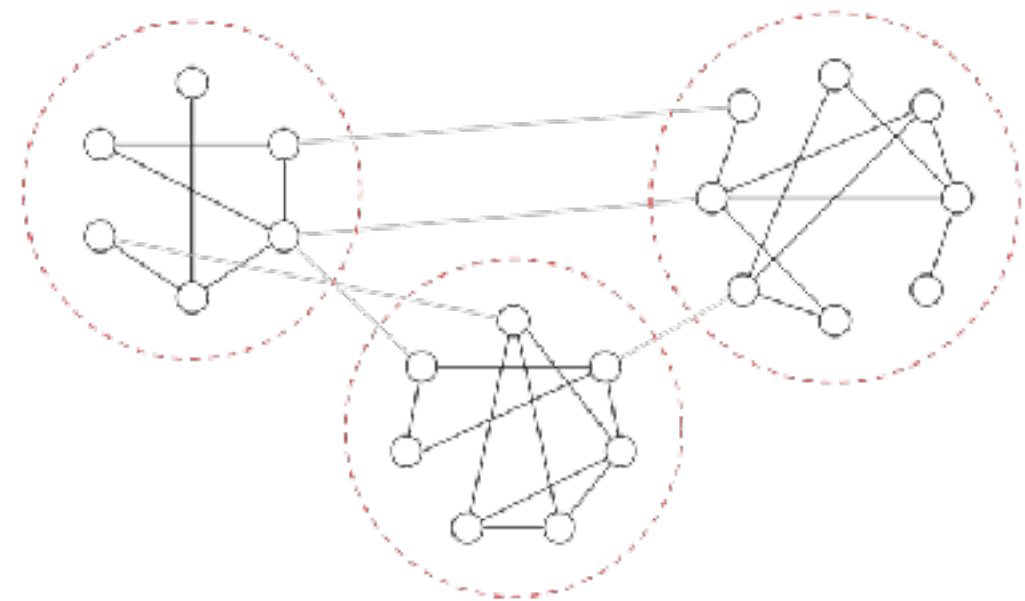


# Gridlock in Networks: The Leximin Method For Hierarchical Community Detection

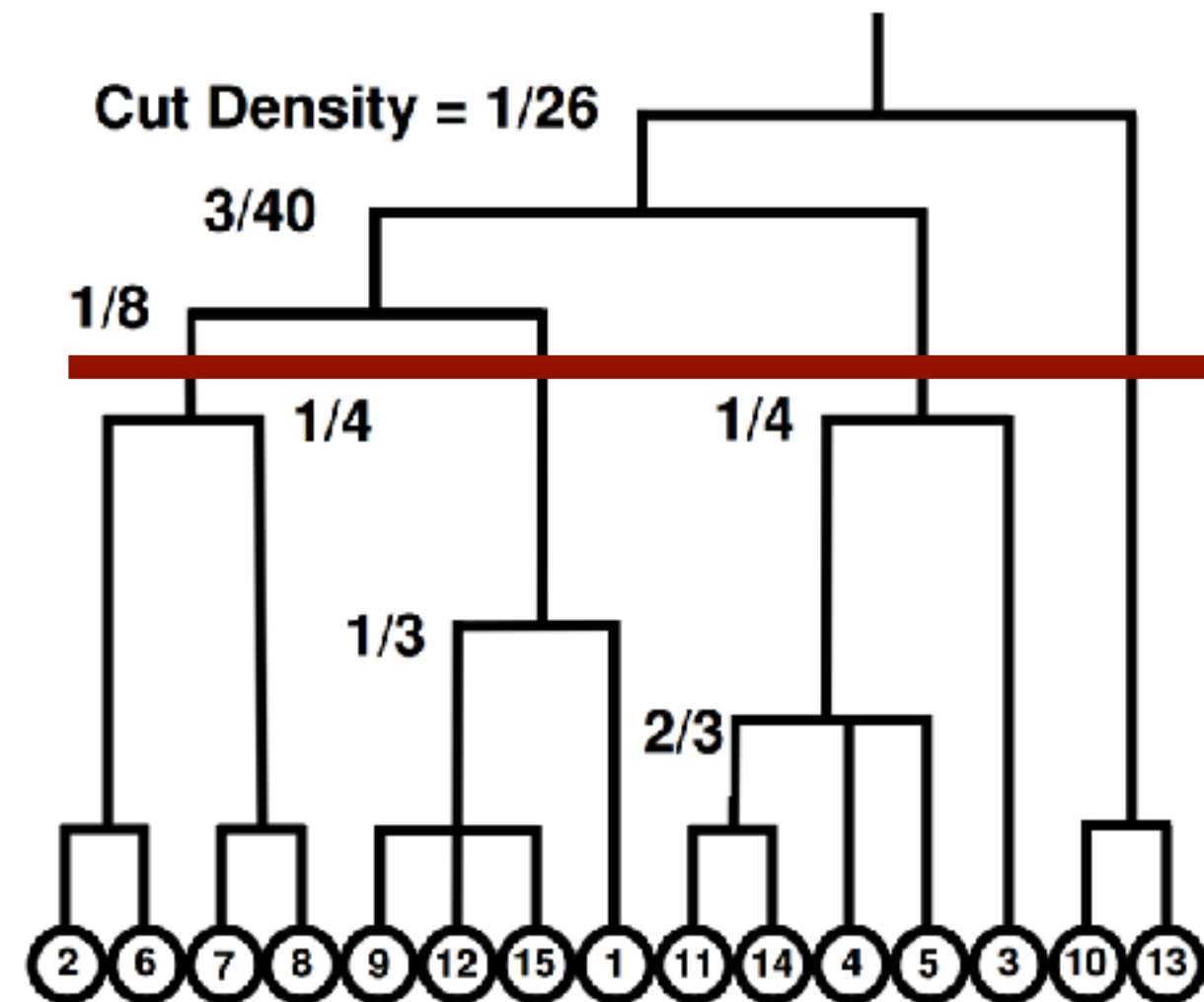
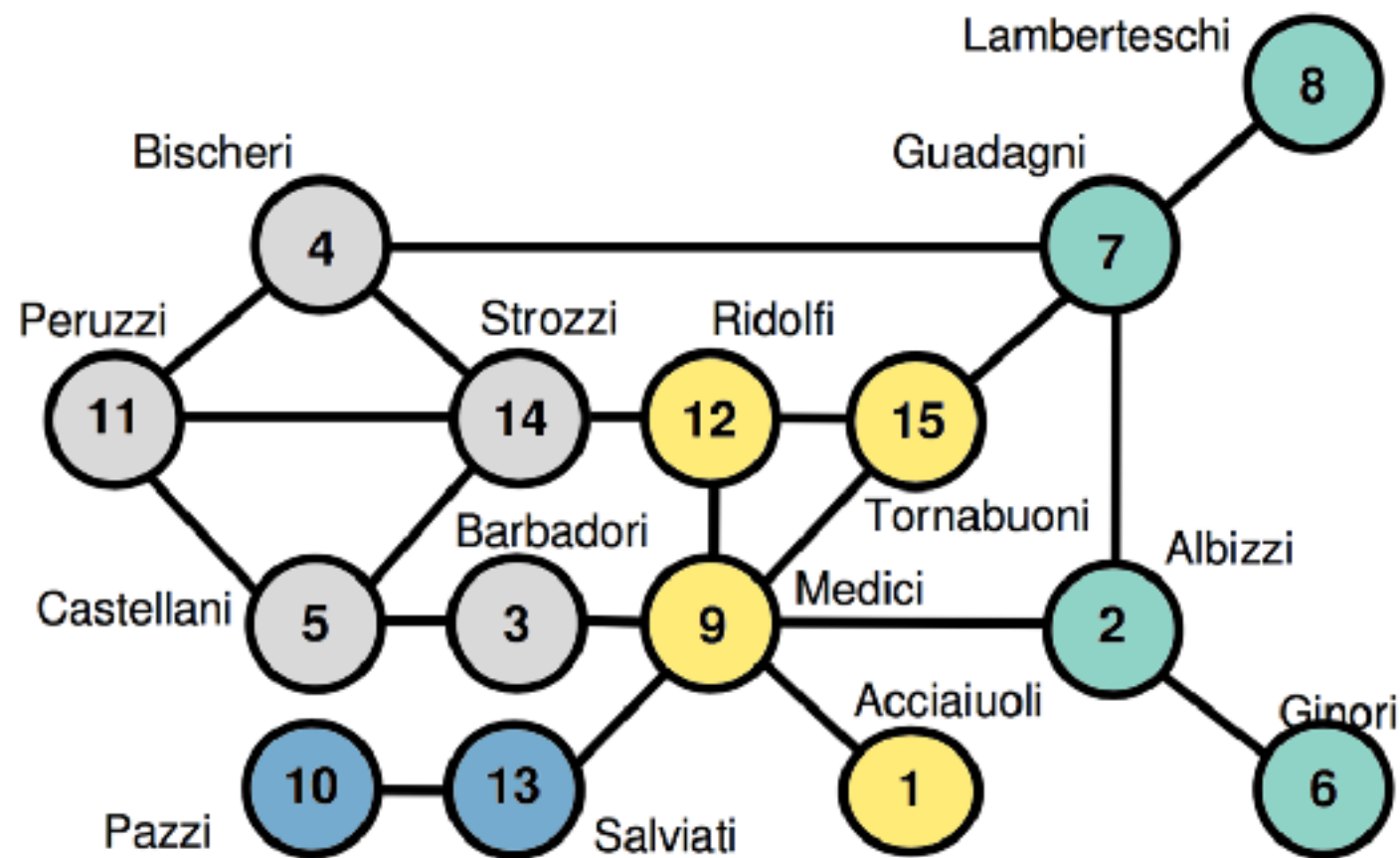
Arya D. McCarthy  
May 9, 2017

# Problem: Community Detection (CD)

- No consensus in field
- Methods are ad hoc



# What if we had groups *in* groups?



# Main Points

- Leximin method achieves competitive score when communities are clear
- Our method produces ties; popular CD methods don't
- Ties are computably a superposition of behaviors
- Complexity of method:  $O(N^{11}) / \log(N)$

# Overview

1. Motivation
2. Technique
3. Robustness and ties
4. Randomness, gridlock, and Menger's Theorem
5. Comparison to other methods

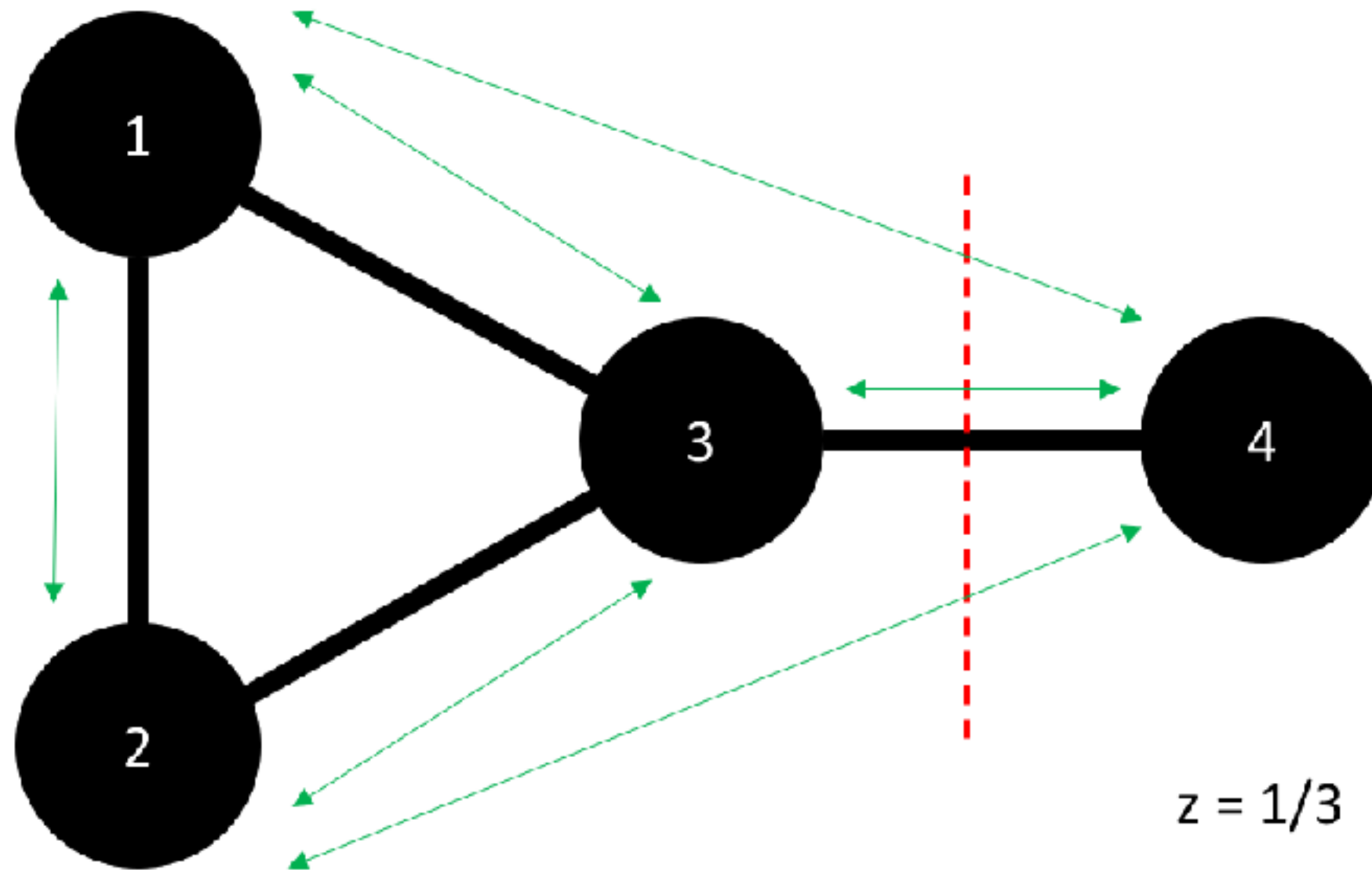
# Leximin for CD

Turtles all the way down

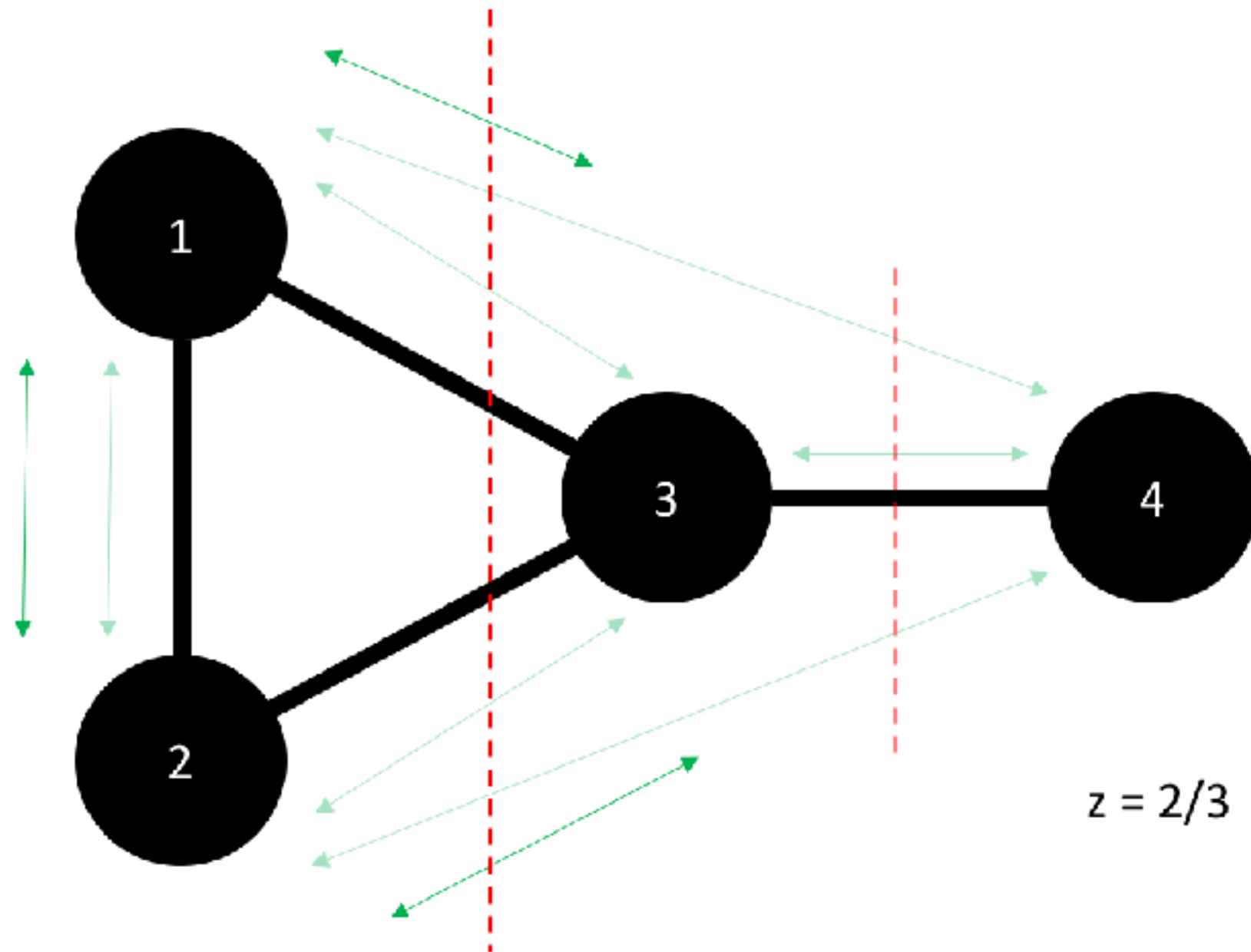
- Formulation as modularity maximization (NP-hard)
- Approximation as min sparsest cut (NP-hard)
- Heuristic through weak dual: lexicographic maximum concurrent flow (P)



# Leximin for CD

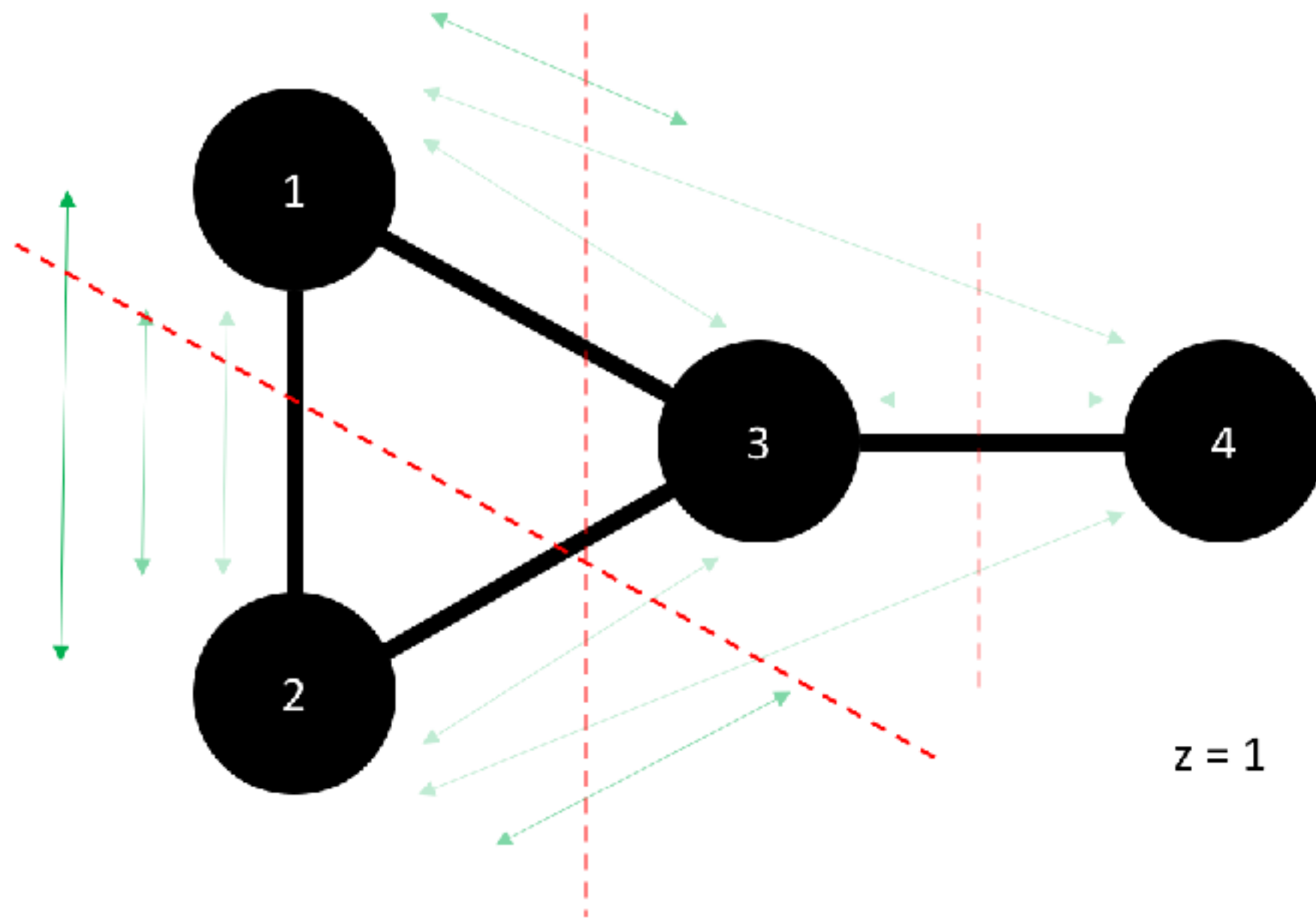


# Leximin for CD

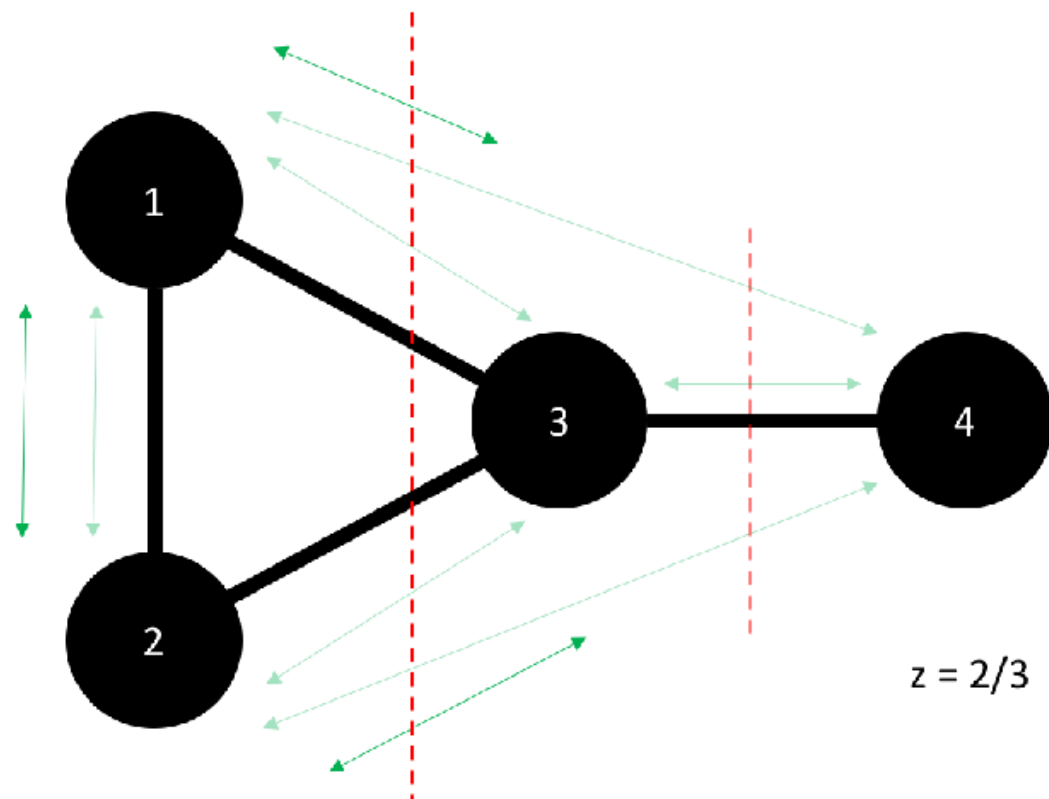




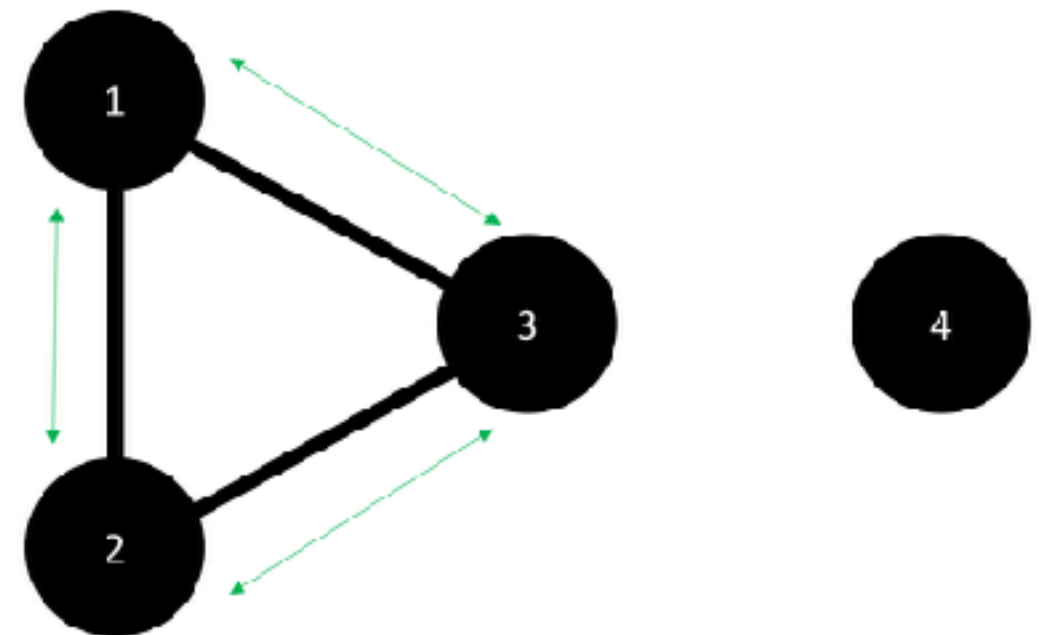
# Leximin for CD



# Aside: Leximin vs MCF Cut



VS



# Complexity

- Linear programming is doable in polynomial time
  - Fastest method:  
 $O(n^3)/\log(n) * L$
- MCFP "triples" formulation:  
 $O(MN)$  variables and  $O(N^2)$  constraints
- Up to  $N-1$  MCFPs are solved
- Result:  **$O(N^{11})$  for sparse graphs**

Length of an LP

$$L = 1 + \lceil \log_2(n) \rceil + \lceil \log_2(m) \rceil + \sum_{j=1}^n \lceil \log_2(|c_j| + 1) \rceil + n + \sum_{i=1}^m \sum_{j=1}^n \lceil \log_2(|a_{ij}| + 1) \rceil + mn + \sum_{i=1}^m \lceil \log_2(|b_i| + 1) \rceil + n.$$

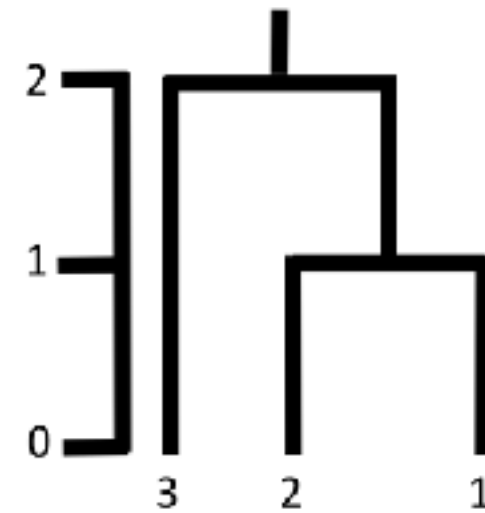
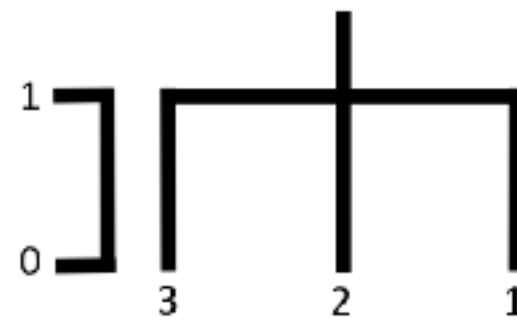
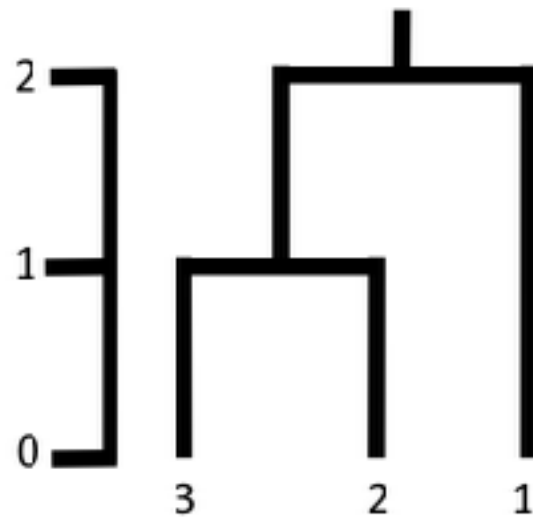
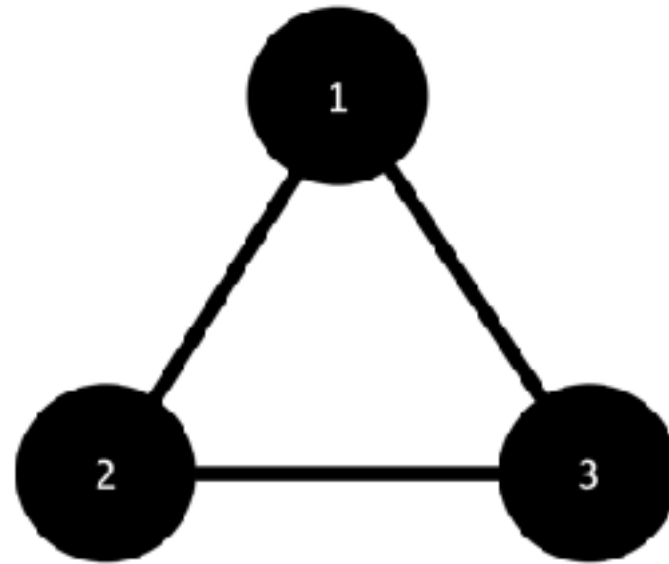


Courtesy of  
Dr. Olinick's course notes

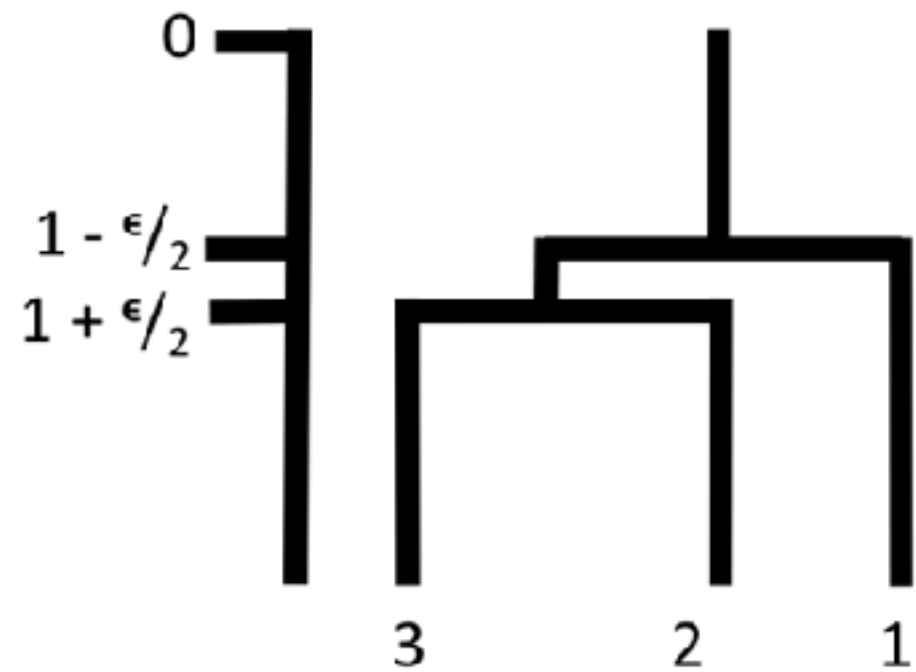
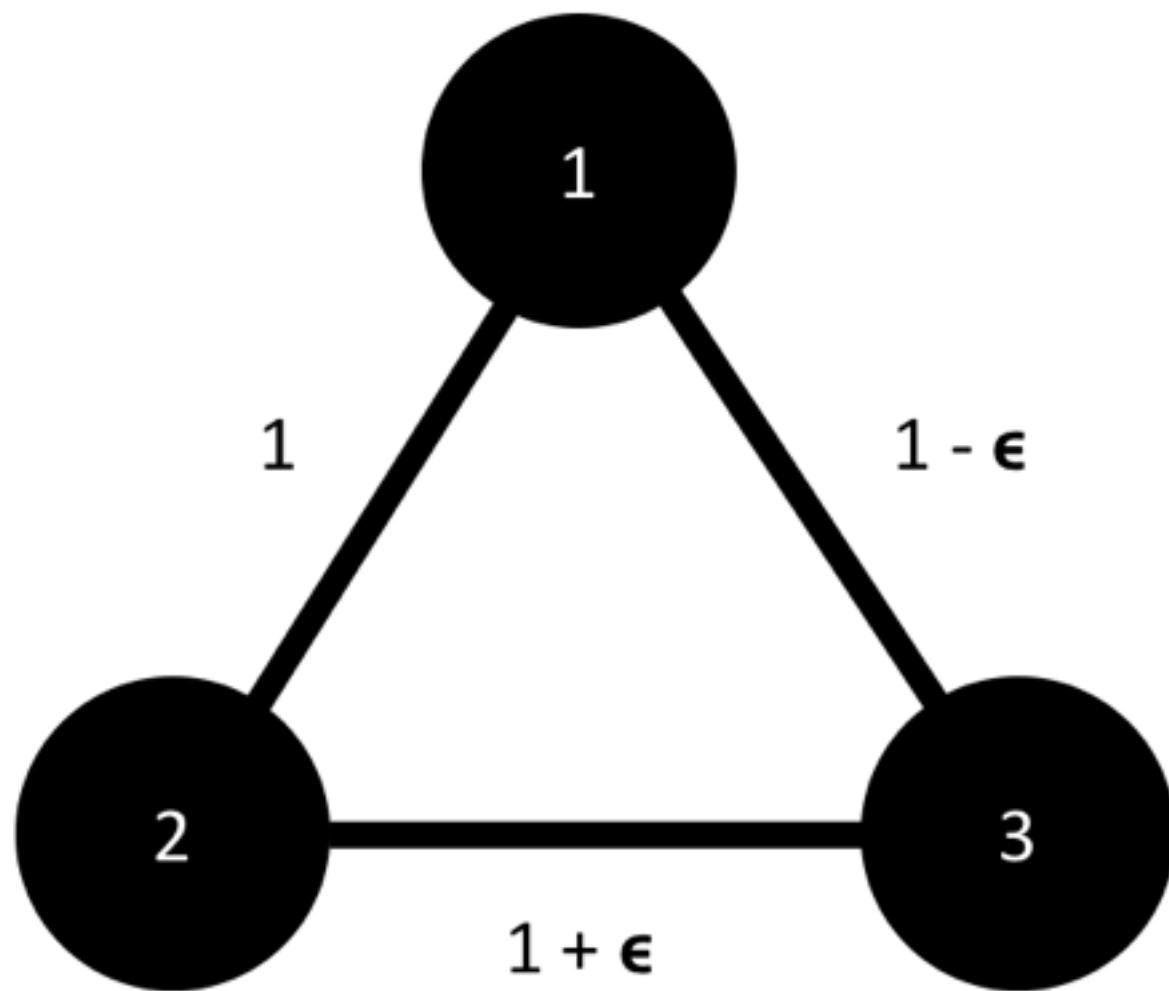
# Ties and Sensitivity

"Please don't make me choose"

# Who should be cut first?



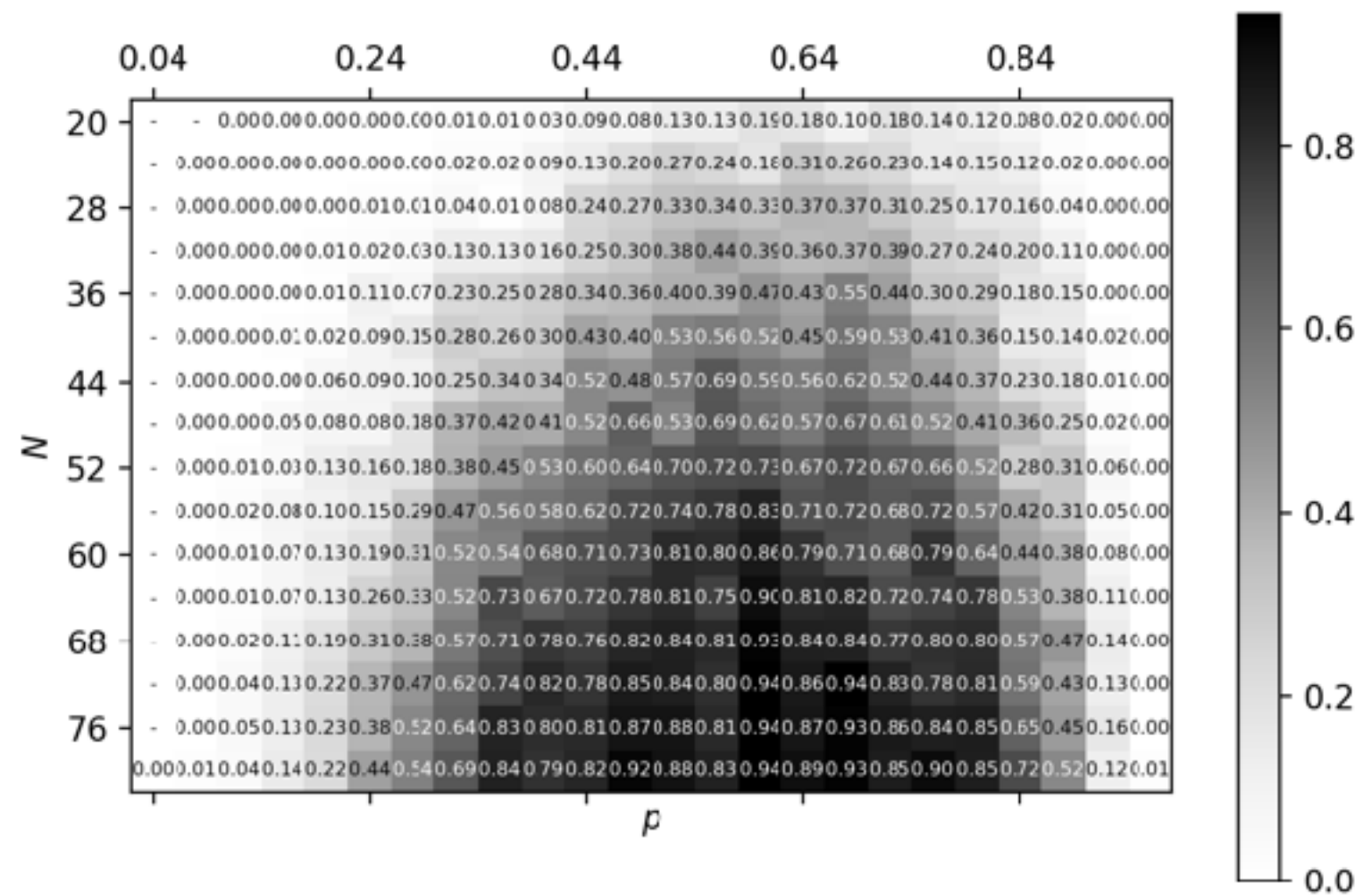
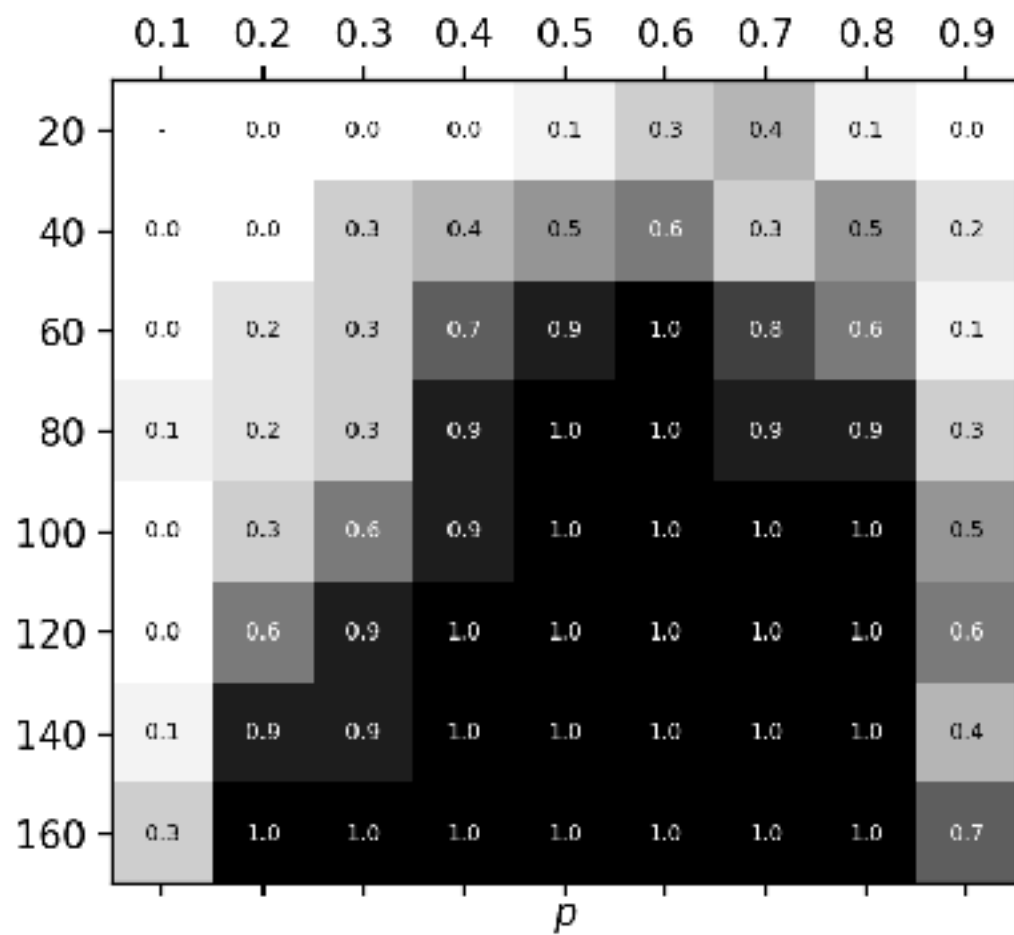
# Sensitivity



# Gridlock

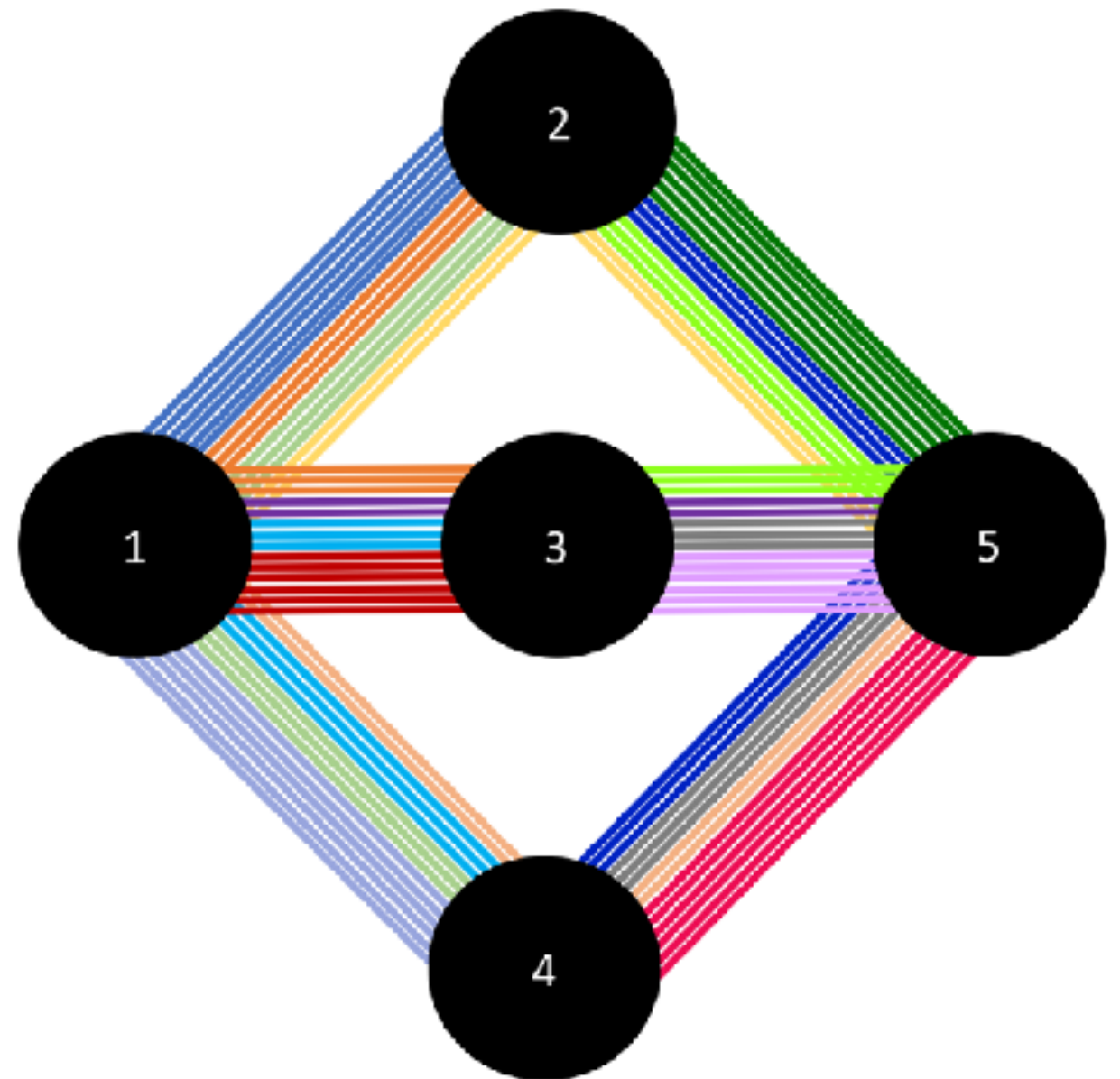
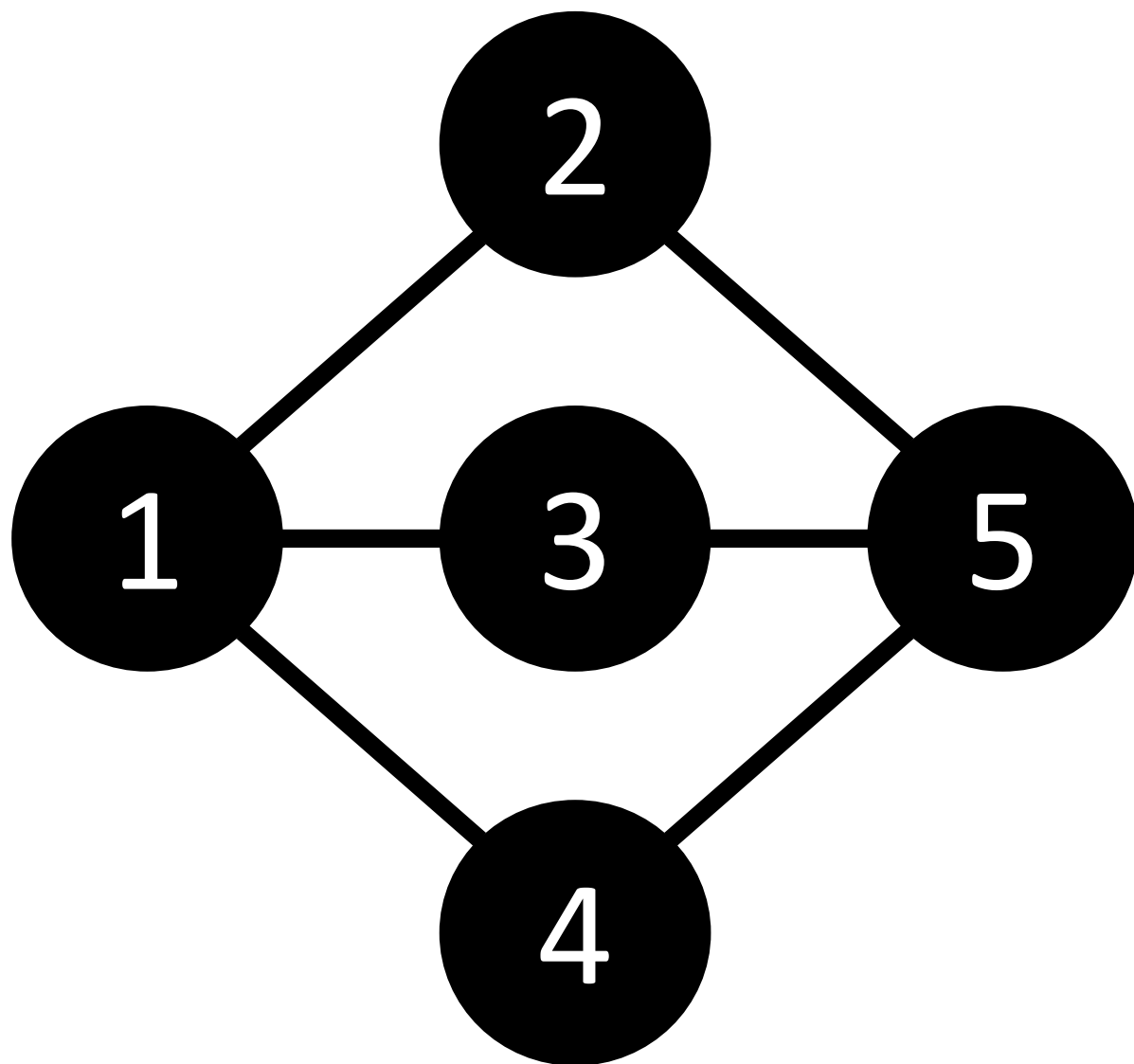
"There's nothing really here, is there?"

# Frequency of gridlock





# $K_{3,2}$ as a multigraph

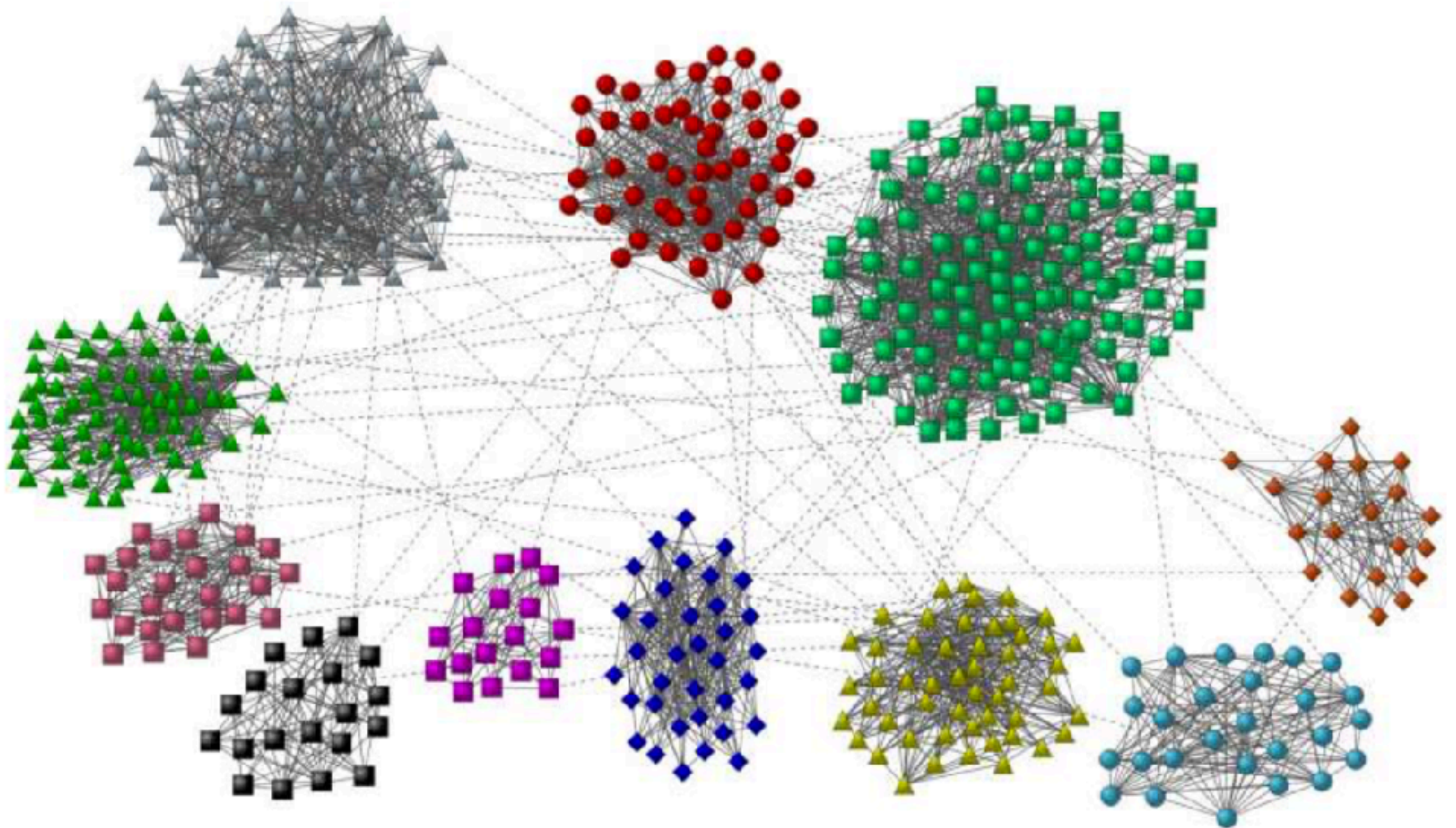


$[N(N-1) - M]$  "Edge-disjoint paths"

# Comparison

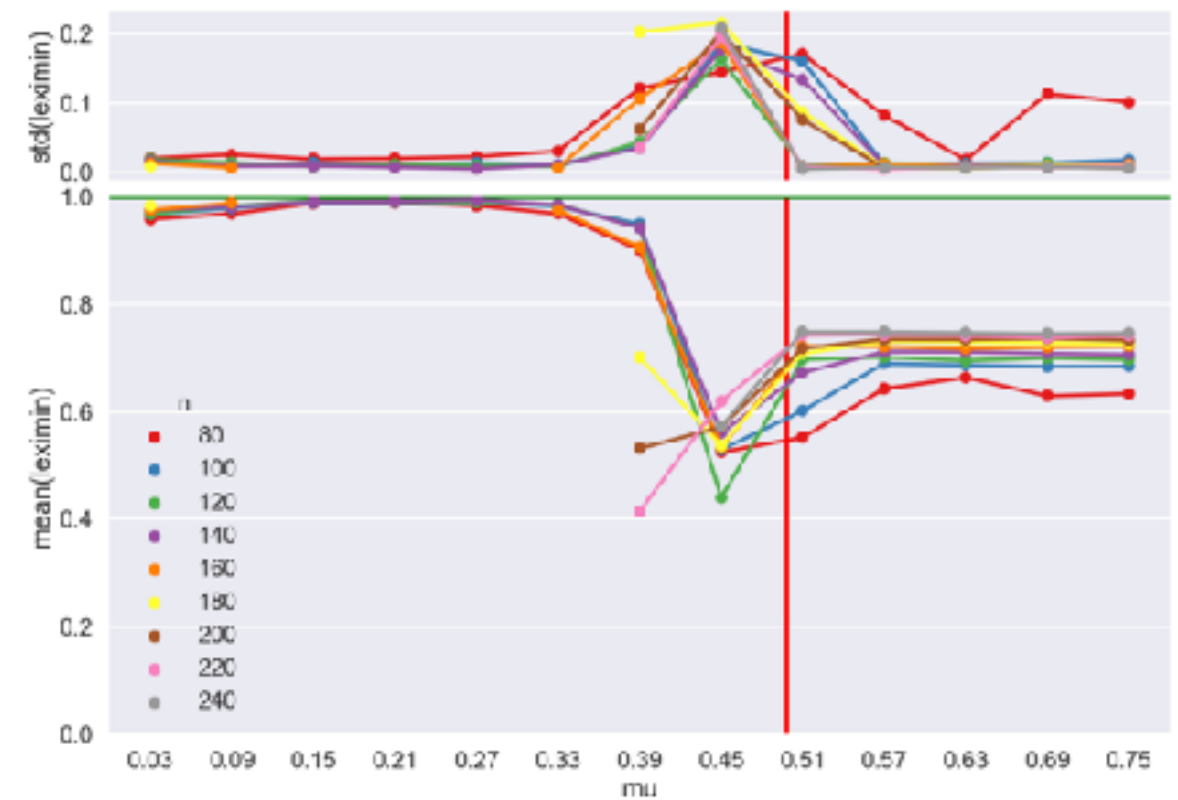
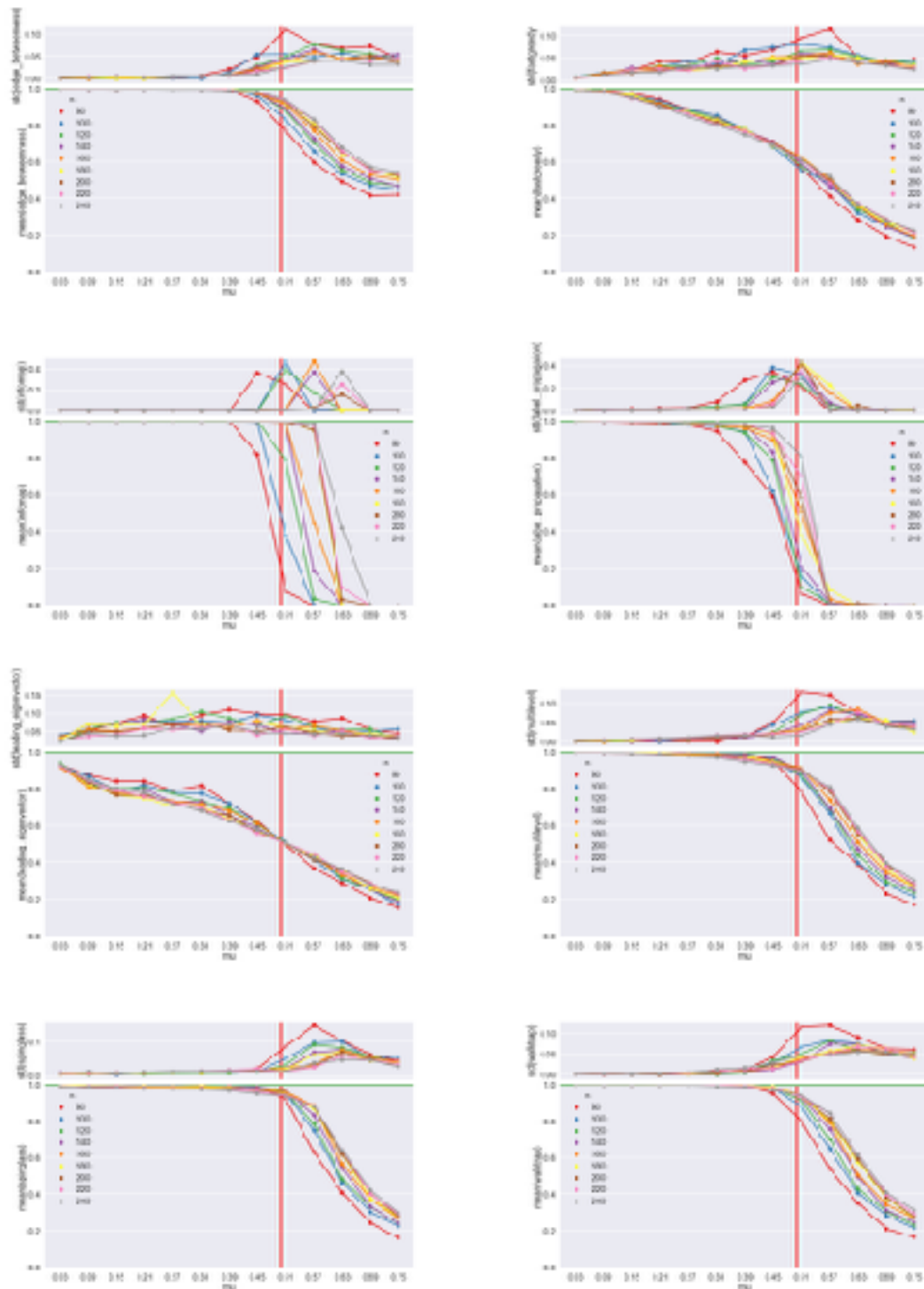
Keeping up with the Joneses  
and Newman  
and Clauset  
and Rosvall  
and Latapy

# LFR Benchmarks

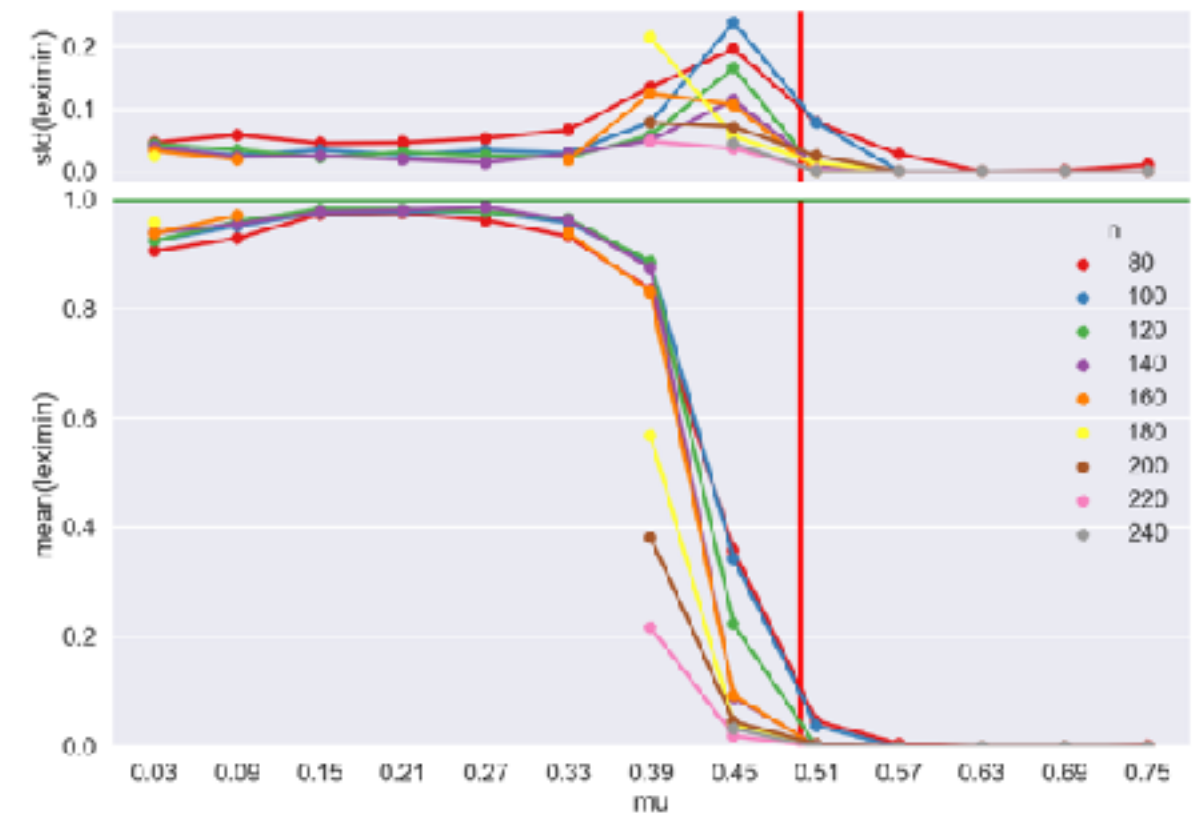
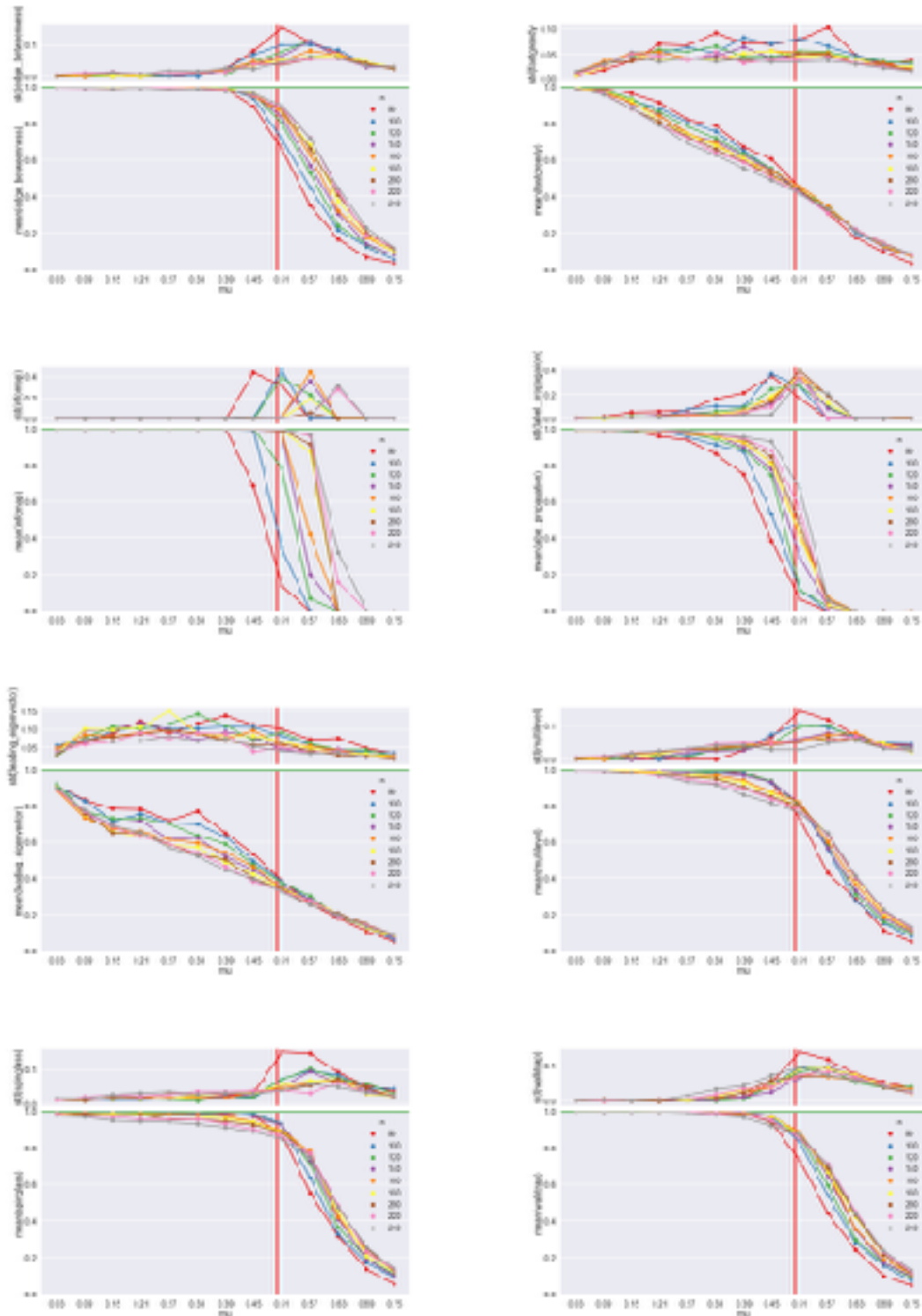




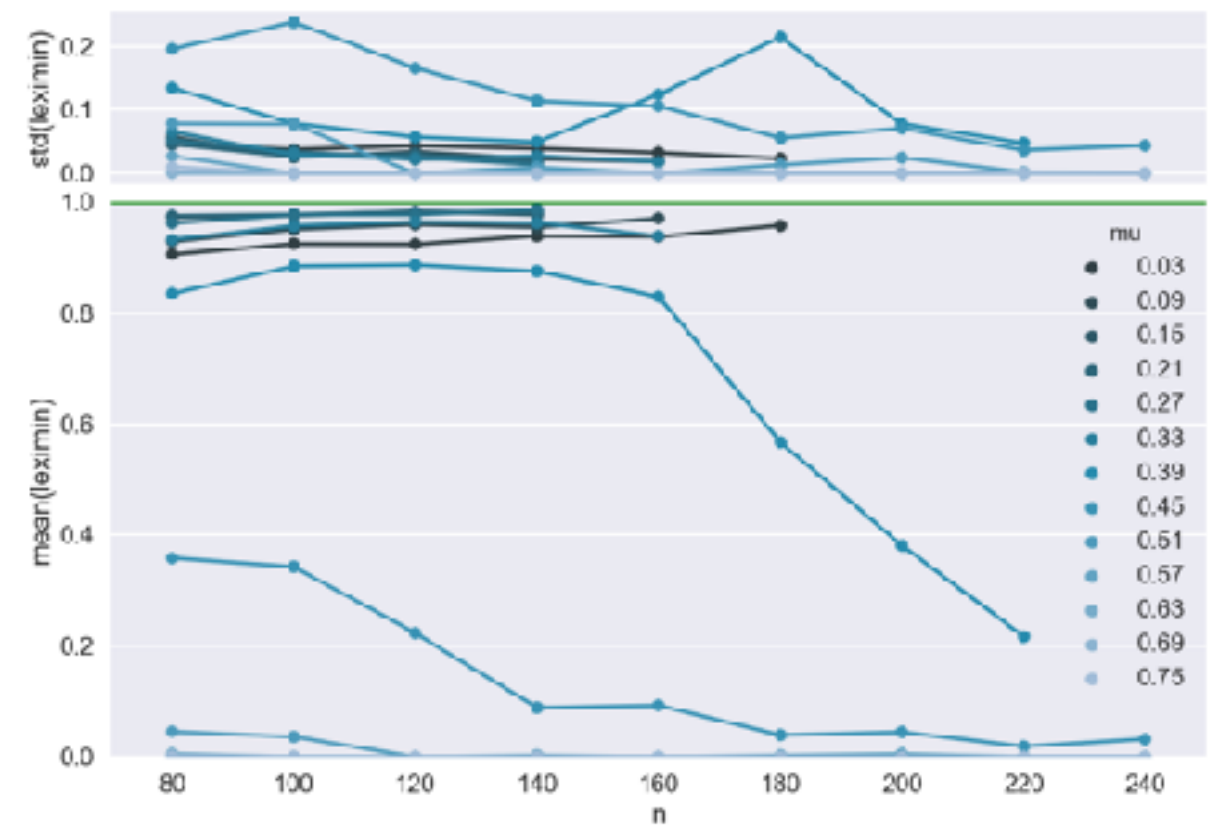
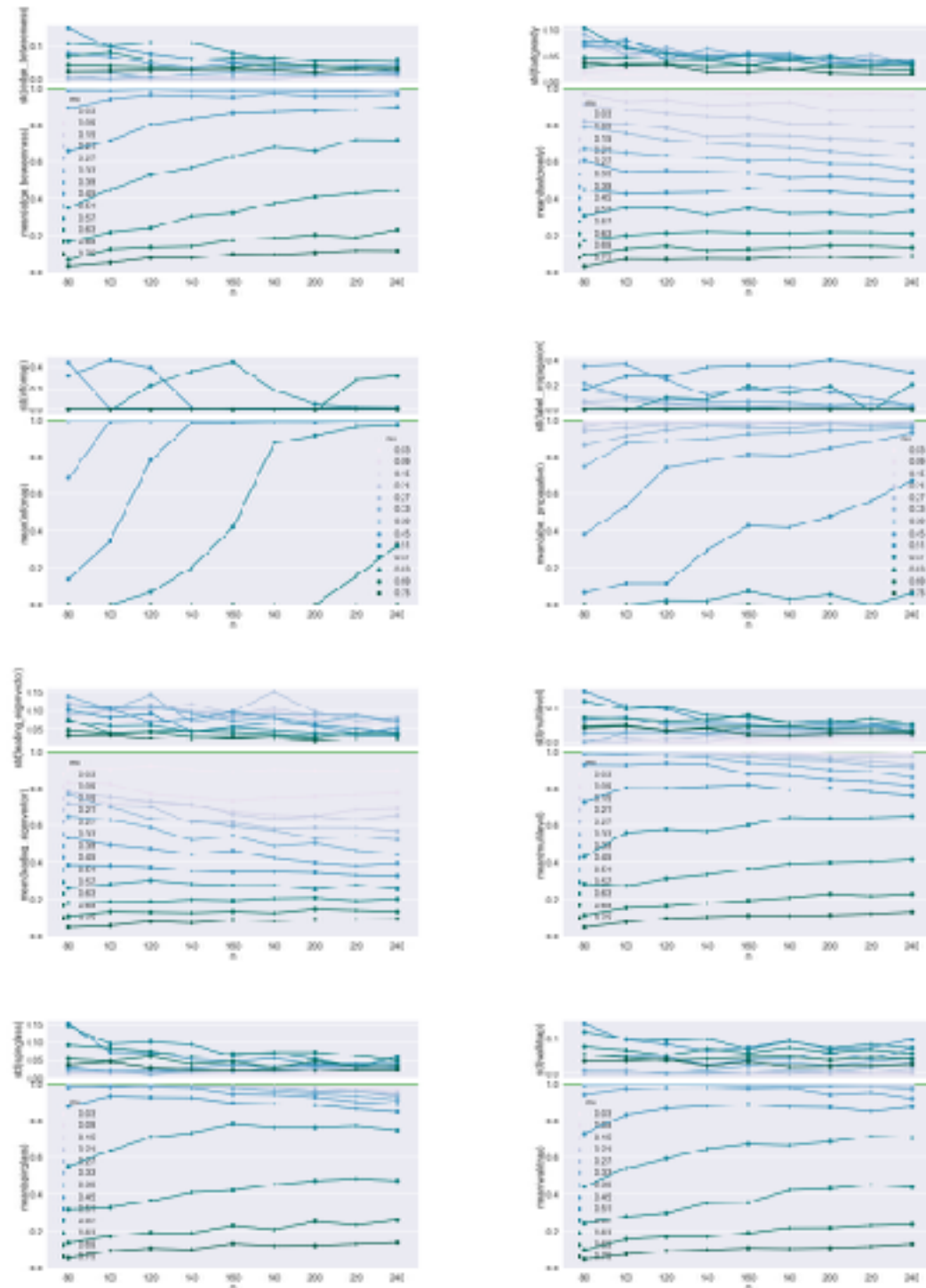
# NMI vs community strength



# AMI vs community strength



# AMI vs network size

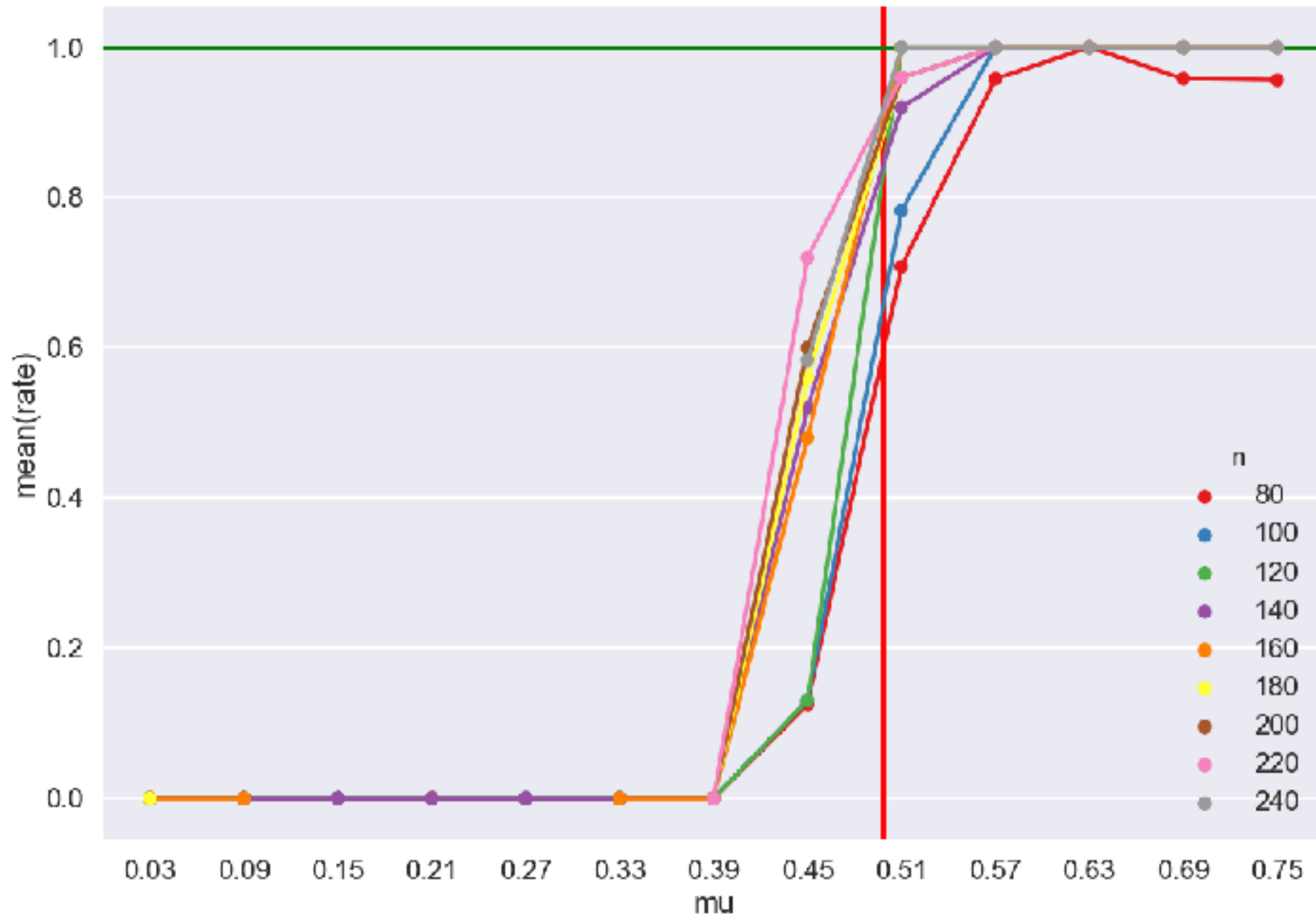


# Lemma 5.2

*If the modularity-maximizing flat clustering produced by the leximin method is into singletons, then gridlock has occurred.*

Leads to "failing fast".

# Gridlock





# Conclusions

- Leximin method achieves competitive score when communities are clear
- AMI is a fairer test than NMI
- Our method produces ties; popular CD methods don't
- Large random graphs lack communities
- Ties are computably a superposition of behaviors
- Complexity of method:  $O(N^{1.1}) / \log(N)$

# Future Directions

- Assessment for overlapping community detection
- AMI of MCFP approximations on LFR
- Prove conjecture about diameter-2 graphs

# Acknowledgments

- Committee:
  - David W. Matula
  - Daniel W. Engels
  - Eric C. Larson
  - Eli V. Olinick
- Creating figures:
  - Stejara Dinulescu
- Discussions:
  - Zizhen Chen
  - Monnie McGee
  - AJ Rao
  - Fernando Vilas
- Implementing the LP
  - Eli V. Olinick, pt. 2

# Conclusions

- Leximin method achieves competitive score when communities are clear
- AMI is a fairer test than NMI
- Our method produces ties; popular CD methods don't
- Large random graphs lack communities
- Ties are computably a superposition of behaviors
- Complexity of method:  $O(N^{11}) / \log(N)$