Statistical Inference for Networks

Hsu Centennial, Beijing, July 2010

Peter Bickel

Statistics Dept. UC Berkeley

(Joint work with Aiyou Chen, *Bell Labs*, E. Levina, *U. Mich*)

(With assistance of Jing Lei)

Outline

- ∗ *1* Networks: types/Examples
- ∗ *2* Networks: questions
  - *a)* Descriptive
  - *b)* Quantitative
- ∗ *3* Modularities
- *4* Statistical issues and selected models
- *5* A nonparametric model for infinite networks and asymptotic theory
- *6* Consistency of modularities and efficient estimation
- *7* The asymptotics of degree distribution and empirical moments
- *8* Other methods of parametric submodel estimation
- *9* Some examples and discussion.

References

1. M.E.J. Newman (2010) Networks: An introduction. Oxford
2. Fan Chung, Linyuan Lu (2004) Complex graphs and networks. CBMS # 107 AMS
3. Eric D. Kolaczyk (2009) Statistical Analysis of Network Data
4. Bela Bollobas, Svante Janson, Oliver Riordan (2007) The Phase Transition in Random Graphs. Random Structures and Algorithms, 31 (1) 3-122
5. B. and A. Chen (2009) A nonparametric view of network models and Newman-Girvan and other modularities, PNAS

Note: We will not discuss dynamically generated models

Examples: Technological Networks

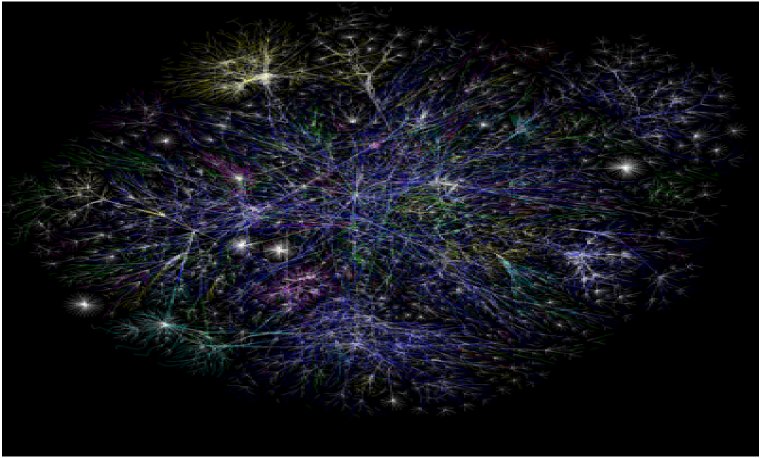


Figure: Internet (The OPTE Project)

Examples: Social Networks

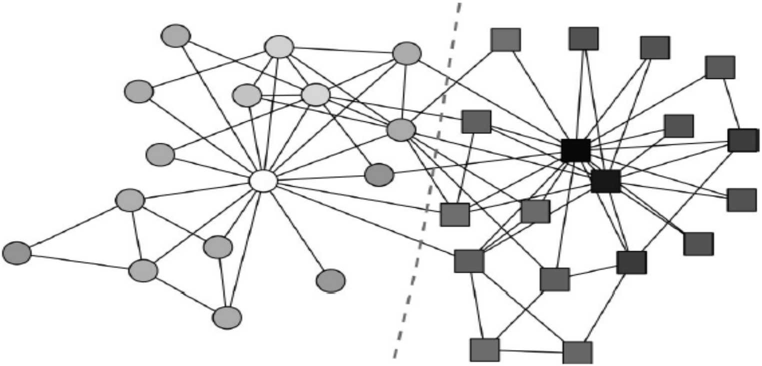


Figure: Karate Club (Newman, PNAS 2006)

Examples: Biological Networks

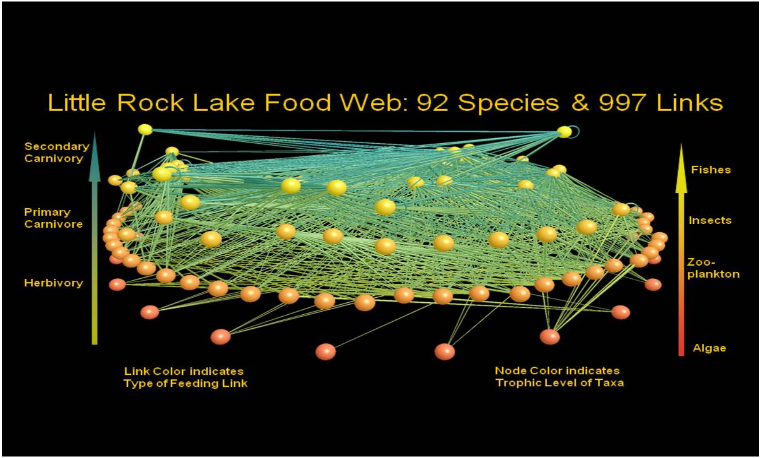


Figure: Food web (Neo Martinez, Berkeley)

Examples: Metabolic Web

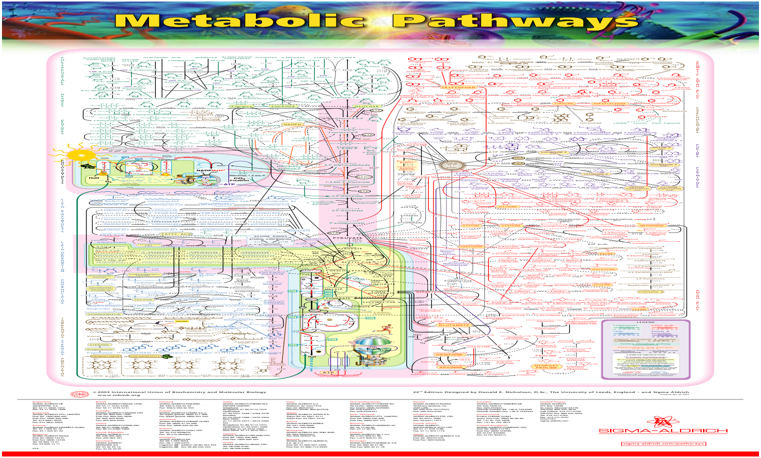


Figure: Metabolic Pathways (IUBMB-Nicholson)

Examples: Information Networks

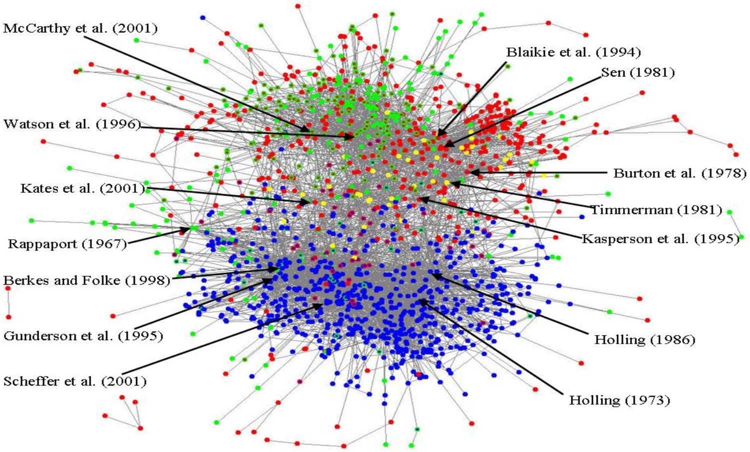


Figure: Paper networks (Marco A. Janssen, ASU)

A Mathematical Formulation

- $G = (V, E)$: undirected graph
- $\{1, \cdots, n\}$: Arbitrarily labeled vertices
- A : adjacency matrix
- $A_{ij} = 1$ if edge between i and j (relationship)
- $A_{ij} = 0$ otherwise

Implications of Mathematical Description

- Undirected: Relations to or from not distinguished.
- Arbitrary labels: individual, geographical information not used.

Descriptive Statistics for Graph Structures

Centrality

- Def: Degree $D_i = \sum_{j \neq i} A_{ij}$

Statistics

- Relative degree $\frac{n}{2L} D_i$: "centrality of vertex"

  $$L \equiv \frac{1}{2} \sum_{i \neq j} A_{ij} = \# \text{ of edges}$$

- Average degree: "centrality of graph"

  $$c = \frac{2L}{n}$$

Graph Structures

Cohesiveness

Def:

- Clique: Maximal fully connected subgraphs
- k-core: Maximal subset of vertices such that each is connected to at least k other members of subset.

Statistics:

- Size of cliques
- Number of k-cores

Clustering

- Transitivity: If i is related to j and j is related to k, then it is likely that i is related to k.
- Global Clustering Coefficient:

$$C = \frac{3 \times \ \# \text{ of } \triangle}{\# \text{ of } \triangle + \# \text{ of } \vee}$$

Chain Structure

Def:

- (Geodesic) Path between i, j: (shortest) set of edges (i, i_1) (i_1, i_2) ... (i_k, j).
- Connected component: Maximal set such that all pairs of vertices are connected by path in set.

Statistics:

- # and size of connected components.

	Network	Type	n	m	c	S	ℓ	α	C	C_{WS}	r	Ref(s).
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980	3.48	2.3	0.20	0.78	0.208	16,323
Social	Company directors	Undirected	7 673	55 392	14.44	0.876	4.60	–	0.59	0.88	0.276	88,253
Social	Math coauthorship	Undirected	253 339	496 489	3.92	0.822	7.57	–	0.15	0.34	0.120	89,146
Social	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838	6.19	–	0.45	0.56	0.363	234,236
Social	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918	4.92	–	0.088	0.60	0.127	234,236
Social	Telephone call graph	Undirected	47 000 000	80 000 000	3.16			2.1				9,10
Social	Email messages	Directed	59 812	86 300	1.44	0.952	4.95	1.5/2.0		0.16		103
Social	Email address books	Directed	16 881	57 029	3.38	0.590	5.22	–	0.17	0.13	0.092	248
Social	Student dating	Undirected	573	477	1.66	0.503	16.01	–	0.005	0.001	−0.029	34
Social	Sexual contacts	Undirected	2 810					3.2				197,198
Information	WWW nd.edu	Directed	269 504	1 497 135	5.55	1.000	11.27	2.1/2.4	0.11	0.29	−0.067	13,28
Information	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914	16.18	2.1/2.7				56
Information	Citation network	Directed	783 339	6 716 198	8.57			3.0/–				280
Information	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977	4.87	–	0.13	0.15	0.157	184
Information	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000		2.7		0.44		97,116
Technological	Internet	Undirected	10 697	31 992	5.98	1.000	3.31	2.5	0.035	0.39	−0.189	66,111
Technological	Power grid	Undirected	4 941	6 594	2.67	1.000	18.99	–	0.10	0.080	−0.003	323
Technological	Train routes	Undirected	587	19 603	66.79	1.000	2.16	–		0.69	−0.033	294
Technological	Software packages	Directed	1 439	1 723	1.20	0.998	2.42	1.6/1.4	0.070	0.082	−0.016	239
Technological	Software classes	Directed	1 376	2 213	1.61	1.000	5.40	–	0.033	0.012	−0.119	315
Technological	Electronic circuits	Undirected	24 097	53 248	4.34	1.000	11.05	3.0	0.010	0.030	−0.154	115
Technological	Peer-to-peer network	Undirected	880	1 296	1.47	0.805	4.28	2.1	0.012	0.011	−0.366	6,282
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996	2.56	2.2	0.090	0.67	−0.240	166
Biological	Protein interactions	Undirected	2 115	2 240	2.12	0.689	6.80	2.4	0.072	0.071	−0.156	164
Biological	Marine food web	Directed	134	598	4.46	1.000	2.05	–	0.16	0.23	−0.263	160
Biological	Freshwater food web	Directed	92	997	10.84	1.000	1.90	–	0.20	0.087	−0.326	209
Biological	Neural network	Directed	307	2 359	7.68	0.967	3.97	–	0.18	0.28	−0.226	323,328

Table 8.1: Basic statistics for a number of networks. The properties measured are: type of network, directed or undirected; total number of vertices n; total number of edges m; mean degree c; fraction of vertices in the largest component S (or the largest weakly connected component in the case of a directed network); mean geodesic distance between connected vertex pairs ℓ; exponent α of the degree distribution if the distribution follows a power law (or "–" if not; in/out-degree exponents are given for directed graphs); clustering coefficient C from Eq. (7.41); clustering coefficient C_{WS} from the alternative definition of Eq. (7.44); and the degree correlation coefficient r from Eq. (7.82). The last column gives the citation(s) for each network in the bibliography. Blank entries indicate unavailable data.

Newman (2010) Networks: an introduction, Oxford

Quartiles (2-4)	Social			Information			Technological			Biological network		
size (K)	59	450	47K	270	780	203K	1	5	24	.3	.8	2
degree	4	14	113	7	8	67	3	4	67	8	9	11
1st.comp	.82	.88	.98	.98	1	1	1	1	1	1	1	1
geo.dist	5	6	16	11	14	16	3	11	19	3	4	7
cluster.cf	.15	.2	.59	.11	.12	.13	.03	.05	.1	.16	.18	.2

Example	size (K)	degree	1st.comp	geo.dist	cluster.cf
math coauthor	253	4	.8	7.57	.15
email	60	1	.9	4.9	*
citation	783	9	*	*	*
Internet	11	6	1.0	3	.03
protein	2	2	.69	6.8	.07
food web	.09	11	1.0	1.9	.2

Community Identification

- $V = V_1 \cup \cdots \cup V_K$
- V_i : communities, $i = 1, \cdots, K$, where K is known.

  V_i highly interiorly, low exteriorly connected.
- Problem: Determine V_j using only A

Approaches to Sub-community Identification: Maximize Modularities

- Newman-Girvan modularity (Phys. Rev. E, 2004) $\mathbf{e} = (e_1, \cdots, e_n)$: $e_i \in \{1, \cdots, K\}$ (community labels)
- The modularity function:

$$Q_N(\mathbf{e}) = \sum_{k=1}^{K} \left(\frac{O_{kk}(\mathbf{e}, A)}{D_+} - \left(\frac{D_k(\mathbf{e})}{D_+} \right)^2 \right),$$

where

$$\begin{aligned} O_{ab}(\mathbf{e}, A) &= \sum_{i,j} A_{ij} \mathbf{1}(e_i = a, e_j = b) \\ &= (\#\text{ of edges between } a \text{ and } b) \quad a \neq b \\ &= 2 \times (\#\text{ of edges between members of } a), \quad a = b \\ D_k(\mathbf{e}) &= \sum_{l=1}^{K} O_{kl}(\mathbf{e}, A) \\ &= \text{sum of degrees of nodes in } k \\ D_+ &= \sum_{k=1}^{K} D_k(\mathbf{e}) = 2 \times (\#\text{ of edges between all nodes }) \end{aligned}$$

Issues

- In principle NP hard
- A relaxation for $K = 2$ leads to method like spectral clustering
- How to compare performance

Stochastic Models

The Erdős-Rényi Model

- Probability distributions on graphs of n vertices.
- P on $\{$Symmetric $n \times n$ matrices of 0's and 1's$\}$.
- E-R (modified): place edges independently with probability c/n ($\binom{n}{2}$ Bernoulli trials).

  $c \approx E(\text{ave degree})$

Qualitative Features of Empirical Graphs vs Qualitative Features of E-R

	E-R	Empirical
Small world	Yes	Yes
Giant component	Yes	Yes
Power-law degree distribution	No	Yes
Communities	No	Yes

Block Models (Holland, Laskey and Leinhardt 1983)

Probability model:

- Community label: $\mathbf{c} = (c_1, \cdots, c_n)$ i.i.d. multinomial $(\pi_1, \cdots, \pi_K) \equiv K$ "communities".
- Relation:

$$\mathbb{P}(A_{ij} = 1 | c_i = a, c_j = b) = P_{ab}.$$

- A_{ij} conditionally independent

$$\mathbb{P}(A_{ij} = 0) = 1 - \sum_{1 \le a,b \le K} \pi_a \pi_b P_{ab}.$$

- $K = 1$: E-R model.

Nonparametric Asymptotic Model for Unlabeled Graphs

Given: P on ∞ graphs

Aldous/Hoover (1983)

$$\mathcal{L}(A_{ij} : i,j \geq 1\} \quad = \quad \mathcal{L}(A_{\pi_i,\pi_j} : i,j \geq 1),$$

for all permutations $\pi \iff$

$$\exists \quad g : [0,1]^4 \to \{0,1\} \text{ such that } A_{ij} \quad = \quad g(\alpha, \xi_i, \xi_j, \eta_{ij}),$$

where

α, ξ_i, η_{ij}, all i, $j \geq i$, i.i.d. $\mathcal{U}(0,1)$, $g(\alpha, u, v, w) = g(\alpha, v, u, w)$,

$\eta_{ij} = \eta_{ji}$.

Ergodic Models

$\mathcal{L}$ is an ergodic probability iff for g with $g(u,v,w) = g(v,u,w)$ $\forall(u,v,w)$,

$$A_{ij} = g(\xi_i, \xi_j, \eta_{ij}).$$

$\mathcal{L}$ is determined by

$$h(u,v) \equiv \mathbb{P}(A_{ij} = 1 | \xi_i = u, \xi_j = v),$$
$$h(u,v) = h(v,u).$$

Notes:

1. K-block models and many other special cases
2. Model (also referred to as threshhold models) also suggested by Diaconis, Janson (2008)
3. More general models (Bollobás, Riordan & Janson (2007))

“Parametrization” of NP Model

- h is not uniquely defined.
- $h\big(\varphi(u), \varphi(v)\big)$, where φ is measure-preserving, gives same model.
- But, h_{CAN} = that $h(\cdot, \cdot)$ in equivalence class such that $P\left[A_{ij} = 1 | \xi_i = z\right] = \int_0^1 h_{\text{CAN}}(z, v) dv \equiv \tau(z)$ with $\tau(\cdot)$ monotone increasing characterizes uniquely.

Asymptotic Approximation

- As given

  Ave. degree $\frac{E(D_+)}{n} = \rho_n(n-1)$

- Broader Approach

$$h_n(u,v) = \rho_n w_n(u,v)$$
$$\rho_n = \mathbb{P}[\text{Edge}]$$
$$w(u,v)dudv = \mathbb{P}\left[\xi_1 \in [u, u+du], \xi_2 \in [v, v+dv] | \text{Edge}\right]$$
$$w_n(u,v) = \min\left\{w(u,v), \rho_n^{-1}\right\}$$
$$\frac{E(D_+)}{n} \equiv \lambda_n = \rho_n(n-1).$$

Approximation

Block model: $\{\rho_n, \pi, W/S\}$

$$\pi \equiv (\pi_1, \ldots, \pi_K)^T$$

$$W_{ab} \equiv \mathbb{P}\left[\xi_1 \in a, \xi_2 \in b | \text{Edge}\right]$$

$$S_{ab} \equiv \frac{\mathbb{P}\left[\text{Edge} | \xi_1 \in a, \xi_2 \in b\right]}{\mathbb{P}[\text{Edge}]}$$

$$W = \pi^D S \pi^D$$

where $\pi^D \equiv \text{diag}(\pi)$

Asymptotic Interpretation h_{CAN}

Suppose $\hat{F}(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}(nD_i/L \le x)$.

Theorem 1

a) (Bollobas et al) If $c = n\rho_n = E(\text{Ave. Degree}) = O(1)$, then $\hat{F} \overset{\text{a.s.}}{\Rightarrow} F$, $Z \sim F$ is d.f. of a mixture of Poisson variable with mixing measure $\tau(\xi)$, $\xi \sim U(0,1)$.

b) If $c \to \infty$, then

$\hat{F}^{-1}(u) \to \tau(u)$, a.e. u

in probability, and therefore

$\hat{F} \Rightarrow \mathcal{L}(\tau(\xi))$

Practical Interpretation

We can replace ξ by $\tau(\xi)$ and think of D_i as measure of "how well i makes friends" (see for example, "Visualizing head-to-tail affinities in large networks", Dyer and Owen 2010).

"Asymptotic" Models: Examples

In spirit of Bollobas et al, Chung and Lu etc

1) Block models

2) $w(u, v) = a(u)a(v)$

$a(u) \propto \int_0^1 w(u, v)dv$

$\therefore$ can take $a(u) = \tau(u) \uparrow$.

3) $w(u, v) = \sum_{j=1}^{p} w_j \phi_j(u)\phi_j(v)$

$|\phi_j| = 1$, $\phi_j \perp \phi_k$, $j \neq k$.

Which Quantitative Properties Can Be Deduced?

1. Small world? Yes. (Bollobas et al for $c = O(1)$, a fortiori in general)
2. Giant component? Yes. with probability $\to 1$ if $c \to \infty$.
3. Degree distribution is approximately power-law? Depends on $\tau(\cdot)$. If $\tau(u) \sim (1-u)^{-\alpha}$, power law.

Community Identification

General Modularity:

- Given Q_n: $K \times K$ positive matrices $\times K$ simplex $\rightarrow \mathbb{R}^+$.
- $Q_n(\mathbf{e}, A) = F_n\left(\frac{O(\mathbf{e},A)}{\mu_n}, \frac{D_+}{\mu_n}, f(\mathbf{e})\right)$.

  $O(\mathbf{e}, A) \equiv ||o_{ab}(\mathbf{e})||$, $\mathbf{f}(\mathbf{e}) \equiv (f_1(\mathbf{e}), \ldots, f_K(\mathbf{e}))^T$, $f_j(\mathbf{e}) \equiv \frac{n_j}{n}$.

  $\hat{\mathbf{c}} \equiv \arg\max Q_n(\mathbf{e}, A)$.

  $\mu_n = E(D_+) = (n-1)\lambda_n$.
- NG: $F_n \equiv F$.

Profile Likelihood

- $\rho > 0$

$$F(M, r, \mathbf{t}) = \sum_{a,b} t_a t_b \tau\left(\frac{\rho M_{ab}}{t_a t_b}\right),$$

$$\tau(x) \equiv x \log x + (1-x)\log(1-x).$$

- $\rho \to 0$

$$F(M, \mathbf{t}) = \sum_{a,b} t_a t_b \sigma\left(\frac{M_{ab}}{t_a t_b}\right),$$

$$\sigma(x) = x \log x - x.$$

Conditions

C1: a) The matrix S has no two rows equal and all elements > 0.

b) $\pi_i > 0$, $i = 1, \ldots, K$. (No two communities have same connection probabilities with others.)

C2: $\mathcal{M} \equiv \{R : R_{ab} \geq 0, \text{ all } a, b, R^T \mathbf{1} = \pi\}$.

$Q(R) \equiv F(RSR^T, 1, R\mathbf{1})$.

$F : \mathcal{M} \times \mathbb{R}^+ \times \mathcal{S} \mapsto \mathbb{R}$, $\mathcal{S} \equiv$simplex, where $\mathbf{1} \equiv (1, 1, \ldots, 1)^T$.

Then $Q(R)$ is uniquely maximized over $\mathcal{M}$ by $R = \pi^D \equiv \text{diag}(\pi)$ for all (π, S) in an open neighborhood Θ of (π_0, S_0). (Unique population maximization)

C3: a) F is Lipschitz in $\mathcal{M}$ in all its arguments.

b) On Θ, F has continuous second directional derivatives and $\frac{\partial Q(\pi^D)}{\partial r_{ab}} < 0$, all $(\pi, S) \in \Theta$. (Local maximization)

Global Consistency

Theorem 1

If C1–3 hold and $\frac{c_n}{\log n} \to \infty$, then

$$\limsup_n c_n^{-1} \log \mathbb{P}\left[\hat{\mathbf{c}} \neq \mathbf{c}\right] \leq -s_Q, \text{ with } s_Q > 0.$$

Extension to $F_n \approx F$ requires simple condition. See also Snijders and Nowicki (1997) J. of Classification.

Corollary

Under the given conditions if

$$\hat{\pi}_a = \frac{1}{n}\sum_{i=1}^{n} 1(\hat{c}_i = a) \equiv \frac{\hat{n}_a}{n},$$
$$\hat{W} = \frac{O(\hat{c}, A)}{D_+},$$

then

$$\sqrt{n}(\hat{\pi} - \pi) \Rightarrow \mathcal{N}(\mathbf{0}, \pi^D - \pi\pi^T),$$
$$\sqrt{n}(\hat{W} - W) \Rightarrow \mathcal{N}\left(\mathbf{0}, \Sigma(\pi, W)\right).$$

These are efficient.

Properties of N-G Modularity

1) NG satisfies C2, C3 if $\mathcal{E}$ has all diagonal entries positive and all nondiagonal entries negative.

2) NG consistency may fail even though $W_{aa} > \sum_{b \neq a} W_{ab}$, $\forall a$.

Degree distributions

Definition

$D_i^\ell \equiv \ell$ degree of i is the number of independent paths of length $\leq \ell$ starting at i.

The Operator

Corresponding to $w_{\mathrm{CAN}} \in L_2(0,1)$ there is operator:

$T : L_2(0,1) \to L_2(0,1)$

$Tf(\cdot) = \int_0^1 w(\cdot, v)dv$

T- Hermitian

Note: $\tau(\cdot) = T(\mathbf{1})(\cdot)$.

Theorem 2

Let $\hat{F}_\ell$ be the empirical distribution of $(D_i, D_i^{(2)}, \ldots, D_i^{(\ell)})$ and F be the joint distribution of $(T(\mathbf{1})(\xi), T^2(\mathbf{1})(\xi), \ldots, T^\ell(\mathbf{1})(\xi))$ where ξ has a $U(0,1)$ distribution.

Theorem 2

If $\rho = c/n$,

1. If c is bounded, then $\hat{F} \Rightarrow G$ in probability, where G is the distribution of a set of independent Poisson variables with parameters $T(\mathbf{1})(\xi), T^2(\mathbf{1})(\xi), \cdots, T^l(\mathbf{1})(\xi)$ given $\xi \sim U(0,1)$;
2. If $c \to \infty$, then $\hat{F} \Rightarrow F$ in probability, where F is the distribution of $(T(\mathbf{1})(\xi), T^2(\mathbf{1})(\xi), \ldots, T^l(\mathbf{1})(\xi))$

Identifiability of NP Model

Theorem 3

The joint distribution $(T(1)(\xi), T^2(1)(\xi), ..., T^m(1)(\xi), ...)$ where $\xi \sim U(0,1)$ determines P

Idea of proof: identify the eigen-structure of T.

Theorem 4

If T corresponds to a K-block model, then, the marginal distributions,

$$\left\{T^k(1)(\xi) : k = 1, ..., K\right\}$$

determine (π, W) uniquely provided that the vectors π, $W\pi$, ..., $W^{K-1}\pi$ are linearly independent.

Three methods of estimation (potentially) yielding $\sqrt{n}$ consistent estimates of block and other parametric submodels

Method of "Moments"

(k, ℓ)-wheel

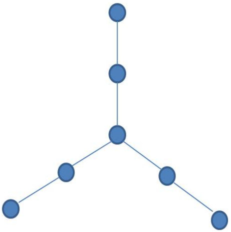
i) A "hub" vertex

ii) l spokes from hub

iii) Each spoke has k connected vertices.

Total # of vertices (order): $k\ell + 1$. Total # of edges (size): $k\ell$.

Eg: a (2,3)-wheel



Definition

Notation:

(i) If $R \subset F_n \equiv \{(i,j) : 1 \leq i < j \leq n\}$

$V(R) \equiv \{i : (i,j) \text{ or } (j,i) \in R, \text{ some } j\}$

$E(R) = R$

A graph G and an edge set are identified if $V(G) = V(R)$ and $E(G) = R$.

(ii) If $R_1, R_2 \subset F_n$, $R_1 \sim R_2$ (isomorphism) iff $|V(R_1)| = |V(R_2)|$ and there exists $\pi : V(R_1) \to V(R_2)$, 1-1, onto, such that

$$E(R_2) = \{(\pi(i), \pi(j)) : (i,j) \in R_1, \text{ or } (j,i) \in R_1, \ \pi(i) < \pi(j)\}.$$

Definitions

Given: $G \sim P$, $G \subset F_n$

For $R \subset F_n$, $\bar{R} \equiv$complement of R in G,

$$P(R) = P[A_{ij} = 1,\ (i,j) \in R,\ A_{ij} = 0,\ (i,j) \in \bar{R}]$$

$$Q(R) \equiv P(A_{ij} = 1,\ (i,j) \in R).$$

Lemma

G generated according to h on F_n.

(1) $P(R) =$

$$E\left[\prod\{h(\xi_i,\xi_j) : (i,j \in R)\}\prod\{(1-h(\xi_i,\xi_j)) : (i,j) \in \bar{R}\}\right]$$

(2) $$P(R) = Q(R) - \sum\{Q(R \cup (i,j)) : (i,j) \in \bar{R}\}$$
$$+\sum\{Q(R \cup \{(i,j),(k,l)\}) : (i,j) \neq (k,l) \in \bar{R}\}$$
$$\cdots \pm Q(G)$$

(3) $Q(R) = \sum\{P(S) : \ S \supset R\}$.

$\sqrt{n}$ Consistency/Asymptotic Normality of "Moments"

Theorem 5

For $R \subset F_n$, $|V(R)| = p$, G generated according to P, let

$\hat{P}(R) = \frac{1}{\binom{n}{p} N(R)} \sum \mathbf{1}(S \sim R : \; S \subset G)$,

$N(R) \equiv |\{S \subset F_n : \; S \sim R\}|$,

$\hat{Q}(R) \equiv \sum \{\hat{P}(S) : S \supset R\}$.

Then

$$\sqrt{n}(\hat{Q}(R) - Q(R)) \Rightarrow N(0, \sigma^2(R, P)).$$

Multivariate normality holds as well.

Extensions

- $|R| = p$ fixed
- $\rho \to 0,\ L \equiv \sum_{i,j} A_{ij}$
- $\tilde{Q}(R) \equiv \rho^{-p} Q(R) \to E\left(\Pi\left\{w(\xi_i, \xi_j) : (i,j) \in R\right\}\right)$
- $\hat{\tilde{Q}}(R) \equiv \left(\frac{L}{n^2}\right)^{-p} \hat{Q}(R)$
- Conclusion of Theorem holds for $\hat{\tilde{Q}}$, $\tilde{Q}$ if $n^2\rho \to \infty$.

Connection With Wheels

Lemma 1

Let G be a random graph generated according to P, $|V(G)| = k\ell + 1$. Then if R is a (k, ℓ)-wheel,

$$Q(R) = E[T^k(1)(\xi_1)]^\ell$$

Fitting by degree distributions

Theorem 2 suggests that

- *For block models*: Do maximum likelihood for l degree distributions $l = 1, \cdots, K$, treating them as independent each a mixture of Poisson with appropriate parameter;
- *In general*, $T = T_\theta, \theta \to T_\theta$ smooth, Fit joint degree distribution as a sample from a mixture of Poisson as in Theorem 2.
- *Conjecture*: Leads to $\sqrt{n}$ consistent estimates.

Pseudo likelihood

Chen-Levina-Bickel (2010) related to Newman-Leicht (2007)

K block models

- Block $i = 1, \cdots, n$ into K blocks arbitrarily, $m = n/K$,

  $B_{10} = \{1, \cdots, m\}, \cdots, B_{K0} = \{(K-1)m + 1, \cdots, n\}$
- Complete data model

  $c_1, \cdots, c_n$ i.i.d. $\mathbb{P}(c_i = a) = \pi_a$, $1 \leq a \leq K$
- Parameter P_{ab}, $1 \leq a, b \leq K$
- A_{ij} independent given c, $\mathbb{P}(A_{ij} = 1|c) = P_{c_i d_j}$ where $d_j = a$ iff $j \in B_{a0}$
- $\text{plhd}(c, A, \pi, P) = \prod_{i=1}^{n} \pi_{c_i} \prod_{i \neq j} P_{c_i d_j}^{A_{ij}} (1 - P_{c_i d_j})^{1-A_{ij}}$

Pseudo likelihood - algorithm

A1. Compute MLE of (π, P), $(\hat{\pi}, \hat{P})$

A2. $\mathbb{P}(c_i = a|\hat{\pi}, \hat{P})$ for any i, a

A3.

$$\hat{c}_i = \arg\max_a \mathbb{P}(c_i = a|\hat{\pi}, \hat{P})$$

Let $B_{a1} = \{i : \hat{c}_i = a\}$, $a = 1, \cdots, K$

A4. Return to A1 with $B_{\cdot 0} = B_{\cdot 1}$.

Simulation

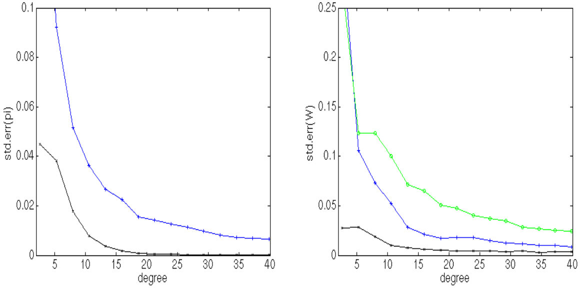


Figure: Estimation of π (left) and W (right) ($K = 2, n = 1000$): pseudo likelihood (black), 1st degree + moment (blue), 1st, 2nd degrees (green)

Statistical Questions For Which These Results Can Be Used

i) Checking "nonparametrically" with p moments whether 2 graphs are same (permutation tests used in social science literature for "block models", e.g., Wasserman and Faust, 1994).

ii) Link prediction: predicting relations to unobserved vertices on the basis of an observed graph.

iii) Model selection for hierarchies (block models).

iv) Error bars on descriptive statistics.

Real Data: Zachary's Karate Club, $K = 2$

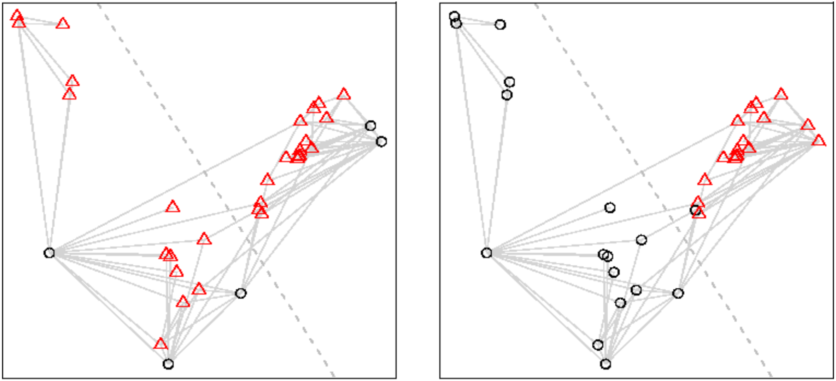


Figure: Left: profile likelihood. Right: Newman-Girvan

Real Data: Zachary's Karate Club, $K = 4$

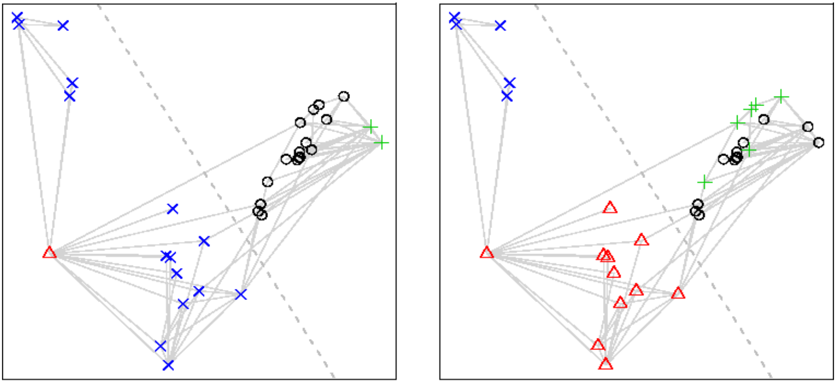


Figure: Left: profile likelihood. Right: Newman-Girvan

Real Data: Private Branch Exchange

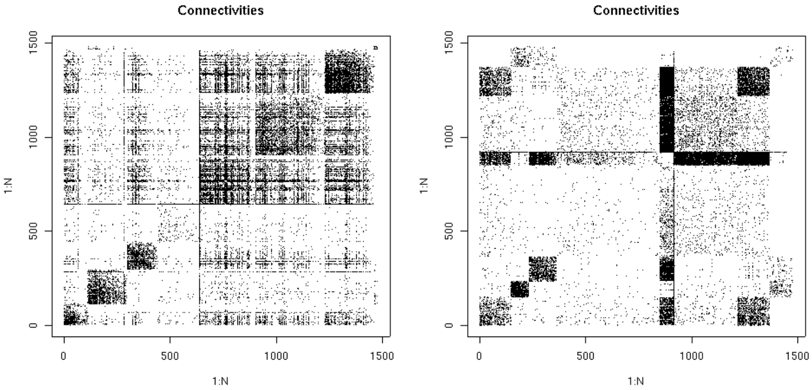


Figure: Different communities formed by NG and profile likelihood

Discussion

Extensions which are theoretically easy, in practice not so

i) Directed graphs

ii) Covariates (edge or vertex information)

Some extensions in progress

iii) Computational issues

iv) Relation of these models to dynamic ones

etc. etc.