The Ohio State University

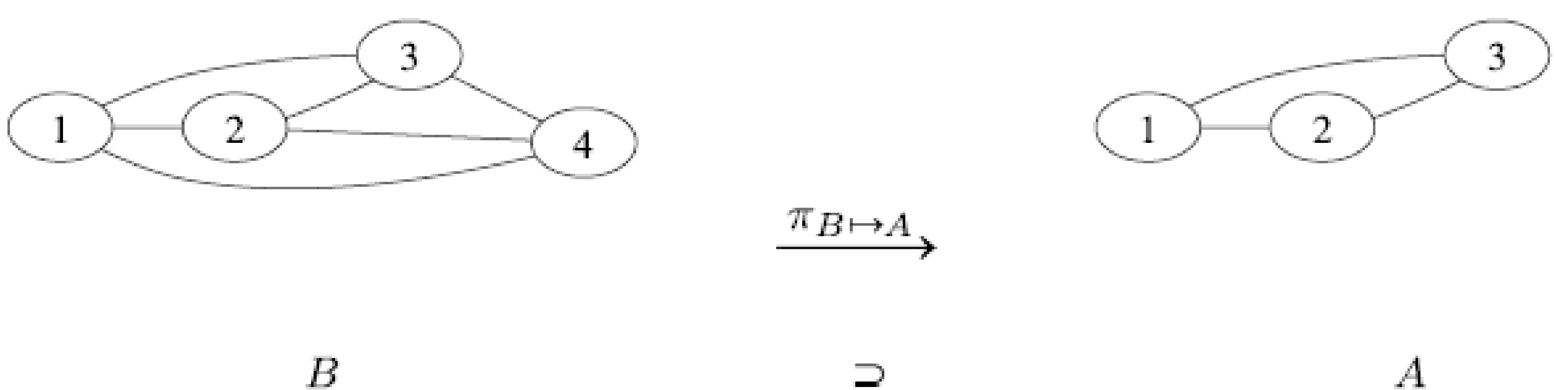
Consistency Under Sampling of Exponential Random Graph Models

Cosma Shalizi and Alessandro Rinaldo

Summary by: Elly Kaizar

Remember ERGMs (Exponential Random Graph Models)

- Exponential family models
- Sufficient statistics count the number of edges, triangles, k-cliques, etc…
- Only observe sub-networks
- Estimate from sub-network, extrapolate to “whole” network



Practical questions

- 1. Are the MLEs from the sub-graph consistent?
  - (What does consistency mean in this setting?)
- 0. "Logically Prior" question: is it "probabilistically consistent to apply the same ERGM, with the same parameters, both to the whole network and its sub-networks?"

Definition 1: Projective

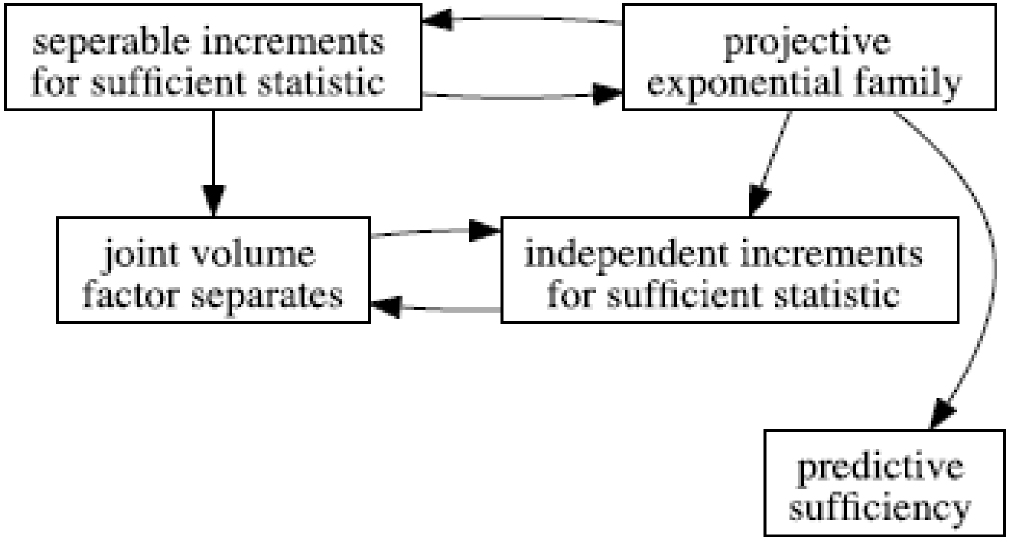
- The family $\{\mathcal{P}_{A,\Theta}\}_{A\in\mathcal{A}}$ is projective when $A \subset B$ implies that $\mathbb{P}_{A,\theta}$ can be recovered by marginalization [over $B\backslash\mathrm{A}$?] over $\mathbb{P}_{B,\theta}$, for all θ.

Definition 2: Separable Increments

- The sufficient statistics of the family $\{\mathcal{P}_{A,\Theta}\}_{A\in\mathcal{A}}$ have separable increments when, for each $A \subset B, x \in \mathcal{X}_A$, the range of possible increments δ is the same for all x, and the conditional volume factor is constant in x, that is, $v_{B\backslash\mathrm{A}|A}(\delta, x) = v_{B\backslash\mathrm{A}}(\delta)$.
- [The number {of super-graphs in B consistent with a subgraph in A and for which the "extra" statistic (increment) = δ} does not depend on the subgraph in A?]
- Note: this "depends only on the functional forms of the sufficient statistics, … and not on the model parameters."

Projective Theorems/Propositions:

- The exponential family $\{\mathcal{P}_{A,\Theta}\}_{A\in\mathcal{A}}$ is projective if and only if the sufficient statistics $\{T_A\}_{A\in\mathcal{A}}$ have separable increments.
- (Predictive Sufficiency) In a projective exponential family, the distribution of $X_{\mathrm{B}\setminus\mathrm{A}}$ conditional on X_A depends on the data only through $T_{B\setminus\mathrm{A}}$.



Consistency Theorems

- Suppose that model $\mathbb{P}_\theta$ (infinite-dimensional A) is projective, and that the log partition function obeys equation (10) for each $A \in \mathcal{A}$. Then the MLE exists and is strongly consistent.
  - wrt a growing sequence of sets A s.t. a function of the measure of the size, $r_{|A|} \to \infty$
  - (10) $\log \mathrm{z}_{\mathrm{A}}(\theta) \equiv a_A(\theta) = r_{|A|} a(\theta)$
  - If instead $\lim_{r_{|A|} \to \infty} \left(\frac{a_A(\theta)}{r_{|A|}} \right) = a(\theta)$, then convergence in probability still holds (theorem 4).
  - If (10) holds only for some θ, then MLE may be consistent only for some θ.
  - All components of T_A must be scaled by the same factor $r_{|A|}$.

Application: ERGM

- Remember that sufficient statistics typically are vectors of counts of edges, triangles, cliques, k-stars, etc.
- Dyadic independence ($t(X) = \sum_{i=1}^{n} \sum_{j<i} t_{ij}(X_{ij}, X_{ji})$) implies projectability via Theorem 1. This applies to the β-model.
- Homophily (friend of a friend is a friend) causes problems…

Application: ERGM (p. 16)

- “Sadly, no statistic which counts triangles, or larger motifs, can have the nice additive form of dyad counts, no matter how we decompose the network. Take, for instance, triangles. Any given edge among the first n nodes *could* be part of a triangle, depending on ties to the next node. Thus to determine the number of triangles among the first n+1 nodes, we need much more information about the sub-graph of the first n nodes than just the number of triangles among them. Indeed, we can go further. The range of possible increments to the number of triangles changes with the number of existing triangles. This is quite incomparable with separable increments, so, by (1), the parameters cannot be projective.”
- Question: Can we look at this problem so one-dimensionally?

Application: ERGM (p.16-17)

- “While these ERGMs are not projective, some of them may … still satisfy equation (14) [and thus the MLEs converge in probability]. For instance, in models where T has two elements, the number of edges and the (normalized) number of triangles or of 2-stars, the log partition function is known to scale like $n(n-1)$ as $n \to \infty$, at least in the parameter regimes where the models behave basically like either very full or very empty Erdős-Rényi networks…Since these models are not projective, however, it is impossible to *improve* parameter estimates by getting more data, since parameters for smaller sub-graphs just cannot be extrapolated to larger graphs (or vice versa)…
- “Such an ERGM may provide a good description of a social network on a certain set of nodes, but it cannot be projected to give predictions on any larger or more global graph from which that one was drawn.”

Application: ERGM (p. 17)

- "If an ERGM is postulated for the whole network, then inference for its parameters must explicitly treat the unobserved portions of the network as missing data (perhaps through an expectation-maximization algorithm), though of course there may be considerable uncertainty about just how much data is missing."
- Question: Does this make sense in light of the last slide?

Solutions?

- Model the evolution of networks over time
  - “Hanneke, Fu and Xing consider situations where the distribution of the network at time t+1 conditional on the network at time t follows an exponential family.”
  - “Even when the statistics in the conditional specification include (say) changes in the number of triangles, the issues raised above do not apply.”
- Give up the exponential family form
  - Won’t work because… “Lauritzen showed that whenever the sufficient statistics for a semi-group, the models must be either ordinary exponential families, or certain generalizations thereof with much the same properties.”

The end… (p 18)

- "…every infinite exchangeable graph distribution is actually a mixture over projective dyadic-independence distributions, though not necessarily ones with a finite-dimensional sufficient statistic. Along any one sequence of sub-graphs from such an infinite graph, in fact, the densities of all motifs approach limiting values which pick out a unique projective dyadic-independence distribution… This suggests that an alternative to parametric inference would be nonparametric estimation of the limiting dyadic-independence model, by smoothing the adjacency matrix; this, too, we pursue elsewhere."

Naïve Thoughts and Questions

- Did I miss the assertion that the ERGM MLEs from sub-graphs are not consistent?
- Is the problem that sampling/structure is considered to be "uniform"? That is, any newly sampled node is equally likely to be connected to all the existing identical nodes (who are currently part of the same motifs)?
- Is there a way to think about a network with 4 nodes as a collection of 4 non-independent networks with 3 nodes? Or 6 networks with 2 nodes? I.e., can we define a dimension that is scientifically meaningful and start there, rather than with all of the nodes? 6 degrees of Kevin Bacon style? Then think of the smaller networks as coming from a mixture of ERGMs?
- Is the problem that we are thinking about binary adjacency matrices? If we extended the matrices to ordinal or continuous measures of connection, would "local" networks make more sense?