Chaos, Complexity, and Inference (36-462)

Lecture 25

Cosma Shalizi

22 April 2008

Inference for Network Models

Matched Random Networks

Hierarchical Random Graphs

Discriminating Network Growth Modes

General reading: Hunter *et al.* (2008)

Matched Random Networks

"So, you think you've found an interesting network structure, do you? Well isn't that *special*!"

Some kinds of network structure follow automatically from others

e.g., assortative $\Rightarrow$ reciprocal, cluster

Is what you are seeing an artifact or does it mean something?

Question of *what does a random network look like*?

But not just any random network, one that is *close* to yours

Basic Algorithm

Observe interesting feature X in your data graph g
Construct a distribution μ over random graphs G that matches g, but doesn't build in X
Draw many samples $G_1, G_2, \ldots G_m$ from μ
See how many of them have feature X

Simple Matched Random Networks

Erdős-Rényi networks are random...
Matching: same number of nodes and same density of edges $=$ expected degree
Very random... in fact, too random
In almost any situation, you know that your network doesn't look like *that*
It would be nice to match some more features than just the size and the density!

Exponential Random Graphs as Matched Random Networks

Pick your functionals on the network = sufficient statistics = T_i, $i \in 1:d$, observe values $t_i = T_i(g)$
Then (as discussed) $\widehat{\theta}_{MLE}$ solves

$$\mathbf{E}_{\widehat{\theta}_{MLE}}[T] = t$$

BUT $\mu = \mathrm{Pr}_{\widehat{\theta}_{MLE}}(G)$ also solves

$$\max H_{\mu}[G] \text{ s.t. } \mathbf{E}_{\mu}[T] = t$$

with H = Shannon entropy (Mandelbrot, 1962).

Maximizing likelihood in the exponential family maximizes entropy over all distributions

Why Should You Care about Maximum Shannon Entropy?

Some people see this as a self-justifying ideal

this is hard to take seriously

Gives distribution closest to independence under the constraint

(Amari, 2001)

but observed $\approx$ expected isn't a universal rule of inference!

unless observation is a big average!

Background problem: picking the right statistics to match

Fixing the Degree Distribution

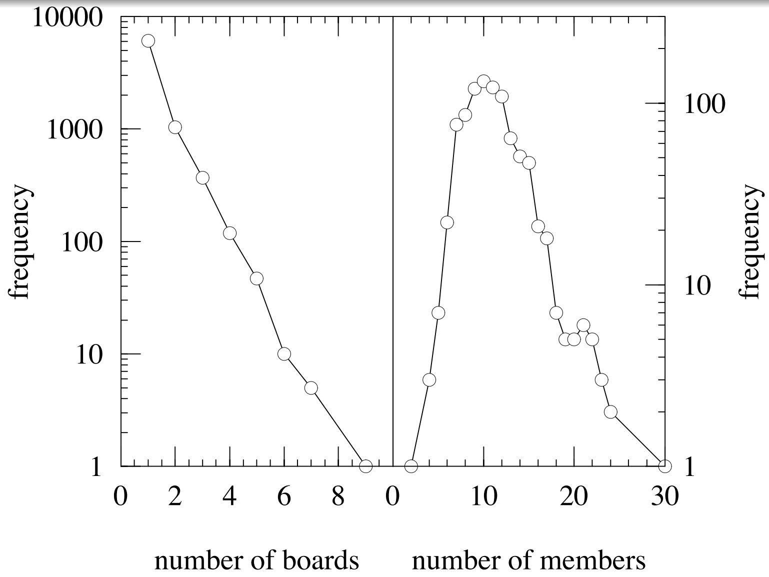
Newman *et al.* (2001)
Fix N
Generate N random numbers K_i from the empirical degree distribution — “stubs”
Choose pairs of free stubs uniformly at random; join them
Equally likely to produce any graph with that degree distribution
Must have even sum-of-degrees but this is not a big issue (if not even, discard and re-simulate)
Modifications required for directed graphs and bipartite graphs to handle summing-up constraints
if sums don’t match, pick one pair, discard their sizes, re-draw; repeat as needed

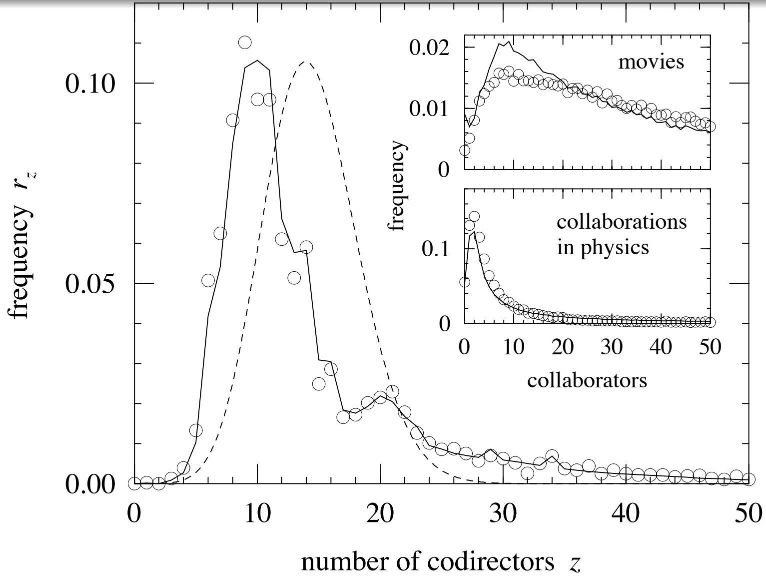
Example: Corporate Boards

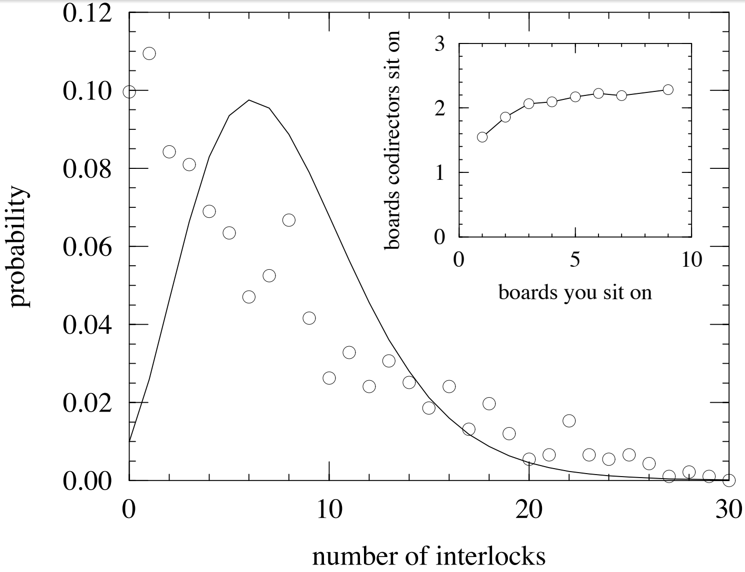
Public corporations have **directors** who represent shareholders and (supposedly) pick the executives
Board members often sit on many boards
This effectively a coordinating mechanism

> *People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.*

Also something that doesn't lead to very good decisions, but does shield rich people from market forces (Khurana, 2002)


10000
1000
100
10
1
frequency
0
2
4
6
8
number of boards
0
10
20
30
number of members
100
10
1
frequency


frequency r_z
0.00
0.05
0.10
0
10
20
30
40
50
number of codirectors z
frequency
movies
0.02
0.01
0
collaborations
in physics
0.1
0
0
10
20
30
40
50
collaborators


probability
number of interlocks
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
boards codirectors sit on
boards you sit on
3
2
1
0
0
5
10

Hierarchical Structure

Clauset *et al.* (2007)

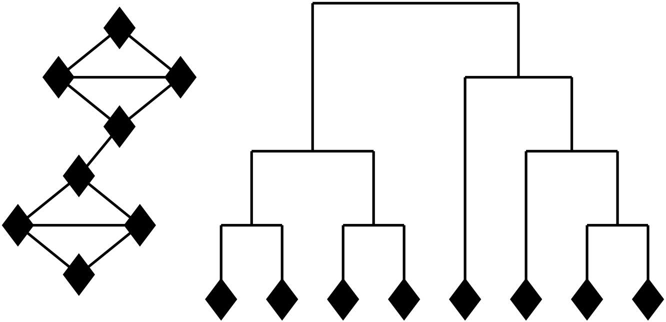
code: `http://www.santafe.edu/~aaronc/randomgraphs/`

Go back to the inhomogeneous Erdős-Rényi model

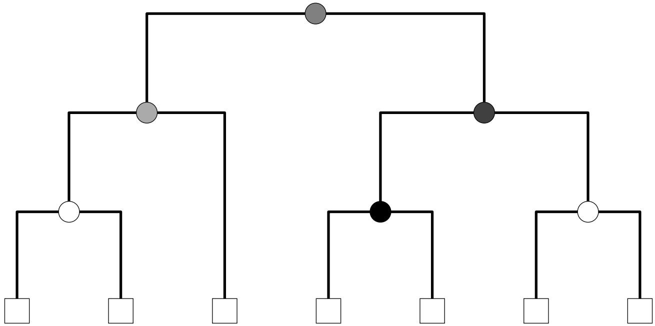
Completely arbitrary mixing matrix suffers a common problem of unconstrained maximum likelihood: over-fit by assigning probability 1 to data

here, 1 type per node

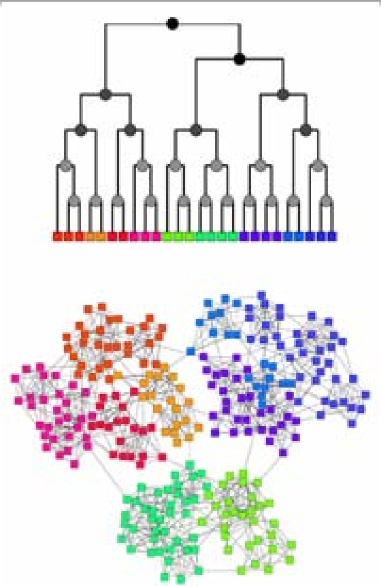
One constraint: hierarchy



Graph, hierarchy on nodes



Linking probability on tree for hierarchy



Probability of within-group linkage for group/type $i = \theta_i$
number of edges within group i, but not any of its sub-groups $= T_i$
number of nodes in left sub-tree, right sub-tree $= L_i, R_i$
Number of possible edges for group i is $L_i R_i$
so log-likelihood is

$$\mathcal{L}(\theta) = \sum_i T_i \log \theta_i + (L_i R_i - T_i) \log (1 - \theta_i)$$

Maximizing $\mathcal{L}$ assumes you know the tree! This is what we want to learn...

Could try all $\approx \sqrt{2}(2N)^{N-1}e^{-N}$ possible trees...

Parameter-counting penalties (like BIC) unhelpful since always $N-1$ parameters

What to do?

Model sampling and model averaging

Pick an initial tree τ however you like; maximized log-likelihood $= \mathcal{L}_\tau$

Randomly perturb it to make a new tree δ, log-likelihood $= \mathcal{L}_\delta$

Accept δ if $\mathcal{L}_\delta \geq \mathcal{L}_\tau$

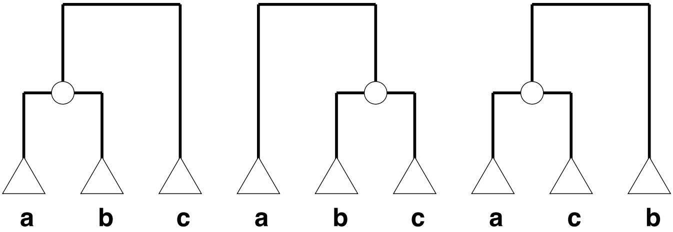
Otherwise, accept with probability equal to $e^{\mathcal{L}_\delta - \mathcal{L}_\tau}$

Gives a sample of trees where $\Pr(\tau) \propto e^{\mathcal{L}_\tau}$

Can average over trees, do weighted average (by likelihood), use common features of many trees...

Randomly perturbing trees

Pick random internal node in tree; has out-group and two in-groups; swap at random



this goes from any tree on N nodes to any other tree on N nodes

Discriminating Network Growth Modes

Middendorf *et al.* (2005)
Given: different models for how a network grew
Wanted: guess as to which one it was
Simulate many networks from each model
Train a **classifier** to reliably discriminate between them
Need **features** (sub-graph census) and **classifiers** (decision trees)
Validate the classifier by showing it has low error rates
Classifier may or may not look at features important in any one model

Amari, Shun-ichi (2001). "Information Geometry on Hierarchy of Probability Distributions." *IEEE Transactions on Information Theory*, **47**: 1701–1711. URL `http://www.islab.brain.riken.go.jp/~amari/pub/IGHI.pdf`.

Clauset, Aaron, Cristopher Moore and Mark E. J. Newman (2007). "Structural Inference of Hierarchies in Networks." In *Statistical Network Analysis: Models, Issues, and New Directions* (Edo Airoldi and David M. Blei and Stephen E. Fienberg and Anna Goldenberg and Eric P. Xing and Alice X. Zheng, eds.), vol. 4503 of *Lecture Notes in Computer Science*, pp. 1–13. New York: Springer-Verlag. URL `http://arxiv.org/abs/physics/0610051`.

Hunter, David R., Steven M. Goodreau and Mark S. Handcock (2008). "Goodness of Fit of Social Network Models." *Journal of the American Statistical Association*, **103**: 248–258. URL `http:`

//www.csss.washington.edu/Papers/wp47.pdf. doi:10.1198/016214507000000446.

Khurana, Rakesh (2002). *Searching for a Corporate Savior: The Irrational Quest for Charismatic CEOs*. Princeton, New Jersey: Princeton University Press.

Mandelbrot, Benoit (1962). "The Role of Sufficiency and of Estimation in Thermodynamics." *Annals of Mathematical Statistics*, **33**: 1021–1038. URL http://links.jstor.org/sici?sici=0003-4851%28196209%2933%3A3%3C1021%3ATROSAO%3E2.0.CO%3B2-N.

Middendorf, Manuel, Etay Ziv and Chris Wiggins (2005). "Inferring Network Mechanisms: The *Drosophila melanogaster* Protein Interaction Network." *Proceedings of the National Academy of Sciences (USA)*, **102**: 3192–3197. URL http://arxiv.org/abs/q-bio/0408010.

Newman, M. E. J., Steven H. Strogatz and Duncan J. Watts (2001). "Random graphs with arbitrary degree distributions and their applications." *Physical Review E*, **64**: 026118. URL `http://arxiv.org/abs/cond-mat/0007235`.