Eric traveling Sept. 27, 30 — no class

Random Graph Models

1. Observe (G, X) graph and attributes
2. observe G wish to infer X
3. observe X infer G

A network takes a value over $\{\mathbb{P}_\theta(G), G \in \mathcal{G} : \theta \in \Theta\}$

Equivalently over adjacency matrices

Good for
- Study mechanisms
- test significance of obs. characteristics
- Several others

Seminal work by Erdös and Renyi.

Emphasis on understanding the implications on

- degree distributions
- centrality
- connectivity

Random Graphs — drawn uniformly from some collection or ensemble $\mathcal{G}$

Erdos-Reyni: $\mathcal{G}_{N_V,N_E} = \{G=(V,E) : |V| = N_V,\ |E| = N_E\}$

$$\mathbb{P}(G) = \binom{N}{N_E}^{-1} \quad N = \binom{N_V}{2}$$

Gilbert: $\mathcal{G}_{N_V,p}$ random edges with prob p

Sorts of asympoptic equivalence.

<u>Seminal resource:</u> Bollobas

Properties: • $G \in \mathcal{G}_{N_V,p}$ $p = \frac{c}{N_V}$ $c > 0$

if $c > 1$ then whp G has one giant component of size $\alpha_c N_V$

all others are size $O(\log N_V)$

if $c < 1$ all components are of size $O(\log N_V)$

• Let $f_d(G)$ be the random proportion of vertices with degree d in G. Then $p = \frac{c}{N_V}$, $c > 0$ whp.

$$(1-\varepsilon)\frac{c^d e^{-c}}{d!} \leq f_d(G) \leq (1+\varepsilon)\frac{c^d e^{-c}}{d!}$$

So $f_d(G) \sim Po(c)$

• Can be sparse $N_E \simeq c$

• Low clustering $cl_T \simeq p = O(N_V^{-1})$

- Small diameter $O(\log N_v)$

Simulation properties – classical RG

- $O(N_v^2)$ for dense
- $O(N_v + N_E)$ for sparse – connects Bernoulli to geometric dists.

Generalized RG

- Equip $G \in \mathcal{G}$ with other characteristics

Most common: fixed degree dist. $\Longrightarrow$ fixed $\bar{d} = \frac{2N_E}{N_V}$ $\Longrightarrow$ fixed # edges

Good resources: Frieze & Karonski, van der Hofstad.

- Similiar properties exist

Simulation

- Matching – create a list of $d_{(i)}$ copies then randomly connect vertices with remaining degrees – accept/reject
- Switching – MCMC then rewire at each step.

Random Graph Models in Statistics

A model needs:
- estimable from data (Plausibility)
- reasonable representation of the data (Inference)
- amenable to model selection (Consistency and goodness of fit)