Random Networks, Graphical Models and Exchangeability

Alessandro Rinaldo
Carnegie Mellon University
joint work with Steffen Lauritzen and Kayvan Sadeghi

October 4, 2015
AMS Central Fall Sectional Meeting
Loyola University

Special Session on Algebraic Statistics and its Interactions with Combinatorics, Computation, and Network Science

Outline

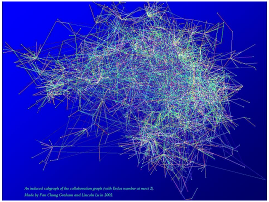
- Exchangeability of (infinite) networks.
- A finite deFinetti theorem and the dissociated property.
- Exchangeable and extendable finite networks are (mixtures of) bidirected graphical models.

Statistical Network (Random Graph) Analysis

- Let $\mathcal{L}_n$ be the set of simple labeled graphs on n nodes: $|\mathcal{L}_n| = 2^{\binom{n}{2}}$.

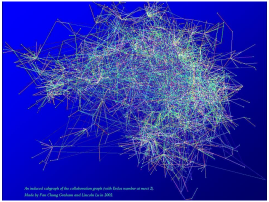
Statistical Network (Random Graph) Analysis

- Let $\mathcal{L}_n$ be the set of simple labeled graphs on n nodes: $|\mathcal{L}_n| = 2^{\binom{n}{2}}$.
- The nodes represent agents in some population of interest and the edges encode the relationships among them.



Statistical Network (Random Graph) Analysis

- Let $\mathcal{L}_n$ be the set of simple labeled graphs on n nodes: $|\mathcal{L}_n| = 2^{\binom{n}{2}}$.
- The nodes represent agents in some population of interest and the edges encode the relationships among them.



Statistical Network Analysis

Pose and estimate probability distributions on $\mathcal{L}_n$
by modeling the joint occurrence of the $\binom{n}{2}$ random edges.

Motivation: asymptotics of networks

- Let $\mathcal{L} = \bigcup_n \mathcal{L}_n$ be the set of all finite (labeled, simple) graphs. A statistical model for $\mathcal{L}$ is a sequence $\{p_n\}_{n \in \mathbb{N}}$ of probability distributions, where p_n is a probability distribution on $\mathcal{L}_n$.
- For $n < m$, let p_m^n denote the marginal of p_m over $\mathcal{L}_n$.

Consistency and Extendability

A statistical model $\{p_n\}_{n \in \mathbb{N}}$ on $\mathcal{L}$ is consistent when, for any pair $n < m$,

$$p_n = p_m^n. \qquad (1)$$

A probability distribution p_n on $\mathcal{L}_n$ is extendable when (1) holds $\forall m > n$.

Motivation: asymptotics of networks

- Let $\mathcal{L} = \bigcup_n \mathcal{L}$, be the set of all finite (labeled, simple) graphs. A statistical model for $\mathcal{L}$ is a sequence $\{p_n\}_{n \in \mathbb{N}}$ of probability distributions, where p_n is a probability distribution on $\mathcal{L}_n$.
- For $n < m$, let p_m^n denote the marginal of p_m over $\mathcal{L}_n$.

Consistency and Extendability

A statistical model $\{p_n\}_{n \in \mathbb{N}}$ on $\mathcal{L}$ is consistent when, for any pair $n < m$,

$$p_n = p_m^n. \quad (1)$$

A probability distribution p_n on $\mathcal{L}_n$ is extendable when (1) holds $\forall m > n$.

Most network models are not consistent!

Consistency via Exchangeability

- Let $\mathcal{L}_\infty$ be the set of (countably) infinite lableled, simple graphs.
  Every probability distrbution on $\mathcal{L}_\infty$ trivially specifies one consistent model!
  We impose one further restriction...

Consistency via Exchangeability

- Let $\mathcal{L}_\infty$ be the set of (countably) infinite lableled, simple graphs.
  Every probability distrbution on $\mathcal{L}_\infty$ trivially specifies one consistent model!
  We impose one further restriction...

Exchangeability

A probability distribution on $\mathcal{L}_\infty$ is exchangeable when all finite isomorphic graphs have the same probabilities.

Consistency via Exchangeability

- Let $\mathcal{L}_\infty$ be the set of (countably) infinite lableled, simple graphs.
  Every probability distrbution on $\mathcal{L}_\infty$ trivially specifies one consistent model!
  We impose one further restriction...

Exchangeability

A probability distribution on $\mathcal{L}_\infty$ is exchangeable when all finite isomorphic graphs have the same probabilities.

- Exchangeability is a most basic form of invariance, suitable to describe the "shape" of networks (large scale property).
- Labeled vs unlabeled. The exchangeability assumption is equivalent to define models on $\mathcal{U}_n$, the set of unlabaled graohs on n nodes, for all n.

Exchangeability and graphons

Analytic representation of exchangeable distributions

- The set of exchangeable distributions, with the topology of weak convergence, is a (Bauer) simplex. Denote its extreme points with $\mathcal{E}_\infty$.
- $p_\infty \in \mathcal{E}_\infty$ if and only if, for every n and $G \in \mathcal{L}_n$

$$p^n_\infty(G) = \int_{[0,1]^n} \prod_{(i,j) \in E(G)} f(z_i, z_j) \prod_{(i,j) \notin E(G)} (1 - f(z_i, z_j)) dz_1 \ldots z_n,$$

  where $f \colon [0,1]^2 \to [0,1]$ is a (measurable) symmetric function, called a graphon.
- Graphons are unique up to measure preserving transformations of $[0,1]$.

Exchangeability and graphons

Analytic representation of exchangeable distributions

- The set of exchangeable distributions, with the topology of weak convergence, is a (Bauer) simplex. Denote its extreme points with $\mathcal{E}_\infty$.
- $p_\infty \in \mathcal{E}_\infty$ if and only if, for every n and $G \in \mathcal{L}_n$

$$p^n_\infty(G) = \int_{[0,1]^n} \prod_{(i,j) \in E(G)} f(z_i, z_j) \prod_{(i,j) \notin E(G)} (1 - f(z_i, z_j)) dz_1 \ldots z_n,$$

  where $f \colon [0,1]^2 \to [0,1]$ is a (measurable) symmetric function, called a graphon.
- Graphons are unique up to measure preserving transformations of $[0,1]$.

- Vast literature: Aldous, Hoover, Kallenberg, Diaconis and Freedman, Chayes, Borgs and Lovász, etc ect...
- Key point: only the finite marginals of $p_\infty \in \mathcal{E}_\infty$ can be realized. General exchangeable models are mixtures of such distributions.

Graphons and homomorphism densities

- For $G \in \mathcal{L}_n$ and $H \in \mathcal{L}_k$ with $k \leq n$, the density homomorphism of H in G is

$$t(H, G) = \frac{|\mathrm{hom}(H, G)|}{n^k}.$$

Graphons and homomorphism densities

- For $G \in \mathcal{L}_n$ and $H \in \mathcal{L}_k$ with $k \leq n$, the density homomorphism of H in G is

$$t(H, G) = \frac{|\hom(H, G)|}{n^k}.$$

Convergence of graph sequences = convergence of marginal probabilities

- A sequence $\{G_n\}_{n \in \mathbb{N}}$ converges if and only if, for some graphon f and each $H \in \mathcal{L}$ with k nodes,

$$\lim_{n \to \infty} t(H, G_n) = \int_{[0,1]^k} \prod_{(i,j) \in E(H)} f(z_i, z_j)\, dz_1 \ldots z_k = \mathbb{P}\left(H \subseteq G'\right),$$

G' a random graph distributed like the p^k_∞, $p_\infty \in \mathcal{E}_\infty$ defined by f.
- The sequence $\{t(H, f)\}_{H \in \mathcal{L}}$ of density homomorphisms uniquely specifies p_∞.

Finite Exchangeability

- But real networks are finite! So, what can be said about the set $\mathcal{P}_n$ of exchangeable distribution on $\mathcal{L}_n$?

Finite Exchangeability

- But real networks are finite! So, what can be said about the set $\mathcal{P}_n$ of exchangeable distribution on $\mathcal{L}_n$?

Finite exchangeability does not yield consistent models

Finite exchangeable probability distributions marginalize to but need not be extendable to exchangeable distributions.

Finite Exchangeability

- But real networks are finite! So, what can be said about the set $\mathcal{P}_n$ of exchangeable distribution on $\mathcal{L}_n$?

Finite exchangeability does not yield consistent models

Finite exchangeable probability distributions marginalize to but need not be extendable to exchangeable distributions.

Our goal

We would like to characterize the distributions in $\mathcal{P}_n$ that are extendable. We seek to establish a parametric (finite dimensional) representation of all the distributions $\{p^n_\infty, p_\infty \in \mathcal{E}_\infty\}$.

The Möius parametrization

- It turns out it is convenient to work with maginal instead of joint probabilities.

Möbius parameters

For any $p_n \in \mathcal{P}_n$, let z_n the vector with entries indexed by subgraphs H of K_n without isolated nodes of the form

$$z_n(H) = \mathbb{P}\left(H \subseteq G_n\right),$$

where G_n is the random graph with distribution p_n. In particular, $z_n(\emptyset) = 1$.

The Möius parametrization

- It turns out it is convenient to work with maginal instead of joint probabilities.

Möbius parameters

For any $p_n \in \mathcal{P}_n$, let z_n the vector with entries indexed by subgraphs H of K_n without isolated nodes of the form

$$z_n(H) = \mathbb{P}\left(H \subseteq G_n\right),$$

where G_n is the random graph with distribution p_n. In particular, $z_n(\emptyset) = 1$.

- Invertible linear transformation:

$$p_n(G) = \sum_{H:E(H)\supseteq E(G)} (-1)^{E(H)-E(G)} z_n(H), \quad \forall G \in \mathcal{L}_n.$$

- By exchangeability, $z_n(H) = z_n(H')$ if H and H are isomorphic.

A finite deFinetti Theorem

- We can describe now the relationships among Möbius parameters of consistent finitely exchangeable distributions.

A deFinetti's theorem for finitely exchageable graphs

Assume $m > n$. Let p_m an exchangeable distribution on $\mathcal{L}_m$ and z_m^n the Möbius parameters corresponding to p_m^n. Then,

$$\max_H \left| z_m^n(H) - \sum_{G \in \mathcal{G}_m} t(H,G) p_m(G) \right| \leq 1 - \frac{(m)_n}{m^n}.$$

A similar guarantee holds for the p_n's.

- See also Matúš for more general statements.

The dissociated property

Corollary (The dissociated property)

- *If $p_n \in \mathcal{P}_n$ is extendable to an (infinite) exchangeable distribution in $\mathcal{E}_\infty$, then it satisfies the dissociated property:*

$$z_\infty^n(H) = z_n(H) = z_n(H_1) z_n(H_2)$$

for all subgraphs $H = H_1 \uplus H_2$ of K_n without isolated nodes.

- *Extendable distributions in $\mathcal{P}_n$ are mixtures of dissociated distributions in $\mathcal{P}_n$.*
- *A distribution p_∞ on $\mathcal{L}_\infty$ is in $\mathcal{E}_\infty$ if and only if p_∞^n satisfies the dissociated property for all n.*

The dissociated property

Corollary (The dissociated property)

- *If $p_n \in \mathcal{P}_n$ is extendable to an (infinite) exchangeable distribution in $\mathcal{E}_\infty$, then it satisfies the dissociated property:*

$$z^n_\infty(H) = z_n(H) = z_n(H_1) z_n(H_2)$$

for all subgraphs $H = H_1 \uplus H_2$ of K_n without isolated nodes.

- *Extendable distributions in $\mathcal{P}_n$ are mixtures of dissociated distributions in $\mathcal{P}_n$.*
- *A distribution p_∞ on $\mathcal{L}_\infty$ is in $\mathcal{E}_\infty$ if and only if p^n_∞ satisfies the dissociated property for all n.*

- Thus, finitely exchangeable distribution must be dissociated in order to be extendable to extremal distributions in $\mathcal{E}_\infty$.
- Result is not new, but derivation via finite exchangeability is.
- ...so what does dissociated distribution in $\mathcal{P}_n$ looks like?

Bidirected Graphical Models for Binary Data

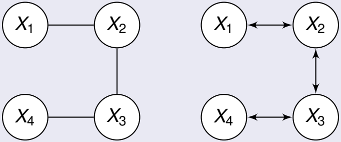
- Graphical models with bidirected edges, where the nodes of the graph represents the variables and lack of (bidirected) edges among nodes signify marginal independence among the corresponding variables.
- See Richardson (2003), Drton and Richardson (2008) and Roverato, Luparelli and LaRocca (2013).

Global Markov property for bidirected (marginal) graphical models

$A \perp\!\!\!\perp B \mid C$ when every path between A and B has a node outside $A \cup B \cup C$. In particular, C may be empty.

Bidirected Graphical Models for Binary Data

Example (Drton and Richardson, 2008)



In the undirected graph (left), the global Markov property expresses, e.g., that

$$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\},$$

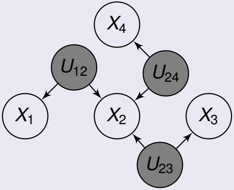
whereas in the bidirected graph (right) the global Markov property expresses, e.g., that

$$X_1 \perp\!\!\!\perp X_4 \quad \{X_1, X_2\} \perp\!\!\!\perp X_4 \quad \text{and} \quad X_1 \perp\!\!\!\perp X_4 \mid X_3.$$

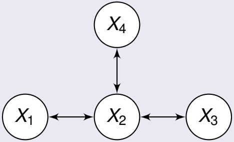
Bidirected Graphical Models for Binary Data

- Bidirected Markov models arise, e.g., as marginals of directed Markov models with unobserved variables.

Example (by S. Lauritzen)



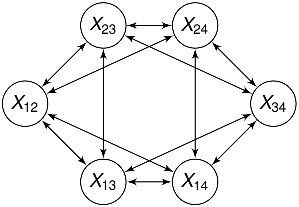
In the graph above, the marginal distribution of (X_1, X_2, X_3) will be bidirected Markov w.r.t. the graph



The canonical model for exchangeable and extendable networks

Dissociated property and bidirected graphical models

A distribution on $\mathcal{L}_n$ is dissociated if and only if it is Markov with respect to the bidirected line graph of K_n.



Contrast this with the Markov graphs of Frank and Strauss (1986) which are Markov w.r.t. the undirected line graph.

The benfits of Möbius parametrization

Using the Möbius parameters are especially convenient because

- are marginalizable: for any $m > n$

$$z_m^n(H) = z_n(H)$$

  for any ubgraph H of K_n without isolated nodes.

- expresses the bidirected Markov property in a simple way:

(From Drton and Richardson, 2008)

A distribution $p_n \in \mathcal{P}_n$ is Markov with respect to the bidirected line graph of K_n if and only if for any $H = H_1 \uplus H_2 \uplus \ldots \uplus H_l \in \mathcal{L}_k$ without isolated nodes,

$$z_n(H) = z_n(H_1) \times \cdots \times z_n(H_l).$$

The Möbius parametrization

Polynomial parametrization

If a probability p_n in $\mathcal{P}_n$ is extendable to some $p_\infty \in \mathcal{E}_\infty$, then

$$p_n(G) = \sum_{U \in \mathcal{U}_n : E(G) \subseteq E(U)} (-1)^{E(U)-E(G)} r(G,U) \prod_{C \in \mathcal{C}(U)} z_n(C), \quad G \in \mathcal{L}_n,$$

where $\mathcal{D}(U)$ denotes the maximal connected components of U and $r(G,U)$ are the number of graphs in $\mathcal{L}_n$ that contain G as a subgraph and are isomorphic to $U \in \mathcal{U}_n$.

- This defines a smooth parametrization, described by a smooth manifold inside $\mathcal{P}_n$ specified by polynomial equations. Its dimension is the number of connected subgraphs of all unlabeled graphs on n nodes.

The curved exponential family parametrization

Exponential parametrization

If a probability p_n in $\mathcal{P}_n$ is extendable to some $p_\infty \in \mathcal{E}_\infty$, then

$$p_n(G;\nu) = \exp\left\{ \sum_{U \in \mathcal{U}_n} \nu_U s(U,G) - \psi(\nu) \right\}, \quad G \in \mathcal{L}_n, \quad \nu \in V \subset \mathbb{R}^{|\mathcal{U}_n|-1},$$

where $s(U,G)$ is the number of non-empty subgraphs of G isomorphic to $U \in \mathcal{U}_n$ and ψ a normalizing constant.

The curved exponential family parametrization

Exponential parametrization

If a probability p_n in $\mathcal{P}_n$ is extendable to some $p_\infty \in \mathcal{E}_\infty$, then

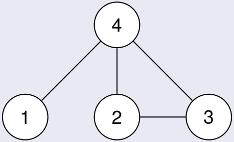
$$p_n(G;\nu) = \exp\left\{ \sum_{U \in \mathcal{U}_n} \nu_U s(U,G) - \psi(\nu) \right\}, \quad G \in \mathcal{L}_n, \quad \nu \in V \subset \mathbb{R}^{|\mathcal{U}_n|-1},$$

where $s(U,G)$ is the number of non-empty subgraphs of G isomorphic to $U \in \mathcal{U}_n$ and ψ a normalizing constant.

- Duality: the mean value parameters are (sums of) the Möbius parameters.
- The natural parameters ν are not free to vary, as they need to enforce the dissociated property. These are defined implicitly!

Example

Suppose we observe the following graph G:



- Under the assumed bidirected model, the likelihood under the Mobiüs parametrization is

$$p(G) = z_{\trianglerighteq} - 2z_{\diamond\!\!\!|} + z_{K_4}.$$

and under the curved exponential model is

$$P(G;\nu) = \exp\{4\nu_{-} + 5\nu_{\wedge} + \nu_{\|} + \nu_{\triangle} + \nu_{\perp} + 2\nu_{\sqcap} + \nu_{\trianglerighteq} - \psi(\nu)\}.$$

Maximum Likelihood Estimation

- Given a observation $G \in \mathcal{L}_n$, the maximum likelihood estimator of p_n is the dissociated point in $\mathcal{P}_n$ with positive coordinates that maximizes the likelihood of G.

Maximum Likelihood Estimation

- Given a observation $G \in \mathcal{L}_n$, the maximum likelihood estimator of p_n is the dissociated point in $\mathcal{P}_n$ with positive coordinates that maximizes the likelihood of G.

Example 1

- For the previous network, the MLE is

$$\hat{z}_{-} = 1/2,\ \hat{z}_{\wedge} = 5/16, \hat{z}_{\|} = 1/4\ \hat{z}_{\triangle} = 3/16,\ \hat{z}_{\perp} = 3/16,\ \hat{z}_{\sqcap} = 1/8,\ \hat{z}_{\trianglerighteq} = 1/16,$$

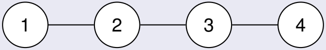
This estimate represents a mixture of the uniform distribution of all networks isomorphic to G (there are 12), and the empty network, with weights 3/4 and 1/16, respectively.

- The MLE does not exist! In fact, we conjecture it never exists.

Maximum Likelihood Estimation

Example 2

- When the observed graph G is



- the likelihood function is maximized for any value of λ satisfying $0 \leq \lambda \leq 1/16$ with

$$\hat{z}_{-} = 1/2,\ \hat{z}_{\wedge} = 3/16,\ \hat{z}_{\|} = 1/4,\ \hat{z}_{\triangle} = 1/16 - \lambda,\ \hat{z}_{\perp} = \lambda,\ \hat{z}_{\sqcap} = 1/16,$$

  and all other z's equal to zero. This corresponds to a random network that has probability 3/4 of being isomorphic to G (12 cases) and the remaining probability mass of 1/4 is distributed arbitrarily between a triangle plus an isolated point (4 cases), and a 3-star (4 cases).

- The MLE does not exist and is not unique!

An open problem

- Does a dissociated exchenagble distribution on $\mathcal{L}_n$ always extend to some $p_\infty \in \mathcal{E}_\infty$?
- No! Example 1 shows this not the case. So the dissociated property is only necessary for extendability.

An open problem

- Does a dissociated exchenagble distribution on $\mathcal{L}_n$ always extend to some $p_\infty \in \mathcal{E}_\infty$?
- No! Example 1 shows this not the case. So the dissociated property is only necessary for extendability.

Open problem

Let $\mathcal{EP}_n \subset \mathcal{P}_n$ the set of exchenagble and extendable distributions on $\mathcal{L}_n$ and $\mathcal{DP}_n \subset \mathcal{P}_n$ the distributions that are exchangeable and dissociated. Then

$$\mathcal{EP}_n \subset \mathcal{DP}_n.$$

What does $\mathcal{DP}_n \setminus \mathcal{EP}_n$ look like?

More open problems...

- What are the algebraic and geometric properties of the proposed bidirected model for networks?

- How do we carry out maximum likelihood estimation in this curved exponential family setting?