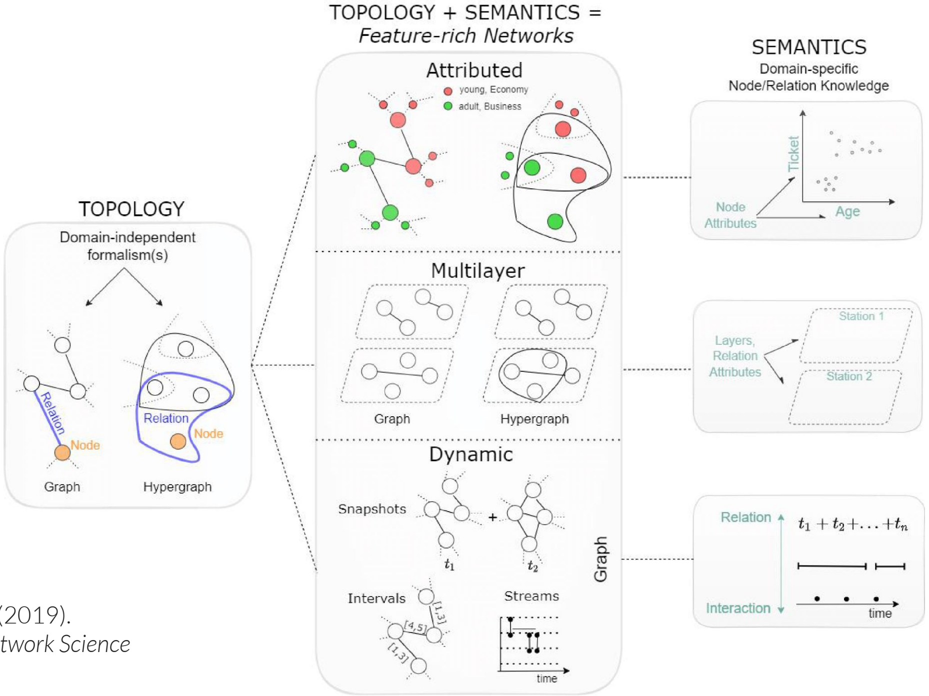
Comin, C. H., Peron, T., Silva, F. N., Amancio, D. R., Rodrigues, F. A., & Costa, L. D. F. (2020). Complex systems: Features, similarity and connectivity. *Physics Reports*

Feature-rich Networks are...

...Models exposing more features in addition to the network topology

Rich information is available on nodes and edges:

- Enhance semantic/domain-specific knowledge;
- Different types of information is represented (e.g., “relations” vs. “interactions” in temporal networks.



Interdonato, R., Atzmueller, M., Gaito, S., Kanawati, R., Largeron, C., & Sala, A. (2019). Feature-rich networks: going beyond complex network topologies. *Applied Network Science*

Feature-rich Networks

From the original paper:

- Several modeling choices…
- …sharing similar characteristics and problems;
- Wide and flexible definition;

…So?

Such augmented representations require new algorithms for extracting patterns and discovering knowledge.

- Attributed graphs, e. g. networks enclosing (vectors of) generic attributes on nodes and edges ("Attributed graphs" section);
- Heterogeneous information networks, e. g. networks modeling heterogeneous node and edge types ("Heterogeneous information networks" section);
- Multilayer networks, e. g. representing different online/offline relations between the same set of users ("Multilayer networks" section);
- Temporal networks, e. g. modeling discrete/continuous time aspects in networked data ("Temporal networks" section);
- Location-aware Networks, e. g. useful for the definition of recommender system (RecSys) applications like itinerary routing and points of interest (PoIs) planning ("Location-aware networks" section);
- Probabilistic networks, e. g. networks modeling uncertain relations, such as sensor networks, or networks inferred from survey data ("Probabilistic networks" section).

Please note that the definition of feature-rich network has been kept intentionally wide and flexible, with the aim to gather under a common denomination a series of network models exhibiting different structures and that were introduced for different needs, but that at the same time show some common characteristics and can lead to similar problems.

from Interdonato, R., Atzmueller, M., Gaito, S., Kanawati, R., Largeron, C., & Sala, A. (2019). Feature-rich networks: going beyond complex network topologies. *Applied Network Science*

Feature-rich Network Mining

A focus on Attributed Community Detection

Community Detection on Networks with Attributes

Identifying **well-connected** and **homogeneous** groups of nodes in complex networks

“homogeneous” w.r.t. **nodes’ attributes**

RECAP: Community Detection

- The aim of Community Detection algorithms is to identify the **meso-scale topologies** hidden within complex network structures;
- Cluster similar nodes relying on **topological information**;
- Ill-posed problem: *what is a community?*
- Not universally shared definitions;
- Algorithms (and taxonomies) based on different properties:
    - Internal Density (Modularity);
    - Distance;
    - Entity Closeness;
    - Link Communities.

Definition 1

"A set of entities where each entity is closer to the other entities within the communities than to the entities outside it"

Definition 2

"A set of entities more tightly connected within each other than with nodes belonging to other sets"

RECAP: Louvain

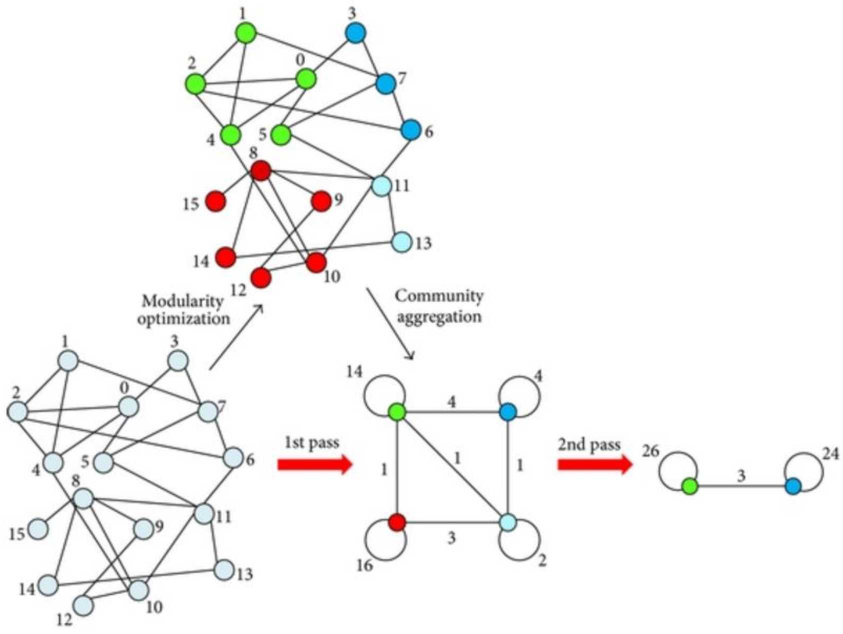
In order to maximize this value efficiently, the Louvain Method has two phases that are repeated iteratively.

Initialization:
Each node in the network is assigned to its own community.

- Phase 1:
Each node is then moved into the adjacent community that guarantee the greatest modularity increase.

- Phase 2:
A new graph is created: its nodes are the updated communities and weighted links connect them accounting for bridges in the original graph.

Phases 1 and 2 are repeated until modularity is maximized



VD Blondel, et al. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment (2008)

RECAP: Modularity

How to assure high density?

General Idea:

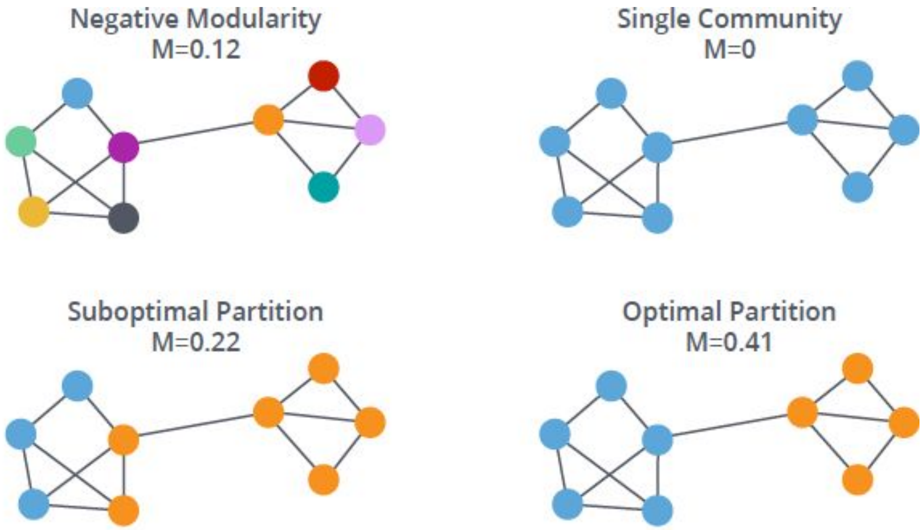
- define a quality function that measures the density of a community and then try to maximize it

Modularity [-1, 1]

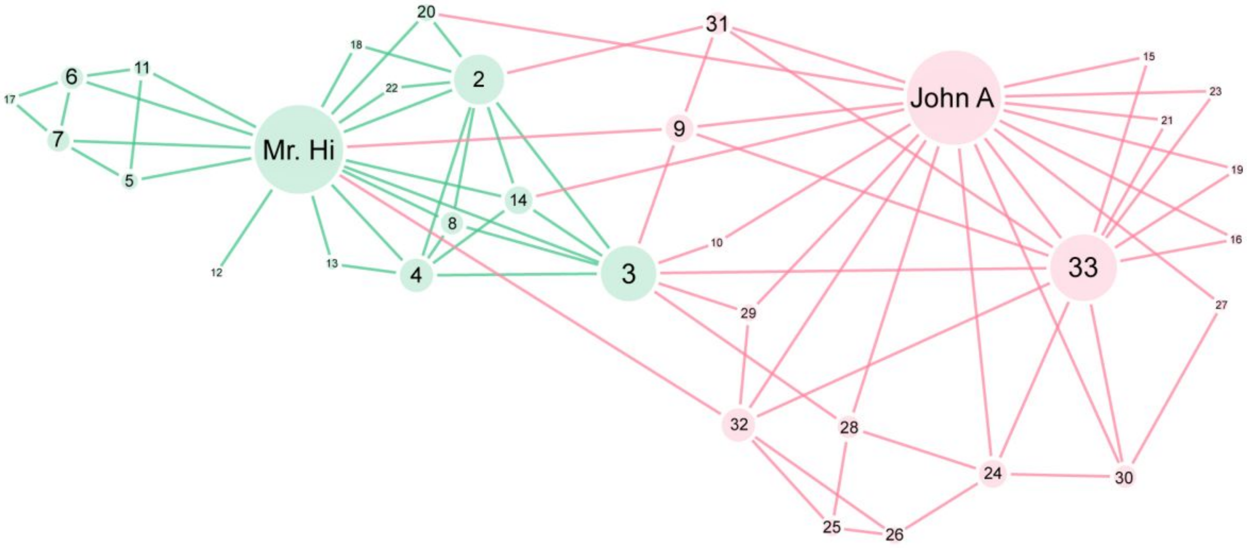
$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

Null Model expected density

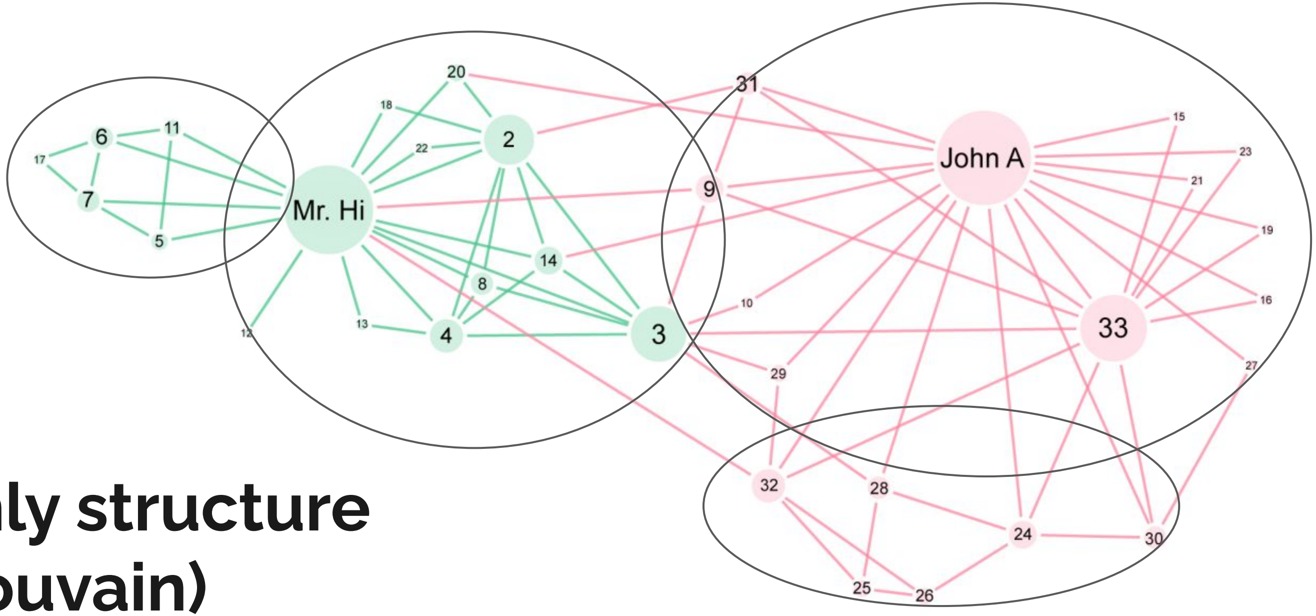
1 if i,j in same community, 0 otherwise



Community Detection

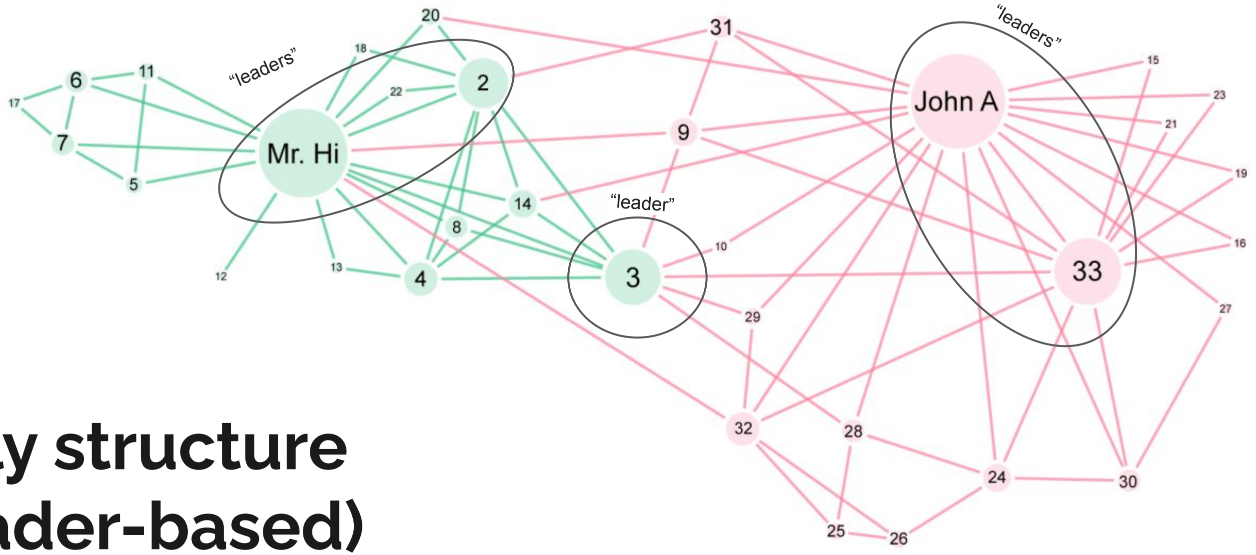


Community Detection



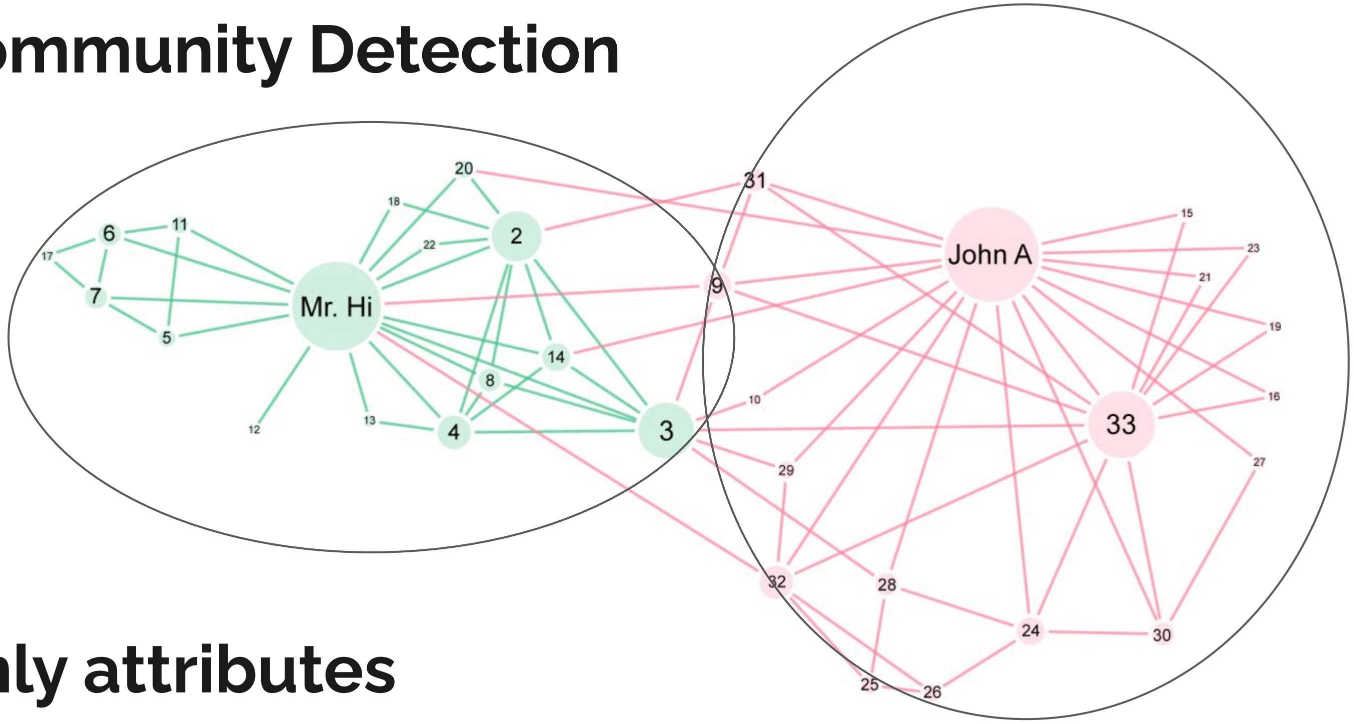
Only structure (Louvain)

Community Detection



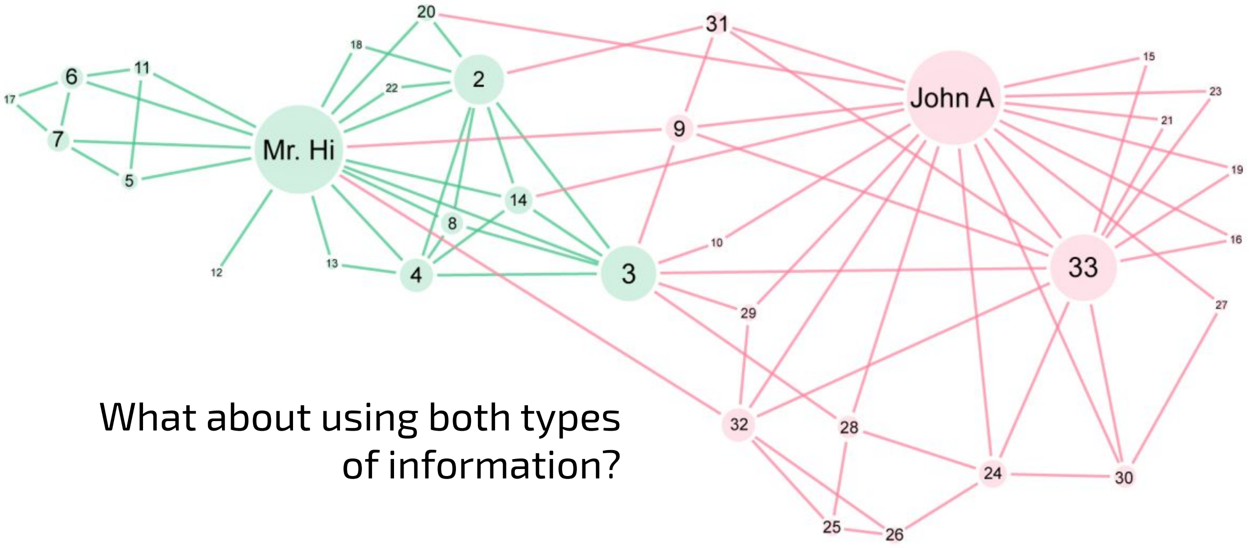
Only structure (leader-based)

Community Detection



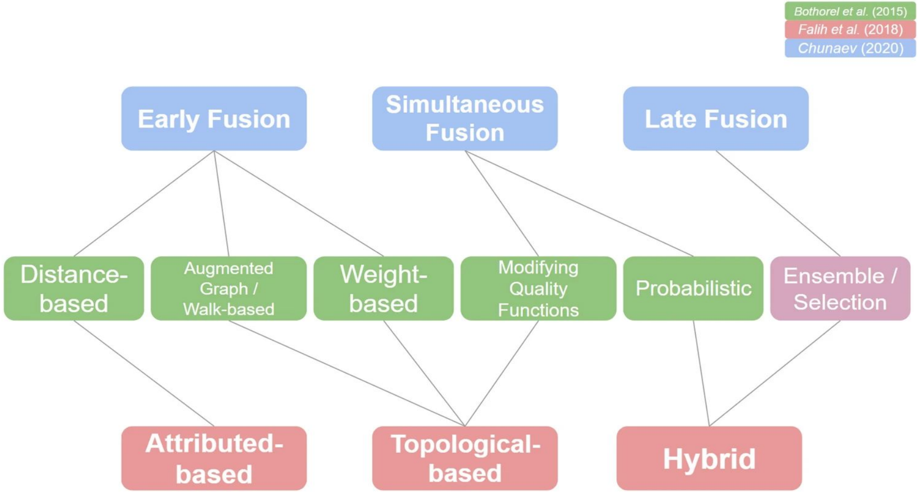
Only attributes

Community Detection



What about using both types of information?

Attributed Community Detection

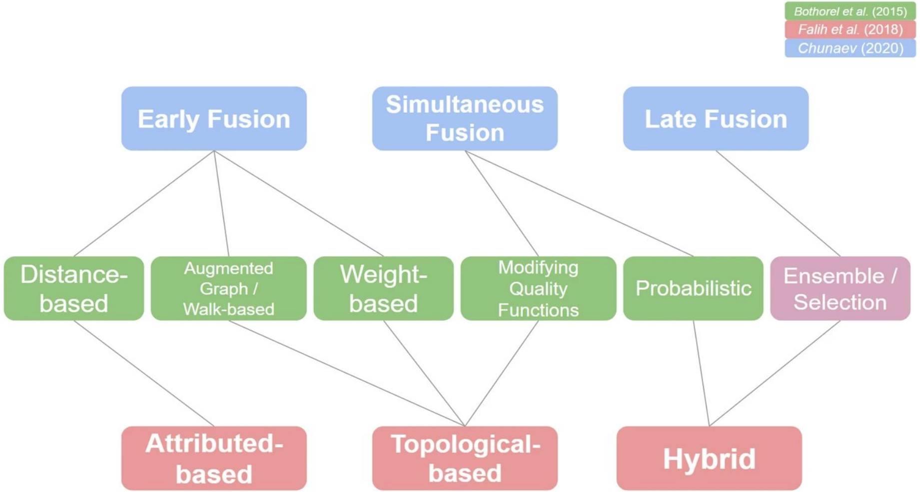


Falih, I., Grozavu, N., Kanawati, R., & Bennani, Y. (2018, April). Community detection in attributed network. In *Companion proceedings of the the web conference 2018* (pp. 1299-1306).

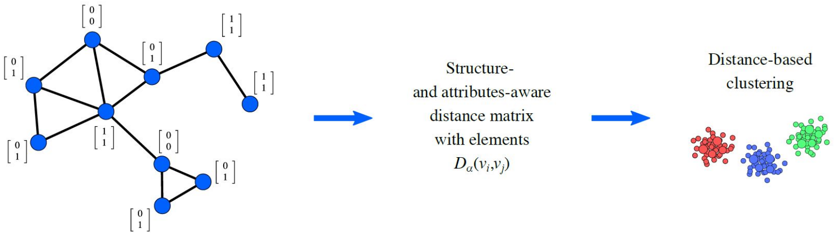
Bothorel, C., Cruz, J. D., Magnani, M., & Micenkova, B. (2015). Clustering attributed graphs: models, measures and methods. *Network Science*, *3*(3), 408-444.

Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, *37*, 100286.

Attributed Community Detection



DISTANCE-BASED, ATTRIBUTED-BASED, EARLY FUSION



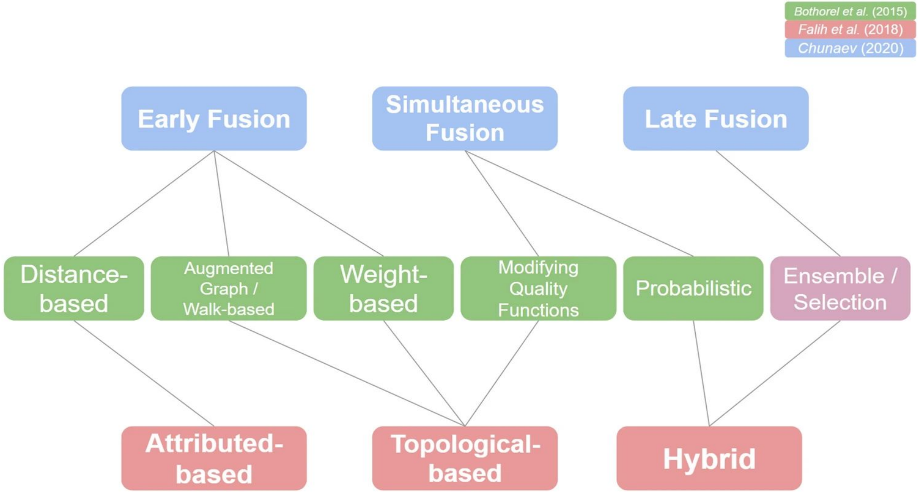
from Chunaev, P. (2020)

Falih, I., Grozavu, N., Kanawati, R., & Bennani, Y. (2018, April). Community detection in attributed network. In *Companion proceedings of the the web conference 2018* (pp. 1299-1306).

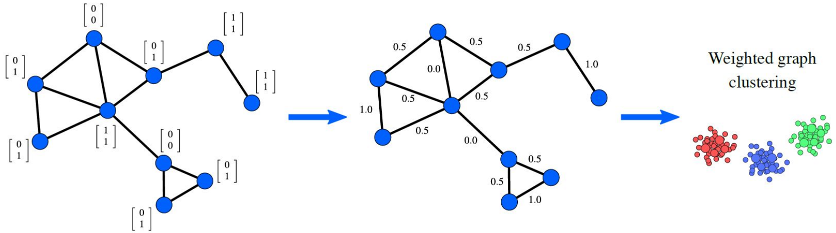
Bothorel, C., Cruz, J. D., Magnani, M., & Micenkova, B. (2015). Clustering attributed graphs: models, measures and methods. *Network Science*, *3*(3), 408-444.

Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, *37*, 100286.

Attributed Community Detection



WEIGHT-BASED, TOPOLOGICAL-BASED, EARLY FUSION



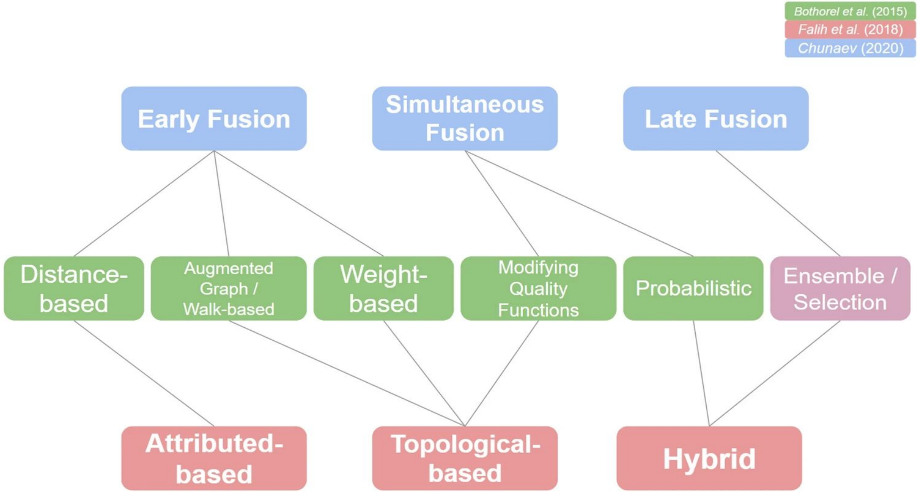
from Chunaev, P. (2020)

Falih, I., Grozavu, N., Kanawati, R., & Bennani, Y. (2018, April). Community detection in attributed network. In *Companion proceedings of the the web conference 2018* (pp. 1299-1306).

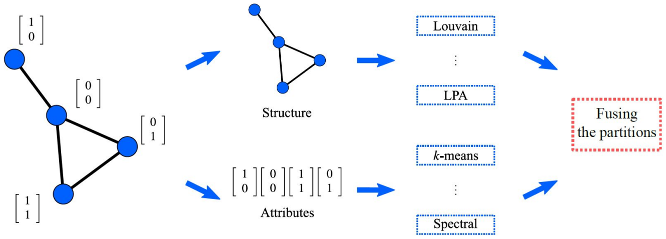
Bothorel, C., Cruz, J. D., Magnani, M., & Micenkova, B. (2015). Clustering attributed graphs: models, measures and methods. *Network Science*, *3*(3), 408-444.

Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, *37*, 100286.

Attributed Community Detection



ENSEMBLE, HYBRID, LATE FUSION



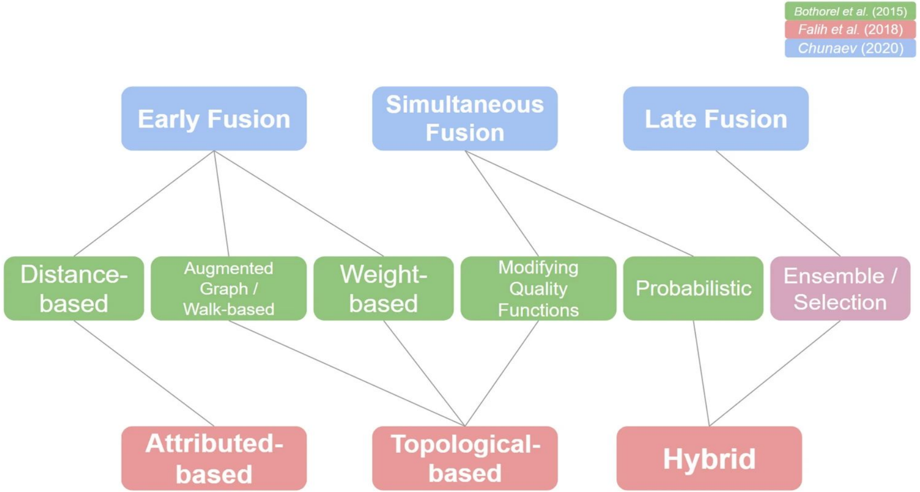
from Chunaev, P. (2020)

Falih, I., Grozavu, N., Kanawati, R., & Bennani, Y. (2018, April). Community detection in attributed network. In *Companion proceedings of the the web conference 2018* (pp. 1299-1306).

Bothorel, C., Cruz, J. D., Magnani, M., & Micenkova, B. (2015). Clustering attributed graphs: models, measures and methods. *Network Science*, *3*(3), 408-444.

Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, *37*, 100286.

Attributed Community Detection



SIMULTANEOUS FUSION

More difficult to generalize:

- methods based on modifying modularity;
- probabilistic methods;
- SBM methods;
- etc.

Falih, I., Grozavu, N., Kanawati, R., & Bennani, Y. (2018, April). Community detection in attributed network. In *Companion proceedings of the the web conference 2018* (pp. 1299-1306).

Bothorel, C., Cruz, J. D., Magnani, M., & Micenkova, B. (2015). Clustering attributed graphs: models, measures and methods. *Network Science*, *3*(3), 408-444.

Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, *37*, 100286.

Attributed Community Detection
Simultaneous Approach (EVA)

How to assure high density and attribute homogeneity?

General Idea:

- define a quality function as a **linear combination** of a structural measure for community density (e.g., modularity) and attribute measure for node homogeneity (e.g., purity) and then maximize it.

- Maintain the same two-phase schema of Louvain

Modularity

$$Q = \frac{1}{(2m)} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{(2m)} \right] \delta(c_v, c_w)$$

Purity

$$P_c = \prod_{a \in A} \frac{\max_{a \in A} \left(\sum_{v \in c} a(v) \right)}{|c|}$$

$$P = \frac{1}{|C|} \sum_{c \in C} P_c$$

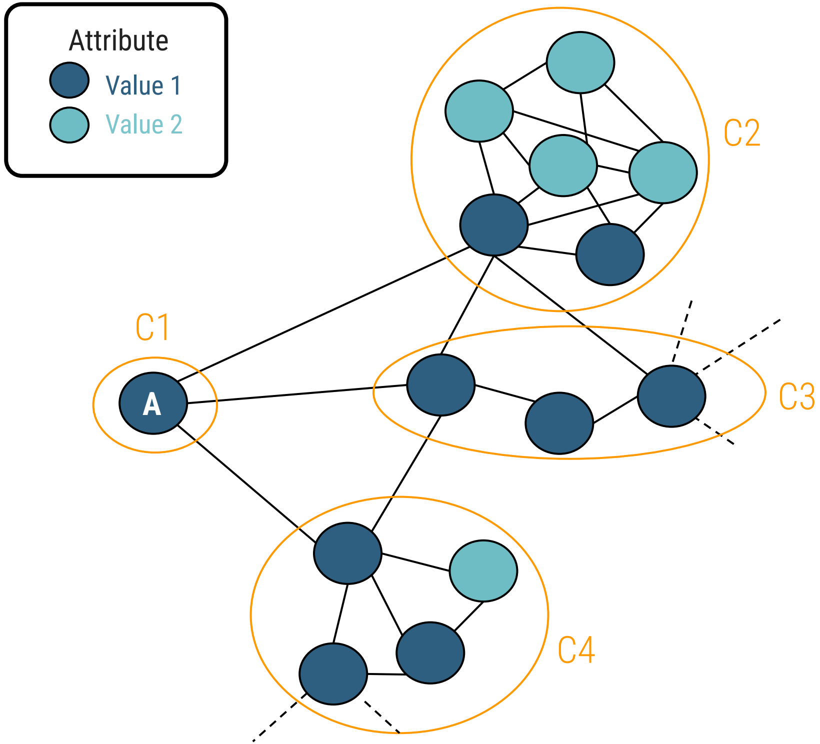
Optimization function

$$Z = \alpha P + (1 - \alpha) Q$$

Citraro, S., & Rossetti, G. (2020). Identifying and exploiting homogeneous communities in labeled networks. *Applied Network Science*, *5*(1), 1-20.

Github: https://github.com/GiulioRossetti/Eva

EVA: (Louvain) “E”xtendend to “V”ertex “A”ttributes



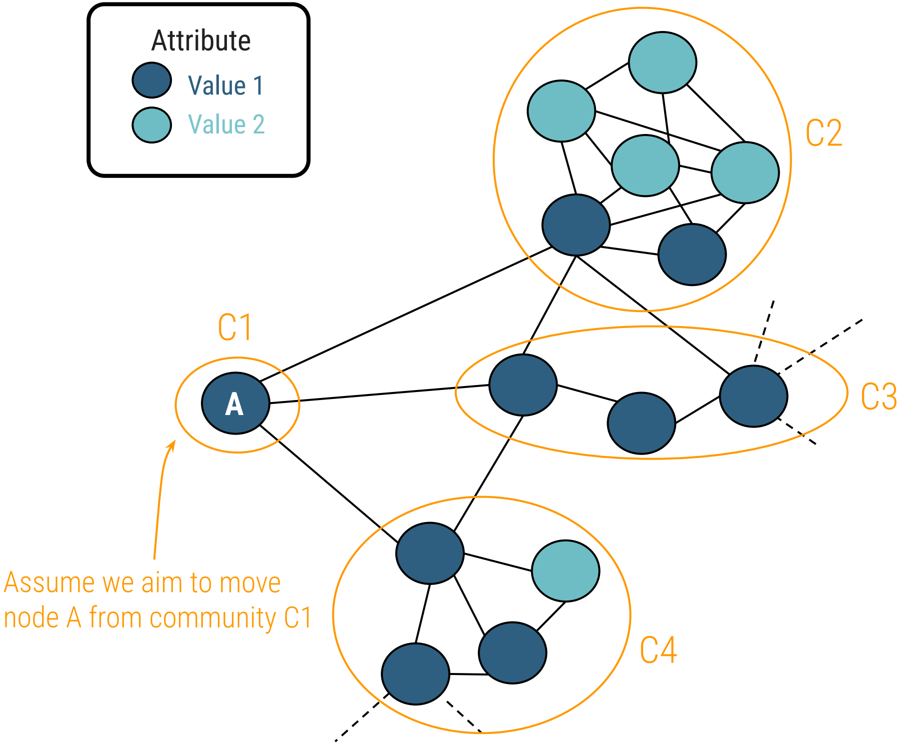
Phase 1

1. Initialize singletons;
2. Repeat until no increases in *Z* occur:
   a. Remove a node from its community;
   b. Move the node to the communities of its neighbors while computing changes in *Z*;
   c. Assign to the node the community label that maximizes *Z* (if tie, move to a bigger one).

Phase 2

1. Transform communities in the nodes of a new graph;
2. Re-apply Phase 1 until no increases in *Z* occur.

EVA: (Louvain) “E”xtendend to “V”ertex “A”ttributes



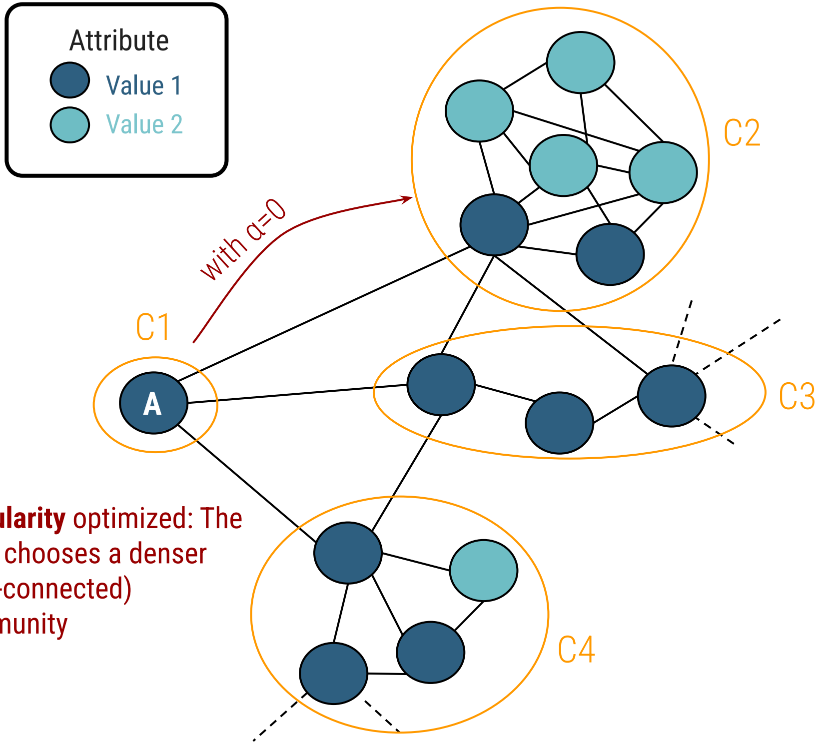
Phase 1

1. Initialize singletons;
2. Repeat until no increases in *Z* occur:
   a. Remove a node from its community;
   b. Move the node to the communities of its neighbors while computing changes in *Z*;
   c. Assign to the node the community label that maximizes *Z* (if tie, move to a bigger one).

Phase 2

1. Transform communities in the nodes of a new graph;
2. Re-apply Phase 1 until no increases in *Z* occur.

EVA: (Louvain) “E”xtendend to “V”ertex “A”ttributes



Modularity optimized: The node chooses a denser (well-connected) community

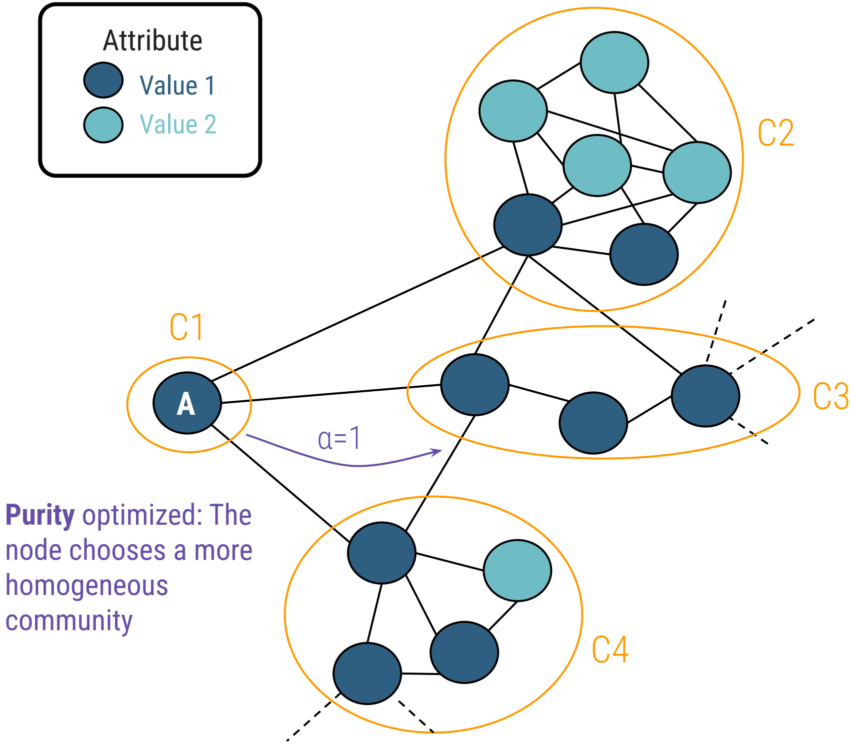
Phase 1

1. Initialize singletons;
2. Repeat until no increases in *Z* occur:
   a. Remove a node from its community;
   b. Move the node to the communities of its neighbors while computing changes in *Z*;
   c. Assign to the node the community label that maximizes *Z* (if tie, move to a bigger one).

Phase 2

1. Transform communities in the nodes of a new graph;
2. Re-apply Phase 1 until no increases in *Z* occur.

EVA: (Louvain) “E”xtendend to “V”ertex “A”ttributes



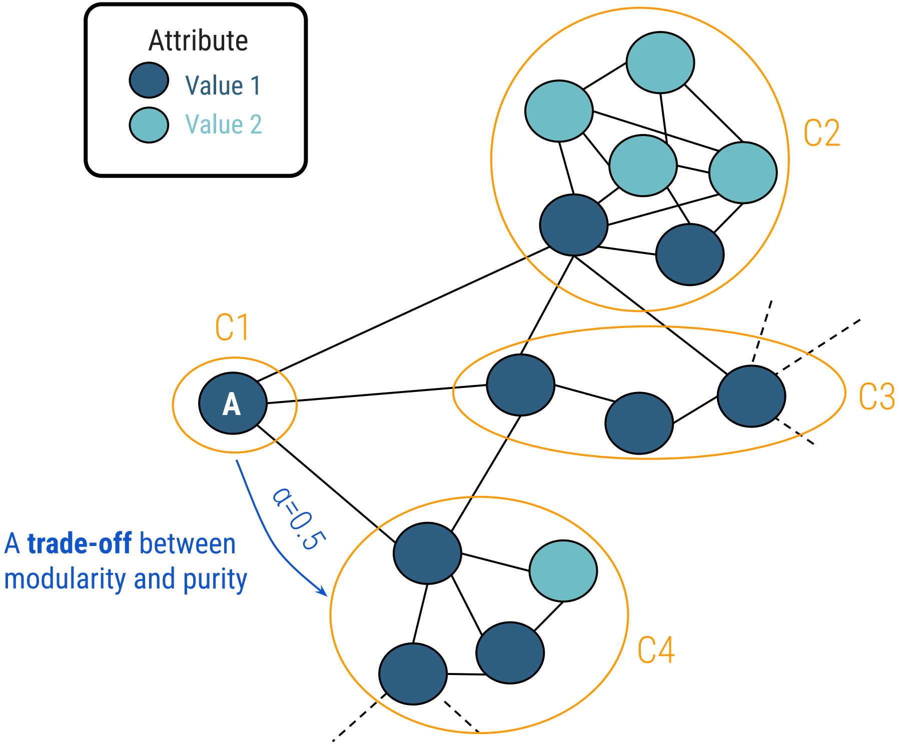
Phase 1

1. Initialize singletons;
2. Repeat until no increases in *Z* occur:
   a. Remove a node from its community;
   b. Move the node to the communities of its neighbors while computing changes in *Z*;
   c. Assign to the node the community label that maximizes *Z* **(if tie, move to a bigger one)**.

Phase 2

1. Transform communities in the nodes of a new graph;
2. Re-apply Phase 1 until no increases in *Z* occur.

EVA: (Louvain) “E”xtendend to “V”ertex “A”ttributes



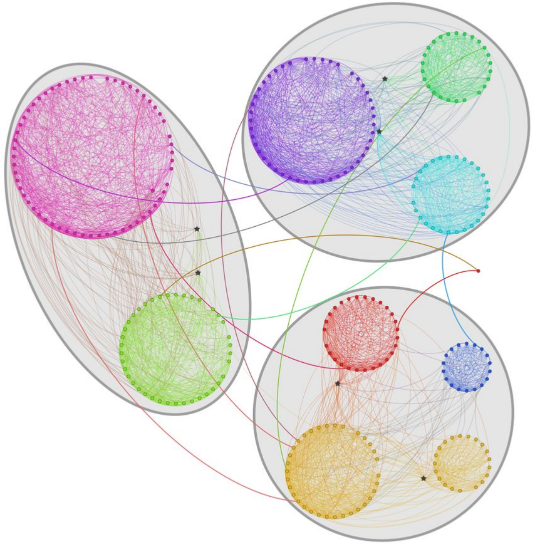
Phase 1

1. Initialize singletons;
2. Repeat until no increases in *Z* occur:
    a. Remove a node from its community;
    b. Move the node to the communities of its neighbors while computing changes in *Z*;
    c. Assign to the node the community label that maximizes *Z* (if tie, move to a bigger one).

Phase 2

1. Transform communities in the nodes of a new graph;
2. Re-apply Phase 1 until no increases in *Z* occur.

RECAP: Evaluating Community Detection Algorithms



Testing **against topological ground truths**

Synthetic graphs with embedded community structure
(e.g., LFR)

- More stable than semantic ground truth partitions
- Community structure depends on the fitness function optimized by the chosen model
- Approximation of real world networks

Lancichinetti, Andrea, Santo Fortunato, and Filippo Radicchi. "Benchmark graphs for testing community detection algorithms." Physical review E 78.4 (2008): 046110.

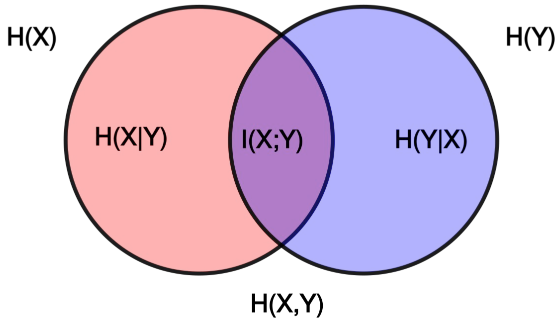
RECAP: NMI

Normalized Mutual Information is a measure of *similarity* borrowed from information theory:

$$NMI(X,Y) = \frac{H(X)+H(Y)-H(X,Y)}{\frac{H(X)+H(Y)}{2}} \in [0,1]$$

- *H(X)* is the entropy of the random variable *X* associated to an identified community,
- *H(Y)* is the entropy of the random variable *Y* associated to a ground truth community,
- *H(X,Y)* is the joint entropy.

The higher the NMI the more similar the compared partitions are



Advantages

- Extensively used in literature

Drawbacks

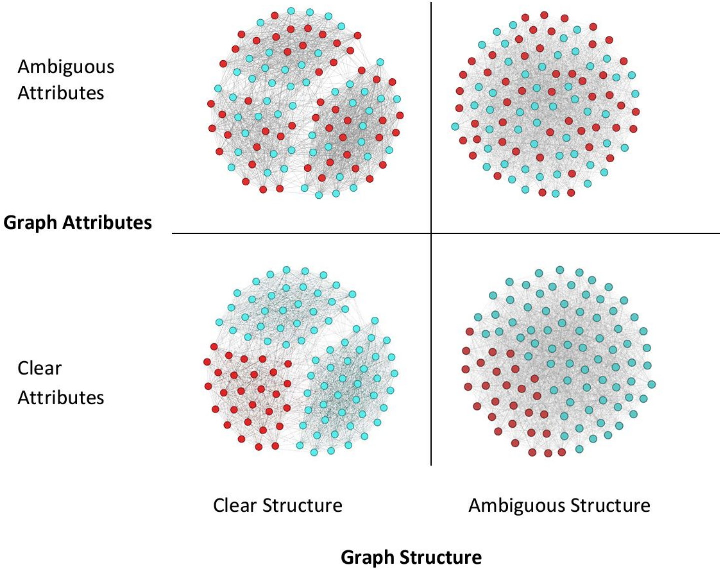
- Computational complexity ~ $O(|C|^2)$ (where C is the community set)
- Needs to be approximated in case of overlapping partitions

Attributed Community Detection
Evaluation

Testing against **topological and homogeneous ground-truths:**

- Synthetic graphs with embedded community structure which are also homogeneous w.r.t. node attributes within the communities:
  - Many LFRs extensions, e.g., **LFR-EA**, combining **structure mixing** and **attribute noise**;

- Common methodology of evaluation:
  - Generate network with communities;
  - Run a node-attributed community detection;
  - Evaluate the ground-truth and the algorithm partition (**NMI, ARI, etc**).



Elhadi, H., & Agam, G. (2013, August). Structure and attributes community detection: comparative analysis of composite, ensemble and selection methods. In *Proceedings of the 7th workshop on social network mining and analysis* (pp. 1-7).

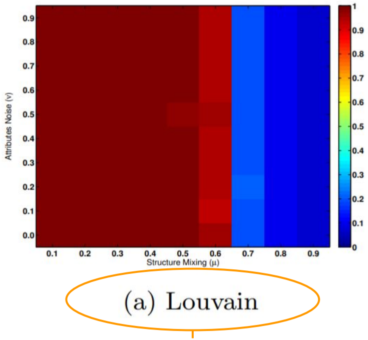
Attributed Community Detection
Evaluation

Testing against **topological and homogeneous ground-truths:**

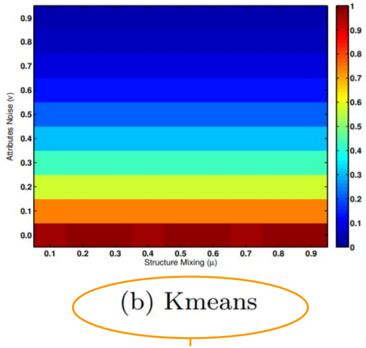
- Synthetic graphs with embedded community structure which are also homogeneous w.r.t. node attributes within the communities:
  - Many LFRs extensions, e.g., **LFR-EA**, combining **structure mixing** and **attribute noise**;

- Common methodology of evaluation:
  - Generate network with communities;
  - Run a node-attributed community detection;
  - Evaluate the ground-truth and the algorithm partition (**NMI, ARI, etc**).

LFR-EA, from *Elhadi and Agam* (2013)



connectivity only



attributes only

Elhadi, H., & Agam, G. (2013, August). Structure and attributes community detection: comparative analysis of composite, ensemble and selection methods. In *Proceedings of the 7th workshop on social network mining and analysis* (pp. 1-7).

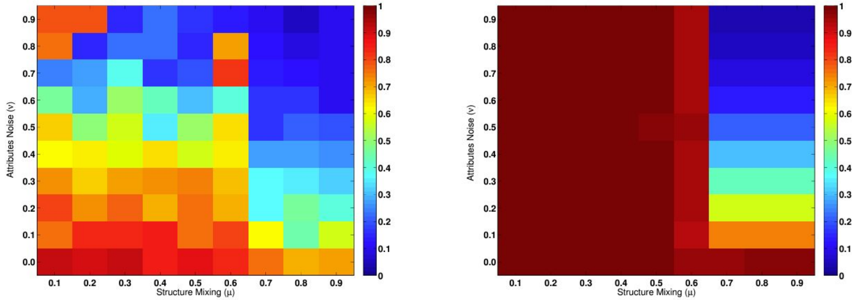
Attributed Community Detection
Evaluation

Testing against **topological and homogeneous ground-truths:**

- Synthetic graphs with embedded community structure which are also homogeneous w.r.t. node attributes within the communities:
    - Many LFRs extensions, e.g., **LFR-EA**, combining **structure mixing** and **attribute noise**;

- Common methodology of evaluation:
    - Generate network with communities;
    - Run a node-attributed community detection;
    - Evaluate the ground-truth and the algorithm partition (**NMI, ARI, etc**).

LFR-EA, from *Elhadi and Agam* (2013)



(g) CSPA (h) Selection

Node-attributed approaches

Elhadi, H., & Agam, G. (2013, August). Structure and attributes community detection: comparative analysis of composite, ensemble and selection methods. In *Proceedings of the 7th workshop on social network mining and analysis* (pp. 1-7).

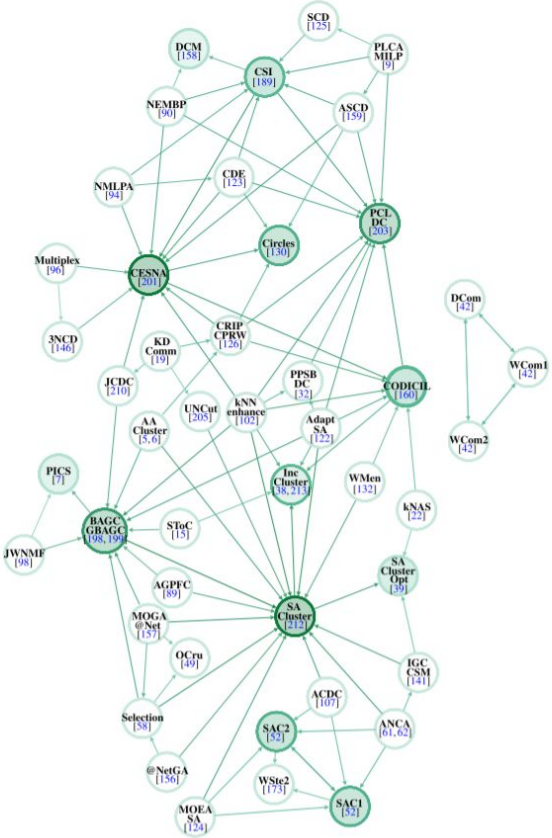
Attributed Community Detection

Final Overview

No “standard”?

A directed graph where:

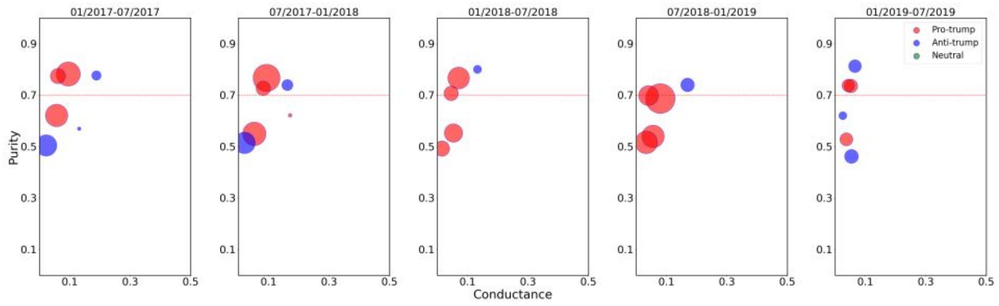
- Nodes are Attributed Community Detection Algorithms;
- Links are algorithms that test another algorithm against the first one (i.e., there is an incoming edge if a method is used as competitor);
- Filled green nodes have the highest PageRank: a **standard?**



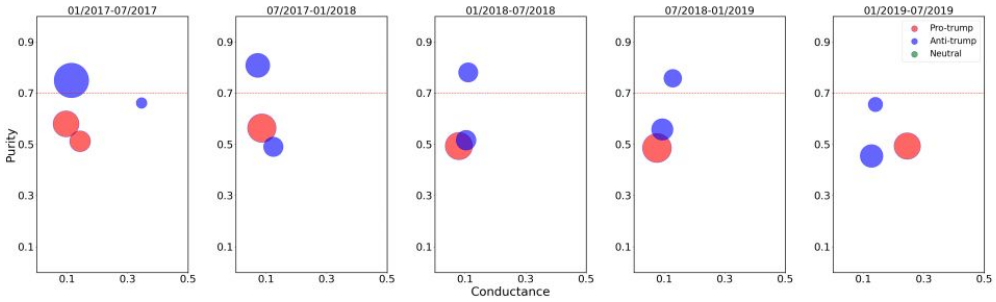
Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, *37*, 100286.

Applications

EVA used in identifying **echo-chambers** in online social networks (ex. from political boards on Reddit)



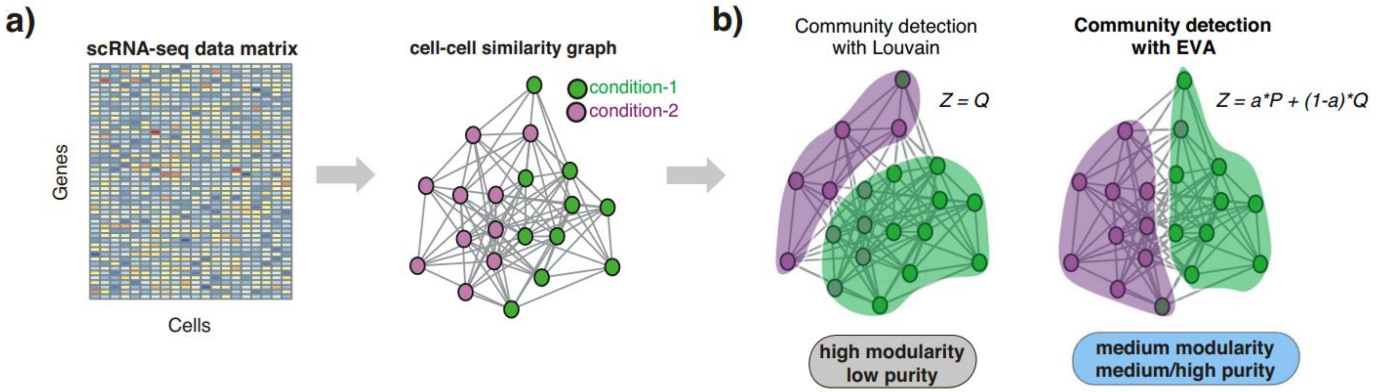
(b) Minorities Discrimination



(c) Political Sphere

Figures from Morini, V., Pollacci, L., & Rossetti, G. (2021). Toward a standard approach for echo chamber detection: Reddit case study. Applied Sciences, 11(12), 5390.

Applications



EVA helped in identifying **differential abundance patterns** in cell-cell complex networks

Figures from Maity, A. K., & Teschendorff, A. E. (2023). Cell-attribute aware community detection improves differential abundance testing from single-cell RNA-Seq data. Nature Communications, 14(1), 3244.

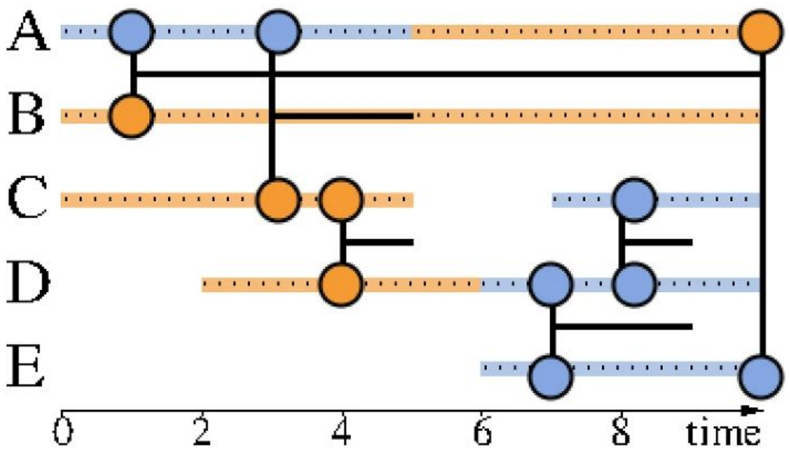
Feature-rich Network Mining

Advanced Network Representations

Attributed Stream Graphs

- Models combined:
    - Dynamic Networks as Stream Graphs;
    - Attributed Networks;

- **Why?**
    - Attributes can change over time;

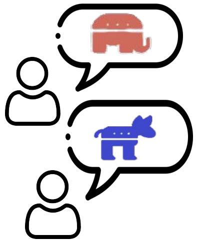


Latapy, M., Viard, T., & Magnien, C. (2018). Stream graphs and link streams for the modeling of interactions over time. *Social Network Analysis and Mining*, *8*, 1-29.

Attributed Stream Hypergraphs (ASHs)

A

Node
Attributes



Correlation between
structure and metadata

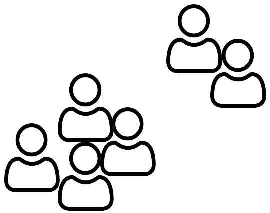
S

Evolving
Topologies

Structures (and metadata)
vary over time

H

High-order
Interactions



Beyond pairwise/dyadic
connectivity patterns

Chapter 7

Conclusion

Take Away Messages

1. Feature-rich Networks involve fusing structure and attributes;
2. They can be used for augmented network mining, .e.g, community detection

Suggested Readings

- Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, *37*, 100286.
- Interdonato, R., Atzmueller, M., Gaito, S., Kanawati, R., Largeron, C., & Sala, A. (2019). Feature-rich networks: going beyond complex network topologies. *Applied Network Science*

What's Next

Chapter 8: How to validate