

On the Separability of Structural Classes of Communities



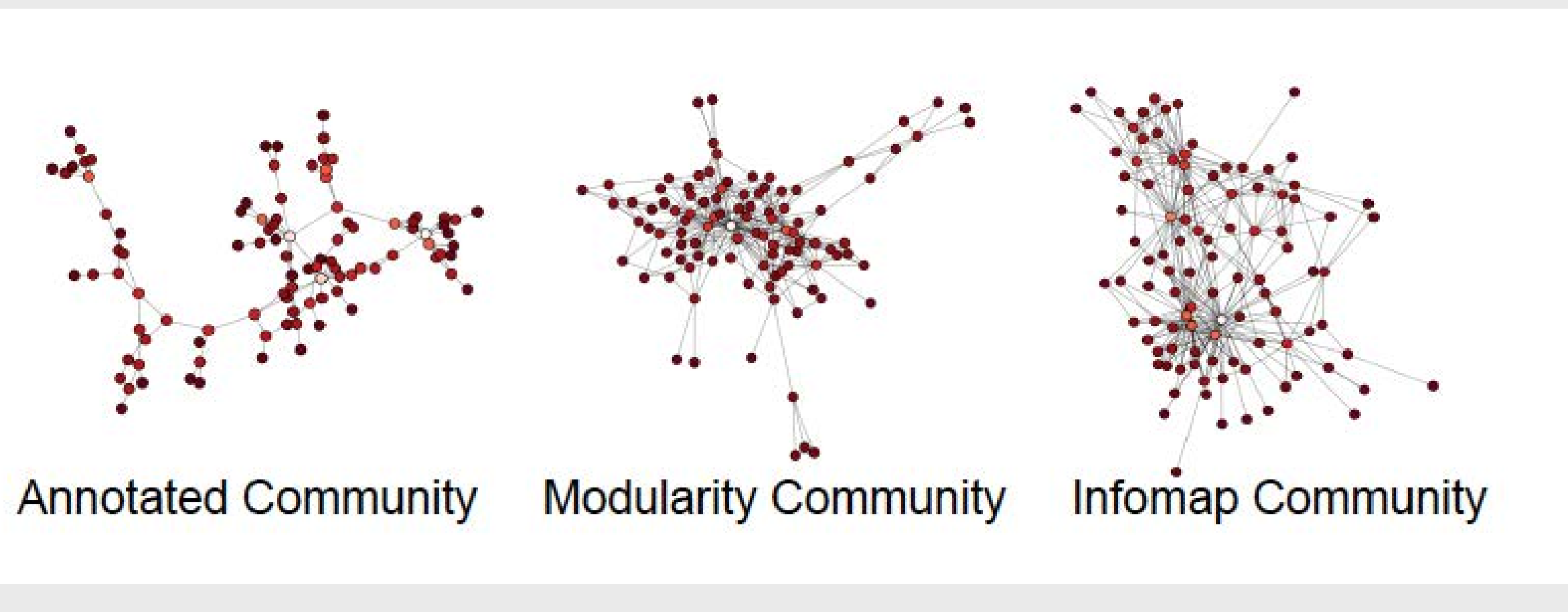
Problem

Motivation:

There are hundreds of community detection algorithms in the literature! How can we understand and categorize their behavior?

Our Problems:

1. How different are the communities identified by different community detection algorithms?
2. Do detected communities reflect the structure of real communities?
3. How can we measure and compare structural profiles of different types of communities?



Overview

Idea:

1. Obtain several classes of communities
2. Characterize communities using structural features
3. Use a classifier to compare structure of different classes of communities

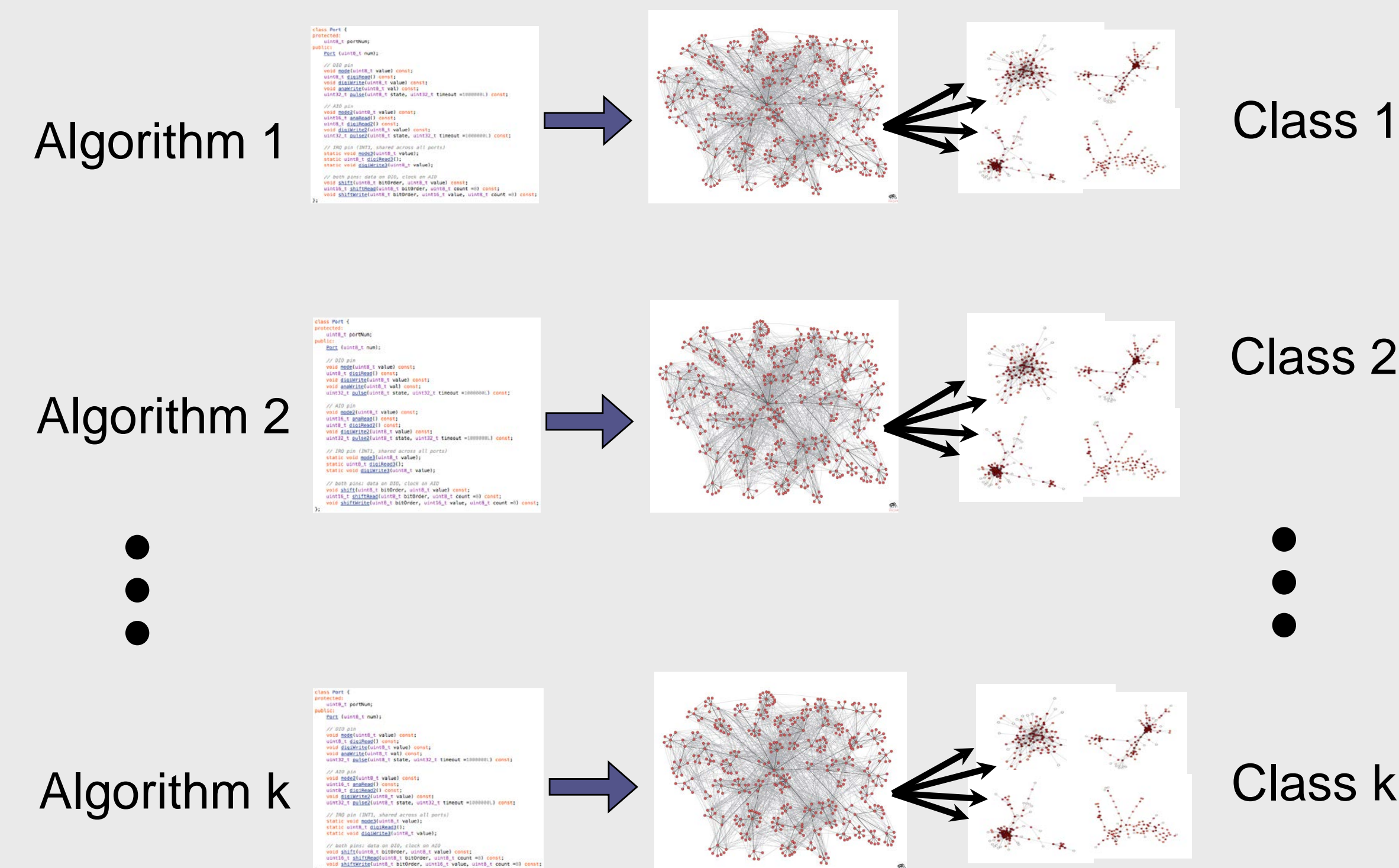
Classes of Communities

1. Extracted communities: Communities found by different algorithms. They can be overlapping or disjoint.
2. Annotated communities: Communities identified using ground-truth metadata (e.g., students in the same major at a university).

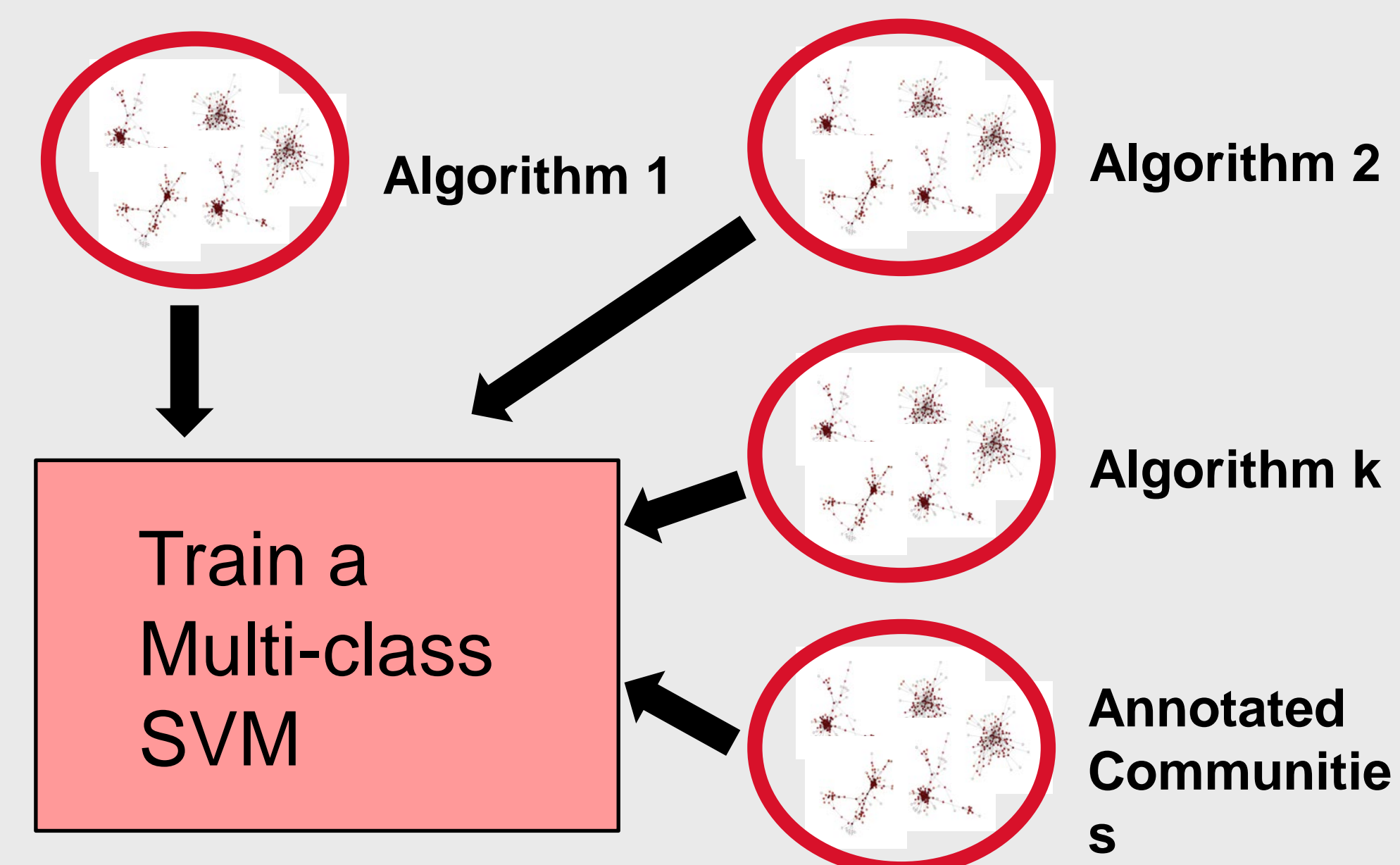
Structural Characteristics:

1. Describe each community with a feature vector containing important structural characteristics

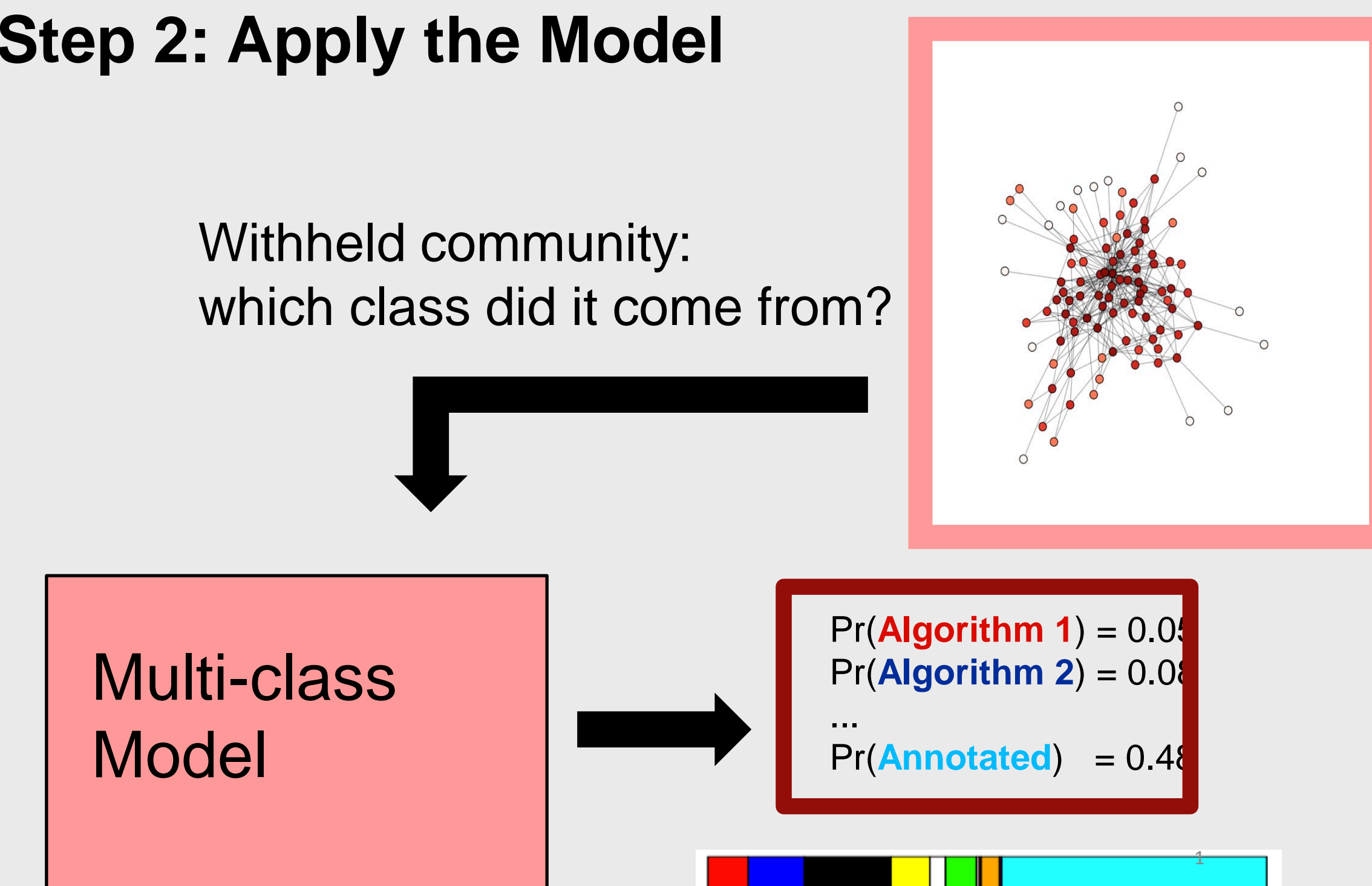
Our Method



Step 1: Train the Model



Step 2: Apply the Model



Big Questions:

1. Can the classifier correctly classify withheld communities? (This means the algorithms are producing different structures!)
2. Which algorithm class does the classifier classify annotated communities as?
3. Which features are most important for distinguishing between classes?

Experiments

Structural Features: Characterizing a Community:

- How big is the community (# nodes, # edges)?
- How compact is the community (diameter, density, etc.)?
- How well connected is the community to the network (conductance, etc.)
- Are there bottlenecks in the community (centrality measures)?

Algorithms:

Baselines

Breadth First Search
Random Walk
Random Walk Restart
 \pm^2 Communities

Comm. Detection Algs:

Infomap
MCL
Metis
Louvain Modularity
Newman Modularity
Link Communities

Annotated Communities:

Real communities from networks:

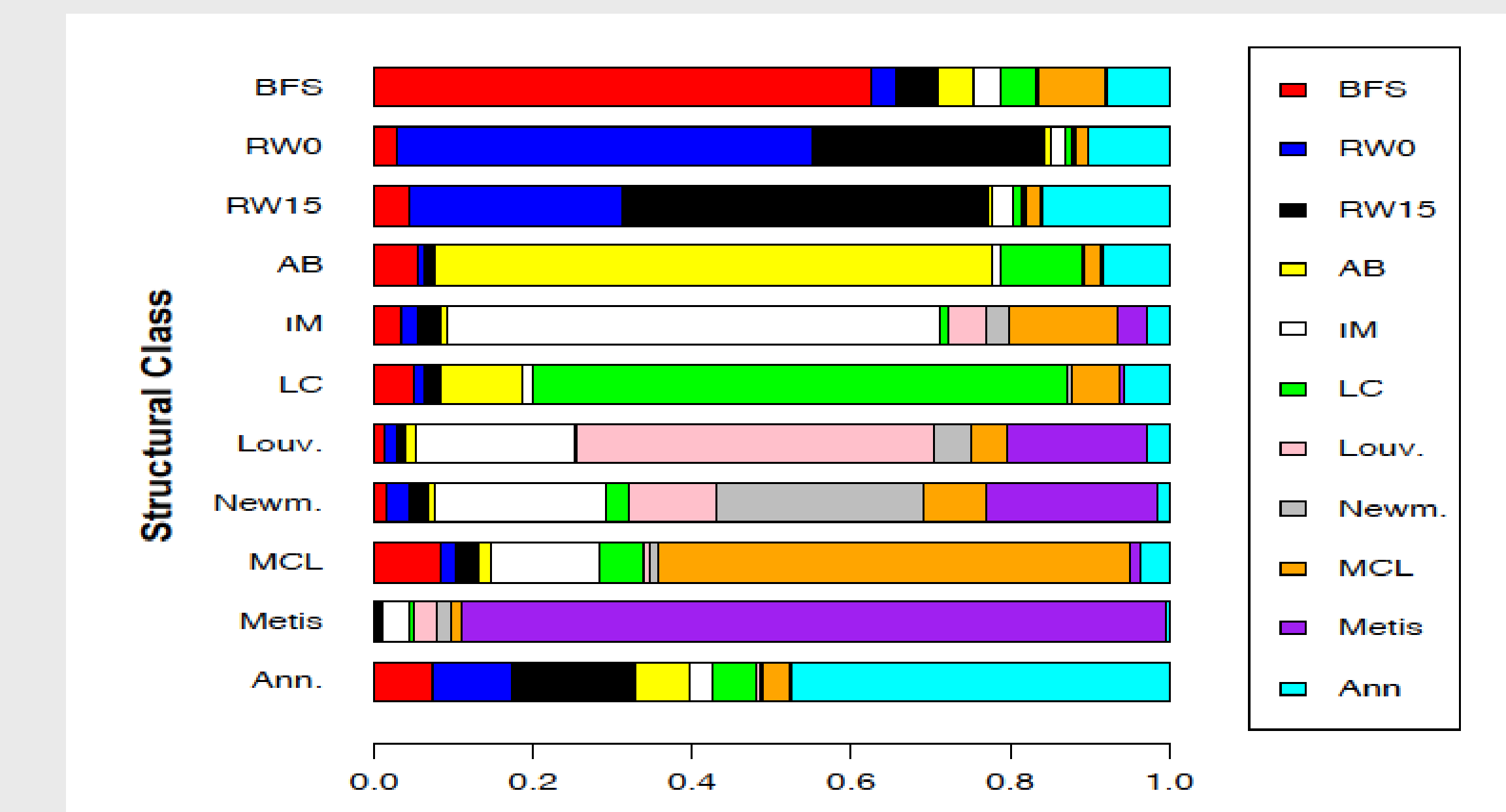
- Students from the same department
- Genes that play similar roles
- Books by the same author
- Bloggers in the same interest group

Datasets:

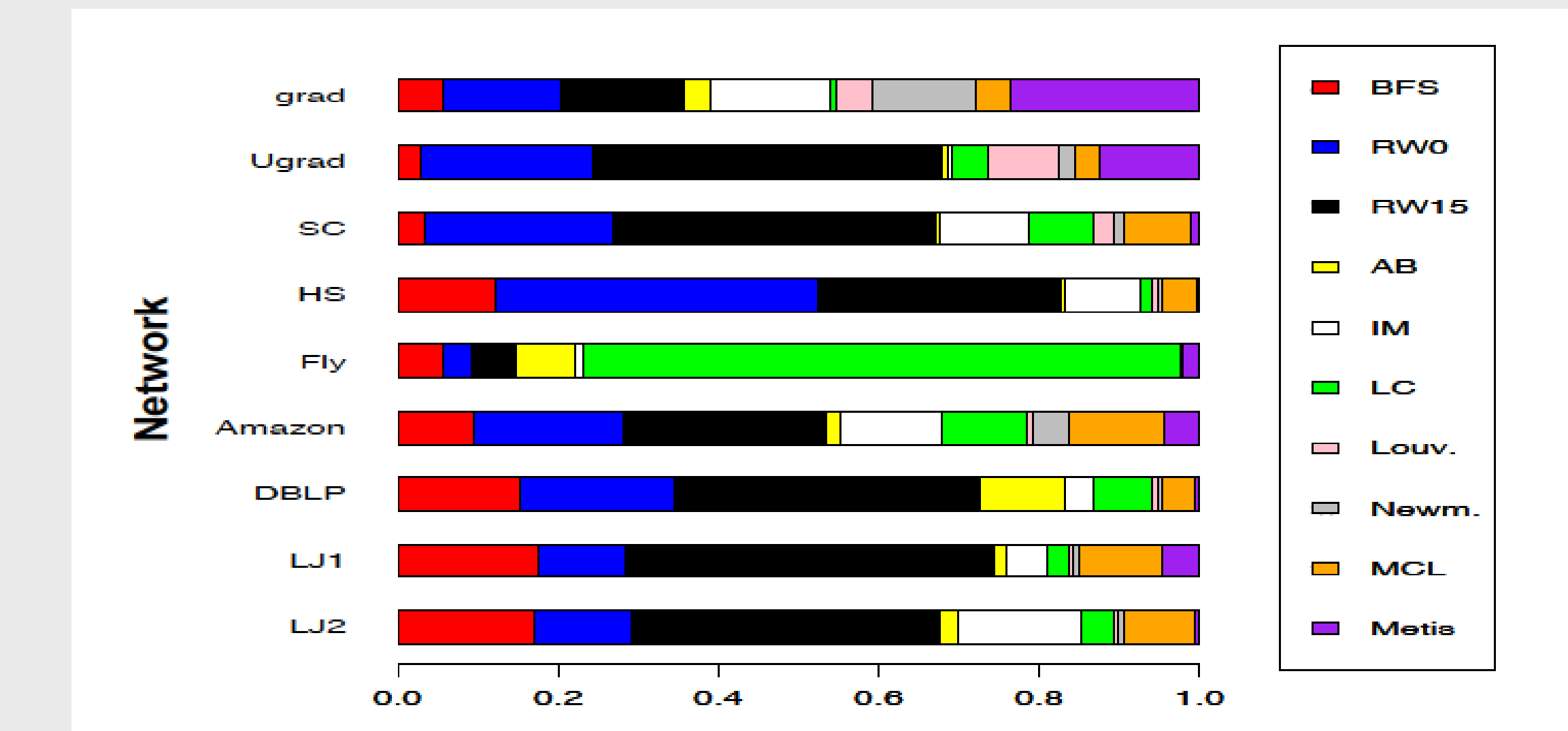


Results

Are the Classes Separable?



Which Algorithm Produces Communities Most Like the Annotated Communities?



Which Features are Useful for Distinguishing?

Algorithm	Conduct.	Diam.	Nd.Betw.	Inf.Cent.
Annotated	Medium	High/Low	High	Low
RW0, RW15	Low	High	High	Low
Louvain	High	High	Low	Medium
Newman	High	High	Low	Medium
AB	Medium	Low	Medium	High
LC	Medium	Low	Medium	High
BFS	Low	Low	Low	High
MCL	Low	Low	Low	Medium
Metis	High	Medium	Medium	Medium
IM	High	Medium	High	Medium

Reference: B. Abrahao, S. Soundarajan, R. Kleinberg, J. Hopcroft. *On the Separability of Structural Classes of Communities*. KDD 2012.

Authors: Bruno Abrahao (abrahao@cs.stanford.edu)
Sucheta Soundarajan (s.soundarajan@cs.rutgers.edu)
Robert Kleinberg (rdk@cs.cornell.edu)
John Hopcroft (jeh@cs.cornell.edu)