Multipodal Phases in Dense Networks

Charles Radin

University of Texas at Austin

May 8, 2014

Joint work with Rick Kenyon, Kui Ren and Lorenzo Sadun

Abstract. We consider asymptotically large, dense, simple graphs constrained in a set of densities, either edges and triangles or edges and one or more k-stars, and study the associated entropy, the goal being to determine the structure of 'exponentially most' graphs with given variable constraints. We will present evidence - proof for edges/stars, and simulation for edges/triangles - that for all parameter values the optimal graphs have very simple 'multipodal' structure, simple modifications of balanced multipartite graphs.

Consider simple graphs G with vertex set $V(G)$ of (labeled) vertices, with $|V(G)| = n$.

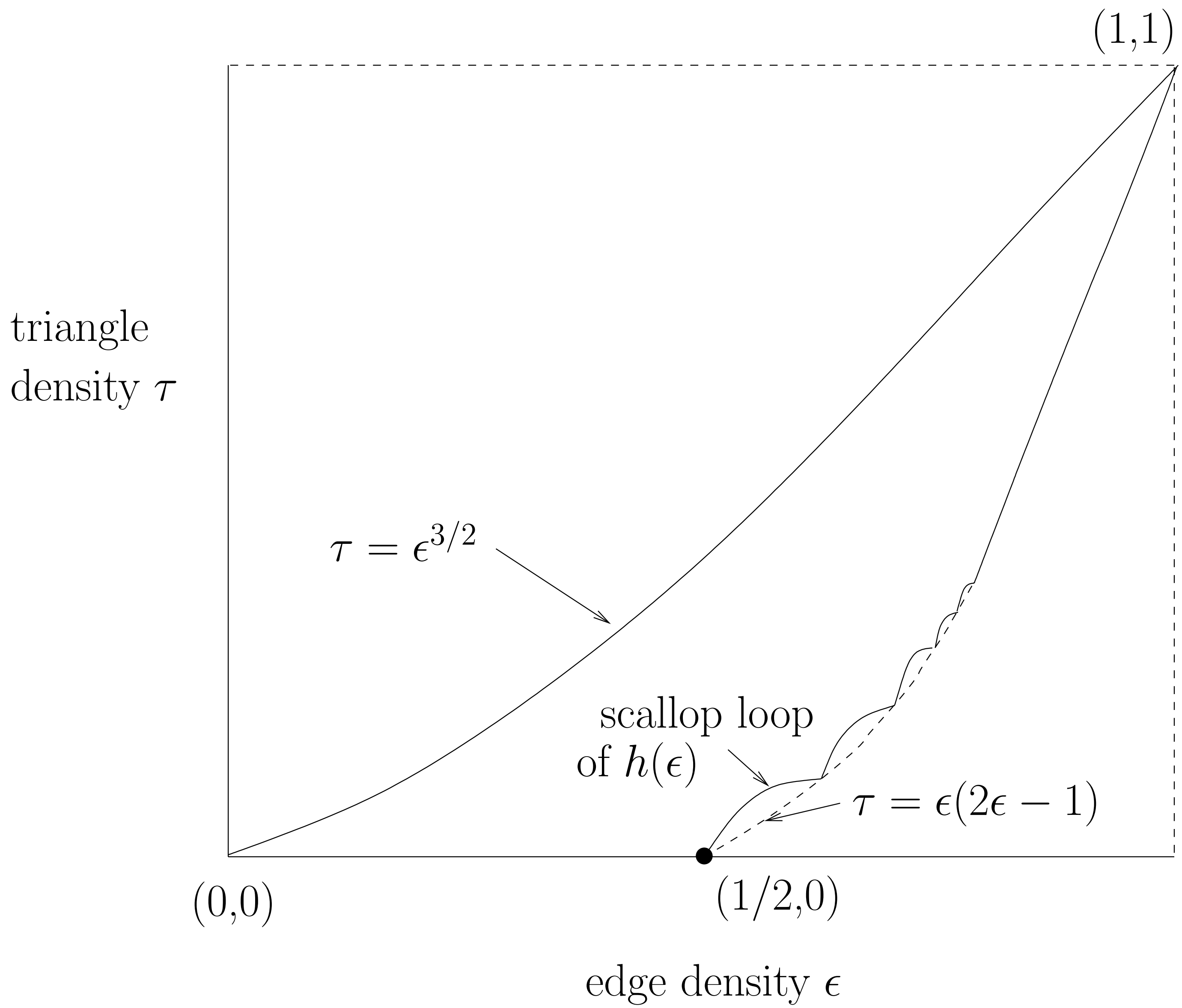
For a simple graph H, its "density" in G is the fraction of all maps, of $V(H)$ into $V(G)$, which preserve edges.

Temporarily specialize constraints to edge density $e(G)$ and triangle densty $t(G)$. Our main tool is $Z^{n,a}_{\epsilon,\tau}$, the number of graphs with densities:

$$e(G) \in (\epsilon - a, \epsilon + a); \quad t(G) \in (\tau - a, \tau + a).$$

By definition $(e(G), t(G)) \in [0, 1]^2$, but in fact the boundary is:

(A. Razborov, Combin. Probab. Comput. 17 (2008) 603-618.)



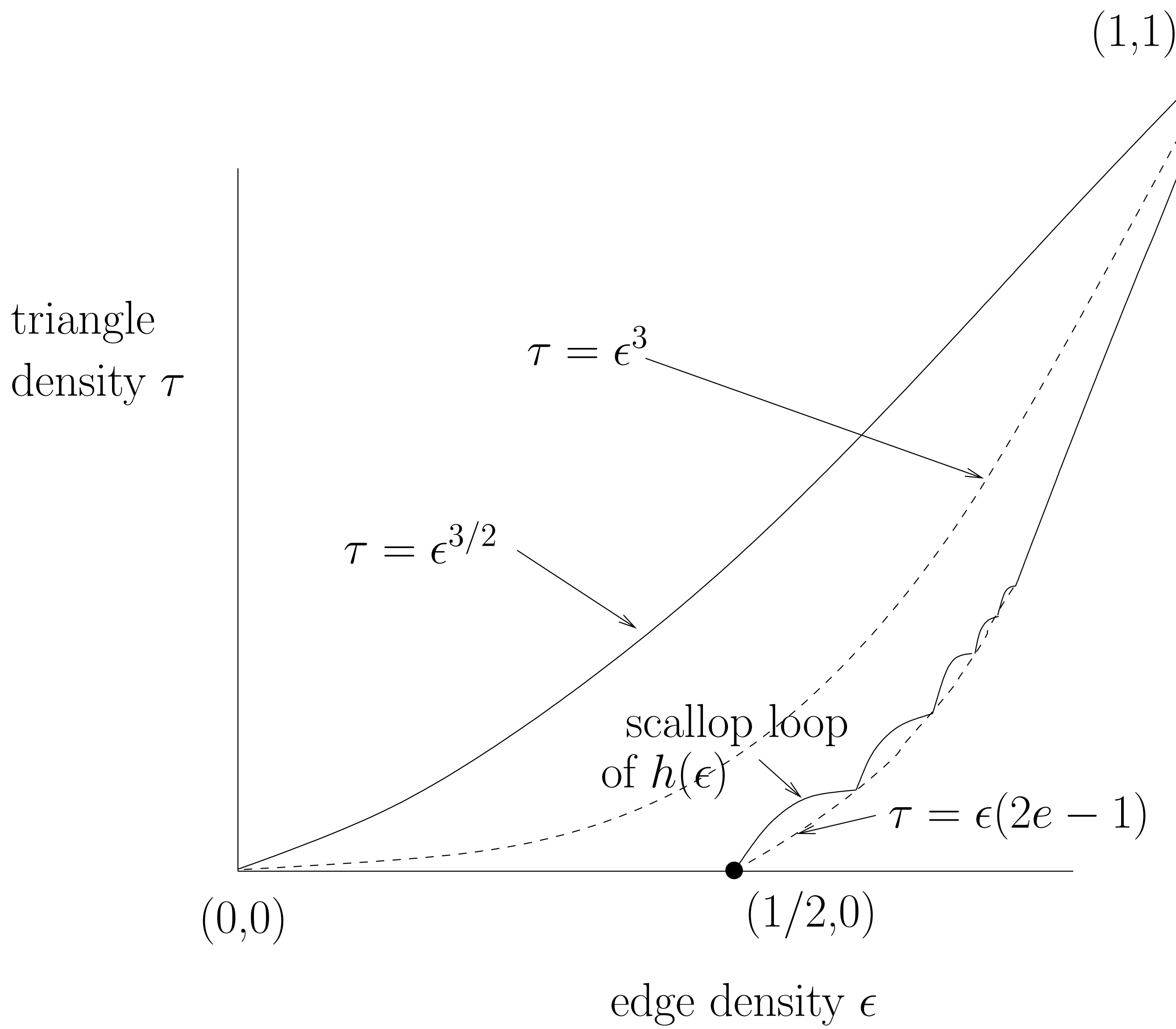
(1,1)
triangle
density τ
$\tau = \epsilon^{3/2}$
scallop loop
of $h(\epsilon)$
$\tau = \epsilon(2\epsilon - 1)$
(0,0)
(1/2,0)
edge density ϵ

Graph for $(\epsilon, \tau) = (0.5, 0)$ is complete, balanced bipartite; it is complete, balanced multipartite for the other vertices, and also known on the rest of the boundary.

This is an example in *extremal graph theory*, the Mantel problem.

We aim to extend to interior, determining 'most' graphs with given constraints.

For instance an optimization principle easily implies that a typical graph with constraints $\tau = \epsilon^3$ has independent edges with probability ϵ.

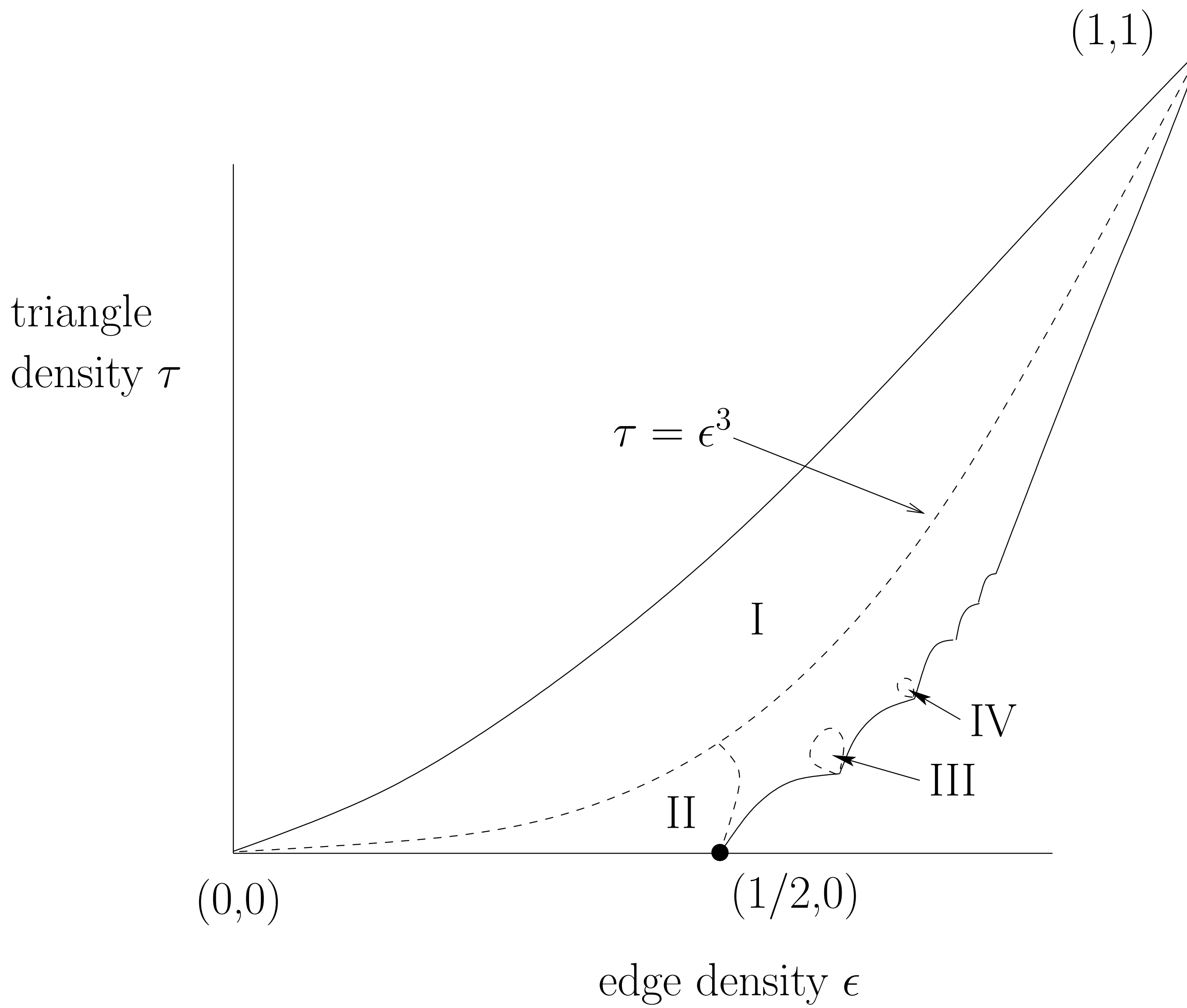

(1,1)
triangle
density τ
$\tau = \epsilon^3$
$\tau = \epsilon^{3/2}$
scallop loop
of $h(\epsilon)$
$\tau = \epsilon(2e-1)$
(0,0)
(1/2,0)
edge density ϵ

Since $Z_{\epsilon,\tau}^{n,a}$ grows like $e^{s(n^2)}$, we normalize as ‘entropy density’:

$$s = s_{\epsilon,\tau}^{n,a} = \frac{\ln(Z_{\epsilon,\tau}^{n,a})}{n^2}, \qquad s(\epsilon,\tau) = \lim_{a\downarrow 0}\lim_{n\to\infty} s_{\epsilon,\tau}^{n,a}.$$

Little is known about $s(\epsilon,\tau)$ beyond existence, but we conjecture it is piecewise smooth in general.

Key definition: A phase is a maximal connected set of parameters where $s(\epsilon,\tau)$ is analytic.


(1,1)
triangle
density τ
$\tau = \epsilon^3$
I
II
III
IV
(0,0)
(1/2,0)
edge density ϵ

Intuitively randomness arises as follows. Start with the assumption that certain kinds of subgraphs $H = (H_1, \ldots, H_m)$ are 'significant' for a large network: edges, triangles, etc. Then determine what most graphs are which have constrained values of the densities $t_H = (t_{H_1}, \ldots, t_{H_m})$ of those type of subgraphs. Large deviations theory gives probabilistic descriptions of such typical graphs through a variational principle for the constrained entropy.

S. Chatterjee, S.R.S. Varadhan, Eur. J. Comb. 32 (2011) 1000-1017

L. Lovász, Large networks and graph limits, AMS, 2012.

Variational Principle (R.-Sadun)

For any (ϵ, τ), $s(\epsilon, \tau) = \max_g[-I(g)]$, where the maximum is over all measurable $g : [0,1]^2 \to [0,1]$, $g(x,y) = g(y,x)$, subject to the constraints

$$e(g) = \int_{[0,1]^2} g(x,y)\,dxdy \ = \ \epsilon$$

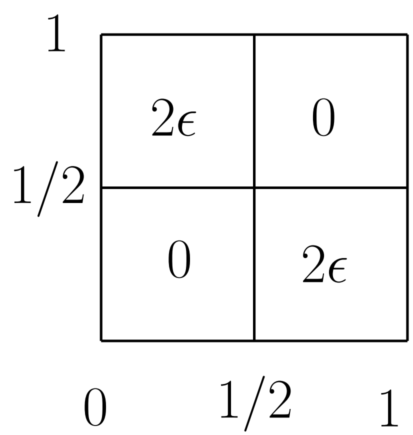
$$t(g) = \int_{[0,1]^3} g(x,y)g(y,z)g(z,x)\,dxdydz \ = \ \tau,$$

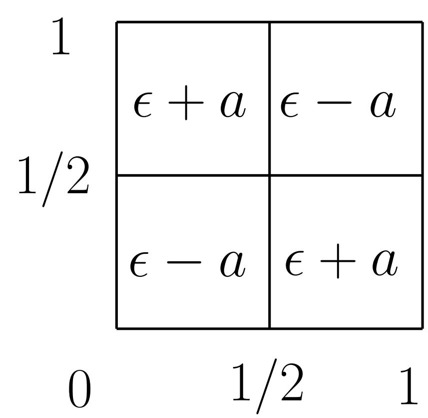
and the 'rate function' $I(g)$ is Shannon entropy:

$$I(g) = \frac{1}{2}\int_{[0,1]^2} g(x,y)\ln[g(x,y)] + [1-g(x,y)]\ln[1-g(x,y)]\,dxdy.$$

Think of points in $[0,1]$ as vertices, and $g(x,y)$ as the probability of an edge between x and y.

Optimizing graphons for phase II (unique up to rearranged vertices):

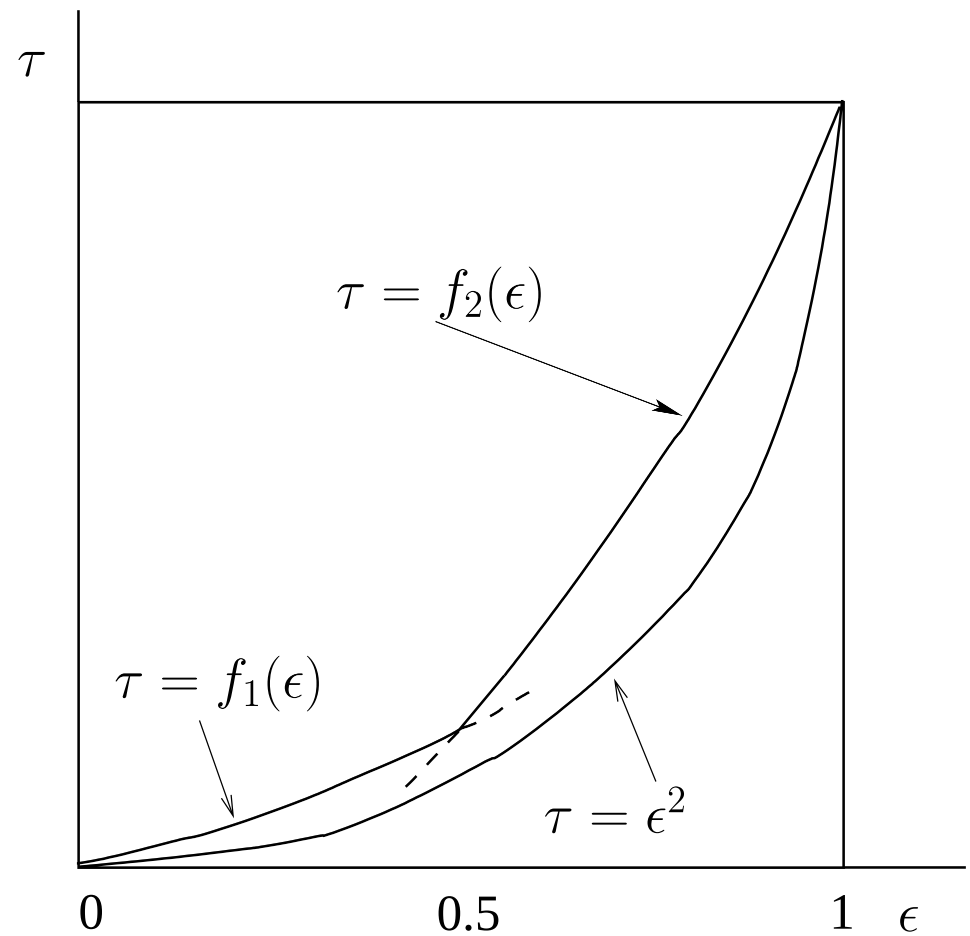




$$a = (\epsilon^3 - \tau)^{1/3}$$

Main result: Simulation suggests for every (ϵ, τ) there is a partition of the vertices into $M < \infty$ subsets $V_1, V_2, \ldots, V_M$, and a set of well-defined probabilities q_{ij} of an edge between any $v_i \in V_i$ and $v_j \in V_j$. We call such states 'multipodal'.

Change constraint from edges/triangles to a finite set of different k-stars, including edges. Phase space for edge/2-star is:



Now the variational principle is to maximize $-I(g)$ subject to

$$t_k(g) = \int_{[0,1]^{k+1}} g(x,y_1)g(x,y_2)\cdots g(x,y_k)\,dxdy_1\cdots dy_k \;=\; \tau_k,$$

for the desired set of $k's$, including $k=1$. (A star model.)

Theorem (Kenyon, R., Ren, Sadun). Every maximizing graphon for a star model is multipodal.

In practice the state is fully described by few variables. We've simulated the edge/triangle model and the 2-star model and so far never seen more than 4 variables, each a function of the constraint values.

Lagrange multipliers; exponential random graphs

Lagrange technique: introduce new variables $\beta = (\beta_1, \cdots, \beta_m)$, one for each constraint, and solve the Euler-Lagrange equations:

$$\delta[-I(g) + \beta \cdot t_H(g)] = 0,$$

together with the constraints $t_H(g) = \tau$, for $s(\epsilon, \tau) = \max_g[-I(g)]$.

Alternate formulation: solve the *unconstrained* problem:

$$\psi(\beta) = \max_g[-I(g) + \beta \cdot t_H(g)],$$

then, if possible, solve for β such that optimizers satisfy constraints.

This optimization problem was first carefully analyzed in:

S. Chatterjee and P. Diaconis, Ann. Statist. 41 (2013) 2428-2461.

However: Any maximizer $\tilde{g}$ of $-I(g) + \beta \cdot t_H(g)$ is automatically a maximizer of $-I(g)$ for some constraints $t_H(g) = \tau'$, namely for $\tau' = t_H(\tilde{g})$. But, as noted by Chatterjee/Diaconis, it is often impossible to find $\beta's$ such that any maximizer of $-I(g)$ with given constraints $t_H(g) = \tau$, will maximize $-I(g) + \beta \cdot t_H(g)$. This is especially true for k-star models; for k-star models all maximizers $\tilde{g}$ of $-I(g) + \beta_1 e(g) + \beta_2 t_k(g)$ satisfy $t_k(\tilde{g}) = e(\tilde{g})^k$, so the two densities cannot be constrained independently.

This phenomenon is known as 'inequivalent ensembles'.