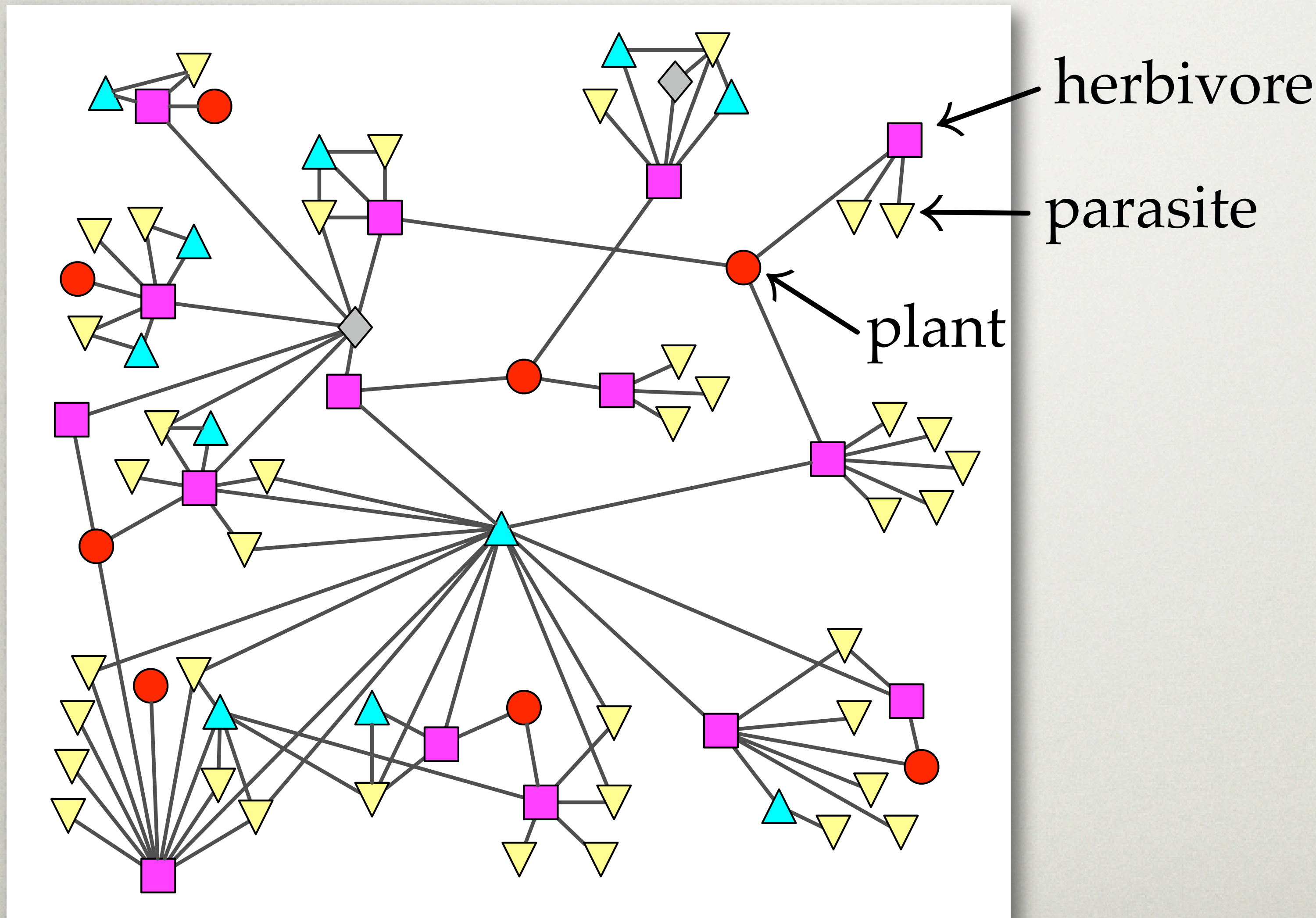
Hierarchically Modular Structure in Complex Networks

Aaron Clauset
Santa Fe Institute

3 November 2008
DIMACS / DyDAn
"Network Models of Biological and Social Contagion"

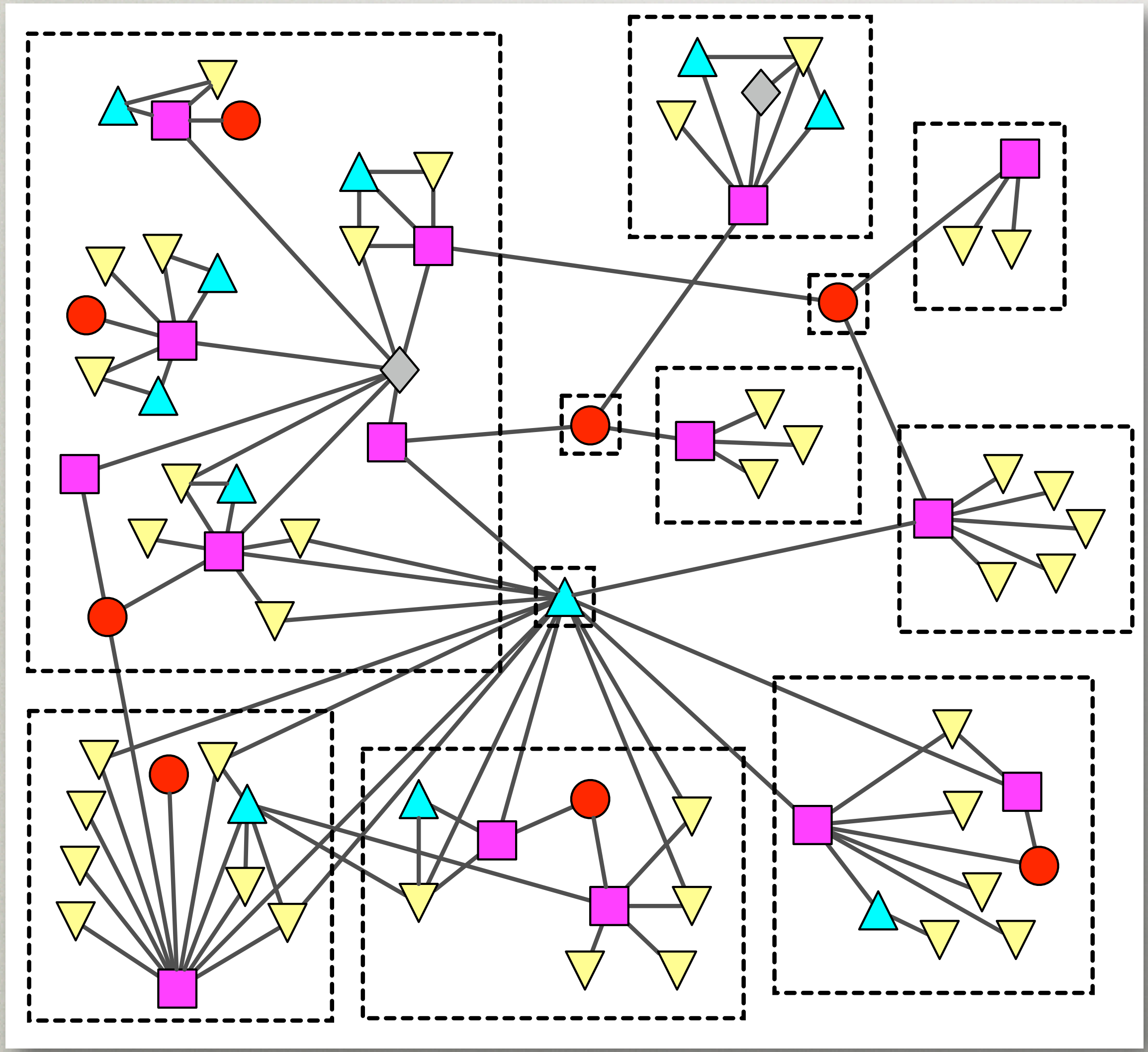
Modular Hierarchies

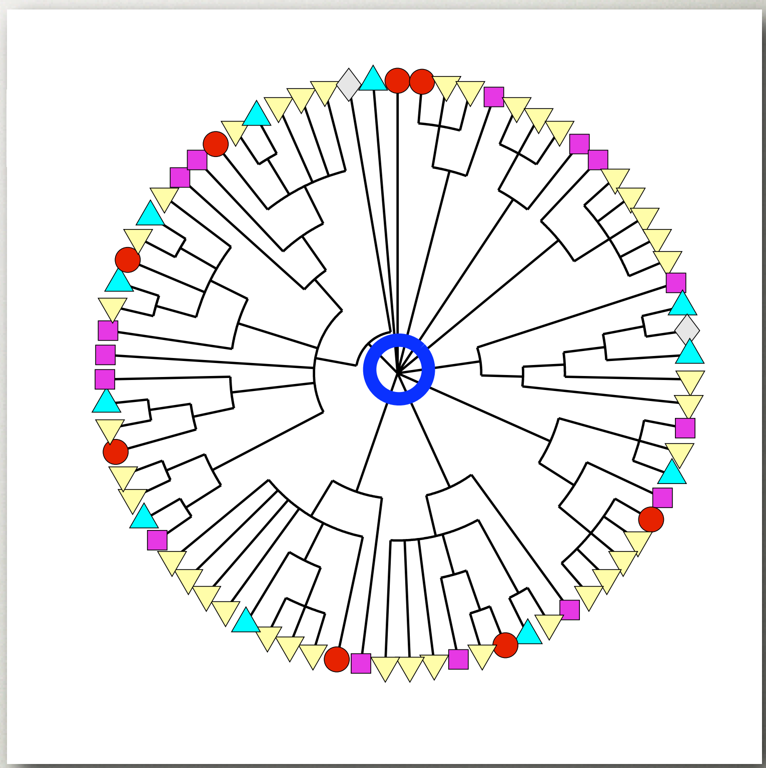


Grassland species*

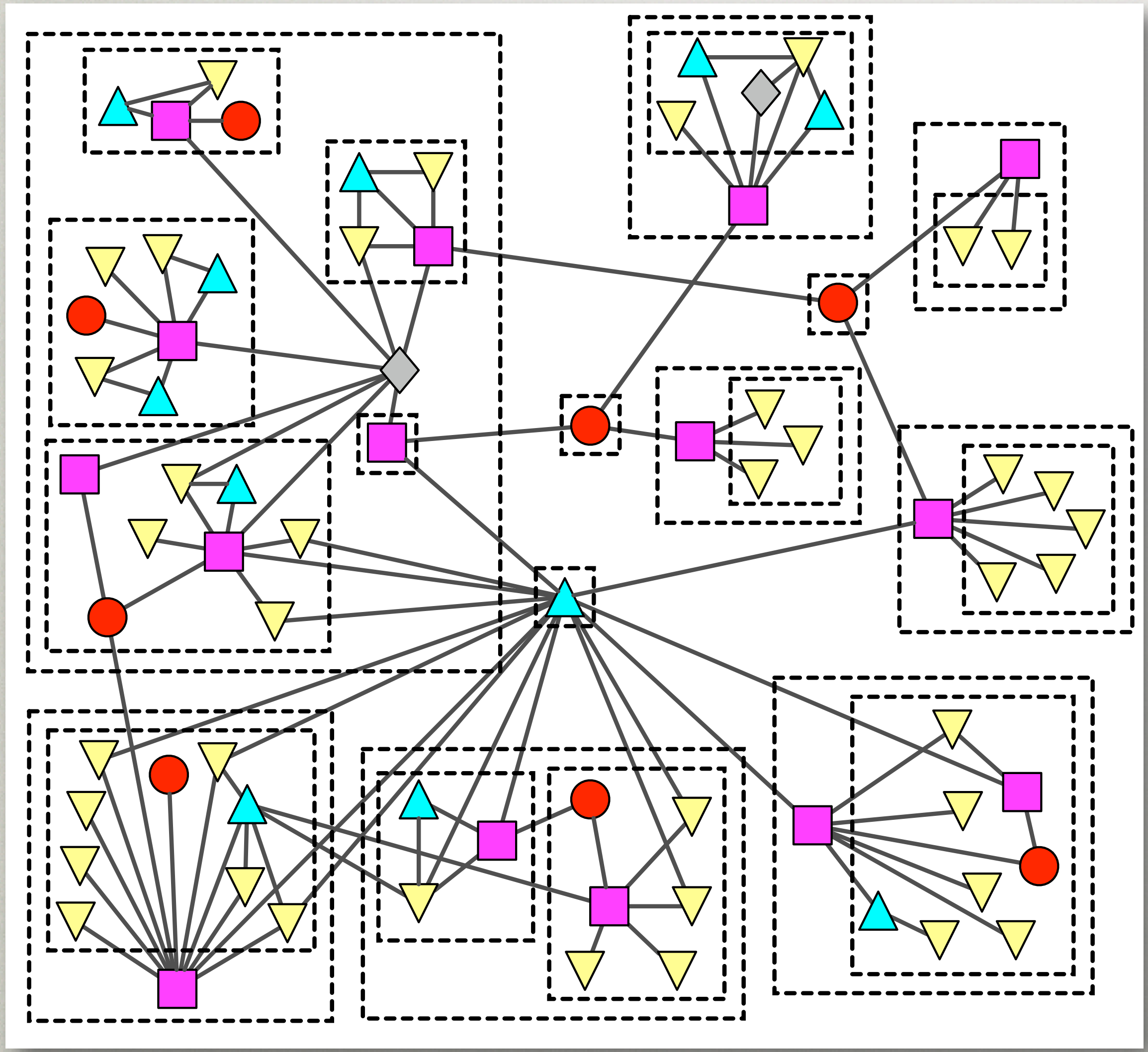
*thank you: Jennifer Dunne

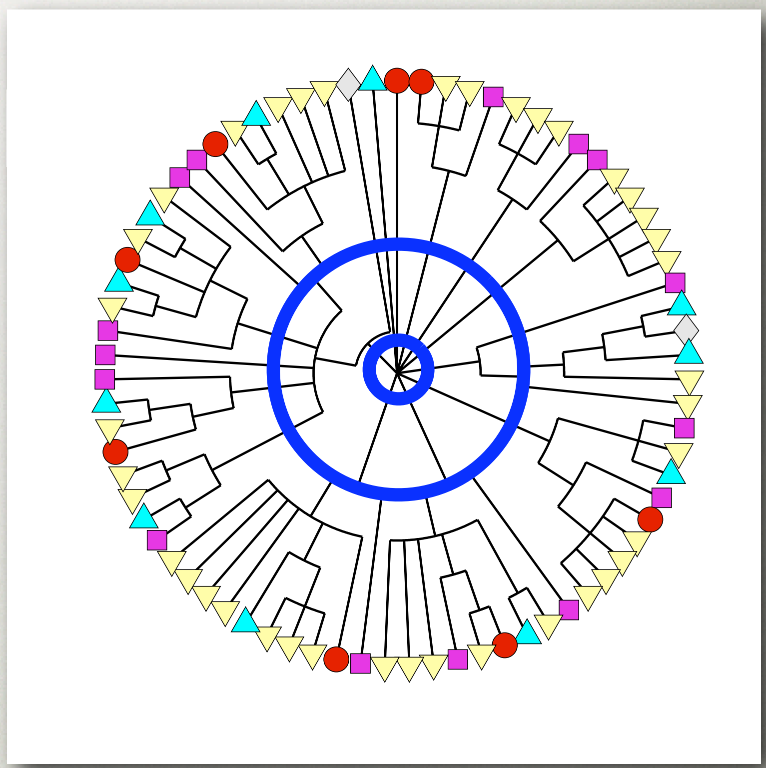
Modular Hierarchies





Modular Hierarchies



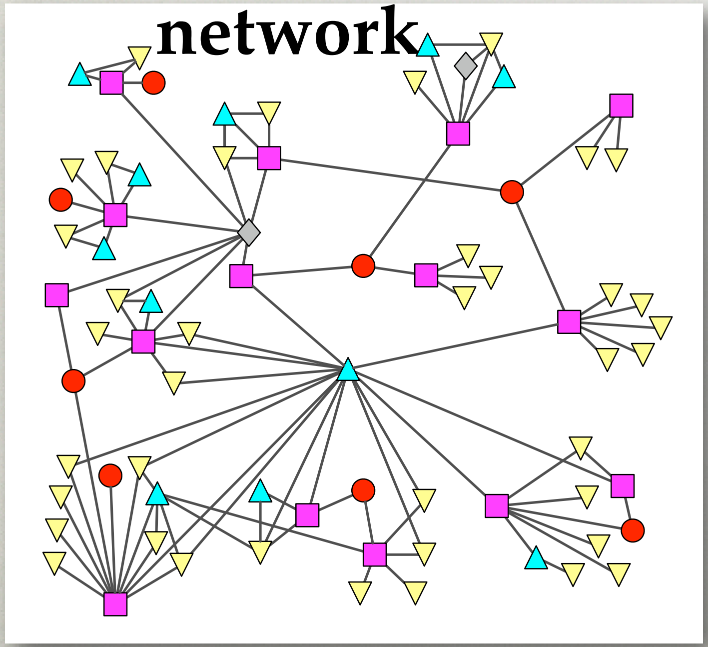


The Task

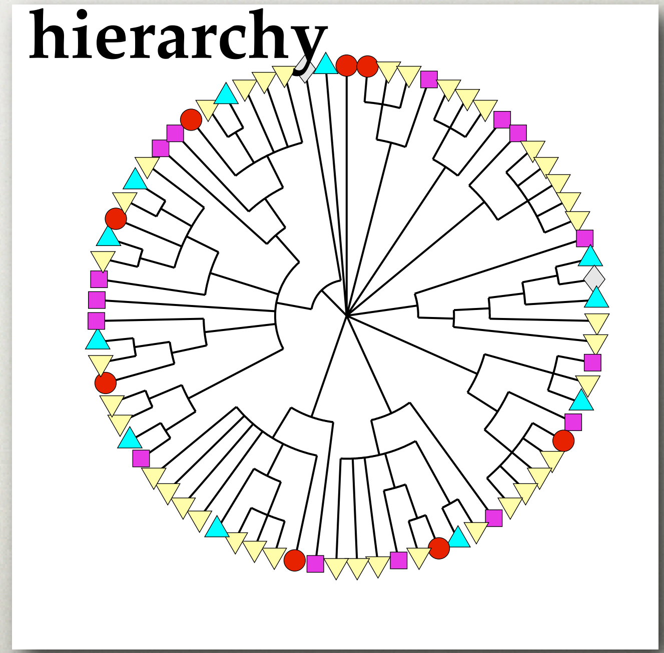
How can we extract

- **this hierarchical (multi-scale) structure**

from complex networks?



? ⟶

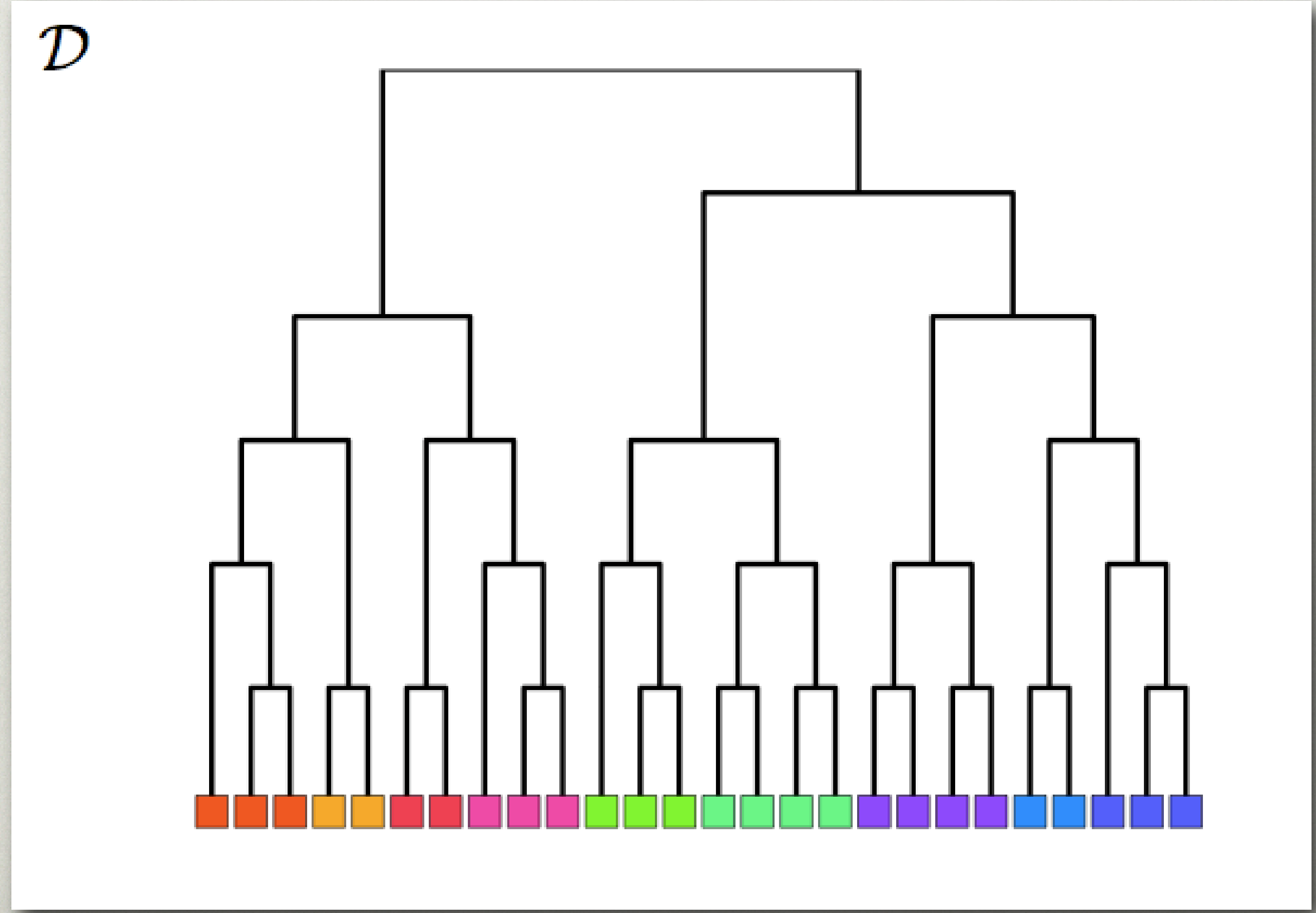


One Approach

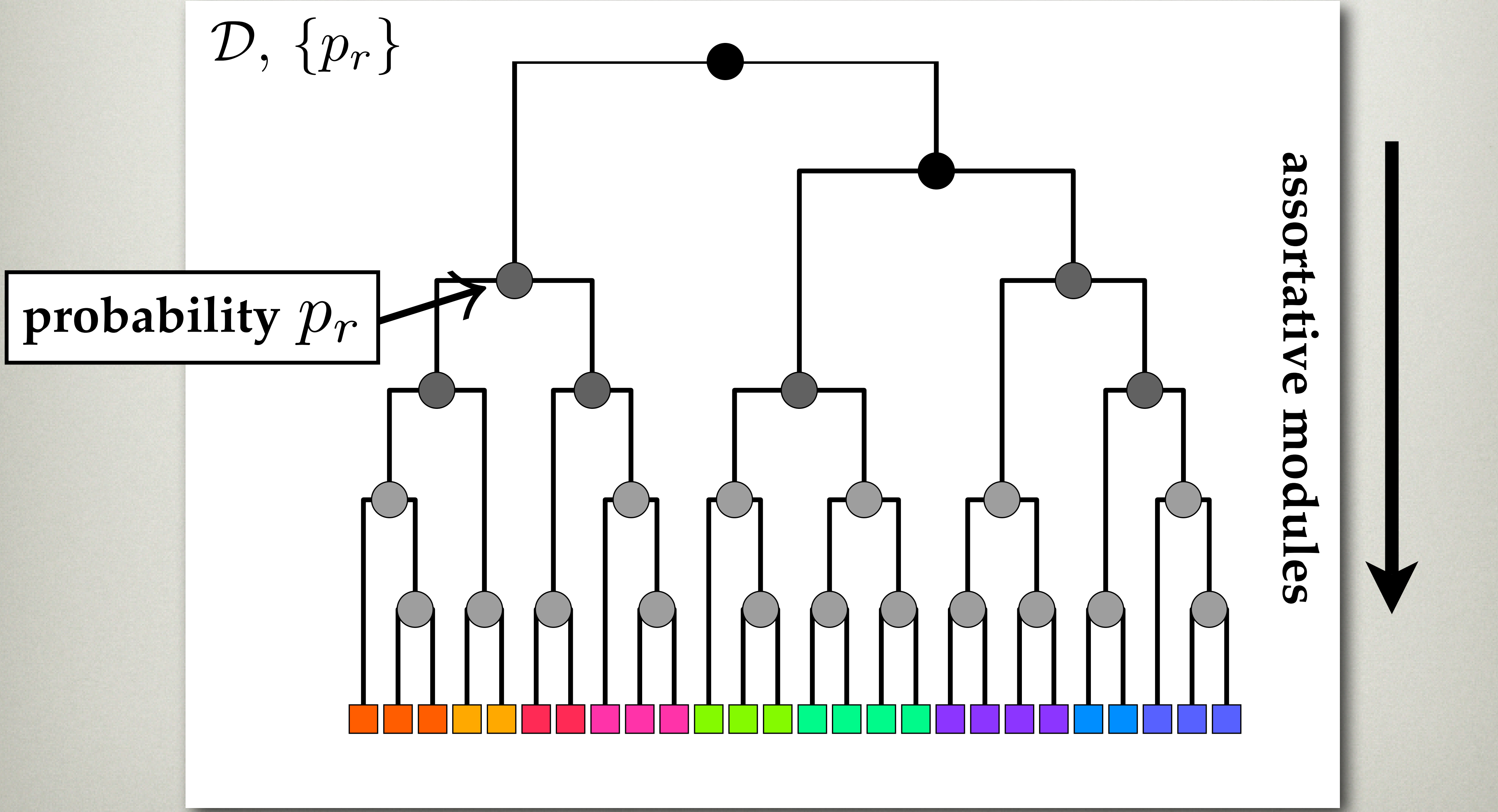
Model-based inference

1. describe how to generate hierarchies (a model)
2. “fit” model to empirical data
3. test “fitted” model
4. extract predictions + insight
5. profit!

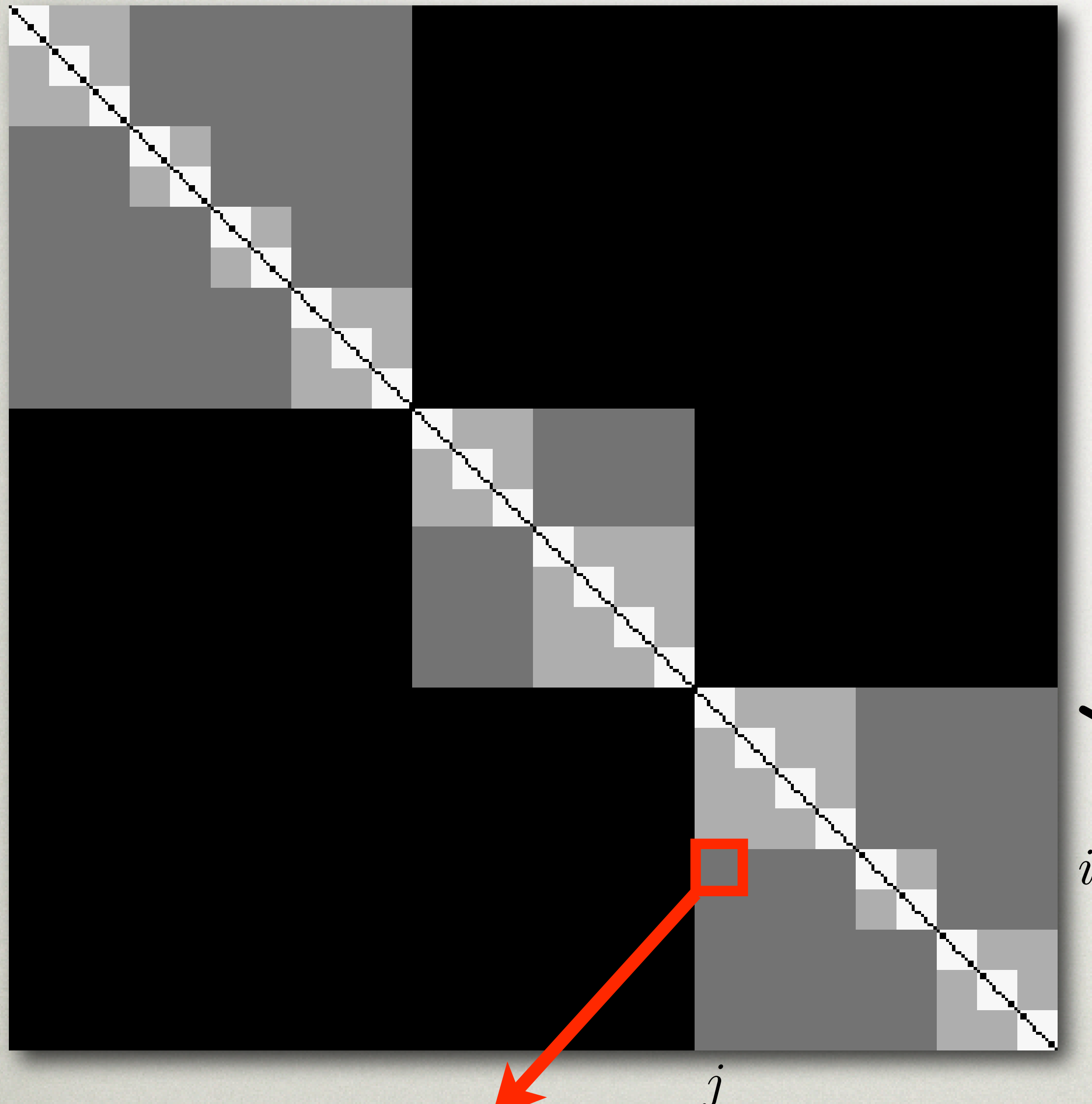
A Model of Hierarchy



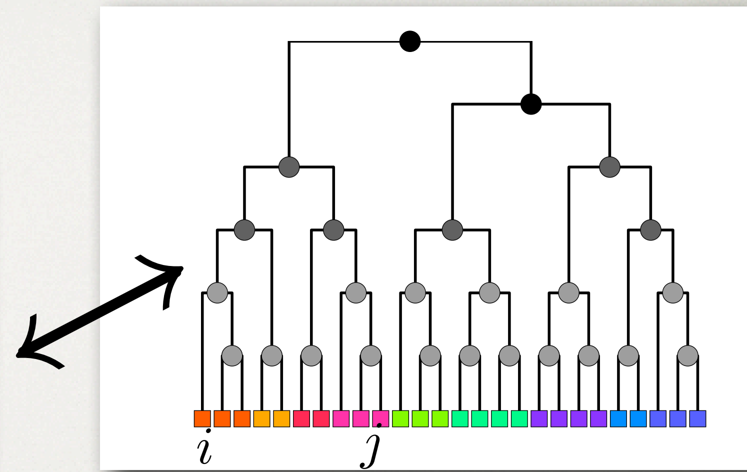
A Model of Hierarchy



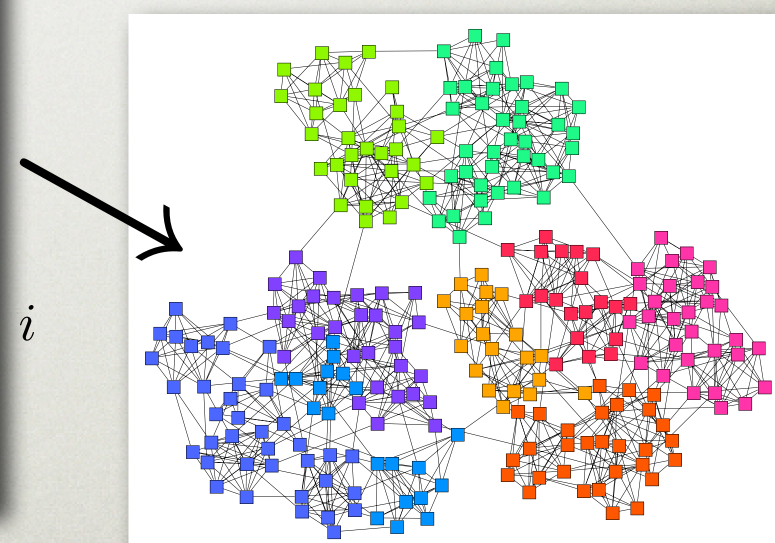
“inhomogeneous” random graph



model



instance



$$\Pr(i, j \text{ connected}) = p_r = p_{(\text{lowest common ancestor of } i,j)}$$

Model Features

- explicit model = explicit assumptions
- very flexible (many parameters)
- captures structure at all scales
- arbitrary mixtures of assortativity, disassortativity
- learnable directly from data

Learning From Data

a direct approach

- **likelihood function** $\mathcal{L} = \Pr(\text{ data } | \text{ model })$

  ($\mathcal{L}$ scores **quality** of model)

- **sample** the **good** models

  via Markov chain Monte Carlo

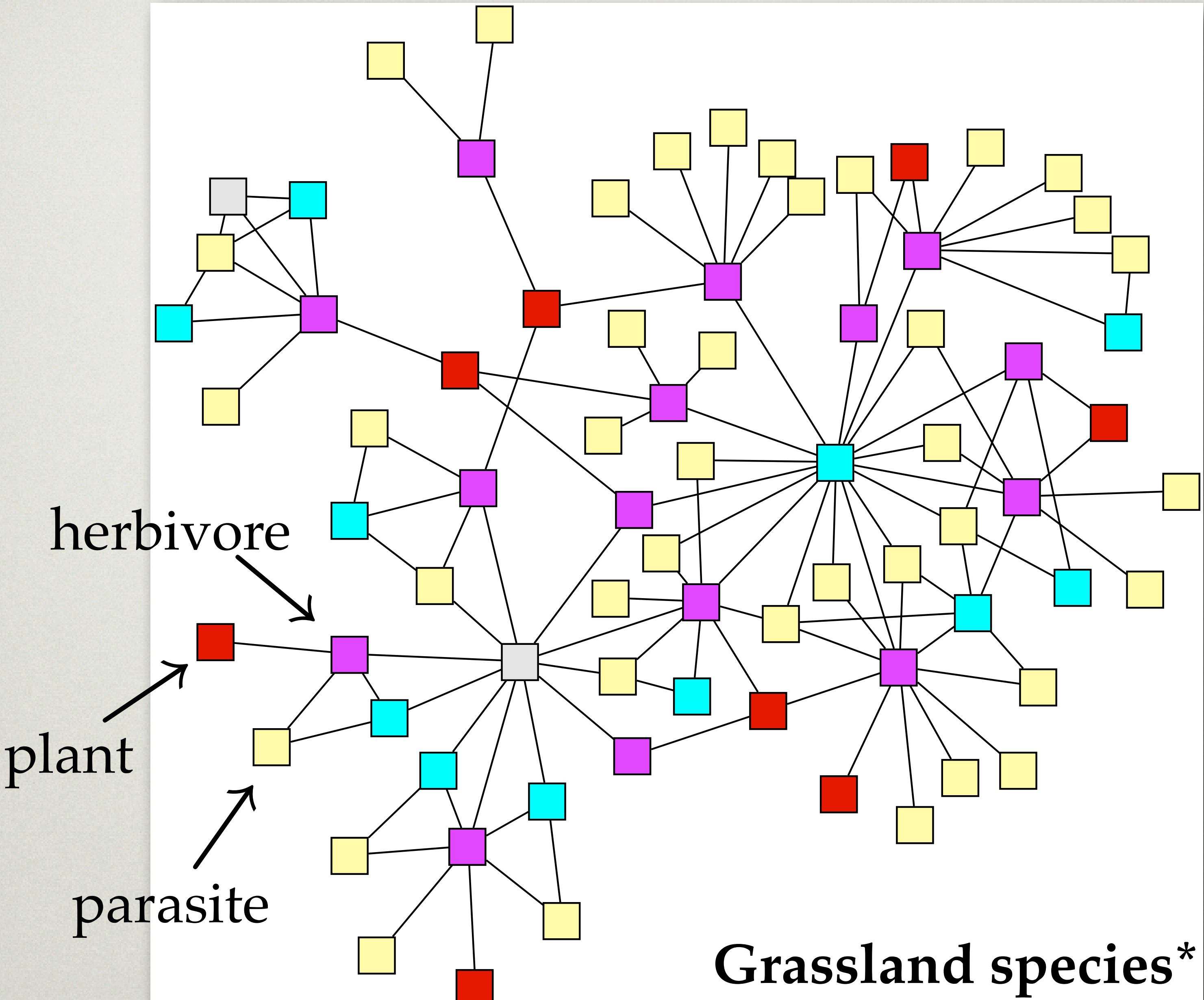
- technical details in arXiv : *physics/0610051*

From Graph to Ensemble

From Graph to Ensemble

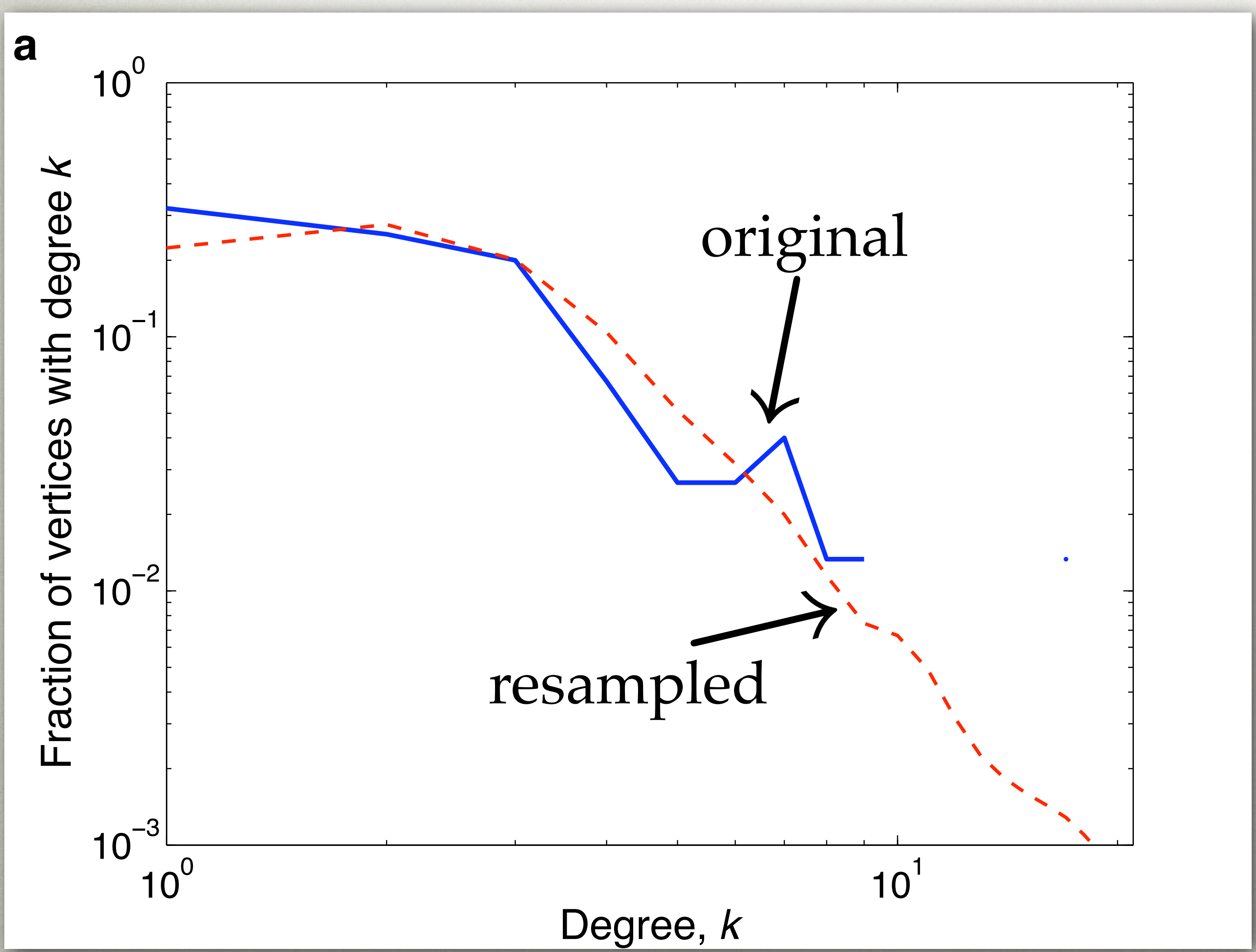
- Given graph G
- run MCMC to equilibrium
- then, for each sampled $\mathcal{D}$, draw a **resampled** graph G' from ensemble

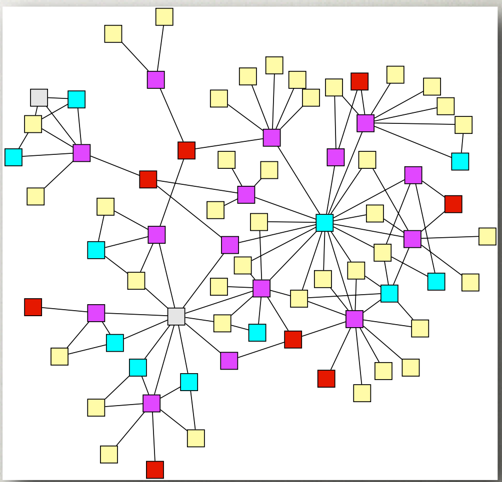
A test: do resampled graphs look like original?



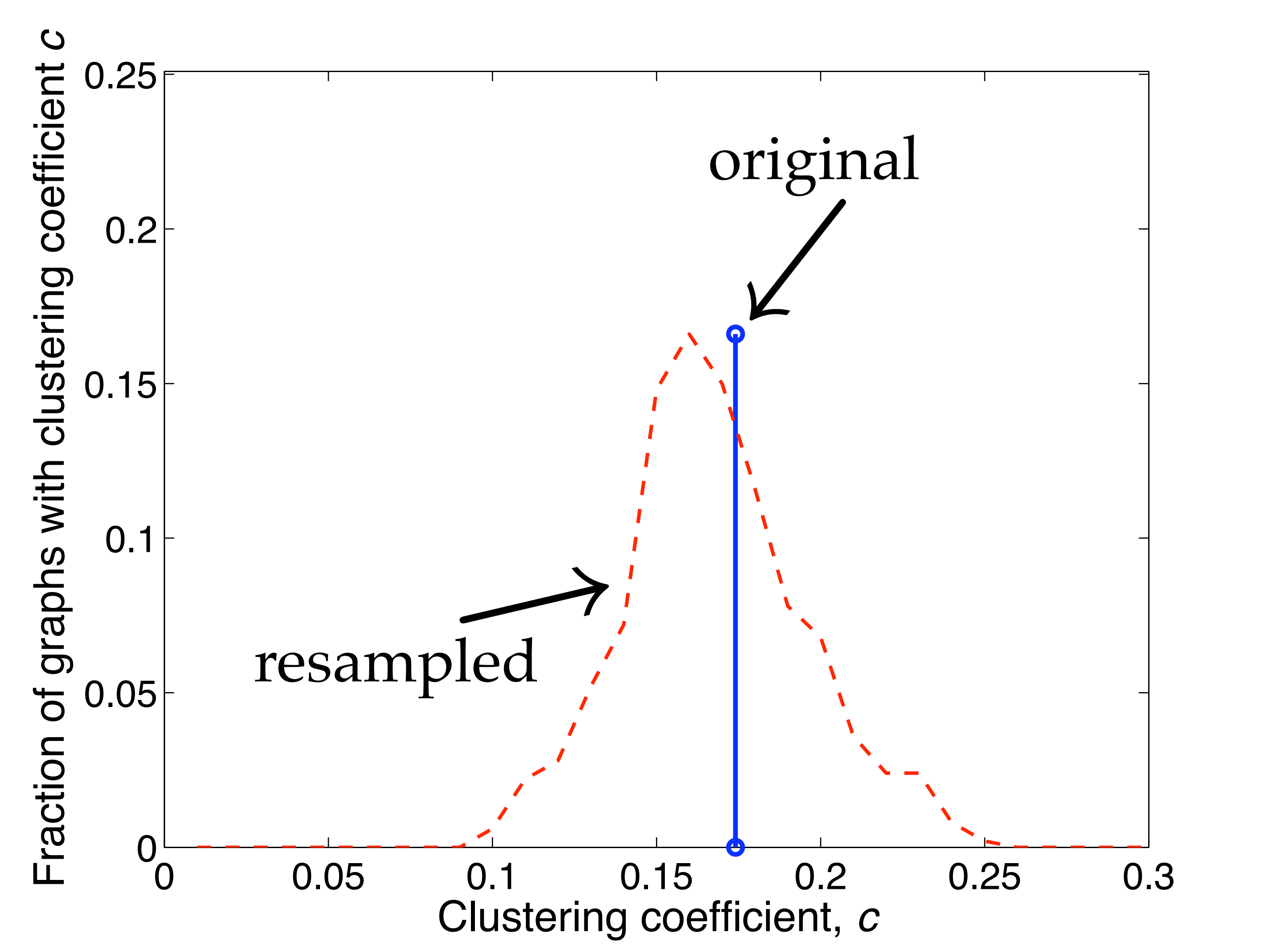
*thank you: Jennifer Dunne

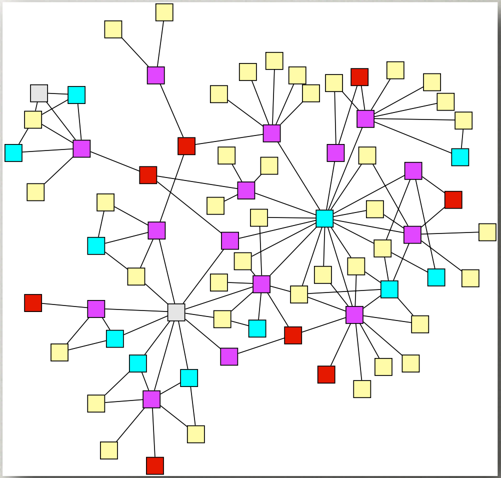
Degree Distribution



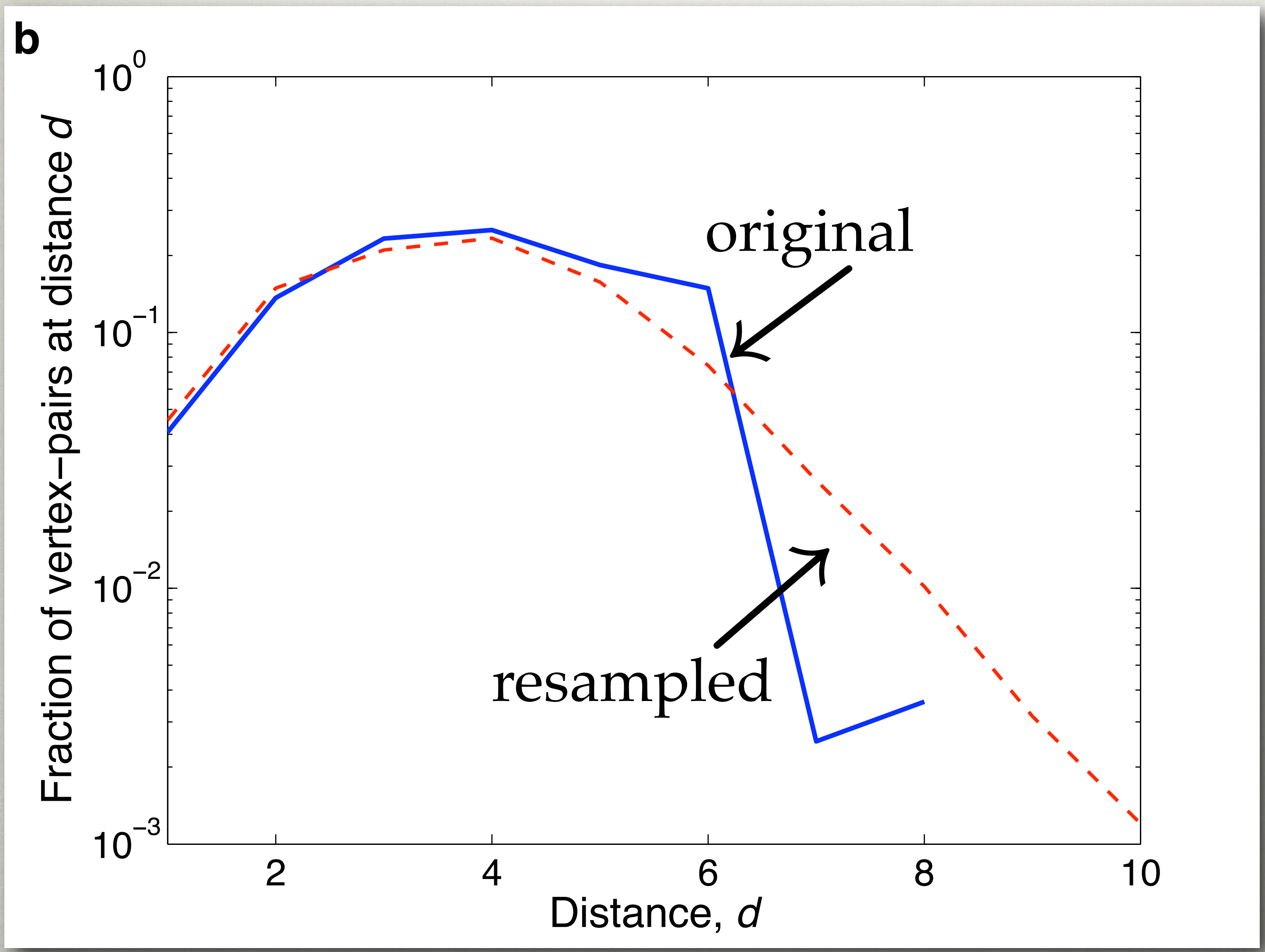


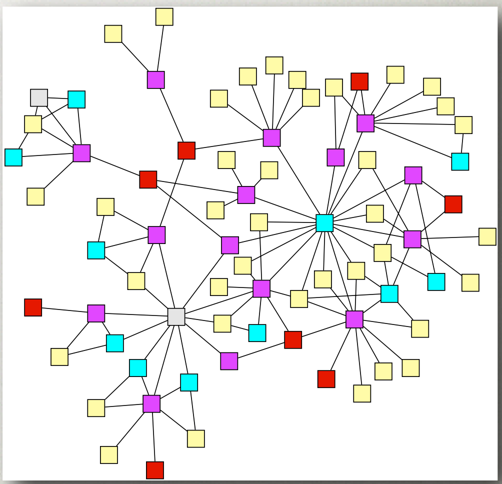
Clustering Coefficient





Distance Distribution





Missing Links

A test: can model predict missing links?

Predicting is Hard

- remove k edges from G
- how easy to guess a missing link?

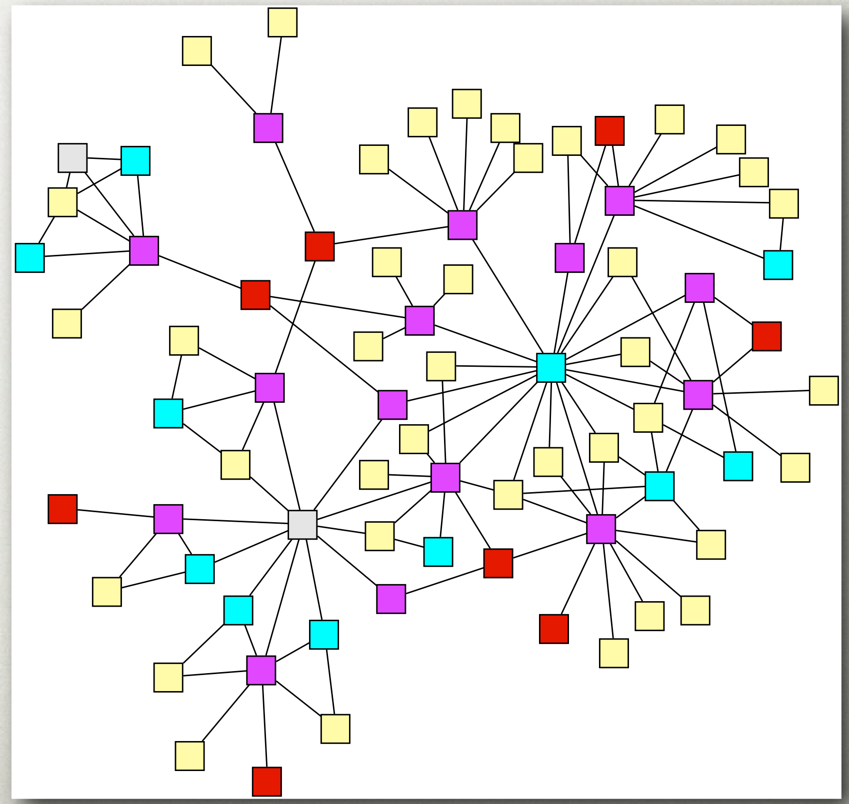
$$p_{\text{guess}} \approx \frac{k}{n^2 - m + k}$$

$$= O(n^{-2})$$

$$n = 75$$

$$m = 113$$

$$p_{\text{guess}} = k/(2662 + k)$$



Predicting Missing Links

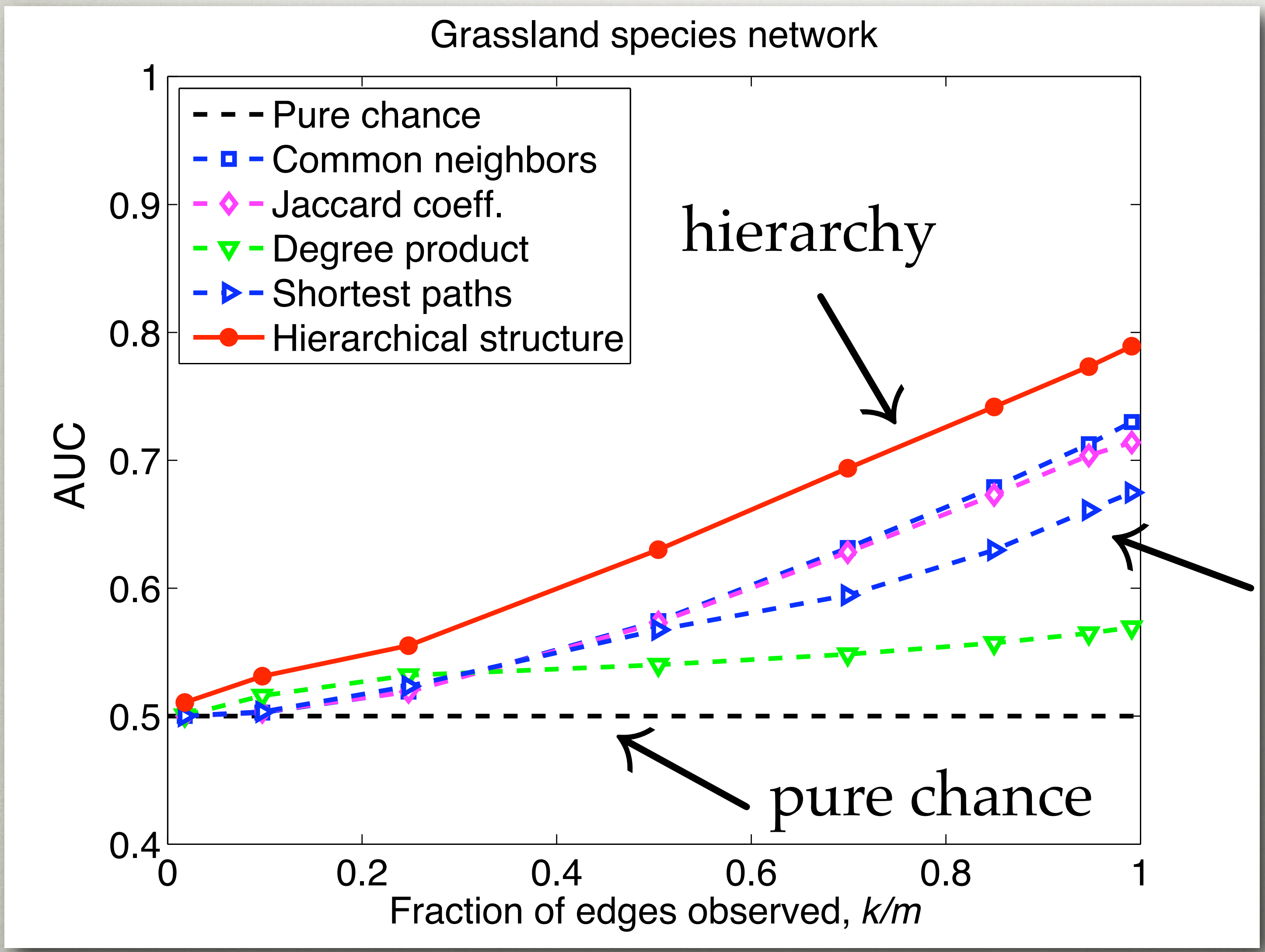
- Given incomplete graph G
- run MCMC to equilibrium
- then, over sampled $\mathcal{D}$, compute average $\langle p_r \rangle$ for links $(i, j) \notin G$
- predict links with high $\langle p_r \rangle$ values are missing

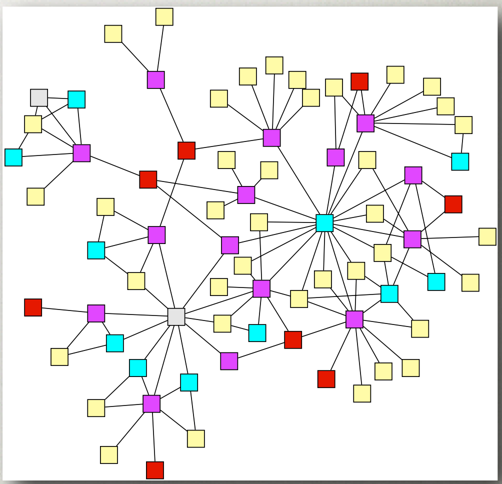
Test idea via leave-*k*-out cross-validation

perfect accuracy: AUC = 1

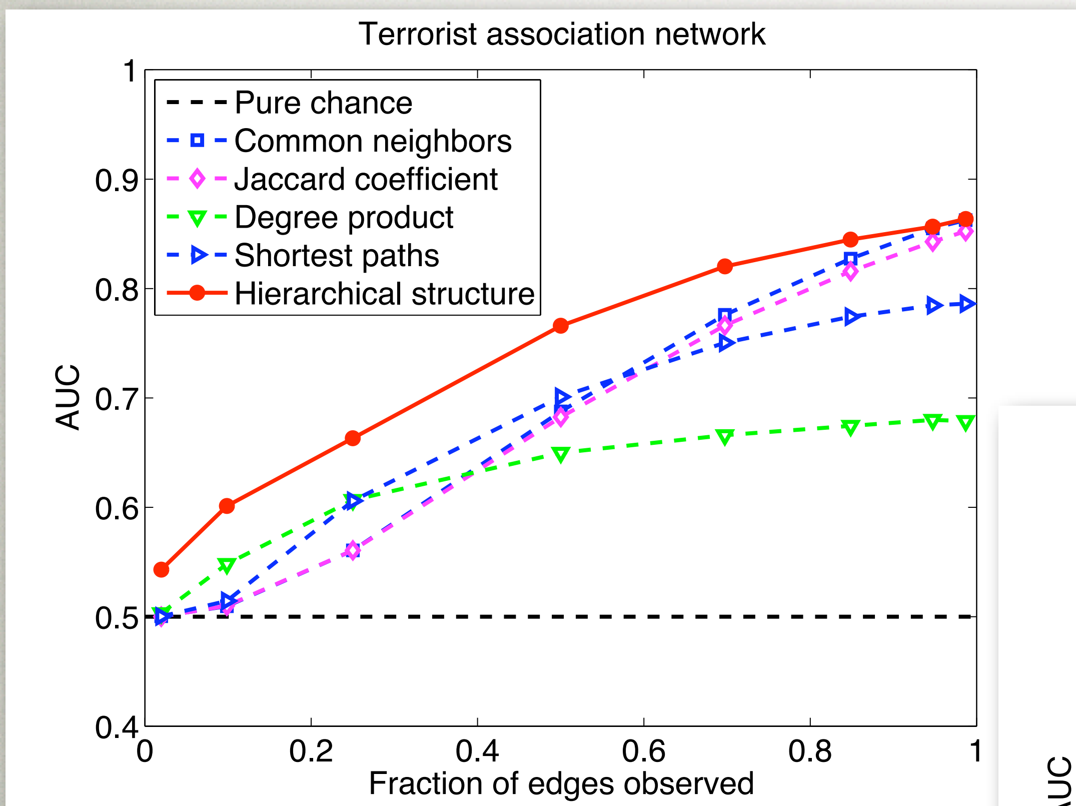
no better than chance: AUC = 1/2

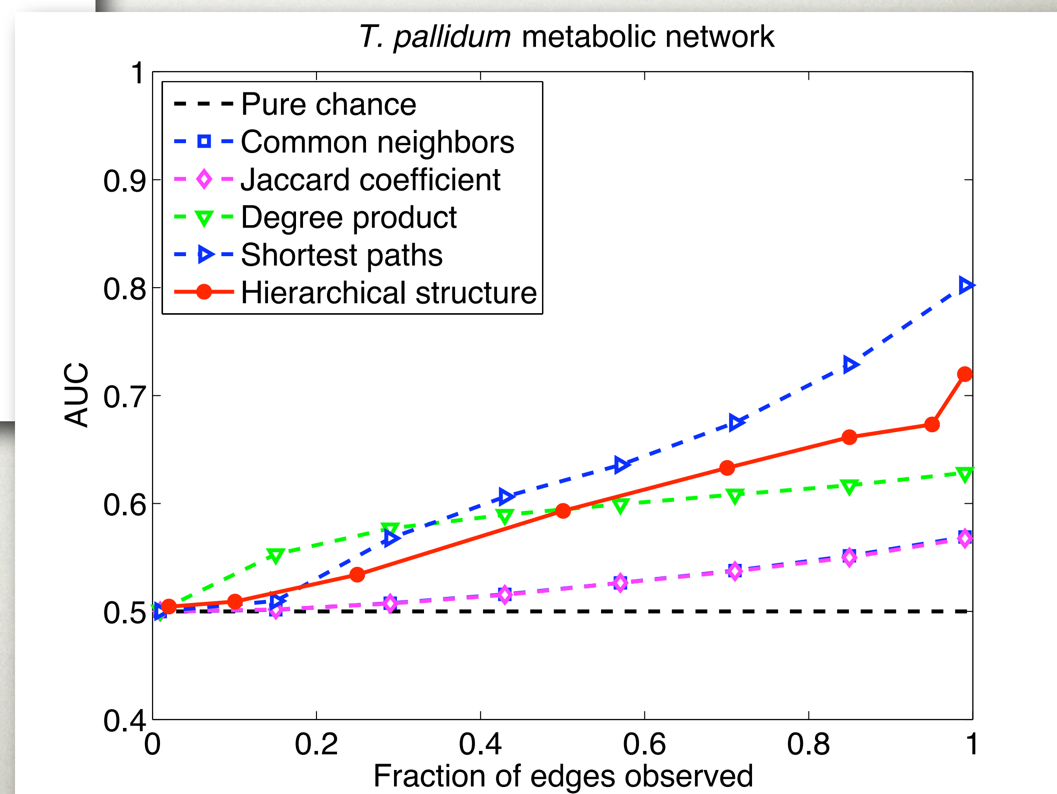
Missing Structure





Other Networks





Summary

- Many real networks are hierarchically modular
- Hierarchies can
  - model multi-scale structure
  - generalize a single network
  - predict missing links
- Model-based inference is very powerful

Acknowledgments:

C. Moore, M.E.J. Newman, C.H. Wiggins, and C.R. Shalizi

Fin