On the Geometry of Discrete Exponential Families with Application to ERG Models

Alessandro Rinaldo
Carnegie Mellon University
joint work with Stephen E. Fienberg and Yi Zhou

SIAM Annual Meeting
July, 13, 2010
Pittsburgh

Outline

- Exponential random graphs (or p^*) models for network data and degeneracy.

- Discrete exponential families, extended exponential families and their geometry.

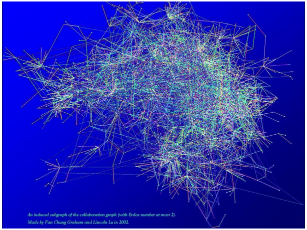
- A geometric characterization of the closure of discrete exponential families. Explanation of degeneracy.

Network Analysis

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

Network Analysis

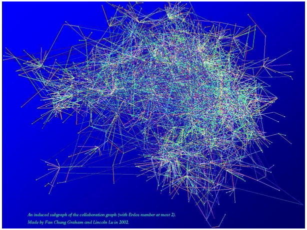
- Let $\mathcal{G}_n$ be the set of simple graphs on n nodes. Thus $|\mathcal{G}_n| = 2^{\binom{n}{2}}$.
- The nodes represent the units of some population of interest. The edges of any $x \in \mathcal{G}_n$ encode a set of static relationships among population units.



Source: Chung and Lu (2006), Complex Graphs and Networks.

Network Analysis

- Let $\mathcal{G}_n$ be the set of simple graphs on n nodes. Thus $|\mathcal{G}_n| = 2^{\binom{n}{2}}$.
- The nodes represent the units of some population of interest. The edges of any $x \in \mathcal{G}_n$ encode a set of static relationships among population units.



Source: Chung and Lu (2006), Complex Graphs and Networks.

Statistical Network Analysis

Construct interpretable and realistic statistical models for $\mathcal{G}_n$.

Exponential Random Graph (ERG) or p^* Models

- Exponential Random Graph (or p^*) models are constructed by specifying a set of informative network statistics on $\mathcal{G}_n$

$$x \mapsto t(x) = (t_1(x), \ldots, t_d(x)) \in \mathbb{R}^d,$$

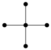
such that the probability of observing x is a function of $t(x)$ *only*.

Exponential Random Graph (ERG) or p^* Models

- Exponential Random Graph (or p^*) models are constructed by specifying a set of informative network statistics on $\mathcal{G}_n$

$$x \mapsto t(x) = (t_1(x), \ldots, t_d(x)) \in \mathbb{R}^d,$$

such that the probability of observing x is a function of $t(x)$ *only*.

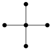
- Some examples:
  - the number of edges $E(x)$ (Erdös-Renyi model); 
  - number of triangles $T(x)$; 
  - the number of k-starts; 
  - the number of nodes with specified degrees (degree statistic).

Exponential Random Graph (ERG) or p^* Models

- Exponential Random Graph (or p^*) models are constructed by specifying a set of informative network statistics on $\mathcal{G}_n$

$$x \mapsto t(x) = (t_1(x), \ldots, t_d(x)) \in \mathbb{R}^d,$$

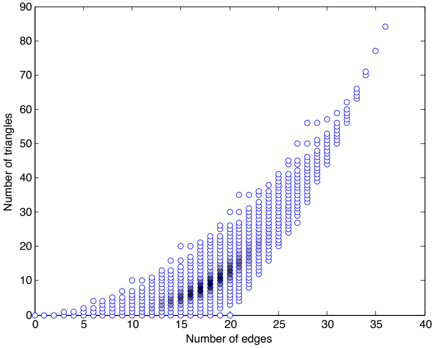
such that the probability of observing x is a function of $t(x)$ *only*.

- Some examples:
  - the number of edges $E(x)$ (Erdös-Renyi model); 
  - number of triangles $T(x)$; 
  - the number of k-starts; 
  - the number of nodes with specified degrees (degree statistic).

- See *Social Networks,* Volume 29, Special Section: Advances in Exponential Random Graph (p*) Models.

The Edge-Triangle (ET) Example

- We consider $\mathcal{G}_9$ with 2-dimensional network statistics $(E(x), T(x))$.
- The number of distinct graphs is 2^{36}, while the number of distinct network statistics is 444.



Exponential Random Graph (ERG) or p^* Models

- (Koopman-Pitman-Darmois Theorem.) Given the choice t of network statistics, construct a family of probability distributions $\{Q_\theta, \theta \in \Theta\}$ on $\mathcal{G}_n$ such that, for a given parameter $\theta \in \Theta$, the probability of observing x is

$$Q_\theta(x) = \exp\left\{\langle \theta, t(x) \rangle - \psi(\theta)\right\},$$

Exponential Random Graph (ERG) or p^* Models

- (Koopman-Pitman-Darmois Theorem.) Given the choice t of network statistics, construct a family of probability distributions $\{Q_\theta, \theta \in \Theta\}$ on $\mathcal{G}_n$ such that, for a given parameter $\theta \in \Theta$, the probability of observing x is

$$Q_\theta(x) = \exp\left\{\langle \theta, t(x) \rangle - \psi(\theta)\right\},$$

where

  - $\psi(\theta) = \log\left(\sum_{x \in \mathcal{G}_n} e^{\langle \theta, t(x) \rangle}\right)$ the log-partition function - a (often intractable) normalizing constant;
  - $\Theta = \{\theta \in \mathbb{R}^d \colon e^{\psi(\theta)} < \infty\}$ is the natural parameter space.

Exponential Random Graph (ERG) or p^* Models

- (Koopman-Pitman-Darmois Theorem.) Given the choice t of network statistics, construct a family of probability distributions $\{Q_\theta, \theta \in \Theta\}$ on $\mathcal{G}_n$ such that, for a given parameter $\theta \in \Theta$, the probability of observing x is

$$Q_\theta(x) = \exp\{\langle \theta, t(x) \rangle - \psi(\theta)\},$$

where

  - $\psi(\theta) = \log\left(\sum_{x \in \mathcal{G}_n} e^{\langle \theta, t(x) \rangle}\right)$ the log-partition function - a (often intractable) normalizing constant;
  - $\Theta = \{\theta \in \mathbb{R}^d : e^{\psi(\theta)} < \infty\}$ is the natural parameter space.

- Two key observations about this model:
  - Invariant with respect to relabeling of the vertices.
  - Redundant: if $t(x') = t(x)$, then $Q_\theta(x) = Q_\theta(x')$, for all θ. For the ET example, the median number of graphs corresponding to a network statistic is $2,741,130$!

Sufficiency Principle

- Let $\mathcal{T} = \{t \colon t = t(x), x \in \mathcal{G}_n\} \subset \mathbb{R}^d$ and $\nu(t) = |\{x \in \mathcal{G}_n \colon t(x) = t\}|$.

- Consider instead the family of probability distributions $\{P_\theta, \theta \in \Theta\}$ on $\mathcal{T}$ such that, for a given parameter $\theta \in \Theta$, the probability of observing t is

$$P_\theta(t) = \exp\left\{\langle \theta, t \rangle - \psi(\theta)\right\} \nu(t),$$

where Θ and $\psi(\theta)$ remain unchanged.

Sufficiency Principle

- Let $\mathcal{T} = \{t \colon t = t(x), x \in \mathcal{G}_n\} \subset \mathbb{R}^d$ and $\nu(t) = |\{x \in \mathcal{G}_n \colon t(x) = t\}|$.

- Consider instead the family of probability distributions $\{P_\theta, \theta \in \Theta\}$ on $\mathcal{T}$ such that, for a given parameter $\theta \in \Theta$, the probability of observing t is

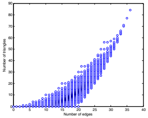
$$P_\theta(t) = \exp\left\{\langle \theta, t \rangle - \psi(\theta)\right\} \nu(t),$$

where Θ and $\psi(\theta)$ remain unchanged.

- In the ET example, instead of

$$\mathcal{G}_9$$

we can work on

$\mathcal{T} =$ 

Inference

- Given one observation x of the network, i.e. given $t = t(x)$,
  - estimate θ;
  - assess whether the ERG model fits the data.

  For the ET example, we want to learn 2 parameters: the edge and triangle parameters.

Inference

- Given one observation x of the network, i.e. given $t = t(x)$,
  - estimate θ;
  - assess whether the ERG model fits the data.

  For the ET example, we want to learn 2 parameters: the edge and triangle parameters.

- The maximum likelihood estimator (MLE) of θ is

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} P_\theta(t).$$

  The MLE is said to be nonexistent when the supremum is not achieved by any point in $\theta \in \Theta$. Nonexistence of the MLE means that there are too many parameters to estimate for the observed statistics t.

Exponential Random Graph Models

Advantages of using exponential families:

- best-behaved class of statistical models built around the notion of sufficiency;
- remarkably good theoretical properties that are well-understood.

Exponential Random Graph Models

Advantages of using exponential families:

- best-behaved class of statistical models built around the notion of sufficiency;
- remarkably good theoretical properties that are well-understood.

Disadvantages of using exponential families for network models:

- MLE almost impossible to compute; MCMLE can be obtained but convergence can be very slow and pseudo-MLE, which are easily computable, may not produce good estimates;
- asymptotic behavior is non-standard (and not well understood);
- degeneracy.

Degeneracy in ERG Models

What constitues degeneracy (Handcock, 2003)?

- *"A random graph model is near degenerate if the model places almost all its probability mass on a small number of graph configurations [...] e.g. empty graph, full graph, an individual graph, no 2-stars"*;
- a degenerrate model is not *"able to represent a range of realistic [networks]"* since only a *"small range of graphs [is] covered as the parameters vary"*;

- the MLE does not exist and/or MCMLE fails to converge;
- the observed network t is very unlikely under the distribution specified by the MLE;
- the model misbehaves...

Degeneracy in the ET Example: Entropy Plot

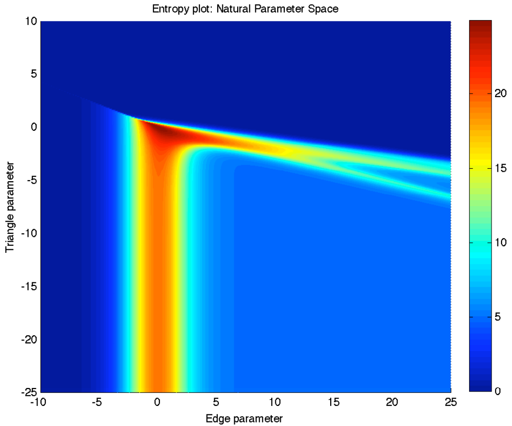
- We capture overall degenerate behavior using Shannon's entropy function

$$\theta \mapsto -\sum_{t \in \mathcal{T}} P_\theta(t) \log_2 \left(\frac{P_\theta(t)}{\nu(t)} \right), \quad \theta \in \Theta.$$

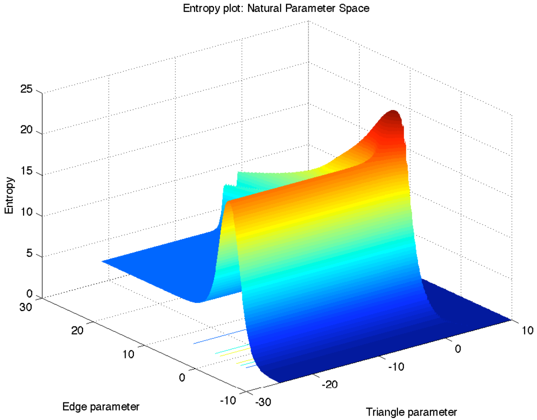
- Rationale: degeneracy occurs when the probability mass is spread over a small number of network statistics, so degenerate distributions will tend to correspond to values of θ for which the entropy function is small.

- Entropy plot: for each point $\theta \in \mathbb{R}^2$ in the natural parameter space, plot the entropy of the corresponding probability distribution.

Degeneracy in the ET Example: Entropy Plot

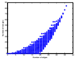


Degeneracy in the ET Example: Entropy Plot



Basics of Discrete Exponential Families

See Barndorff-Nielsen (1974) or Brown (1986)

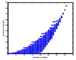
- Let T be a random vector taking values in a finite set $\mathcal{T}$, for example 

  The distribution of T belongs to the exponential family $\mathcal{E} = \{P_\theta, \theta \in \Theta\}$, with

$$P_\theta(t) = \exp\{\langle \theta, t \rangle - \psi(\theta)\}\, \nu(t), \quad \theta \in \Theta = \mathbb{R}^d.$$

Basics of Discrete Exponential Families

See Barndorff-Nielsen (1974) or Brown (1986)

- Let T be a random vector taking values in a finite set $\mathcal{T}$, for example 

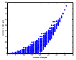
  The distribution of T belongs to the exponential family $\mathcal{E} = \{P_\theta, \theta \in \Theta\}$, with

  $$P_\theta(t) = \exp\{\langle\theta, t\rangle - \psi(\theta)\}\,\nu(t), \quad \theta \in \Theta = \mathbb{R}^d.$$

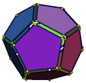
- The set $\mathrm{P} = \mathrm{convhull}(\mathcal{T})$ is called the convex support.

Basics of Discrete Exponential Families

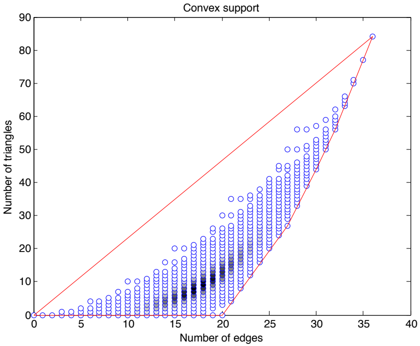
See Barndorff-Nielsen (1974) or Brown (1986)

- Let T be a random vector taking values in a finite set $\mathcal{T}$, for example 

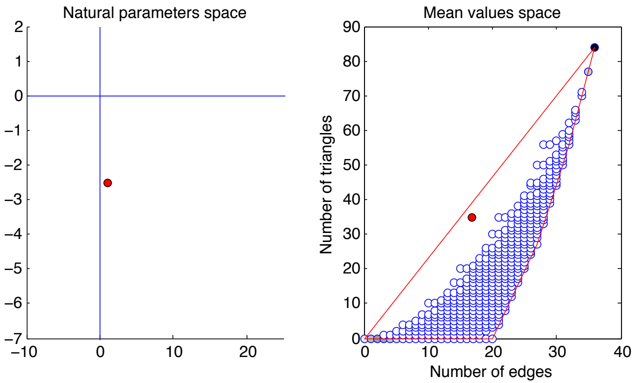
  The distribution of T belongs to the exponential family $\mathcal{E} = \{P_\theta, \theta \in \Theta\}$, with
  $$P_\theta(t) = \exp\{\langle \theta, t \rangle - \psi(\theta)\}\, \nu(t), \quad \theta \in \Theta = \mathbb{R}^d.$$

- The set $\mathrm{P} = \mathrm{convhull}(\mathcal{T})$ is called the convex support.
  - It is a polytope. 
  - $\mathrm{int}(\mathrm{P}) = \{\mathbb{E}_\theta[T], \theta \in \Theta\}$ is precisely the set of all possible expected values of T: mean value space.
  - $\mathrm{int}(\mathrm{P})$ and Θ are homeomorphic: we can represent the exponential family using $\mathrm{int}(\mathrm{P})$ instead of Θ: mean value parametrization.

- Convex support for the ET example.



- One-to-one correspondence between Θ and $\mathrm{int}(\mathrm{P})$.

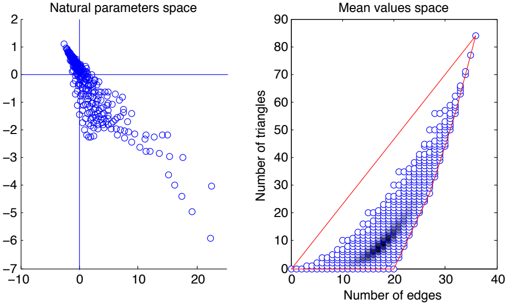


Basics of Exponential Families

- Existence of the MLE

  the MLE exists if and only if $t \in \mathrm{int}(\mathrm{P})$.

- In the ET example, MLE exists for 415 of 444 possible network statistics.

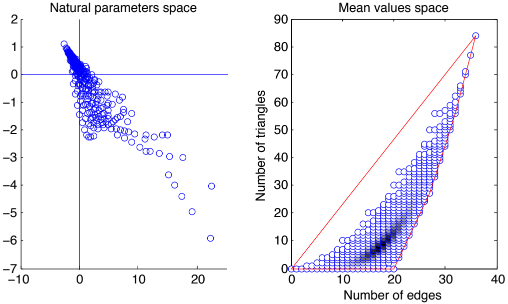


Basics of Exponential Families

- Existence of the MLE

  the MLE exists if and only if $t \in \mathrm{int}(\mathrm{P})$.

- In the ET example, MLE exists for 415 of 444 possible network statistics.



- How do we model the boundary of P (where the MLE does not exist)? Solution: compute the closure of $\mathcal{E}$.

Extended Exponential Families: Geometric Construction

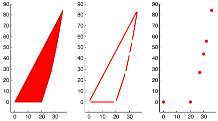
- For every face F of P, construct the exponential family of distributions $\mathcal{E}_F$ for the points in F with convex support F. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.

Extended Exponential Families: Geometric Construction

- For every face F of P, construct the exponential family of distributions $\mathcal{E}_F$ for the points in F with convex support F. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.

- The extended exponential family is

$$\overline{\mathcal{E}} = \mathcal{E} \cup \{\mathcal{E}_F \colon F \text{ is a face of } \mathrm{P}\}$$

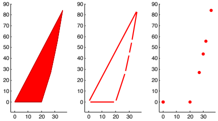


Extended Exponential Families: Geometric Construction

- For every face F of P, construct the exponential family of distributions $\mathcal{E}_F$ for the points in F with convex support F. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.

- The extended exponential family is

$$\overline{\mathcal{E}} = \mathcal{E} \cup \{\mathcal{E}_F \colon F \text{ is a face of } \mathrm{P}\}$$



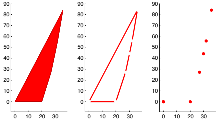
- Within $\overline{\mathcal{E}}$, the MLE always exists.

Extended Exponential Families: Geometric Construction

- For every face F of P, construct the exponential family of distributions $\mathcal{E}_F$ for the points in F with convex support F. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.

- The extended exponential family is

$$\overline{\mathcal{E}} = \mathcal{E} \cup \{\mathcal{E}_F \colon F \text{ is a face of } \mathrm{P}\}$$

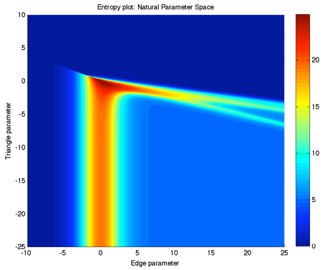


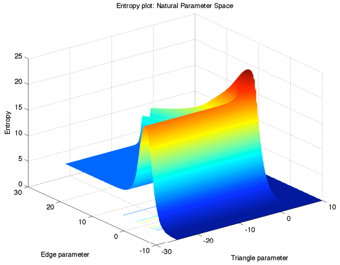
- Within $\overline{\mathcal{E}}$, the MLE always exists.

Extended Exponential Family

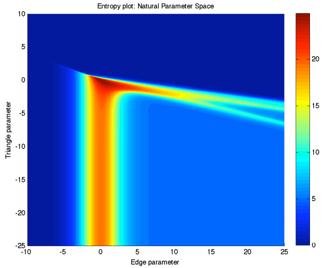
The extended exponential family is the closure of the original family. Geometrically, this corresponds to including the boundary of the convex support P, i.e. to taking the closure of the mean value space.

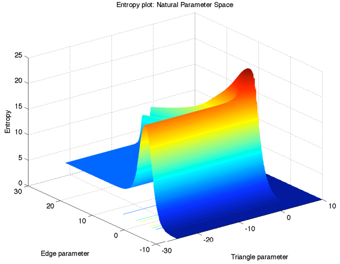
Back to Degeneracy

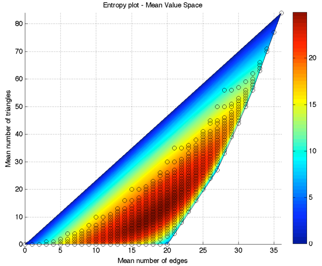


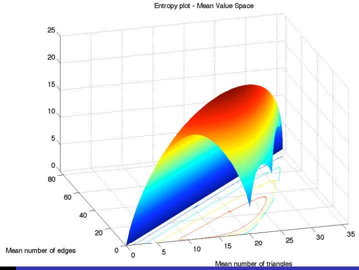


Back to Degeneracy









Extended Exponential Families: Dual Geometric Construction

- We derive an alternative geometric construction of the extended exponential families using the natural parameter space and not the mean value space.

Extended Exponential Families: Dual Geometric Construction

- We derive an alternative geometric construction of the extended exponential families using the natural parameter space and not the mean value space.

- Let P be a full-dimensional polytope in $\mathbb{R}^d$. The normal cone to a face F is the polyhedral cone

$$N_F = \left\{ c \in \mathbb{R}^k \colon F \subset \{ x \in \mathrm{P} \colon \langle c, x \rangle = \max_{y \in \mathrm{P}} \langle c, y \rangle, \} \right\}$$

consisting of all the linear functionals on P that are maximal over F.

Extended Exponential Families: Dual Geometric Construction

- We derive an alternative geometric construction of the extended exponential families using the natural parameter space and not the mean value space.

- Let P be a full-dimensional polytope in $\mathbb{R}^d$. The normal cone to a face F is the polyhedral cone

$$N_F = \left\{ c \in \mathbb{R}^k \colon F \subset \{ x \in \mathrm{P} \colon \langle c, x \rangle = \max_{y \in \mathrm{P}} \langle c, y \rangle, \} \right\}$$

  consisting of all the linear functionals on P that are maximal over F. The set of cones

$$\mathcal{N}(\mathrm{P}) = \{ N_F, F \text{ is a face of } \mathrm{P} \}$$

  is called the normal fan of P.

Extended Exponential Families: Dual Geometric Construction

- We derive an alternative geometric construction of the extended exponential families using the natural parameter space and not the mean value space.

- Let P be a full-dimensional polytope in $\mathbb{R}^d$. The normal cone to a face F is the polyhedral cone

$$N_F = \left\{ c \in \mathbb{R}^k \colon F \subset \{ x \in \mathrm{P} \colon \langle c, x \rangle = \max_{y \in \mathrm{P}} \langle c, y \rangle, \} \right\}$$

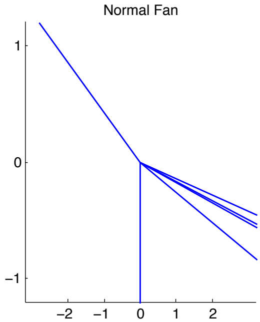
  consisting of all the linear functionals on P that are maximal over F. The set of cones

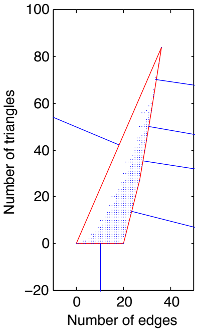
$$\mathcal{N}(\mathrm{P}) = \{ N_F, F \text{ is a face of } \mathrm{P} \}$$

  is called the normal fan of P.

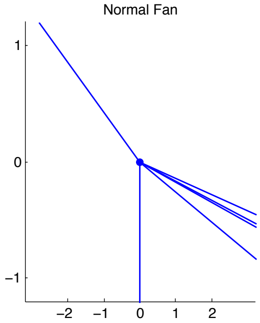
- Key properties:
  - The (relative interiors of the) cones in $\mathcal{N}(\mathrm{P})$ partition $\mathbb{R}^d$.
  - $\dim(N_F) = d - \dim(F)$.

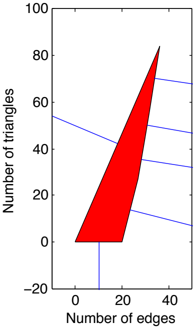
The Normal Fan: ET Example



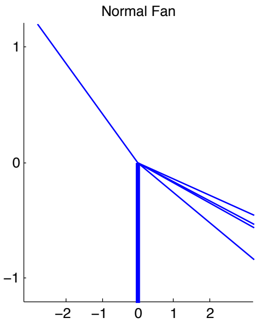


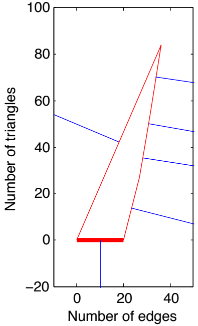
The Normal Fan: ET Example



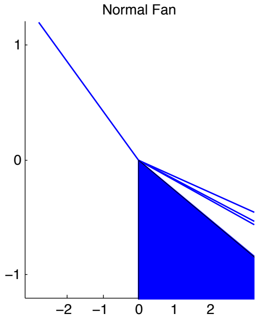


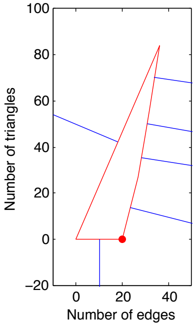
The Normal Fan: ET Example





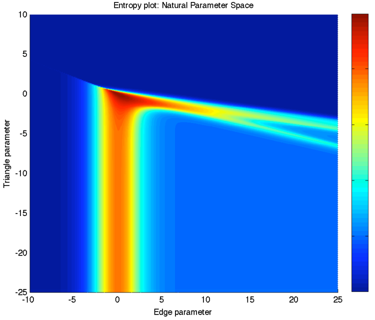
The Normal Fan: ET Example

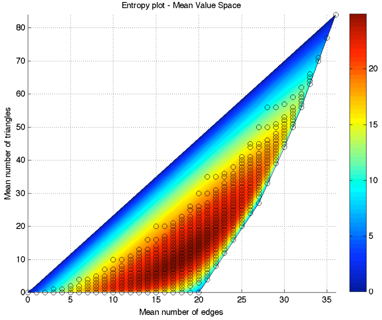




Main Result (Graphical Form)

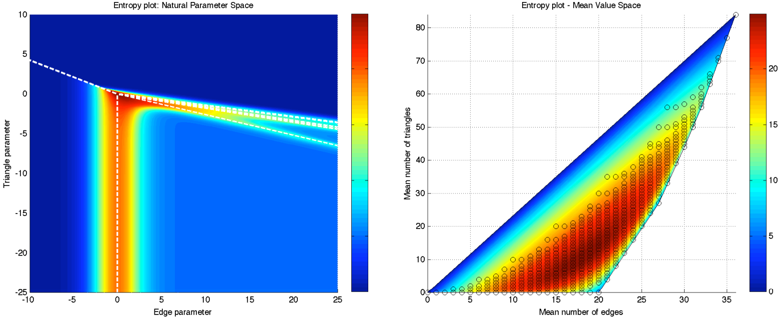
- Entropy plots of the natural space and mean value spaces.





Main Result (Graphical Form)

- Entropy plots of the natural space space with superimposed the normal fan and of the mean value space.



Main Result (Colloquial Form)

- Pick any $\theta \in \Theta$, any face F of P and any direction $d \neq 0$ in the interior of the normal cone N_F.
- For a sequence of positive numbers $\rho_n \to \infty$, let $\theta_n = \theta + \rho_n d \subset \Theta$
- Let μ_n be the mean value parameter corresponding to θ_n (i.e. $\mu_n = \mathbb{E}_{\theta_n}[T]$ and the sequence $\{\mu_n\}$ is contained in the interior of P).
- Then, $\lim_n \mu_n$ is a point on the boundary of P and in the interior of F, which depends only on θ and d.
- Conversely, any point on the boundary of the convex support can be obtained in this way.

Main Result (Colloquial Form)

- Pick any $\theta \in \Theta$, any face F of P and any direction $d \neq 0$ in the interior of the normal cone N_F.
- For a sequence of positive numbers $\rho_n \to \infty$, let $\theta_n = \theta + \rho_n d \subset \Theta$
- Let μ_n be the mean value parameter corresponding to θ_n (i.e. $\mu_n = \mathbb{E}_{\theta_n}[T]$ and the sequence $\{\mu_n\}$ is contained in the interior of P).
- Then, $\lim_n \mu_n$ is a point on the boundary of P and in the interior of F, which depends only on θ and d.
- Conversely, any point on the boundary of the convex support can be obtained in this way.

Normal Fan and Extended Exponential Families

The normal fan realizes geometrically the closure of the family inside the natural parameter space.

See Rinaldo, Fienberg and Yi (2009) for the full statement.

Conclusions

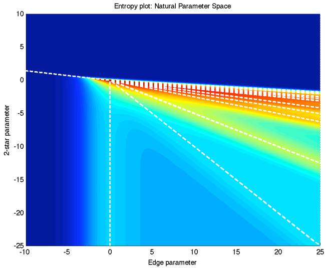
- Degeneracy can be explained in terms of closures of the exponential families. In particular, nonexistence of the MLE and "near-boundary points" capture essential features of ERG models.
- The normal fan to the convex support plays a central role in the geometry of discrete extended exponential families. The result is general and applies to all discrete exponential families with polyhedral convex support, for instance to graphical models.
- Network analysis is gaining popularity and momentum very quickly. There are many open problems that could be approached and perhaps solved using algebraic statistics.

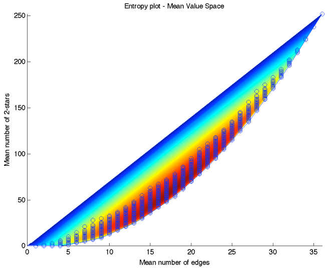
Conclusions

Thank you

Conclusions

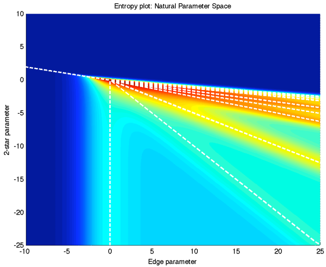
- $\mathcal{G}_9$. Network statistics: number of edges versus number of 2-stars 

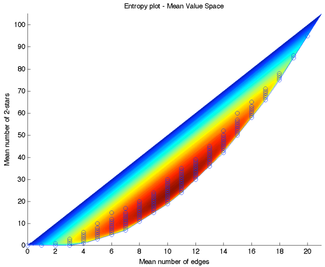




Conclusions

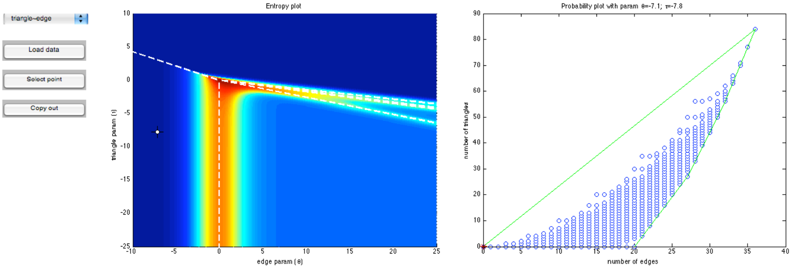
- $\mathcal{G}_7$. Network statistics: number of edges versus number of 2-stars 





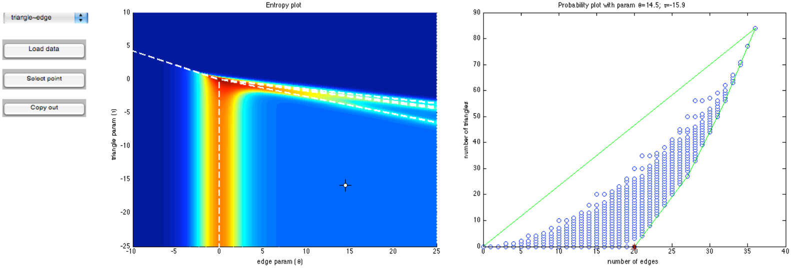
Conclusions

- ET Example.



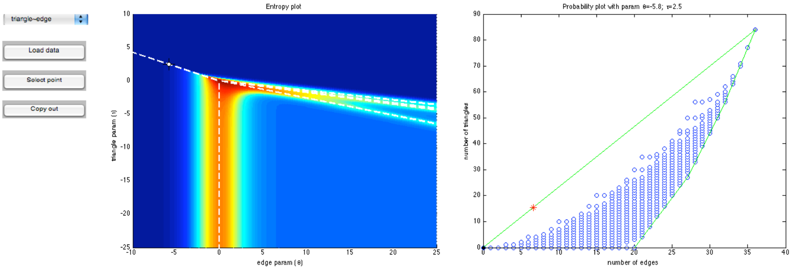
Conclusions

- ET Example.



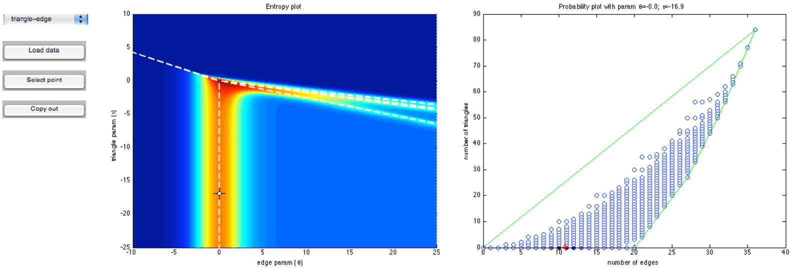
Conclusions

- ET Example.



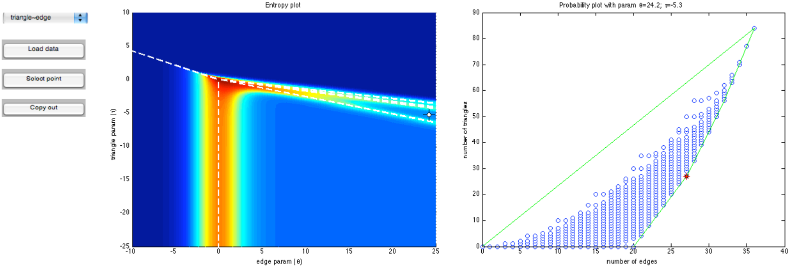
Conclusions

- ET Example.



Conclusions

- ET Example.



Conclusions

- ET Example.