Testability of the minimum k-way cut density

Marianna Bolla, Katalin Friedl, Tamás Kói,
and **András Krámli**
Budapest University of Technology and Ecomomics
and *University of Szeged*
marib@math.bme.hu

Fete of Combinatorics and Computer Science
Keszthely, August 12, 2008

Motivation

- To classify large data sets into homogeneous parts, cf. Bolla and Tusnády, Discrete Math., 1994. Statistical parameters whose values are approximated by taking a smaller sample.
- To investigate the testability of different kinds of minimum multiway cut densities emerging in classification problems.
- For this purpose we generalize a theorem of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs I, 2006 to formulate equivalent statements for the testability of weighted graph parameters.
- We use some theorems of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs II, 2007 to prove testability of special constrained versions of minimum multiway cut densities.
- To investigate effects of random perturbations on the weights, and the cut-norm of the graphon assigned to the so-called Wigner-noise, cf. Bolla, Lin. Alg. Appl., 2005.

Motivation

- To classify large data sets into homogeneous parts, cf. Bolla and Tusnády, Discrete Math., 1994. Statistical parameters whose values are approximated by taking a smaller sample.
- To investigate the testability of different kinds of minimum multiway cut densities emerging in classification problems.
- For this purpose we generalize a theorem of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs I, 2006 to formulate equivalent statements for the testability of weighted graph parameters.
- We use some theorems of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs II, 2007 to prove testability of special constrained versions of minimum multiway cut densities.
- To investigate effects of random perturbations on the weights, and the cut-norm of the graphon assigned to the so-called Wigner-noise, cf. Bolla, Lin. Alg. Appl., 2005.

Motivation

- To classify large data sets into homogeneous parts, cf. Bolla and Tusnády, Discrete Math., 1994. Statistical parameters whose values are approximated by taking a smaller sample.
- To investigate the testability of different kinds of minimum multiway cut densities emerging in classification problems.
- For this purpose we generalize a theorem of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs I, 2006 to formulate equivalent statements for the testability of weighted graph parameters.
- We use some theorems of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs II, 2007 to prove testability of special constrained versions of minimum multiway cut densities.
- To investigate effects of random perturbations on the weights, and the cut-norm of the graphon assigned to the so-called Wigner-noise, cf. Bolla, Lin. Alg. Appl., 2005.

Motivation

- To classify large data sets into homogeneous parts, cf. Bolla and Tusnády, Discrete Math., 1994. Statistical parameters whose values are approximated by taking a smaller sample.
- To investigate the testability of different kinds of minimum multiway cut densities emerging in classification problems.
- For this purpose we generalize a theorem of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs I, 2006 to formulate equivalent statements for the testability of weighted graph parameters.
- We use some theorems of Borgs, Chayes, Lovász, Sós, Vesztergombi, Convergent sequences of dense graphs II, 2007 to prove testability of special constrained versions of minimum multiway cut densities.
- To investigate effects of random perturbations on the weights, and the cut-norm of the graphon assigned to the so-called Wigner-noise, cf. Bolla, Lin. Alg. Appl., 2005.

Notation

$G = G_n$: weighted graph on the node set $[n] = \{1, \dots, n\} = V(G)$.
Edge-weights: $\beta_{ij} \in \mathbb{R}$, $\beta(G) = (\beta_{ij}) \in \mathrm{Sym}_n$ (strength of the interaction between the nodes).
For randomization purposes suppose that $\beta_{ij} \in [0, 1]$ (0=no edge).
Node-weights: $\alpha_i > 0$, $i = 1, \dots, n$ (individual weights of the nodes).
Let $\mathcal{G}$ denote the set of such weighted graphs.
$\alpha_G := \sum_{i=1}^{n} \alpha_i$ (volume of G)
$\alpha_U := \sum_{i \in U} \alpha_i$ (volume of the node-set $U \subset V(G)$)

$$e_G(U, T) := \sum_{u \in U} \sum_{t \in T} \alpha_u \alpha_t \beta_{ut}, \qquad U, T \subset V = V(G)$$

$\mathcal{P}_k$: set of k-partitions $P = (V_1, \dots, V_k)$ of V.

Minimum k-way cut densities

Let $k < n$ be a fixed positive integer.

$$f_k(G) := \min_{P \in \mathcal{P}_k} \frac{1}{\alpha_G^2} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} e_G(V_i, V_j)$$

minimum k-way cut density of G.
Let $c \leq 1/k$ be a fixed positive real number.
$\mathcal{P}_k^c$: set of k-partitions of V such that $\frac{\alpha_{V_i}}{\alpha_G} \geq c$ $(i = 1, \ldots, k)$, or equivalently, $c \leq \frac{\alpha_{V_i}}{\alpha_{V_j}} \leq \frac{1}{c}$ $(i \neq j)$.

$$f_k^c(G) := \min_{P \in \mathcal{P}_k^c} \frac{1}{\alpha_G^2} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} e_G(V_i, V_j)$$

minimum c-balanced k-way cut density of G.

Let $\mathbf{a} = \{a_1, \ldots, a_k\}$ be a probability distribution on $[k]$.
$\mathcal{P}_k^{\mathbf{a}}$: set of k-partitions of V such that

$$\left(\frac{\alpha_{V_1}}{\alpha_G}, \ldots, \frac{\alpha_{V_k}}{\alpha_G}\right)$$

is approximately $\mathbf{a}$-distributed.

$$f_k^{\mathbf{a}}(G) := \min_{P \in \mathcal{P}_k^{\mathbf{a}}} \frac{1}{\alpha_G^2} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} e_G(V_i, V_j)$$

minimum $\mathbf{a}$-balanced k-way cut density of G.

Minimum weighted k-way cut densities

We want to penalize cluster volumes that wildly differ. Historically,

$$\mu_k(G) := \min_{P \in \mathcal{P}_k} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{1}{\alpha_{V_i} \cdot \alpha_{V_j}} \cdot e_G(V_i, V_j)$$

minimum weighted k-way cut density of G.

$$\mu_k^c(G) := \min_{P \in \mathcal{P}_k^c} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{1}{\alpha_{V_i} \cdot \alpha_{V_j}} \cdot e_G(V_i, V_j)$$

minimum weighted c-balanced k-way cut density of G, where $0 < c \leq 1/k$.

Remark:

$$\mu_k(G) = \min_{G/P \in \hat{S}_k(G)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij}(G/P),$$

where the weighted graph G/P is the k-quotient of G with respect to P, and $\hat{S}_k(G)$ is the set of k-quotients.

Testability of weighted graph parameters

Definition

A weighted graph parameter f is testable if for every $\varepsilon > 0$ there is a positive integer k such that if $G \in \mathcal{G}$ satisfies

$$\max_i \frac{\alpha_i(G)}{\alpha_G} \leq \frac{1}{k},$$

then

$$\mathbb{P}(|f(G) - f(\xi(k, G))| > \varepsilon) \leq \varepsilon,$$

where $\xi(k, G)$ is a random simple graph on k nodes randomized "appropriately" from G.

The randomization procedures will be discussed later.

Remarks:

- $|V(G)| \geq k$ follows from the node-condition.
- In the proof of the subsequent theorem we use the graphon-randomization procedure introduced in Section 4.4 of Borgs et al. I, where a random simple graph is randomized out of the step-function graphon W_G assigned to the weighted graph G.
- To be testable, f must be invariant under scaling the node-weights.

Equivalent statements of testability

Theorem

Equivalent statements for the testability of the bounded weighted graph parameter f.

- *For every $\varepsilon > 0$ there is a positive integer k such that for every weighted graph $G \in \mathcal{G}$ satisfying the node-condition $\max_i \alpha_i(G)/\alpha_G \leq 1/k$, $|f(G) - \mathbb{E}(f(\xi(k, G)))| \leq \varepsilon$.*
- *For every left-convergent weighted graph sequence (G_n) with $\max_i \alpha_i(G_n)/\alpha_{G_n} \to 0$, $f(G_n)$ is also convergent $(n \to \infty)$.*
- *f can be extended to graphons such that $\tilde{f}(W)$ is continuous in the cut-norm and $\tilde{f}(W_{G_n}) - f(G_n) \to 0$, whenever $\max_i \alpha_i(G_n)/\alpha_{G_n} \to 0$ $(n \to \infty)$.*
- *For every $\varepsilon > 0$ there is an $\varepsilon_0 > 0$ real and an $n_0 > 0$ integer such that if G_1, G_2 are weighted graphs satisfying $\max_i \alpha_i(G_1)/\alpha_{G_1} \leq 1/n_0$, $\max_i \alpha_i(G_2)/\alpha_{G_2} \leq 1/n_0$, and $\delta_\square(G_1, G_2) < \varepsilon_0$, then $|f(G_1) - f(G_2)| < \varepsilon$.*

About the proof

With minor modifications the proof of Theorem 6.1. of Borgs et al. I is followed. The left-convergence means the convergence of the homomorphism density

$$t(F,G) = \frac{1}{(\alpha_G)^k} \sum_{\Phi:V(F)\to V(G)} \prod_{i=1}^{k} \alpha_{\Phi(i)} \prod_{ij\in E(F)} \beta_{\Phi(i)\Phi(j)}$$

for any simple graph F $(k = |V(F)|)$. We consider mainly injective homomorphisms $\Phi \in \mathrm{Inj}(F,G)$ and use the notation:

$$\alpha_\Phi = \prod_{i=1}^{k} \alpha_{\Phi(i)}, \qquad inj_\Phi(F,G) = \prod_{ij\in E(F)} \beta_{\Phi(i)\Phi(j)},$$

$$ind_\Phi(F,G) = \prod_{ij\in E(F)} \beta_{\Phi(i)\Phi(j)} \prod_{ij\in E(\bar{F})} (1-\beta_{\Phi(i)\Phi(j)}),$$

$$inj(F,G) = \sum_{\Phi \in \mathrm{Inj}(F,G)} \alpha_\Phi \cdot inj_\Phi(F,G),$$

$$ind(F,G) = \sum_{\Phi \in \mathrm{Inj}(F,G)} \alpha_\Phi \cdot ind_\Phi(F,G).$$

As for any $\Phi \in \mathrm{Inj}(F,G)$

$$inj_\Phi(F,G) = \sum_{F' \supseteq F} ind_\Phi(F',G),$$

$$inj(F,G) = \sum_{F' \supseteq F} ind(F',G)$$

also holds, and therefore the convergence of $t(F,G_n)$ implies the convergence of $t_{\mathrm{ind}}(F,G_n) = ind(F,G_n)/(\alpha_{G_n})^k$, that roughly equals $\mathbb{P}(\xi(k,G_n) = F)$.

Randomization procedures

A simple graph on k nodes is randomized out of the weighted graph G. For large $|V(G)| = n$ the following procedures give similar results, as far as $\mathbb{P}(\xi(k, G) = F)$ is concerned.

1. k vertices are chosen with replacement with respective probabilities $\alpha_i(G)/\alpha(G)$ $(i = 1, \ldots, n)$. Given the node-set $\{\Phi(1), \ldots, \Phi(k)\}$, the edges come into existence conditionally independently, with probabilities of the edge-weights. $\xi_1(k, G)$ is the resulting random graph.

$$\mathbb{P}(\xi_1(k, G) = F \,|\, \Phi \in \mathrm{Inj}(F, G)) = \sum_{\Phi \in \mathrm{Inj}(F,G)} \frac{\alpha_\Phi}{(\alpha_G)^k} \cdot ind_\Phi(F, G).$$

(By graphon-randomization we may get back F even if Φ is not injective. If $k \ll n$, then most of the homomorphisms are injective.)

2. k vertices are chosen without replacement one after the other, etc. $\xi_2(k,G)$ is the resulting random graph.

$$\mathbb{P}(\xi_2(k,G)=F)=\sum_{\Phi\in \mathrm{Inj}(F,G)} \frac{\alpha_\Phi}{\prod_{i=1}^k \sum_{j\notin\{\Phi(1),\dots,\Phi(i-1)\}} \alpha_j} \cdot ind_\Phi(F,G),$$

where $\{\Phi(1),\dots,\Phi(i-1)\}=\emptyset$, if $i=1$.

3. k vertices are chosen at once, etc. $\xi_3(k,G)$ is the resulting random graph.

$$\mathbb{P}(\xi_3(k,G)=F)=\sum_{\Phi\in \mathrm{Inj}(F,G)} \frac{\alpha_\Phi}{k!(\alpha)_k} \cdot ind_\Phi(F,G),$$

where $(\alpha)_k$ is the elementary symmetric polynomial of degree k of the variables $\alpha_1,\dots,\alpha_n$.

Testability of the minimum k-way cut densities

$f_k(G)$ is testable, though $f_k(G_n) \to 0$ if there is no dominant node-weight. So, this is of not much use.
(See example: $f_k(G_n) \le \frac{\alpha_{max}(G_n)}{\alpha_{G_n}} \cdot \frac{\alpha_{G_n} - \alpha_{max}(G_n)}{\alpha_{G_n}} \to 0$.)
$f_k^c(G)$ is testable for any $c \le 1/k$.
The proof is based on the 3rd equivalent statement of the Theorem.

$$f_k(G) = \min_{P \in \mathcal{P}_k} \frac{1}{\alpha_G^2} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} e_G(V_i, V_j) =$$

$$= \min_{P \in \mathcal{P}_k} f_k(G; V_1, \dots, V_k),$$

where the minimum is taken over k-partitions $P = (V_1, \dots, V_k)$ of the vertices.

$f_k(G)$ is extended to graphons:

$$\tilde{f}_k(W) := \inf_{Q \in \mathcal{Q}_k} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \iint_{S_i \times S_j} w(x,y)\, dx\, dy =$$

$$= \inf_{Q \in \mathcal{Q}_k} \tilde{f}_k(W; S_1, \dots, S_k),$$

where the infimum is taken over the Lebesgue-measurable k-partitions $Q = (S_1, \dots, S_k)$ of [0,1] ($\sum_{i=1}^{k} \lambda(S_i) = 1$), and $0 \le w(x,y) \le 1$ is the two-variable symmetric function assigned to the graphon W.

Similarly, for a given $0 < c \leq 1/k$

$$f_k^c(G) = \min_{P \in \mathcal{P}_k^c} \frac{1}{\alpha_G^2} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} e_G(V_i, V_j) = \min_{P \in \mathcal{P}_k^c} f_k(G; V_1, \dots, V_k),$$

where the minimum is taken for the c-balanced k-partitions $P = (V_1, \dots, V_k)$ of the vertices.
$f_k^c(G)$ is extended to graphons:

$$\tilde{f}_k^c(W) := \inf_{Q \in \mathcal{Q}_k^c} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \iint_{S_i \times S_j} w(x,y)\,dx\,dy = \inf_{Q \in \mathcal{Q}_k^c} \tilde{f}_k(W; S_1, \dots, S_k),$$

where the infimum is taken over the Lebesgue-measurable k-partitions $Q = (S_1, \dots, S_k)$ of $[0,1]$ such that $\lambda(S_i) \geq c$ $(i = 1, \dots, k)$.

- First we show that both $\tilde{f}_k(W)$ and $\tilde{f}_k^c(W)$ are continuous in the cut-norm. It follows trivially by

$$0 \le \sup_{S\subset[0,1]} \Big| \iint_{S\times\bar{S}} w(x,y)\,dx\,dy \Big| \le$$

$$\le \sup_{S,T\subset[0,1]} \Big| \iint_{S\times T} w(x,y)\,dx\,dy \Big| = \|W\|_{\square}.$$

- Next we show that $\tilde{f}_k(W_{G_n}) - f_k(G_n) \to 0$ and $\tilde{f}_k{}^c(W_{G_n}) - f_k^c(G_n) \to 0$ whenever $\max_i \alpha_i(G_n)/\alpha_{G_n} \to 0$. The local infima of $\tilde{f}_k(W; S_1, \ldots, S_k)$ are taken over the k-partitions of [0,1] measurable with respect to the algebra generated by $I_1, \ldots, I_n$, where $\lambda(I_j) = \alpha_j/\alpha_G$. The global infima cannot differ too much.

Relation to the ground state energies

$$f_k(G) = \min_{\Phi: V(G)\to[k]} \mathcal{E}_\Phi(G, \mathbf{J}, \mathbf{0})$$

where the magnetic field is $\mathbf{0}$ and $\mathbf{J} \in \mathrm{Sym}_k$ is the following: $J_{ii} = 0$ ($i = 1, \ldots, k$) and $J_{ij} = -1/2$ ($i \neq j$). By Theorem 2.15 of Borgs et al. II the left-convergence of (G_n) implies the convergence of the ground-state energies, that is the testability of f_k.

$$f_k^{\mathbf{a}}(G) = \min_{\Phi\in\Omega_{\mathbf{a}}(G)} \mathcal{E}_\Phi(G, \mathbf{J}, \mathbf{0})$$

with the above J. By Theorem 2.14 of Borgs et al. II the left-convergence of (G_n) is equivalent to the convergence of the microcanonical ground-state energies (for any magnetic field, $\mathbf{J}$, and $\mathbf{a}$) that – with this special $\mathbf{J}$ – also implies the testability of $f_k^{\mathbf{a}}$ for any distribution $\mathbf{a}$ over $[k]$.

Testability of the weighted minimum k-way cut densities

μ_k is not testable:
We can show an example where $\mu_k(G_n) \to 0$, but randomizing a sufficiently large part of G_n, the weighted minimum k-way cut density of that part is constant.
The testability of μ_k^c can be proved the same way as that of f_k^c, making use of the fact that that, due to

$$\frac{1}{\alpha_{V_i}} \leq \frac{1}{c \cdot \alpha_G},$$

the integrand is bounded from above.

Blown-up weight matrices

From now on, the node-weights are 1, and a weighted graph G on n vertices is identified with its $n \times n$ symmetric weight matrix $\mathbf{A}$. $G_{\mathbf{A}}$ denotes the weighted graph with unit node-weights and edge-weights in $\mathbf{A}$.

Definition

The $n \times n$ symmetric random matrix $\mathbf{W}$ is a Wigner-noise if its entries w_{ij} $(1 \le i \le j \le n)$ are independent random variables, $\mathbb{E}(w_{ij}) = 0$, the w_{ij}'s are uniformly bounded, and there is a constant $\sigma > 0$ (that won't change by n) such that $\mathrm{var}\,(w_{ij}) \ge \sigma^2$, $\forall i, j$.

Though, the main results of this paper can be extended to w_{ij}'s with any light-tail distribution (especially to Gaussian distributed w_{ij}'s), our almost sure results are based on the assumptions of this definition.

Definition

The $n \times n$ symmetric real matrix $\mathbf{B}$ is a blown-up matrix, if there is a $k \times k$ symmetric so-called pattern matrix $\mathbf{P}$ with entries $0 \leq p_{ij} \leq 1$, and there are positive integers $n_1, \dots, n_k$ with $\sum_{i=1}^{k} n_i = n$, such that the matrix $\mathbf{B}$ can be divided into $k \times k$ blocks, where block (i, j) is an $n_i \times n_j$ matrix with entries equal to p_{ij} $(1 \leq i, j \leq n)$.

Such schemes are sought for in microarray analysis and they are called chess-board patterns, cf. Kluger et al., Genome Research, 2003.

Fix $\mathbf{P}$, blow it up to an $n \times n$ matrix $\mathbf{B}_n$, and consider the noisy matrix $\mathbf{A}_n := \mathbf{B}_n + \mathbf{W}_n$.
As $\operatorname{rank}(\mathbf{B}_n) = k$ and $\|\mathbf{W}_n\| = \mathcal{O}(\sqrt{n})$ almost surely ($n \to \infty$), the noisy matrix $\mathbf{A}_n$ almost surely has k protruding eigenvalues (of order n), and all the other eigenvalues are of order $\sqrt{n} \longrightarrow$ spectral gap between the k largest and the other eigenvalues.
$\mathbf{X}_n := (\mathbf{x}_1, \dots, \mathbf{x}_k)$ $n \times k$: eigenvectors of $\mathbf{A}_n$.
Rows of $\mathbf{X}_n$: $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^k$ vertex representatives of G_{A_n}.

$$S_k^2(\mathbf{X}_n) := \sum_{i=1}^{k} \sum_{j \in V_i} \|\mathbf{x}^j - \bar{\mathbf{x}}^i\|^2, \quad \text{where} \quad \bar{\mathbf{x}}^i = \frac{1}{n_i} \sum_{j \in V_i} \mathbf{x}^j .$$

k-variance of the representatives.

Theorem

$$S_k^2(\mathbf{X}_n) = \mathcal{O}\left(\frac{1}{n}\right)$$

almost surely, under the growth condition $n_i/n \geq c$ $(i = 1, \dots, k)$.

Cut-norm of the Wigner-noise

In the other direction: For sufficiently large n, under some conditions, we can separate an $n \times n$ symmetric "error-matrix" $\mathbf{E}$ from $\mathbf{A}$, such that $\|\mathbf{E}\| = \mathcal{O}(\sqrt{n})$ and the remaining matrix $\mathbf{A} - \mathbf{E}$ is a blown-up matrix $\mathbf{B}$ of "low rank" $\longrightarrow$ $G_{\mathbf{B}}$ is a weighted graph with homogeneous edge-densities within the clusters (determined by the blow-up).

It resembles to the weak Szemerédi-partition, but the error-term is bounded in spectral norm, instead of the cut-norm.

However, by large deviations, we can prove that

$$\|\mathbf{W}_{G_{W_n}}\|_{\square} \to 0 \quad \text{almost surely as} \quad n \to \infty,$$

and hence, if $\mathbf{A}_n = \mathbf{B}_n + \mathbf{W}_n$, then

$$G_{\mathbf{A}_n} \to G_{\mathbf{P}} \quad \text{almost surely as} \quad n \to \infty$$

(left-convergence), where $\mathbf{P}$ is the $k \times k$ pattern matrix.