

Statistical models: An introduction to random graph distributions

©Department of Psychology, University of Melbourne

Outline

Bernoulli (Erdos-Renyi) random graph distribution

Properties of the Bernoulli distribution

Simulation

Conditional random graph distributions

$U | L$

$U | M, A, N$

$U | \{i_a\}, U | \{o_a\}, U | \{i_a\}, \{o_a\}$

Other conditional random graph distributions

Conditional random multigraph distributions

Quadratic Assignment Procedure (QAP)

The p_1 distribution

Fitting the p_1 distribution

©Department of Psychology, University of Melbourne

Bernoulli random graphs

Basic definitions

Fixed node set: $N = \{1, 2, \dots, n\}$

Variable edges: X_{ij} (asks *the question*: is there an edge from i to j ?)

Realised edge values: x_{ij} (1 if there is an edge from i to j ; 0 otherwise)

Edge probability: We write $p_{ij} = \Pr(X_{ij} = 1)$

Bernoulli random graph distribution

Assume that edges are independent:

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{ij} p_{ij}^{x_{ij}} (1 - p_{ij})^{(1-x_{ij})}$$

Simple Bernoulli (or Erdos-Renyi) random graph distribution

If we also assume that $p_{ij} = p$, for all i and j :

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{ij} p^{x_{ij}} (1 - p)^{(1-x_{ij})}$$

©Department of Psychology, University of Melbourne

Some properties of Bernoulli graphs

Expected number of edges: $pn(n-1)/2$

Expected density: p

Expected average degree: $p(n-1)$

Probability that a node has degree k (for large n):

$\approx (\alpha)^k e^{-\alpha} / k!$ where $\alpha = 2L/n$ *Poisson degree distribution*

Random graph theory (e.g., Bollobas, 1985)

examines graph “properties” as a function of p and n

finds that many properties emerge “suddenly” as a function of p : there is a critical probability $p_c(n)$, below which the property is very rare, above which it is very common

©Department of Psychology, University of Melbourne

Some critical probabilities

<i>For</i>	<i>critical probability* $p_c(n)$</i>
Subgraphs:	
A tree of order k	$cn^{-k/(k-1)}$
A cycle of order k	cn^{-1}
A complete graph of order k	$cn^{-2/(k-1)}$
A subgraph with k nodes and l edges	$cn^{-k/l}$
A giant component	$cn^{-2/(k-1)}$
A giant cluster	$1/(n-1)$
The graph to be connected	$(\ln n)/(n-1)$

**c is a constant*

©Department of Psychology, University of Melbourne

Expected graph structure as a function of p

$p < 1/(n-1)$ *isolated trees*

$p > 1/(n-1)$ *giant cluster*

$p > (\ln n)/(n-1)$ *completely connected*

©Department of Psychology, University of Melbourne

Simulating a random graph distribution

Very easy

$n(n-1)/2$ independent draws from a binomial distribution with probability p to obtain adjacency matrix

repeat many times

Can use

Almost any statistical package or programming language

Pnet, with just an *edge parameter*, set to $\ln(p/1-p)$

©Department of Psychology, University of Melbourne

Using Bernoulli random graphs as a null model

Bernoulli models

poor models for almost all social systems

but it is often valuable to compare the property of an observed network to the distribution of the property in a random graph with the same expected tie probability p

For example

we observe a given triad census (or some other property) in a graph
we compute the distribution of the triad census (or other property) for a random graph with the same expected tie probability

if the observed value(s) are much higher (or much lower) than that expected for a random graph distribution, we can infer that the null model is inadequate *and* that the triad census (or other property) needs to be explained in another way

©Department of Psychology, University of Melbourne

The uniform distribution U is equivalent to a Bernoulli distribution with $p = 0.5$

The *uniform random graph distribution* U assigns equal probability to all graphs on n nodes:

$$\Pr(\mathbf{X} = \mathbf{x}) = 1/[2^{n(n-1)/2}]$$

This is identical to the Bernoulli random graph model with $p = 1/2$:

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{ij} p^{x_{ij}}(1-p)^{(1-x_{ij})} = (1/2)^{n(n-1)/2}$$

©Department of Psychology, University of Melbourne

Conditional uniform random graph distributions

Suppose Q is a *property* of a graph, such as:

- Having L edges
- Having T triangles
- Having a fixed degree d for each node
- Having degree sequence (d_1, d_2, \dots, d_n)
- Having degree distribution $(k_0, k_1, \dots, k_{n-1})$
- Etc, etc.....

Let Q be the set of all graphs on n nodes with the property Q

The *conditional uniform random graph distribution* $U|Q$ assigns:
equal probability to all graphs in Q (ie all graphs with the property Q) and
zero probability to all other graphs (not in Q)

©Department of Psychology, University of Melbourne

Some conditional uniform random graph and directed graph distributions

Conditional uniform random graph (and directed graph) distributions have been used as more “sophisticated” null models, since they allow one to assess the distribution of a feature of interest in a way that conditions on other features of a (directed) graph, often regarded as “lower order” properties.

Some examples

Q	U Q
number of edges	$U L$
outdegrees	$U \{X_{i+}\}$
the dyad census	$U MAN$
outdegrees, no of mutual ties	$U \{X_{i+}\}, M$
in-, out-degrees, no of mutual ties	$U \{X_{i+}\}, \{X_{+i}\}, M$

©Department of Psychology, University of Melbourne

Simulating conditional uniform random graph distributions

Many of these conditional uniform random graph distributions can be simulated using permutation distributions:

U | L randomly permute above-diagonal entries of adjacency matrix (graphs) or off-diagonal entries (directed graph)

U | $\{X_{i+}\}$ independently within each row, randomly permute non-diagonal entries

U | MAN randomly permute dyads associated with upper diagonal, and then randomly assign 1 or 0 to above diagonal for asymmetric ties

etc

Stocnet:

Z0 simulates the distribution $U | \{i_a\}, \{o_a\}$ (Snijders, 1991)

PACNET uses conditional uniform random graph distributions to assess path structures (Pattison et al, 2000)

©Department of Psychology, University of Melbourne

Applications using the triad census

Triad census:

count of subgraph types of order 3

$\mathbf{T} = \{T_u\}$ where T_u is the no of triads of type u

Distribution of triad census in certain conditional uniform random graph distributions:

mean $\boldsymbol{\mu}_T$ and covariance matrix $\boldsymbol{\Sigma}_T$ of triad census:

$U | MAN$ (Holland & Leinhardt, 1975)

$U | \{X_{i+}\}$ (Wasserman, 1977) distributions

Compare frequency of each triad type in an observed network with expected distribution

©Department of Psychology, University of Melbourne

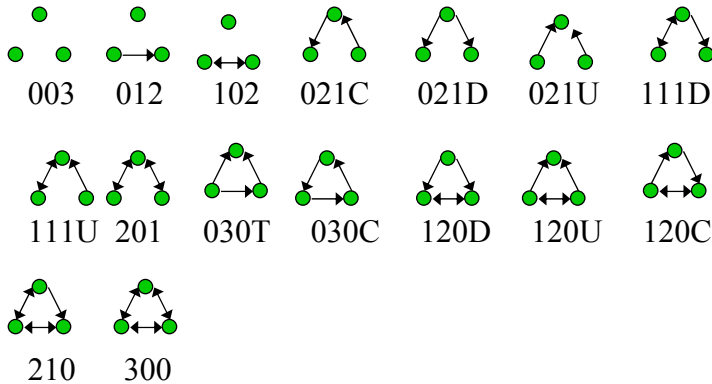
Test statistics for any linear combination of triad counts

$$\tau_w = \frac{\mathbf{w}'\mathbf{T} - \mathbf{w}'\boldsymbol{\mu}_T}{\sqrt{(\mathbf{w}'\boldsymbol{\Sigma}_T\mathbf{w})}}$$

Can use to assess transitivity, clustering, hierarchy, closure, etc.

©Department of Psychology, University of Melbourne

Triad census for a directed graph



©Department of Psychology, University of Melbourne

Application to models for networks of positive affect (Johnsen, 1986)

-
- {300,102} *balance model*
- {300,102,003} *clustering model*
- {300,102,120D,120U,030T,021D,021U,003} *ranked cluster model*
- {300,102,120D,120U,030T,021D,021U,003,012} *transitivity model*
- {300,102,120D,120U,030T,021D,021U,003,012,210} *hierarchical cliques model*
- {300,102,120D,120U,030T,021D,021U,003,012,210} *39+ model*

©Department of Psychology, University of Melbourne

Quadratic Assignment Procedure (QAP)

Given two observed graphs, what is the relationship between them?

Compute a measure of association between the two graphs (eg correlation)

Consider the uniform conditional random *multiple* graph distribution for the pair of (directed) graphs in which every possible permutation of node labels in the second graph is equally likely (a model of *no association* between graphs)

Compute a measure of association between the two graphs for each permutation to obtain a permutation distribution of correlation measures

Compare the observed measure to the distribution: if the observed measure is extreme in the conditional uniform random graph distribution, the model of no association is questionable

©Department of Psychology, University of Melbourne

QAP on UCINET (Tools/statistics/matrix(QAP)/QAP corr)

```
Observed matrix:      FADGB
Structure matrix:    FADGM
# of Permutations:    2500
Random seed:         233
```

Univariate statistics

```
FADGB  FADGM  -----
1  Mean  0.125  0.167
2  Std Dev  0.331  0.373
3  Sum  30.000  40.000
4  Variance  0.109  0.139
5  SSQ  30.000  40.000
6  MCSSQ  26.250  33.333
7  Euc Norm  5.477  6.325
8  Minimum  0.000  0.000
9  Maximum  1.000  1.000
10 N of Obs  240.000  240.000
```

Hubert's gamma: 16.000

1	2	3	4	5	6	7				
			Value	Signif	Avg	SD	P(Large)	P(Small)	NPerm	
1	Pearson Correlation:		0.372	0.000	0.003	0.092	0.000	1.000	2500.000	
2	Simple Matching:		0.842	0.000	0.751	0.027	0.000	1.000	2500.000	
3	Jaccard Coefficient:		0.296	0.000	0.081	0.046	0.000	1.000	2500.000	
4	Goodman-Kruskal Gamma:		0.797	0.000	-0.054	0.379	0.000	1.000	2500.000	
5	Hamming Distance:		38.000	0.000	59.770	5.588	1.000	0.000	2500.000	

©Department of Psychology, University of Melbourne

p_1 model (Holland & Leinhardt, 1981)

Assume *independent dyads* $D_{ij} = (X_{ij}, X_{ji})$ with

$$\Pr(D_{ij}=(1,1)) = m_{ij}, i < j$$

$$\Pr(D_{ij}=(1,0)) = a_{ij}, i \neq j$$

$$\Pr(D_{ij}=(0,1)) = n_{ij}, i < j$$

where $m_{ij} + a_{ij} + a_{ji} + n_{ij} = 1$, for all $i < j$

©Department of Psychology, University of Melbourne

The p_1 model

Then

$$p_1(\mathbf{x}) = \Pr(\mathbf{X}=\mathbf{x}) = K \exp[\sum_{i < j} \rho_{ij} X_{ij} X_{ji} + \sum_{ij} \theta_{ij} X_{ij}]$$

where:

$$\rho_{ij} = \log \{m_{ij} n_{ij} / (a_{ij} a_{ji})\}, i < j$$

$$\theta_{ij} = \log \{a_{ij} / n_{ij}\}, i, j$$

$$K = \prod_{i < k} 1/k_{ij}$$

$$k_{ij} = 1 + \exp(\theta_{ij}) + \exp(\theta_{ji}) + \exp(\rho_{ij} + \theta_{ij} + \theta_{ji})$$

©Department of Psychology, University of Melbourne

A common form of the p_1 model

If we add restrictions:

$$\rho_{ij} = \rho \quad i < j$$

$$\theta_{ij} = \theta + \alpha_i + \beta_j \quad i \neq j$$

then

$$p_1(\mathbf{x}) = K \exp[\rho M + \theta L + \sum_i \alpha_i X_{i+} + \sum_i \beta_i X_{+i}]$$

Reciprocity

density

expansiveness

popularity

©Department of Psychology, University of Melbourne

Aside: p_1 as a loglinear model

p_1 can be expressed as a standard log-linear model on the array \mathbf{Y} defined by

$$Y_{ijkl} = \begin{cases} 1 & \text{if } X_{ij} = k \text{ and } X_{ji} = l \\ 0, & \text{otherwise} \end{cases}$$

p_1 corresponds to the *log-linear model on \mathbf{Y} with all two-way interactions*

©Department of Psychology, University of Melbourne

Fitting p_1 using UCINET

In UCINET

Tools

Statistics

p_1

Enter data file

©Department of Psychology, University of Melbourne

The p_1 model for the Doreain network

```
G-Square      DF
-----
99.16         63
```

```
Theta = -0.3981
Rho   = -0.1991
```

Expansiveness and Popularity Parameters

		1	2
		Alpha	Beta
		-----	-----
1	Executive	0.038	0.038
2	Auditor	0.038	0.038
3	Sheriff	0.243	0.243
4	Council 1	0.433	0.433
5	Council 2	-0.463	-0.463
6	Council 3	0.243	0.243
7	Council 4	-0.189	-0.189
8	President	0.243	0.243
9	Council 5	0.038	0.038
10	Council 6	0.038	0.038
11	Fr Council	-0.872	-0.872
12	Fr Pres	0.433	0.433
13	City Mayor	0.243	0.243
14	Prosecutor	-0.463	-0.463

©Department of Psychology, University of Melbourne

Fitted values

	Exec	Audi	Sher	Coun	Coun	Coun	Coun	Pres	Coun	Coun	Fr	C	Fr	P	City	Pros
1 Executive	0.00	0.30	0.39	0.49	0.14	0.39	0.21	0.39	0.30	0.30	0.07	0.49	0.39	0.14		
2 Auditor	0.30	0.00	0.39	0.49	0.14	0.39	0.21	0.39	0.30	0.30	0.07	0.49	0.39	0.14		
3 Sheriff	0.39	0.39	0.00	0.59	0.19	0.49	0.29	0.49	0.39	0.39	0.10	0.59	0.49	0.19		
4 Council 1	0.49	0.49	0.59	0.00	0.26	0.59	0.38	0.59	0.49	0.49	0.13	0.68	0.59	0.26		
5 Council 2	0.14	0.14	0.19	0.26	0.00	0.19	0.09	0.19	0.14	0.14	0.02	0.26	0.19	0.05		
6 Council 3	0.39	0.39	0.49	0.59	0.19	0.00	0.29	0.49	0.39	0.39	0.10	0.59	0.49	0.19		
7 Council 4	0.21	0.21	0.29	0.38	0.09	0.29	0.00	0.29	0.21	0.21	0.04	0.38	0.29	0.09		
8 President	0.39	0.39	0.49	0.59	0.19	0.49	0.29	0.00	0.39	0.39	0.10	0.59	0.49	0.19		
9 Council 5	0.30	0.30	0.39	0.49	0.14	0.39	0.21	0.39	0.00	0.30	0.07	0.49	0.39	0.14		
10 Council 6	0.30	0.30	0.39	0.49	0.14	0.39	0.21	0.39	0.30	0.00	0.07	0.49	0.39	0.14		
11 Fr Council	0.07	0.07	0.10	0.13	0.02	0.10	0.04	0.10	0.07	0.07	0.00	0.13	0.10	0.02		
12 Fr Pres	0.49	0.49	0.59	0.68	0.26	0.59	0.38	0.59	0.49	0.49	0.13	0.00	0.59	0.26		
13 City Mayor	0.39	0.39	0.49	0.59	0.19	0.49	0.29	0.49	0.39	0.39	0.10	0.59	0.00	0.19		
14 Prosecutor	0.14	0.14	0.19	0.26	0.05	0.19	0.09	0.19	0.14	0.14	0.02	0.26	0.19	0.00		

©Department of Psychology, University of Melbourne

Residual values

	Exec	Audi	Sher	Coun	Coun	Coun	Coun	Pres	Coun	Coun	Fr	C	Fr	P	City	Pros
1 Executive	0.00	-0.30	0.61	0.51	-0.14	-0.39	0.79	-0.39	-0.30	-0.30	-0.07	-0.49	0.61	-0.14		
2 Auditor	-0.30	0.00	-0.39	-0.49	-0.14	-0.39	-0.21	0.61	0.70	0.70	-0.07	0.51	-0.39	-0.14		
3 Sheriff	0.61	-0.39	0.00	0.41	-0.19	0.51	0.71	-0.49	-0.39	-0.39	-0.10	-0.59	-0.49	0.81		
4 Council 1	0.51	-0.49	0.41	0.00	0.74	0.41	0.62	-0.59	-0.49	-0.49	-0.13	0.32	-0.59	-0.26		
5 Council 2	-0.14	-0.14	-0.19	0.74	0.00	0.81	-0.09	-0.19	-0.14	-0.14	-0.02	-0.26	-0.19	-0.05		
6 Council 3	-0.39	-0.39	0.51	0.41	0.81	0.00	-0.29	-0.49	-0.39	-0.39	-0.10	0.41	-0.49	0.81		
7 Council 4	0.79	-0.21	0.71	0.62	-0.09	-0.29	0.00	-0.29	-0.21	-0.21	-0.04	-0.38	-0.29	-0.09		
8 President	-0.39	0.61	-0.49	-0.59	-0.19	-0.49	-0.29	0.00	0.61	0.61	-0.10	0.41	0.51	-0.19		
9 Council 5	-0.30	0.70	-0.39	-0.49	-0.14	-0.39	-0.21	0.61	0.00	0.70	-0.07	-0.49	0.61	-0.14		
10 Council 6	-0.30	0.70	-0.39	-0.49	-0.14	-0.39	-0.21	0.61	0.70	0.00	-0.07	-0.49	0.61	-0.14		
11 Fr Council	-0.07	-0.07	-0.10	-0.13	-0.02	-0.10	-0.04	-0.10	-0.07	-0.07	0.00	0.87	-0.10	-0.02		
12 Fr Pres	-0.49	0.51	-0.59	0.32	-0.26	0.41	-0.38	0.41	-0.49	-0.49	0.87	0.00	0.41	-0.26		
13 City Mayor	0.61	-0.39	-0.49	-0.59	-0.19	-0.49	-0.29	0.51	0.61	0.61	-0.10	0.41	0.00	-0.19		
14 Prosecutor	-0.14	-0.14	0.81	-0.26	-0.05	0.81	-0.09	-0.19	-0.14	-0.14	-0.02	-0.26	-0.19	0.00		

©Department of Psychology, University of Melbourne

References

- Albert, R., & Barabási, A-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47-97.
- Bollobás, B. (1985). *Random graphs*. London: Academic
- Holland, P., & Leinhardt, S. (1975). Local structure in social networks. In D. Heise (Ed.) *Sociological Methodology 1976* (pp. 1-45). San Francisco: Jossey-Bass.
- Holland, P., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76, 33-50.
- Pattison, P., Wasserman, S., Robins, G., & Kanfer, A. (2000). Statistical evaluation of algebraic constraints for social networks. *Journal of Mathematical Psychology*, 44, 536-568.
- Snijders, T. A. B. (1991). Enumeration and simulation models for 0-1 matrices with given marginals. *Psychometrika*, 56, 397-417.