Bernoulli Graph Bounds for General Random Graph Models

Carter T. Butts

Department of Sociology and

Institute for Mathematical Behavioral Sciences

University of California, Irvine

buttsc@uci.edu

UCI MURI AHM, 05/25/10

This work was supported in part by ONR award N00014-08-1-1015.

Understanding Random Graph Models

- Substantial progress has been made on modeling networks (see, e.g. Wasserman and Robins, 2005; Hunter and Handcock, 2006)
  - ERG form provides a lingua franca for model specification
  - Once in ERG form, can simulate, perform inference (if not too big)
- Problem: little theory on the behavior of general random graphs
  - We can write down a model, but what does it do?
  - Simulation is an option, but only for small N
  - Few if any analytical results
- Today: one approach to this issue
  - Use some simple ideas from stochastic process theory to bound general random graphs by Bernoulli graphs
  - Can then use the (large) body of knowledge on Bernoulli graphs to study more general models
  - Has both methodological and theoretical uses (as we shall see)

Notational Note

- Assume $G = (V, E)$ to be the graph formed by edge set E on vertex set V
  - Here, will take $|V| = N$ to be fixed, and assume elements of V to be uniquely identified
  - E may be random, in which case $G = (V, E)$ is a *random graph*
  - Adjacency matrix $Y \in \{0, 1\}^{N \times N}$ (may also be random); for G random, will use notation y for adjacency matrix of realization g of G
  - Graph/adjacency matrix sets denoted by $\mathcal{G}, \mathcal{Y}$; set of all graphs/adjacency matrices of order N denoted $\mathcal{G}_N, \mathcal{Y}_N$
- Additional matrix notation
  - y^+_{ij},y^-_{ij} denote matrix y with i, j cell set to 1 or 0 (respectively)
  - y^c_{ij} denotes all cells of matrix y other than y_{ij}
  - Can be applied to random matrices, as well

General Random Graphs in ERG Form

- For order-N random graph G w/adjacency matrix Y on support $\mathcal{Y}_N$, can write pmf in ERG form by

$$\Pr(Y = y|\theta, t, \mathcal{Y}_N, X) = \frac{\exp\left(\theta^T t(y, X)\right)}{\sum_{y' \in \mathcal{Y}_N} \exp\left(\theta^T t(y', X)\right)} \mathbb{I}_{\mathcal{Y}_N}(y) \quad (1)$$

- $\theta^T \mathbf{t}$: linear predictor
  - $\mathbf{t} : \mathcal{Y}, X \to \mathbb{R}^m$: vector of sufficient statistics
  - $\theta \in \mathbb{R}^m$: vector of parameters
  - $\sum_{y' \in \mathcal{Y}_N} \exp\left(\theta^T \mathbf{t}(y', X)\right)$: normalizing factor (aka partition function, Z)
- Fully general framework, but analytically difficult
  - Z reduces to a tractable form in few cases (generally involving independent edges/dyads)
  - Even simulation is hard when differences of the form $t(y) - t(y')$ are expensive; requires MCMC, often prohibitive for large N

Taking Things Apart

- Instead of working with Y as a whole, can factor into "hierarchical" form

$$\Pr(Y=y|\theta,t,\mathcal{Y}_N,X) = \Pr(Y_{11}=y_{11}|\theta,t,\mathcal{Y}_N,X)\Pr(Y_{12}=y_{12}|\theta,t,\mathcal{Y}_N,X,Y_{11}=y_{11})\ldots \times \Pr(Y_{NN}=y_{NN}|\theta,t,\mathcal{Y}_N,X,Y_{11}=y_{11},\ldots,Y_{N(N-1)}=y_{N(N-1)}) \tag{2}$$

$$= \prod_{i=1}^{N}\prod_{j=1}^{N}\Pr(Y_{ij}=y_{ij}|\theta,t,\mathcal{Y}_N,X,Y_{<ij}=y_{<ij}), \tag{3}$$

  - (where $y_{<ij}$ denotes the sequence $y_{11},y_{12},\ldots,y_{i(j-1)}$ (for $j>1$) or $y_{11},y_{12},\ldots,y_{(i-1)N}$ (otherwise))
- Since we can't compute the marginals of Y, this seems pretty useless; but there is a workaround....

Some Random Observations

- Let $X = (X_1, \ldots, X_n)$ be a set of discrete random variables on joint support $\mathcal{S}$
  - Can write joint pmf as $\Pr(X = x) = \Pr(X_1 = x_1) \ldots \Pr(X_n = x_n | X_1 = x_1, \ldots, X_{n-1} = x_{n-1})$
  - By definition, $\Pr(X_i = x_i | X_{<i} = x_{<i}) = \sum_{x' \in \mathcal{S}: x'_{<i} = x_{<i}} \Pr(X_i = x_i | X_{-i} = x'_{-i}) \Pr(X_{-i} = x'_{-i} | X_{<i} = x_{<i})$
    - Thus, the ith marginal is a convex combination of the full conditionals of X_i
- An implication: full conditionals can be used to bound marginals
  - $\min_{x' \in \mathcal{S}} \Pr(X_i = x_i | X_{-i} = x'_{-i}) \leq \Pr(X_i = x_i | X_{<i} = x_{<i}) \leq \max_{x' \in \mathcal{S}} \Pr(X_i = x_i | X_{-i} = x'_{-i})$
  - Follows immediately from convexity condition
  - Where bounding full conditionals is easy, this can be very useful....

Why Is This Helpful? Conditionals in ERGs

- Well-known advantage of ERG form: full conditionals are easy to obtain
  - If B is the Bernoulli pmf, the full conditional pmf for Y_{ij} is given by

$$\Pr(Y_{ij} = y_{ij} | \theta, t, \mathcal{Y}_N, X, Y^c_{ij} = y^c_{ij}) = B\left(Y_{ij} = y_{ij} \middle| \left[1 + \exp\left[\theta^T \left(t\left(y^-_{ij}, X\right) - t\left(y^+_{ij}, X\right)\right)\right]\right]^{-1}\right) \quad (4)$$

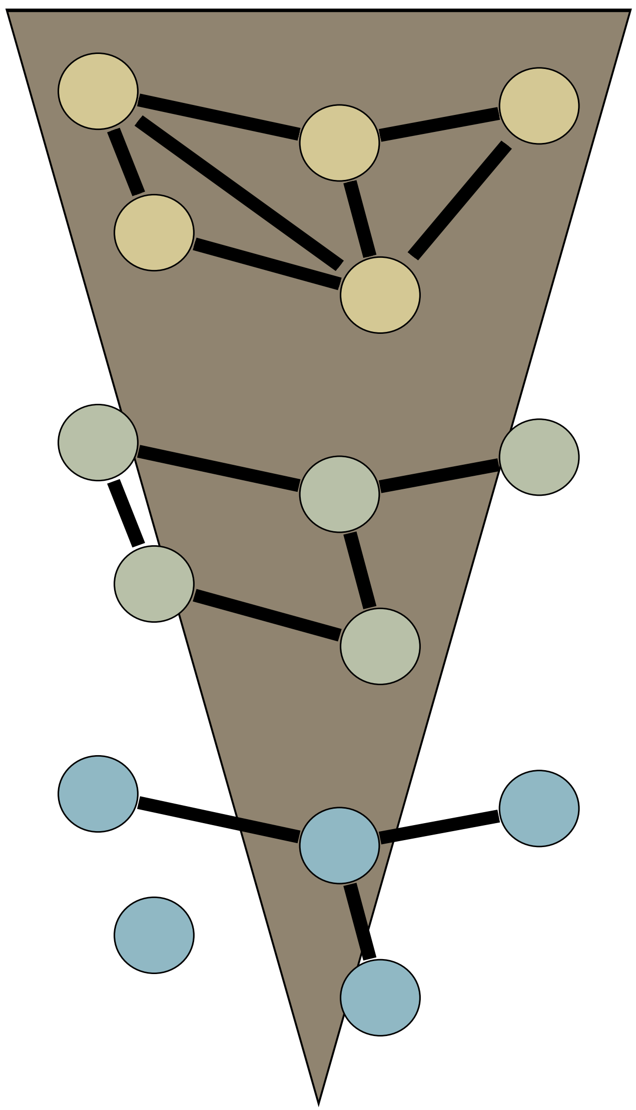
  - Equivalently, $\Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y^c_{ij} = y^c_{ij})$ is the inverse logit of $\theta^T \left(t\left(y^+_{ij}, X\right) - t\left(y^-_{ij}, X\right)\right)$
- Note: extrema of full conditionals (or bounds thereon) can often be found analytically
  - Just need knowledge of $t\left(y^+_{ij}, X\right) - t\left(y^-_{ij}, X\right)$ over $\mathcal{Y}_N$
  - Provides a way out of the “numerical trap”

From Bounding Distributions to Bounding Processes

- Place bounds on the edgewise marginals, using the previous result
  - Define $\Lambda, \Psi \in [0,1]^{\{N \times N\}}$ such that $\Lambda_{ij} \leq \min_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y^c_{ij} = y'^c_{ij})$ and $\Psi_{ij} \geq \max_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y^c_{ij} = y'^c_{ij})$
  - For convenience, define $\Gamma_{ij} \equiv \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{<ij} = y_{<ij})$
    - Note that Γ is itself a random matrix, but that Γ_{ij} depends only on $\Gamma_{<ij}$
- From the above, we can construct a pair of *bounding processes* for Y
  - Let $R = (R_{11}, \ldots, R_{NN})$ be an iid uniform vector on [0,1]
  - Define matrices $L, U \in \{0,1\}^{N \times N}$ such that $L_{ij} = 1$ iff $R_{ij} < \Lambda_{ij}$, and $U_{ij} = 1$ iff $R_{ij} < \Psi_{ij}$; let $Y_{ij} = 1$ iff $R_{ij} < \Gamma_{ij}$
  - L and U are said to be *Bernoulli bounding graphs* for Y (in adjacency matrix form)

Aside: The Subgraph Relation

- Given graphs G, H, G is a *subgraph* of H (denoted $G \subseteq H$) if $V(G) \subseteq V(H)$ and $E(G) \subseteq E(H)$
  - If y and y' are the adjacency matrices of G and H, $G \subseteq H$ implies $y_{ij} \leq y'_{ij}$ for all i, j
  - We use $y \subseteq y'$ to denote this condition
- f such that $G \subseteq H$ implies $f(G) \leq f(H)$ (or $f(G) \geq f(H)$) is *subgraph monotone*
  - Includes density, connectedness, degree scores, maximum component size, mean geodesic distance, etc.



Bernoulli Bounding Graphs for ERGs

- L and U bound Y in the following sense:

Theorem 1. *Let l, u, and y be realizations of L, U, and Y as defined above, for common realization vector $r_{11}, \ldots, r_{NN}$ of R. Then $l \subseteq y \subseteq u$.*

  - The proof is fairly immediate from the construction of L, U, and Y, given coupling through common realizations of R

- This has an important consequence:

Corollary 1. *Let f be a graph statistic that is monotone in $\subseteq$. Then $\Pr(f(L) \leq x) \geq \Pr(f(Y) \leq x) \geq \Pr(f(U) \leq x)$ for all x if f is monotone increasing in $\subseteq$, with $\Pr(f(L) \leq x) \leq \Pr(f(Y) \leq x) \leq \Pr(f(U) \leq x)$ otherwise.*

  - Thus, all raw moments of f for Y are bounded by the corresponding moments on L and U
  - Taking f to be an indicator variable likewise allows L and U to bound the probability of any monotone event (e.g., being fully connected) in Y

Bernoulli Graph Bounds on ERG Behavior

- From Corollary 1, we can bound the behavior of an arbitrary ERG by examining the behavior of its associated Bernoulli graphs
  - We refer to bounds derived in this way as *Bernoulli graph bounds*
- Bernoulli graph bounds can be obtained for any subgraph-monotone property
  - Includes many favorites, e.g. connectedness, density/mean degree, individual degrees, maximum component size, closeness, etc.
- Bounds can be either analytical or numerical
  - Often, can use classical random graph theory to study properties of the bounding graphs
  - Ease of exact simulation for Bernoulli graphs facilitates use even without analytical results

Summarizing the Approach

- General outline of the required procedure:
  1. Write the pmf of Y in ERG form
  2. Derive Λ and Ψ using the full conditionals
  3. Examine the desired (monotone) property, f, on L and U
  4. Use the behavior of f on L and U to describe the behavior of f on Y
- Bottom line: converts problems involving general random graphs (hard) to problems involving Bernoulli graphs (less hard)

The Value of Setting Boundaries

- Several important applications of Bernoulli graph bounds
  - Identifying potentially pathological ERGMs (and verifying that others are "safe")
  - Studying asymptotic behavior of graph processes
  - Identifying necessary/sufficient conditions for the emergence of particular structural features
  - Robustness testing for network models (e.g., bounding the impact of omitted effects)
  - Model approximation (e.g., substituting a Bernoulli model for a general ERG)
- Bounds will often be very loose, but can still provide useful results – we demonstrate by way of example

Example: Specification Robustness of Spatial Bernoulli Graph Models

- Need to be able to model networks on scales of cities, counties, etc.
  - Important to allow simulation of information diffusion, disease propagation, etc.
  - Also important to predict the effects of interventions (e.g., new technology) on the above
- Problem: how do we simulate networks for very large (e.g., $1 \times 10^4 - 1 \times 10^6$) populations?
  - MCMC is too slow to be practical here
  - Core idea: rely on the potential predictive power of geography (Butts, 2003) to pin down structure

Spatial Bernoulli Graph Models

- Spatial Bernoulli graphs provide a possible solution
  - Special case of the inhomogeneous Bernoulli family
  - Scale well, relatively easy to fit/simulate
- Basic idea: given distance matrix, D, $\Pr(Y_{ij} = 1|\theta, D) = \mathcal{F}(D_{ij}, \theta)$ (with all edges independent given D)
  - $\mathcal{F}$ is a *spatial interaction function* (SIF), e.g. $\mathcal{F}(D_{ij}, \theta) = p_b/(1 + \alpha D_{ij})^\gamma$ w/$\theta = (p_b, \alpha, \gamma)$
  - Can express in ERG form via edgewise pmf:

$$\Pr(Y_{ij} = 1|\theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}'^c) = \left[1 + \exp\left[-\log\left(\frac{(1 + \alpha D_{ij})^\gamma}{p_b} - 1\right)\right]\right]^{-1} \quad (5)$$

- Example: extrapolative simulation of friendship ties in a US county
  - $\mathcal{F}$ as defined above, with $\theta = (0.529, 0.031423, 2.768)$ from analysis by Butts (2002) from data of Festinger et al. (1950)
  - Test region is Choctaw, MS (pop 9,785) with population placed using block-level data from the year 2000 US Census

Population Distribution, Choctaw, MS

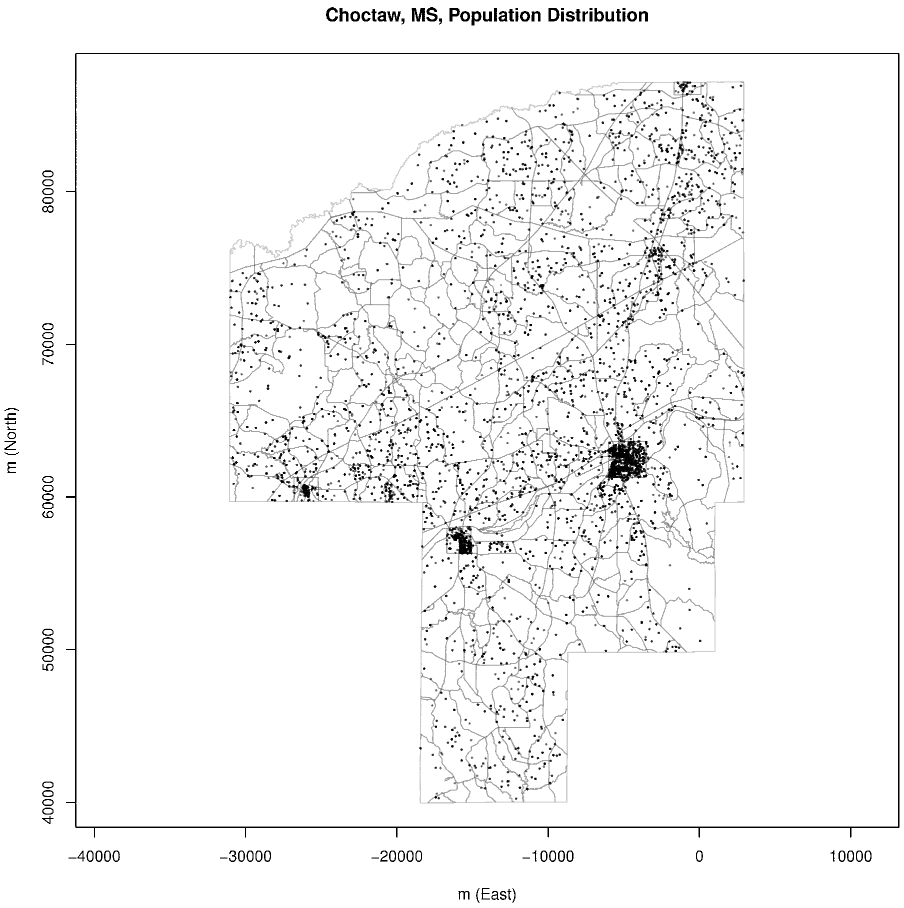


Figure 1: Population distribution for Choctaw County, MS. Points are placed uniformly by census block ($N = 9,758$).

Typical Model Realization, Choctaw, MS

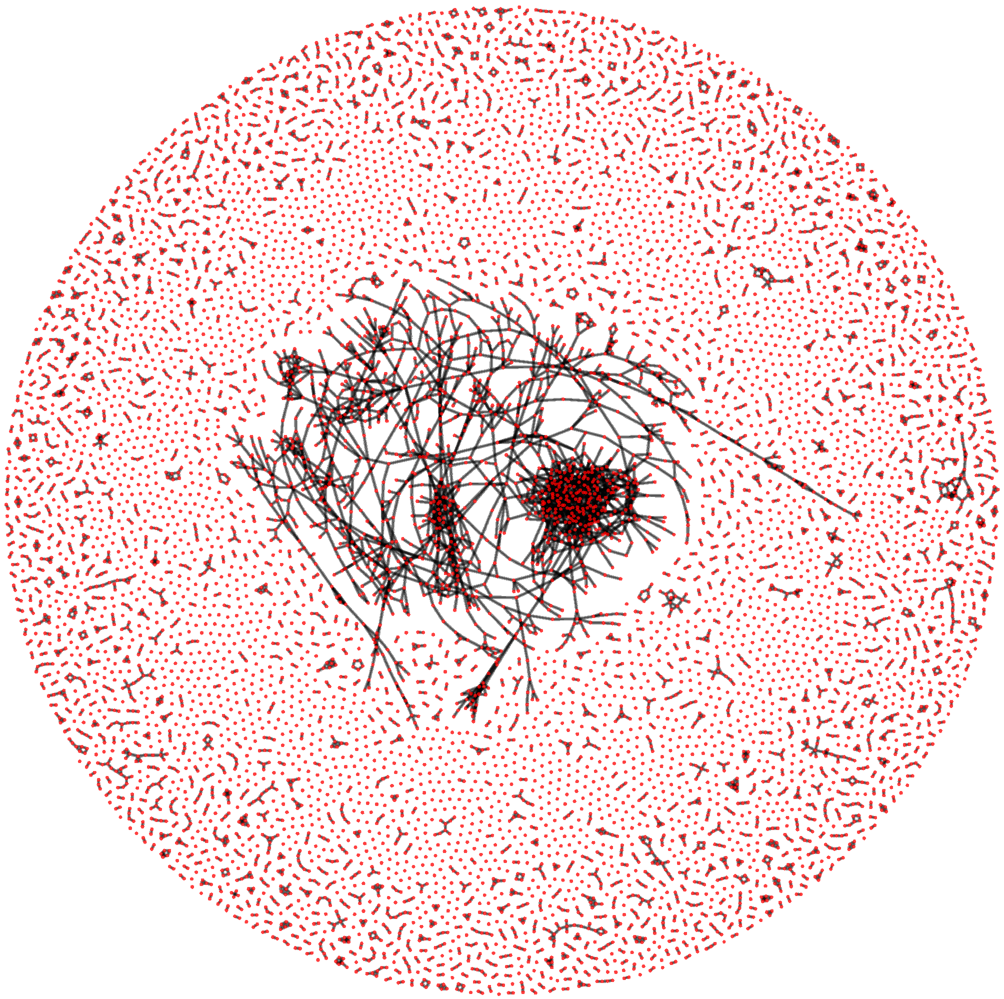


Figure 2: Typical realization of the social friendship model for Choctaw, MS.

Testing Our Limits

- Dyadic independence makes life easy, here, but is it safe?
  - Real network could have dependence mechanisms (e.g., endogenous clustering)
  - Would like to assess extent to which model is robust to omission of these factors
- Example mechanism: local triangulation
  - Define the *local triangle statistic* to be $t_{t_\ell}(y, D, \tau) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N} I(D_{ij} < \tau) I(D_{jk} < \tau) I(D_{ik} < \tau) y_{ij} y_{jk} y_{ik}$
    - Intuitively, can use to express idea that shared partners increase tie probability, but only when all parties are within radius τ of each other
  - Adding to ERG conditional gives us

$$\Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y_{ij}^c = y_{ij}'^c) = \left[1 + \exp\left[-\log\left(\frac{(1+\alpha D_{ij})^\gamma}{p_b} - 1\right) - \theta_t t_{t_\ell}(y, D, \tau)\right]\right]^{-1} \quad (6)$$

- How much impact could local triangulation have on the model?
  - Here, focus on diffusion: how many people could be reached by a rumor introduced to a randomly selected member of the population?
  - Question: how large must θ_t, τ be before its behavior changes substantially?

Bounding the Expanded Model

- We are interested in the increased diffusion resulting from excess clustering, and so focus on the upper bound (U) in the $\theta_t > 0$ regime
  - Note that the maximum conditional edge probability occurs when Y^c_{ij} is complete
  - From Equation 6, this gives us

$$\Psi_{ij} = \max_{y' \in \mathcal{Y}_N} \Pr(Y_{ij} = 1 | \theta, t, \mathcal{Y}_N, X, Y^c_{ij} = y'^c_{ij}) \tag{7}$$

$$= \left[1 + \exp\left[\log\left[\frac{(1 + \alpha D_{ij})^\gamma}{p_b} - 1\right] - \theta_t I(D_{ij} < \tau) \sum_{k \neq i,j} I(D_{jk} < \tau) I(D_{ik} < \tau)\right]\right]^{-1} \tag{8}$$

  - U is then a Bernoulli graph with parameter matrix Ψ
- To examine the impact of θ_t, τ on diffusion, we employ simulation
  - Draw from U, compute diffusion potentials (weighted distribution of component sizes)
  - Examine range of θ_t, τ over which there is little change in diffusion potential in U; this provides an upper bound on the true impact of local triangulation

Results: Mean Diffusion Size

	Mean Number Reached									
θ	0.0000	0.0020	0.0039	0.0078	0.0156	0.0312	0.0625	0.1250	0.2500	0.5000
$\tau = 0$m	103.02	103.31	103.04	104.04	103.67	103.47	103.70	103.76	103.29	103.93
3m	103.29	103.89	103.06	103.48	102.96	103.66	104.31	104.45	104.62	107.07
9m	103.45	104.63	104.18	104.15	104.42	106.11	108.36	112.27	118.59	125.94
27m	104.47	103.62	103.98	104.48	105.75	107.35	112.34	118.70	127.97	138.63
81m	104.25	105.29	107.07	111.53	119.09	129.35	149.81	171.89	197.90	224.19
243m	103.27	122.46	139.58	165.49	197.61	241.84	292.11	339.01	364.75	380.68
729m	102.68	225.80	284.04	331.03	363.88	394.30	441.22	474.45	554.09	800.55
2187m	103.10	409.02	460.34	513.53	590.84	995.69	4779.88	7270.19	8947.38	9651.42

Table 1: Upper bounds for the mean number of persons reachable from a randomly chosen seed as a function of θ and τ, based on 500 simulated networks.

Results: Median Diffusion Size

	Median Number Reached									
θ	0.0000	0.0020	0.0039	0.0078	0.0156	0.0312	0.0625	0.1250	0.2500	0.5000
$\tau = 0$m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
3m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
9m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00
27m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00
81m	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00	4.00
243m	3.00	3.00	3.00	3.00	3.00	3.00	4.00	4.00	4.00	5.00
729m	3.00	3.00	3.00	3.00	4.00	4.00	5.00	6.00	26.00	139.00
2187m	3.00	4.00	4.00	5.00	7.00	553.00	6825.00	8383.00	9345.00	9706.00

Table 2: Upper bounds for the median number of persons reachable from a randomly chosen seed as a function of θ and τ, based on 500 simulated networks.

Summary

- Often, we need to study general random graphs that are difficult to simulate (or we simply want analytical expressions for model behavior)
- Bernoulli graph bounds provide one approach to this problem
  - Leads to a pair, L, U of graphs such that $L \subseteq Y \subseteq U$ for a given set of random inputs
  - For subgraph monotone properties, features of L and U bound the features of Y
  - Since L, U are Bernoulli graphs, can use classical random graph theory and/or simulation to study
- Sample application: evaluating robustness of spatial Bernoulli models to local clustering
  - In the studied case, can show that fairly strong/long range clustering is required to change diffusion behavior
  - Results suggest that this model (in this setting, at least) is not sensitive to assumption of no excess clustering, at least for diffusion

Further Developments

- Approximate ERG sampling in fixed worst-case time
  - Given set of random "coins," some elements of Y will be fixed by their bounds
  - Can iteratively update bounds on Y_i marginals by previously fixed edge variables; when no progress can be made, perturb by selecting first unfixed Y_i and randomly fixing (approximating the true marginal by a random guess)
  - Clearly $O(N^4)$ in worst case, but can approach $O(N^2)$ when dependence is weak; haven't yet implemented, so not clear how well it works
- "Pseudo-marginalized" likelihoods
  - The pseudo-likelihood approximates all marginals by their full conditionals, effectively approximating marginal sum by a point mass
  - Why not try marginalizing by summing over draws from a (Bernoulli) baseline model (conditioning as we go)?
  - Could lead to more robust estimator than the MPLE with similar computational cost

1 References

Butts, C. T. (2002). *Spatial Models of Large-scale Interpersonal Networks*. Doctoral Dissertation, Carnegie Mellon University.

Butts, C. T. (2003). Predictability of large-scale spatially embedded networks. In Breiger, R., Carley, K. M., and Pattison, P., editors, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. National Academies Press, Washington, D.C.

Festinger, L., Schachter, S., and Back, K. (1950). *Social Pressures in Informal Groups*. Stanford University Press, Stanford, California.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15:565–583.

Pattison, P. and Robins, G. (2002). Neighborhood-based models for social networks. *Sociological Methodology*, 32:301–337.

Wasserman, S. and Robins, G. (2005). An introduction to random graphs, dependence graphs, and $p*$. In Carrington, P. J., Scott, J., and Wasserman, S., editors, *Models and*

Methods in Social Network Analysis, chapter 10, pages 192–214. Cambridge University Press, Cambridge.