Exponential random graphs

Keith Briggs
Keith.Briggs@bt.com
http://keithbriggs.info

BT

Unilever 2005 Nov 22 1400
Shi Zhou and Raúl Mondragón

★ Accurately modeling the internet topology

★ Network parameters:

- number of nodes, number of links, average degree, exponent of power law, rich-club connectivity, maximum degree, degree distribution, characteristic path length, average triangle coefficient, maximum triangle coefficient, average quadrangle coefficient, maximum quadrangle coefficient, average $k_{mn}$, average betweenness, maximum betweenness...
- girth, spectrum, ...
Motivation

- we use many random graph models in network applications...
- but rarely specify the statistical ensemble precisely...
- so even the averages we compute are suspect...
- and even the famous Barabási-Albert scale-free model has known problems...

we need a unified, rigorous framework

related ideas in earlier literature:
- Markov random fields
- \( p^* \) models of social networks
- Ising-type models in physics
- agricultural field trials
- image processing
- ...
Dependency graphs (Frank, Strauss, Besag, . . . )

★ consider a random vector $X = (X_1, X_2, \ldots, X_m)$

★ $P(x) = \exp(Q(x))/\sum \exp Q(x) \iff Q(x) = \log P(x) + \text{const}$
  ▶ only restriction $P(x) > 0 \ \forall x$

★ let $D$ be the dependency graph of $X$; i.e. $i \sim j \iff x_i$ not independent of $x_j$
  ▶ e.g. all $x_i$ independent: empty graph
  ▶ e.g. Markov chain: line graph
  ▶ e.g. multivariate Gaussian: complete graph (generically)

★ inclusion-exclusion principle $Q(x) = \sum_{s \subseteq \{1,2,\ldots,m\}} \lambda_s(x_s)$
  ▶ $x_s \equiv$ components of $x$ corresponding to elements of $s$
  ▶ $\Pr[\cap_i A_i] = \sum_i \Pr[A_i] - \sum_{i<j} \Pr[A_i \cup A_j] + \ldots$

★ Hammersley-Clifford theorem: $\lambda_s \equiv 0$ unless $s$ is a clique of $D$
  ▶ a clique is a complete subgraph
Markov graphs

★ to apply to a graph $g$ with edge dependencies, let $X$ be the edge indicator functions.

★ this defines the dependency graph $D(g)$ of $g$: $D(g)$ contains an edge $(i, j)$ if $X_i$ and $X_j$ ($i \neq j$) are dependent.

★ definition: $g$ is Markov if $D(g)$ contains no edge between edges which are disjoint in $E(g)$.

★ in other words, edges can only ‘interact’ if they share a common end-point.
Markov graph example \((n = 4, m = 5)\)

\[
\begin{array}{c}
1 \\
\downarrow a \\
\downarrow c \\
\downarrow b \\
\downarrow d \\
\downarrow e \\
2 \\
\downarrow g \\
3 \\
\downarrow a \\
\downarrow b \\
\downarrow d \\
\downarrow e \\
4 \\
\downarrow g \\
\end{array}
\]

\[D(g)\]

\[\begin{align*}
\text{a: } & 1-2 \\
\text{b: } & 1-3 \\
\text{c: } & 1-4 \\
\text{d: } & 2-3 \\
\text{e: } & 3-4 \\
\end{align*}\]

\[\text{cliques: } \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{a, b\}, \{b, c\}, \{a, c\}, \{d, e\}, \{a, b, c\}\}\]

\[Q(x) = \lambda_a(x_a) + \lambda_b(x_b) + \lambda_c(x_c) + \lambda_d(x_d) + \lambda_e(x_e) + \lambda_{ab}(x_a, x_b) + \lambda_{bc}(x_b, x_c) + \lambda_{ac}(x_a, x_c) + \lambda_{de}(x_d, x_e) + \lambda_{abc}(x_a, x_b, x_c)\]
if we require all isomorphic graphs to have the same probability, then a further simplification results: 

let $t(g)$ be the number of triangles in $g$

let $s_k(g)$ be the number of $k$-stars in $g$

then $P(g)$ can only depend on $t(g)$ and $s_k(g)$, in the form

$$P_{\beta}(g) = \frac{1}{Z(\beta)} \exp \left[ \beta_0 t(g) + \sum_{k=1}^{n-1} \beta_k s_k(g) \right]$$

where $\beta_i$ are fixed parameters

here $Z(\beta) = \sum_g \exp \left[ \beta_0 t(g) + \sum_{k=1}^{n-1} \beta_k s_k(g) \right]$
alternatively, we may use \( d_j \), the number of nodes of degree \( j \) \( (s_k(g) \equiv \sum_{j \geq k} \binom{j}{k} d_j(g)) \)

and let \( \theta_k(g) \equiv \sum_{k \leq j} \binom{j}{k} \beta_k \); then

\[
P_{\theta}(g) = \frac{1}{Z(\theta)} \exp \left[ \theta_0 t(g) + \sum_{j=1}^{n-1} \theta_j d_j(g) \right]
\]

in other words, the Hamiltonian can only be a linear function of the number of triangles and \( k \)-stars

note: if \( A \) if the adjacency matrix of \( g \), then \( m(g) = d_1(g) = \text{tr} \left( A^2 \right) / 2 \) is the number of edges and \( t(g) = \text{tr} \left( A^3 \right) / 6 \)
Exponential random graphs

★ fix a number of nodes $n$ ★

★ consider the set $G(n)$ of all graphs on $n$ nodes ★

★ we will assign to each $g \in G(n)$ a probability $P(g)$ ★

★ let $x = \{x_1, x_2, \ldots \}$ be a set of functions on $G(n)$ representing properties we are interested in, for example

- $x_1(g)$ = number of edges
- $x_2(g)$ = number of nodes of degree 3
- $x_3(g)$ = number of triangles ★

★ we then assign the probabilities $P$ by

$$P_\theta(g) = \frac{1}{Z(\theta)} \exp (\theta_1 x_1 + \theta_2 x_2 + \ldots)$$

where $Z(\theta) = \sum_{g \in G(n)} \exp (\theta_1 x_1 + \theta_2 x_2 + \ldots)$
Statistical mechanics

★ Hamiltonian: \( H(\theta, g) = \sum_i \theta_i x_i(g) \)

★ probability of \( g \): \( P_{\theta}(g) = \exp (H(\theta, g))/Z(\theta) \)

★ partition function: \( Z(\theta) = \sum_g \exp (H(\theta, g)) \)

★ entropy \( S(\theta) = -\sum_g P_{\theta}(g) \log(P_{\theta}(g)) \)

★ \( S \) is maximized by our choice of \( P \)

★ free energy: \( F(\theta) = \log(Z(\theta)) \)

★ \( E[x_i] = \frac{dF(\theta)}{d\theta_i} \)
Exactly soluble example - Bernoulli model \( G\{n, p\} \)

★ \( g \in G\{n, p\} \) has \( n \) nodes and each possible edge appears independently with fixed probability \( p \)

★ let \( x(g) = \) number of edges in graph \( g \)

★ \( H(\theta, g) = \theta m(g) \)

★ \( Z(\theta) = (1+\exp(-\theta))^n \binom{n}{2} \)

★ \( p = 1/(1+\exp(\theta)) \)

★ \( F(\theta) = \binom{n}{2} \log(1+\exp(-\theta)) \)

★ which gives \( P_p(g) = \binom{n}{m(g)} p^{m(g)} (1-p)^{\binom{n}{2}-m(g)} \) as expected

★ \( E[x] = m = \binom{n}{2} p \)
Likelihood of parameters, given a graph: $L(\theta|g) \propto P_\theta(g)$. (Loglikelihood: $l(\theta|g) \equiv \log P_\theta(g) + \text{const.}$)

Each figure shows the likelihood for one of the 34 graphs, the parameters corresponding to the number of nodes of degrees one and two. i.e. $H_\theta(g) = \theta_1 d_1(g) + \theta_2 d_2(g)$
Metropolis simulation

- works by defining a random walk in $G$ which has equilibrium distribution equal to our desired $P$

- typically we choose a pair of nodes, and then flip its state depending on whether the flip is energetically favourable

1. choose a proposal dyad $i, j \in N(g)$ uniformly at random

2. compute the energy change $\delta H$ that would occur if the dyad $(i, j)$ were flipped

3. if $\delta H$ or $u < \exp(\delta H)$, $u \sim U(0, 1)$, then accept the proposal; i.e. flip the edge

4. go to (1)

- estimate loglikelihood by (where $x$ is the vector of graph statistics, $\theta_{\text{ref}}$ a reference value of graph parameters (hopefully close to the true ones) and $x_{\text{data}}$ the statistics from the data):

$$l(\theta) - l(\theta_{\text{ref}}) \approx - \log \langle \exp[(l(\theta) - l(\theta_{\text{ref}}))(x(t) - x_{\text{data}})] \rangle$$
Metropolis simulation example

- 18 nodes; graph shows fluctuations in $m(g)$
Metropolised independence sampling (MIS)

Consider the target distribution $\pi_\theta(j) = \exp(j \log \theta)/Z(\theta)$, $Z(\theta) \equiv (1-\theta^n)/(1-\theta)$ on the set $\{0, 1, 2, \ldots, n-1\}$.

An MIS scheme is:

1. Start at $x = n-1$.
2. Choose a proposal $y \in X$ uniformly at random.
3. If $y < x$ or $u < \theta^{y-x}$, $u \sim U(0, 1)$, then accept the proposal; i.e. set $x = y$.
4. Go to (1).

Note that this scheme ignores the current position $x$ and assumes no knowledge of $\pi$. In general, we can do better.
for this example, it can be proven \cite{Diaconis1998} that

\[ 4 \left| \left| M^k - \pi \right| \right| \leq \frac{1}{(1 - \theta)(\theta^n - 1)} \left( 1 - \frac{1}{n} \right)^{2k} \]
What next?

★ directed graphs
★ perturbation theory (around Bernoulli model?)
★ more rapidly converging sampling schemes
★ parameter estimation for real examples by maximum likelihood (e.g. internet AS graph)
★ . . . ?
Some references (amongst many)


Z Burda & J Correia & A Krzywicki *Statistical ensemble of scale-free random graphs* http://xxx.soton.ac.uk/abs/cond-mat/0104155

J Berg & M Lässig *Correlated random networks* http://xxx.soton.ac.uk/abs/cond-mat/0205589


K M Briggs *graphlib-1.0* http://keithbriggs.info/graphlib.html