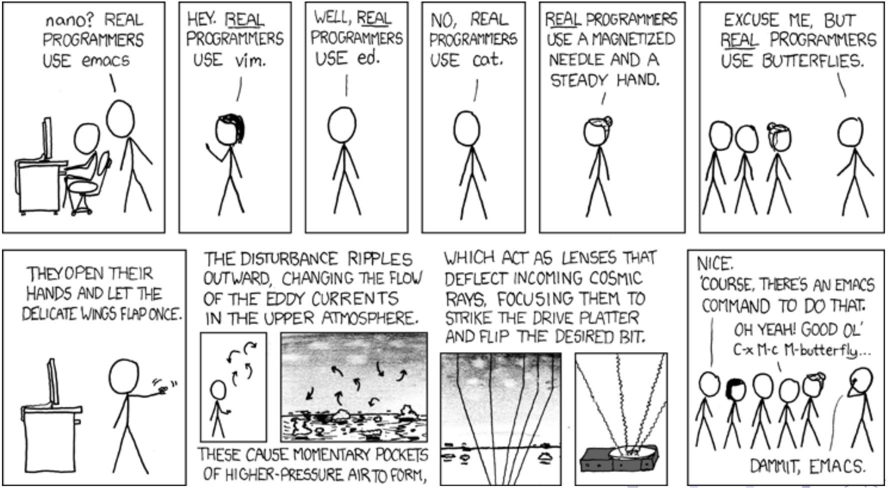
Chaos, Complexity, and Inference (36-462)

Lecture 21: More Networks: Models and Origin Myths

Cosma Shalizi

31 March 2009

New Assignment: Implement Butterfly Mode in R



Real Agenda: Models of Networks, with Origin Myths

Erdős-Rényi Again

n nodes, edges are IID binary variables with probability p
Degree of node i = K_i

$$K_i \sim \mathrm{Binom}(n-1, p) \rightsquigarrow \mathrm{Pois}(np)$$

Problems

Degree distribution Not Poisson

Reciprocity $\Pr\left(A_{ji} = 1 | A_{ij} = 1\right) \neq p$

Transitivity $\Pr\left(A_{ik} = 1 | A_{ij} = A_{jk} = 1\right) \neq p$

Homophily/Assortativeness $\Pr\left(A_{ij} = 1 | \mathrm{type}_i = \mathrm{type}_j\right) \neq p$

Inhomogeneous E-R Models

Give each node a type, $1, \ldots k$, T_i
mixing matrix P_{ab} = probability of link from type a to type b
Edges are still independent *given type*
Edges are *not* independent ignoring type
Example: $k = 2$, types uniform and independent

$$P = \left[\begin{array}{cc} 0.9 & 0.1 \\ 0.1 & 0.9 \end{array} \right]$$

Obviously gives homophily

$$\begin{aligned} p = \Pr\left(A_{ij} = 1\right) & \\ = & \ 0.9\Pr\left(T_i = T_j = 1\right) + 0.1\Pr\left(T_i = 1, T_j = 2\right) \\ & +0.1\Pr\left(T_i = 2, T_j = 1\right) + 0.9\Pr\left(T_i = T_j = 2\right) \\ = & \ 0.9 \times 0.25 + 0.1 \times 0.25 + 0.1 \times 0.25 + 0.9 \times 0.25 = 0.5 \end{aligned}$$

Also gives reciprocity:

$$
\begin{aligned}
\Pr\left(A_{ji=1} = 1, A_{ij} = 1\right) \\
&= 0.81\Pr\left(T_i = T_j = 1\right) + 0.01\Pr\left(T_i = 1, T_j = 2\right) \\
&\quad + 0.01\Pr\left(T_i = 2, T_j = 1\right) + 0.81\Pr\left(T_i = T_j = 2\right) \\
&= 0.41 \\
\Pr\left(A_{ji=1} = 1 | A_{ij} = 1\right) \\
&= \frac{\Pr\left(A_{ji} = 1, A_{ij} = 1\right)}{\Pr\left(A_{ij} = 1\right)} \\
&= 0.82 > 0.5
\end{aligned}
$$

EXERCISE: Show that this model has transitivity of edges as well

One direction for extending this: **block models** (“block” = type), indicating “type A gets links from type B, gives links to type C, never gets links from D or E. . . ”

Community structure or **modularity** is a limiting case of this, where mixing matrix has big diagonal entries, small off-diagonal ones

References: Reichardt and White (2007) for discovering block models; Clauset *et al.* (2007) for discovering hierarchies of modules; `http://bactra.org/notebooks/community-discovery.html` for references on community structure and community discovery

Watts-Strogatz "Small World" Graphs

Watts and Strogatz (1998)
Regular lattices have a lot of reciprocity and transitivity/clustering
but are "large worlds", in d dimensions diameter $= O(n^{1/d}) \gg O(\log n)$
Somehow interpolate between lattices and E-R graphs to get all three properties
but work with undirected graphs for simplicity

Solution: start with regular lattice, add “long-range shortcuts” at random

First approach: For each edge, with probability ρ, re-wire one edge to a uniformly random new node (avoiding self-loops)

As $\rho \rightarrow 0$, go to regular lattice

As $\rho \rightarrow 1$, go to E-R graph with same density as lattice

can create disconnected graphs

Second approach: add random edges *without* removing old ones

easier to manipulate, doesn't quite go to E-R as $\rho \rightarrow 1$

Will do more with this in the EXERCISES

Exponential Family Random Graphs

Measure graph properties like density, reciprocity, transitivity; specify graph probabilities in terms of them
Exponential families are the easiest way to do this

$$\begin{aligned}
\Pr(X = x) &= \frac{h(x) \exp\left\{\sum_{i=1}^{d} \theta_i T_i(x)\right\}}{\int dx\ h(x) \exp\left\{\sum_{i=1}^{d} \theta_i T_i(x)\right\}} \\
&= \frac{h(x) \exp\left\{\sum_{i=1}^{d} \theta_i T_i(x)\right\}}{Z(\theta)}
\end{aligned}$$

T_i are **sufficient statistics**, θ_i are **natural parameters**
Acronym: ERGM, Exponential family Random Graph Model ("err-gim" or "err-gum")

E-R model is an exponential family:

$$\begin{aligned}
\Pr(A = a) &= \prod_{i=1}^{n} \prod_{j \neq i} p^{a_{ij}} (1-p)^{(1-a_{ij})} \\
&= p^{\sum_{ij} a_{ij}} (1-p)^{n(n-1) - \sum_{ij} a_{ij}} \\
&= (1-p)^{n(n-1)} \left(\frac{p}{1-p} \right)^{\sum_{ij} a_{ij}} \\
&= (1-p)^{n(n-1)} \exp\left\{ (\log p/(1-p)) \sum_{ij} a_{ij} \right\}
\end{aligned}$$

so $T = \sum_{ij} a_{ij}$, $\theta = \log p/(1-p)$, $Z(\theta) = (1-p)^{-n(n-1)}$

Exponential family models are easy to fit by maximum likelihood, *if* you can find $Z(\theta)$ or $\mathbf{E}_{\theta}\left[T_i(x)\right]$

$$\frac{\partial \log \Pr\left(X = x\right)}{\partial \theta_i}$$

$$= \frac{\partial}{\partial \theta_i} \log h(x) + \frac{\partial}{\partial \theta_i} \sum_{j=1}^{d} \theta_j T_j(x) - \frac{\partial}{\partial \theta_i} \log Z(\theta)$$

$$= 0 + T_i(x) - \frac{1}{Z(\theta)} \frac{\partial Z(\theta)}{\partial \theta_i}$$

The last term is worth a look:

$$
\begin{aligned}
\frac{1}{Z(\theta)}\frac{\partial Z(\theta)}{\partial \theta_i} &= \frac{1}{Z(\theta)}\frac{\partial}{\partial \theta_i}\int dx\ h(x)\exp\left\{\sum_{j=1}^{d}\theta_j T_j(x)\right\} \\
&= \frac{1}{Z(\theta)}\int dx\ \frac{\partial}{\partial \theta_i} h(x)\exp\left\{\sum_{j=1}^{d}\theta_j T_j(x)\right\} \\
&= \frac{1}{Z(\theta)}\int dx\ h(x)\exp\left\{\sum_{j\neq i}\theta_j T_j(x)\right\}\frac{\partial}{\partial \theta_i}\exp\left\{\theta_i T_i(x)\right\} \\
&= \frac{1}{Z(\theta)}\int dx\ h(x)\exp\left\{\sum_{j\neq i}\theta_j T_j(x)\right\} T_i(x)\exp\{\theta_i T_i(x)
\end{aligned}
$$

continued:

$$\begin{aligned}\frac{1}{Z(\theta)}\frac{\partial Z(\theta)}{\partial \theta_i} &= \int dx\ T_i(x)\frac{h(x)\exp\left\{\sum_{i=1}^{d}\theta_i T_i(x)\right\}}{Z(\theta)} \\ &= \mathbf{E}_\theta\left[T_i(X)\right]\end{aligned}$$

Go back to the likelihood equation:

$$\begin{aligned}\frac{\partial \log \Pr\left(X=x\right)}{\partial \theta_i} &= T_i(x) - \frac{1}{Z(\theta)}\frac{\partial Z(\theta)}{\partial \theta_i} \\ &= T_i(x) - \mathbf{E}_\theta\left[T_i(X)\right]\end{aligned}$$

The derivatives are zero at the MLE $\hat{\theta}$:

$$T_i(x) = \mathbf{E}_{\hat{\theta}}\left[T_i(X)\right]$$

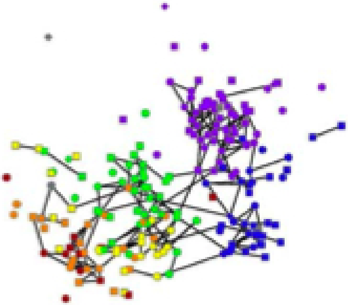
For E-R model, $\mathbf{E}_{\theta}\left[\sum_{ij} A_{ij}\right] = n(n-1)p$
so

$$\widehat{p}_{MLE} = \frac{\sum_{ij} a_{ij}}{n(n-1)}$$

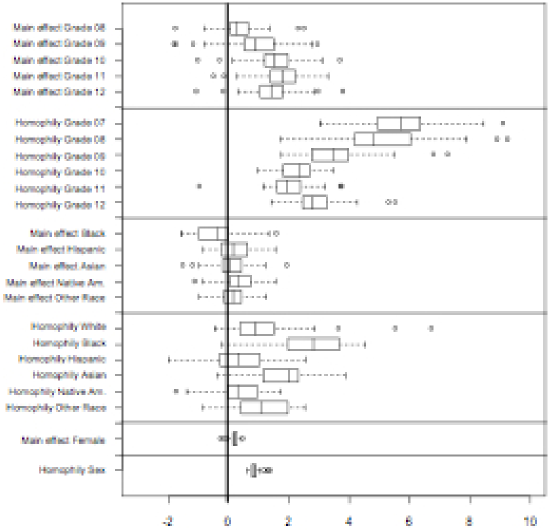
What about more complicated ERGMs?
"p_1 model": sufficient statistics are total number of edges, and total number of *reciprocal* edges
Not so easy to solve but can be done (Wasserman and Faust, 1994; Hunter *et al.*, 2008)
p^*: general ERGM, can add more features, homophily as such vs. reciprocity or transitivity as such...

Example of ERGMs Working

High school friendship network (Goodreau *et al.*, 2005)



Fit model including homophily by sex, grade, race; also different over all probability of forming edges ("main effect")


Main effect Grade 08
Main effect Grade 09
Main effect Grade 10
Main effect Grade 11
Main effect Grade 12
Homophily Grade 07
Homophily Grade 08
Homophily Grade 09
Homophily Grade 10
Homophily Grade 11
Homophily Grade 12
Main effect Black
Main effect Hispanic
Main effect Asian
Main effect Native Am.
Main effect Other Race
Homophily White
Homophily Black
Homophily Hispanic
Homophily Asian
Homophily Native Am.
Homophily Other Race
Main effect Female
Homophily Sex
-2
0
2
4
6
8
10

Best R package: `statnet` (on CRAN) — see special issue (vol. 24) of the *Journal of Statistical Software*, `http://www.jstatsoft.org/v24`

Generally *not* possible to solve

Use simulation to approximate $Z(\theta)$ and/or $\mathbf{E}_{\theta}[T(X)]$ (Hunter and Handcock, 2006)

even then there can be pathologies from bad choice of model (e.g. model say probability of these network statistics is 10^{-50})

Some Important Weaknesses of ERGMs

1. Possible pathologies in fitting
2. "Statistics convenient for us to measure" $\neq$ "important causal variables"
3. Matching some statistics doesn't mean matching others (Hunter *et al.*, 2008)
4. No origin myth/generative model (typically)

Some Generative Models

E-R model edges appear and disappear independently *over time* (works whether or not homogeneous)

p_1 model Markov chain, edge in one direction makes *adding* edge more likely, *losing* one edge makes other tend to go away

Watts-Strogatz Models See Clauset and Moore (2003) for a semi-plausible story about adaptive re-wiring

E-R again Add nodes one by one, each node adds links to existing nodes independently with probability p

Preferential attachment Graphical version of Yule-Simon process

Preferential Attachment

Made famous by Barabási and Albert (1999); Albert and Barabási (2002)

At each time-step a new node arrives

With probability ρ, new node *i* makes edge to old node *j*, picking $j \propto k_j$, degree of *j*

With probability $1 - \rho$, *i* links to a completely random node

This is *exactly* the Yule-Simon process that produces power law tails (Bornholdt and Ebel, 2001)

Apparently first applied to networks by Price (1965)

Will see more in the EXERCISES

Albert, Réka and Albert-László Barabási (2002). "Statistical Mechanics of Networks." *Reviews of Modern Physics*, **74**: 47–97. URL http://arxiv.org/abs/cond-mat/0106096.

Barabási, Albert-László and Réka Albert (1999). "Emergence of Scaling in Random Networks." *Science*, **286**: 509–512. URL http://arxiv.org/abs/cond-mat/9910332.

Bornholdt, Stefan and Holger Ebel (2001). "World-Wide Web scaling exponent from Simon's 1955 model." *Physical Review E*, **64**: 035104. URL http://arxiv.org/abs/cond-mat/0008465.

Clauset, Aaron and Cristopher Moore (2003). "How Do Networks Become Navigable?" *Physical Review Letters*, **submitted**. URL http://www.arxiv.org/abs/cond-mat/0309415.

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`.

Goodreau, Steven M., David R. Hunter and Martina Morris (2005). *Statistical Modeling of Social Networks: Practical Advances and Results*. Tech. Rep. 05-01, Center for Studies in Demography and Ecology, University of Washington. URL `http://csde.washington.edu/downloads/05-01.pdf`.

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.

Hunter, David R. and Mark S. Handcock (2006). "Inference in curved exponential family models for networks." *Journal of Computational and Graphical Statistics*, **15**: 565–583. URL http://www.stat.psu.edu/%7Edhunter/papers/cef.pdf.

Price, Derek J. de Solla (1965). "Networks of Scientific Papers." *Science*, **149**.

Reichardt, Jörg and Douglas R. White (2007). "Role models for complex networks." E-print, arxiv.org, 0708.0958. URL http://arxiv.org/abs/0708.0958.

Wasserman, Stanley and Katherine Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge, England: Cambridge University Press.

Watts, Duncan J. and Steven H. Strogatz (1998). "Collective Dynamics of "Small-World" Networks." *Nature*, **393**: 440–442.