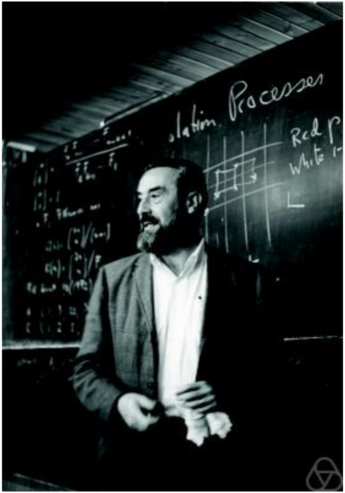
Spin models, Random graphs, and Semidefinite relaxations

Andrea Montanari

Stanford University

May 15, 2015

Seymour Sherman (1917-1977)



Worked in

- Statistical mechanics
- Probability theory
- Information theory
  ('Non-mean-square error criteria'
  1958)

I knew well one result of his! GHS inequality

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 3 MARCH 1970

Concavity of Magnetization of an Ising Ferromagnet in a Positive External Field

ROBERT B. GRIFFITHS*
Institute for Theoretical Physics,† *State University of New York at Stony Brook, Stony Brook, New York* 11790

C. A. HURST
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia 5001

AND

S. SHERMAN‡
Department of Mathematics, Indiana University, Bloomington, Indiana 47401
and
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia 5001

(Received 23 June 1969)

I knew well one result of his! GHS inequality

II. NOTATION AND PRINCIPAL THEOREM

Consider an Ising model with spins σ_i, $i = 1, 2 \cdots n$, which take the values ± 1, and a Hamiltonian

$$\mathcal{H} = -\sum_{1 \le i < j \le n}\sum J_{ij}\sigma_i\sigma_j - H\sum_{i=1}^{n}\sigma_i, \tag{2.1}$$

where H is the external magnetic field and we assume that

$$J_{ij} \geq 0, \tag{2.2}$$

that is, all exchange interactions are ferromagnetic. Let

$$s = \sigma_1 + \sigma_2 + \cdots + \sigma_n \tag{2.3}$$

and let

$$M = \langle s \rangle / n \tag{2.4}$$

be the average magnetization per spin. Here angular brackets denote a thermal average. It follows from general convexity considerations[4] [for which (2.2) is not necessary] that

$$\frac{\partial M}{\partial H} \geq 0 \tag{2.5}$$

for all H. We wish to establish the following:

Theorem 1:

$$\frac{\partial^2 M}{\partial H^2} \leq 0, \quad \text{for all} \quad H \geq 0, \tag{2.6}$$

This talk

$$\mathcal{H}(\sigma) = \frac{1}{2} \sum_{i,j=1}^{n} B_{ij}\sigma_i\sigma_i = \frac{1}{2}\langle \sigma, B\sigma \rangle$$

$$\mu_{\beta,B}(\sigma) = \frac{1}{Z(\beta, B)}\, e^{\beta\mathcal{H}(\sigma)}$$

Will consider $B = B(G)$, G sparse random graphs

This talk

$$\mathcal{H}(\sigma) = \frac{1}{2}\sum_{i,j=1}^{n} B_{ij}\sigma_i\sigma_i = \frac{1}{2}\langle\sigma, B\sigma\rangle$$

$$\mu_{\beta,B}(\sigma) = \frac{1}{Z(\beta, B)}\, e^{\beta\mathcal{H}(\sigma)}$$

Will consider $B = B(G)$, G sparse random graphs

This talk

$$\mathcal{H}(\sigma) = \frac{1}{2} \sum_{i,j=1}^{n} B_{ij} \sigma_i \sigma_i = \frac{1}{2} \langle \sigma, B\sigma \rangle$$

$$\mu_{\beta,B}(\sigma) = \frac{1}{Z(\beta, B)} e^{\beta \mathcal{H}(\sigma)}$$

Will consider $B = B(G)$, G sparse random graphs

Outline

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

- $|V| = n$ vertices
- Each edge present with probability γ/n

Random regular graph $G = (V, E) \sim \mathcal{G}^{\text{reg}}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

- $|V| = n$ vertices
- Each edge present with probability γ/n

Random regular graph $G = (V, E) \sim \mathcal{G}^{\text{reg}}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

- $|V| = n$ vertices
- Each edge present with probability γ/n

Random regular graph $G = (V, E) \sim \mathcal{G}^{\text{reg}}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

- $|V| = n$ vertices
- Each edge present with probability γ/n

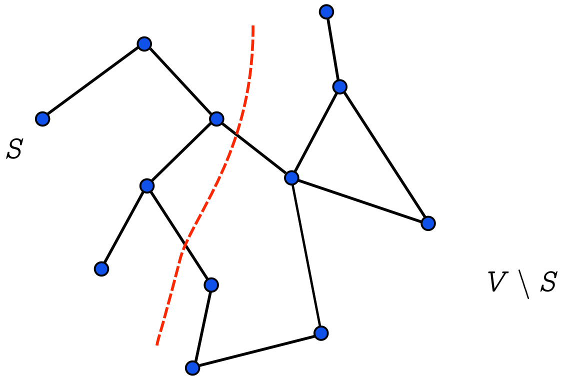
Random regular graph $G = (V, E) \sim \mathcal{G}^{\rm reg}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Extremal cuts

Cuts



$$\mathrm{cut}_G(S) \equiv \Big|\{(i,j) \in E :\ i \in S,\, j \in V \setminus S\}\Big|$$

Interesting for many reasons

- Clustering similarity matrices
- Graph layout
- Community structure in social networks
- ...

Extremal cuts

Minimum bisection

$$\mathsf{mcut}(G) = \min\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\ \right\}$$

Maximum bisection

$$\mathsf{MCUT}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\ \right\}$$

Maximum Cut

$$\mathsf{MaxCut}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\ \right\}$$

Extremal cuts

Minimum bisection

$$\mathsf{mcut}(G) = \min\left\{\text{cut}_G(S)\;:\; S\subseteq V\,,\;|S| = n/2\right\}$$

Maximum bisection

$$\mathsf{MCUT}(G) = \max\left\{\text{cut}_G(S)\;:\; S\subseteq V\,,\;|S| = n/2\right\}$$

Maximum Cut

$$\mathsf{MaxCut}(G) = \max\left\{\text{cut}_G(S)\;:\; S\subseteq V\right\}$$

Extremal cuts

Minimum bisection

$$\mathsf{mcut}(G) = \min\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\ \right\}$$

Maximum bisection

$$\mathsf{MCUT}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\ \right\}$$

Maximum Cut

$$\mathsf{MaxCut}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\ \right\}$$

A long history

- Bollobas, 1984
- Alon, 1997
- Coppersmith, Gamarnik, Hajiaghayi, G. Sorkin, 2004
- Díaz, Serna, and Wormald, 2007
- Daudé, Martínez, Rasendrahasina, Ravelomanana, 2012
- Gamarnik, Li, 2014
- ...

Typical result:
If $G \sim \mathcal{G}(n, \gamma/n)$ then, with high probability

$$\frac{n\gamma}{4} + C_1 n\sqrt{\gamma} \le \mathsf{MCUT}(G) \le \frac{n\gamma}{4} + C_2 n\sqrt{\gamma}$$

A long history

- Bollobas, 1984
- Alon, 1997
- Coppersmith, Gamarnik, Hajiaghayi, G. Sorkin, 2004
- Díaz, Serna, and Wormald, 2007
- Daudé, Martínez, Rasendrahasina, Ravelomanana, 2012
- Gamarnik, Li, 2014
- ...

Typical result:
If $G \sim \mathcal{G}(n, \gamma/n)$ then, with high probability

$$\frac{n\gamma}{4} + C_1 n\sqrt{\gamma} \le \mathsf{MCUT}(G) \le \frac{n\gamma}{4} + C_2 n\sqrt{\gamma}$$

Insights from statistical physics

Conjectures

- Fu, Anderson, 1986
- Mézard, Parisi, 2001
- Zdéborova, Boettcher, 2010

Theorem (Franz, Leone, Toninelli, 2003, Bayati, Gamarnik, Tetali, 2009)

The following limit exists, for $G_n \sim \mathcal{G}(n, \gamma/n), \mathcal{G}^{\rm reg}(n, \gamma)$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{MaxCut}(G_n)$$

Subadditivity: no limit value.

Insights from statistical physics

Conjectures

- Fu, Anderson, 1986
- Mézard, Parisi, 2001
- Zdéborova, Boettcher, 2010

Theorem (Franz, Leone, Toninelli, 2003, Bayati, Gamarnik, Tetali, 2009)

The following limit exists, for $G_n \sim \mathcal{G}(n, \gamma/n), \mathcal{G}^{\rm reg}(n, \gamma)$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{MaxCut}(G_n)$$

Subaddittivity: no limit value.

Insights from statistical physics

Conjectures

- Fu, Anderson, 1986
- Mézard, Parisi, 2001
- Zdéborova, Boettcher, 2010

Theorem (Franz, Leone, Toninelli, 2003, Bayati, Gamarnik, Tetali, 2009)

The following limit exists, for $G_n \sim \mathcal{G}(n, \gamma/n), \mathcal{G}^{\rm reg}(n, \gamma)$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{MaxCut}(G_n)$$

Subaddittivity: no limit value.

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\mathrm{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\mathrm{cut}(S) \geq \mathbb{E}\mathrm{cut}(S) + \Delta\Big\} \leq \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2} \mathrm{cut}(S) \geq \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \leq 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\text{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\text{cut}(S) \geq \mathbb{E}\text{cut}(S) + \Delta\Big\} \leq \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2}\text{cut}(S) \geq \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \leq 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\text{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\text{cut}(S) \geq \mathbb{E}\text{cut}(S) + \Delta\Big\} \leq \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2} \text{cut}(S) \geq \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \leq 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\text{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\text{cut}(S) \geq \mathbb{E}\text{cut}(S) + \Delta\Big\} \leq \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2} \text{cut}(S) \geq \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \leq 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Hence, with high probability

$$\mathsf{MCUT}(G) \le \frac{n\gamma}{4} + C_2 n\sqrt{\gamma}$$

In other words

The $n\gamma/4$ term is 'trivial'...

A new result

Theorem (Dembo, Montanari, Sen, 2015)

Assume $G \sim \mathcal{G}(n, \gamma/n)$ *or* $G \sim \mathcal{G}^{\rm reg}(n, \gamma/n)$. *Then, with high probability,*

$$\frac{1}{n}\mathsf{mcut}(G) = \frac{\gamma}{4} - \mathsf{P}_*\sqrt{\frac{\gamma}{4}} + o_\gamma(\sqrt{\gamma})\,,$$

$$\frac{1}{n}\mathsf{MCUT}(G) = \frac{\gamma}{4} + \mathsf{P}_*\sqrt{\frac{\gamma}{4}} + o_\gamma(\sqrt{\gamma})\,,$$

$$\frac{1}{n}\mathsf{MaxCut}(G) = \frac{\gamma}{4} + \mathsf{P}_*\sqrt{\frac{\gamma}{4}} + o_\gamma(\sqrt{\gamma})\,.$$

where $\mathsf{P}_* = \dots$ *(wait a minute).*

Remarks

- The $\sqrt{\gamma}$ term has a well defined coefficient.
- The coefficient has a formula.
- It is the same for 3 problems, 2 graph models.
- $\mathsf{P}_* \approx 0.7632$

What is P_*?

GOE random matrix:

$$J \in \mathbb{R}^{n\times n}, \quad J = J^{\mathsf{T}}, \quad (J_{ij})_{i<j} \sim \mathrm{N}(0, 1/n), \quad J_{ii} \sim \mathrm{N}(0, 2/n).$$

Sherrington-Kirkpatrick spin-glass model

$$\mathcal{H}_{\mathrm{SK}}(\sigma) \equiv \frac{1}{2}\langle \sigma, J\sigma\rangle .$$

Finally

$$\mathrm{P}_* \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E} \max_{\sigma\in\{+1,-1\}^n} \mathcal{H}_{\mathrm{SK}}(\sigma).$$

What is P_*?

GOE random matrix:

$$J \in \mathbb{R}^{n\times n}, \quad J = J^{\mathsf{T}}, \quad (J_{ij})_{i<j} \sim \mathrm{N}(0, 1/n), \quad J_{ii} \sim \mathrm{N}(0, 2/n).$$

Sherrington-Kirkpatrick spin-glass model

$$\mathcal{H}_{\rm SK}(\sigma) \equiv \frac{1}{2}\langle \sigma, J\sigma\rangle .$$

Finally

$$P_* \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E} \max_{\sigma\in\{+1,-1\}^n} \mathcal{H}_{\rm SK}(\sigma) .$$

What is P_*?

GOE random matrix:

$$J \in \mathbb{R}^{n\times n}\,, \quad J = J^{\mathsf{T}}\,, \quad (J_{ij})_{i<j} \sim \mathrm{N}(0, 1/n), \quad J_{ii} \sim \mathrm{N}(0, 2/n)\,.$$

Sherrington-Kirkpatrick spin-glass model

$$\mathcal{H}_{\mathrm{SK}}(\sigma) \equiv \frac{1}{2}\langle \sigma, J\sigma\rangle\,.$$

Finally

$$\mathsf{P}_* \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E} \max_{\sigma\in\{+1,-1\}^n} \mathcal{H}_{\mathrm{SK}}(\sigma)\,.$$

What is P_*?

$$\mathsf{P}_* \equiv \lim_{n\to\infty} \frac{1}{n} \mathbb{E} \max_{\sigma\in\{+1,-1\}^n} \mathcal{H}_{\rm SK}(\sigma)\, .$$

- The limit exists (consequence of Guerra, Toninelli, 2002)
- Given by *'Parisi's formula'* (Talagrand, 2006)
- Clarifies why 'standard' combinatorial methods were unsuccessful

What is P_*?

$$\mathsf{P}_* \equiv \lim_{n\to\infty} \frac{1}{n} \mathbb{E} \max_{\sigma\in\{+1,-1\}^n} \mathcal{H}_{\rm SK}(\sigma)\,.$$

- The limit exists (consequence of Guerra, Toninelli, 2002)
- Given by *'Parisi's formula'* (Talagrand, 2006)
- Clarifies why 'standard' combinatorial methods were unsuccessful

(Informal) Implication

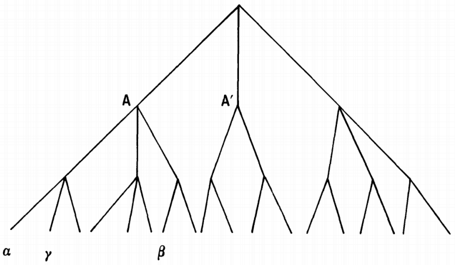


Fig. 1. — The tree of the states. The different states α, β, γ... are the extremities of the branches of the tree. The distance between two states is a monotonic function of the number of steps one has to climb along the tree to find a common ancestor.

MaxCut, max-bisection, min-bisection have ∞-RSB structure

Proof strategy

1. Proof for mcut, MCUT with $G \sim \mathcal{G}(n, \gamma/n)$:
   Interpolation

2. Extend to $G \sim \mathcal{G}^{\rm reg}(n, \gamma)$:
   Coupling between $\mathcal{G}(n, \gamma/n)$ and $\mathcal{G}^{\rm reg}(n, \gamma)$
   (Tricky: G' and G'' differ in about $n\sqrt{\gamma}$ edges)

3. Prove that MaxCut$(G) \approx$ MCUT(G)
   (Tricky)

Proof strategy

1. Proof for mcut, MCUT with $G \sim \mathcal{G}(n, \gamma/n)$:
   Interpolation

2. Extend to $G \sim \mathcal{G}^{\rm reg}(n, \gamma)$:
   Coupling between $\mathcal{G}(n, \gamma/n)$ and $\mathcal{G}^{\rm reg}(n, \gamma)$
   (Tricky: G' and G'' differ in about $n\sqrt{\gamma}$ edges)

3. Prove that MaxCut$(G) \approx$ MCUT(G)
   (Tricky)

Proof strategy

1. Proof for mcut, MCUT with $G \sim \mathcal{G}(n, \gamma/n)$:
   Interpolation

2. Extend to $G \sim \mathcal{G}^{\rm reg}(n, \gamma)$:
   Coupling between $\mathcal{G}(n, \gamma/n)$ and $\mathcal{G}^{\rm reg}(n, \gamma)$
   (Tricky: G' and G'' differ in about $n\sqrt{\gamma}$ edges)

3. Prove that MaxCut$(G) \approx$ MCUT(G)
   (Tricky)

Proof strategy

1. Proof for mcut, MCUT with $G \sim \mathcal{G}(n, \gamma/n)$:
   Interpolation

2. Extend to $G \sim \mathcal{G}^{\rm reg}(n, \gamma)$:
   Coupling between $\mathcal{G}(n, \gamma/n)$ and $\mathcal{G}^{\rm reg}(n, \gamma)$
   (Tricky: G' and G'' differ in about $n\sqrt{\gamma}$ edges)

3. Prove that MaxCut$(G) \approx$ MCUT(G)
   (Tricky)

Proof strategy

1. Proof for mcut, MCUT with $G \sim \mathcal{G}(n, \gamma/n)$:
   Interpolation

2. Extend to $G \sim \mathcal{G}^{\rm reg}(n, \gamma)$:
   Coupling between $\mathcal{G}(n, \gamma/n)$ and $\mathcal{G}^{\rm reg}(n, \gamma)$
   (Tricky: G' and G'' differ in about $n\sqrt{\gamma}$ edges)

3. Prove that $\mathsf{MaxCut}(G) \approx \mathsf{MCUT}(G)$
   (Tricky)

Interpolation (mcut)

Concentration $\Rightarrow$ Work with expectations

Interpolation (mcut)

Hamiltonian $\mathcal{H}_G : \Omega_n \to \mathbb{R}$

$$\mathcal{H}_G(\sigma) = \frac{1}{\sqrt{\gamma}} \sum_{(i,j)\in E} \sigma_i \sigma_j \, ,$$

$$\Omega_n = \Big\{ \sigma \in \{+1,-1\}^n : \; \sum_{i=1}^{n} \sigma_i = 0 \Big\} \, .$$

$$\text{mcut}(G) = \min_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \Big(\frac{1-\sigma_i\sigma_j}{2}\Big)$$

$$= \frac{1}{2}|E| - \frac{1}{2} \max_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \sigma_i\sigma_j$$

$$= \frac{n\gamma}{4} - \sqrt{\frac{\gamma}{4}} \max_{\sigma\in\Omega_n} \mathcal{H}_G(\sigma) \, .$$

Interpolation (mcut)

Hamiltonian $\mathcal{H}_G : \Omega_n \to \mathbb{R}$

$$\mathcal{H}_G(\sigma) = \frac{1}{\sqrt{\gamma}} \sum_{(i,j)\in E} \sigma_i \sigma_j \, ,$$

$$\Omega_n = \Big\{ \sigma \in \{+1,-1\}^n : \; \sum_{i=1}^{n} \sigma_i = 0 \Big\} \, .$$

$$\begin{aligned} \mathrm{mcut}(G) &= \min_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \Big(\frac{1-\sigma_i\sigma_j}{2} \Big) \\ &= \frac{1}{2}|E| - \frac{1}{2} \max_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \sigma_i\sigma_j \\ &= \frac{n\gamma}{4} - \sqrt{\frac{\gamma}{4}} \max_{\sigma\in\Omega_n} \mathcal{H}_G(\sigma) \, . \end{aligned}$$

Interpolation (mcut)

We want to show

$$\frac{1}{n}\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_G(\sigma) = \mathsf{P}_* + o_\gamma(1) + o_n(1)\,.$$

We know

$$\frac{1}{n}\mathbb{E}\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_{\rm SK}(\sigma) = \mathsf{P}_* + o_n(1)\,.$$

Idea: Interpolation ($\sim$ smart path method, $\sim$ Lindeberg method, ...)

Two steps

Step 1: For $t \in [0,1]$, $G(t) = (V, E(t)) \sim \mathcal{G}(n, \gamma\, t/n)$

$$\mathcal{H}_t(\sigma) = \frac{1}{\sqrt{\gamma}} \sum_{(i,j)\in E(t)} \sigma_i\sigma_j + \frac{1}{2}\sqrt{1-t}\sum_{i,j=1}^{n} J_{ij}\sigma_i\sigma_j$$

Step 2: Move from $\max_{\sigma\in\Omega_n}(\,\cdots\,)$ to $\max_{\sigma\in\{+1,-1\}^n}(\,\cdots\,)$.

Step 1: Key trick

Free energy density ('smooth max')

$$\phi_n(\beta; t) \equiv \frac{1}{n}\mathbb{E}\log\Big\{\sum_{\sigma\in\Omega_n} e^{\beta\mathcal{H}_t(\sigma)}\Big\}$$

Lemma

$$\Big|\frac{1}{\beta}\phi_n(\beta; t) - \frac{1}{n}\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_t(\sigma)\Big| \le \frac{\log 2}{\beta}\,.$$

Step 1: Key trick

Free energy density ('smooth max')

$$\phi_n(\beta;t) \equiv \frac{1}{n}\mathbb{E}\log\Big\{\sum_{\sigma\in\Omega_n} e^{\beta\mathcal{H}_t(\sigma)}\Big\}$$

Lemma

$$\Big|\frac{1}{\beta}\phi_n(\beta;t) - \frac{1}{n}\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_t(\sigma)\Big| \le \frac{\log 2}{\beta}\,.$$

Step 1: Key trick

Lemma

$$\left|\frac{\partial \phi_n}{\partial \beta}(\beta; t)\right| \le C\,\frac{\beta^3}{\gamma^{1/2}}$$

Proof.

$$\frac{\partial \phi(\beta; t)}{\partial t} = I + II,$$

where

$$I = \beta\sqrt{\gamma}\frac{1}{n^2}\sum_{ij}\mathbb{E}[\mu_{\beta;t}(\langle\sigma_i,\sigma_j\rangle)] + \frac{\beta^2}{n^2}\sum_{ij}\mathbb{E}[\mu_{\beta;t}(\langle\sigma_i,\sigma_j\rangle^2) - \mu_{\beta;t}(\langle\sigma_i,\sigma_j\rangle)$$

$$II = -\frac{\gamma}{2n^2}\sum_{ij}\mathbb{E}\left[\log\mu_{\beta;t}\Big(\exp(\frac{2\beta}{\sqrt{\gamma}}\langle\sigma_i,\sigma_j\rangle)\Big)\right] + o(1)\,.$$

Step 1: Key trick

Lemma

$$\left| \frac{\partial \phi_n}{\partial \beta}(\beta; t) \right| \leq C \, \frac{\beta^3}{\gamma^{1/2}}$$

Proof.

$$\frac{\partial \phi(\beta; t)}{\partial t} = I + II,$$

where

$$I = \beta\sqrt{\gamma}\frac{1}{n^2}\sum_{ij}\mathbb{E}[\mu_{\beta;t}(\langle\sigma_i,\sigma_j\rangle)] + \frac{\beta^2}{n^2}\sum_{ij}\mathbb{E}[\mu_{\beta;t}(\langle\sigma_i,\sigma_j\rangle^2) - \mu_{\beta;t}(\langle\sigma_i,\sigma_j\rangle)$$

$$II = -\frac{\gamma}{2n^2}\sum_{ij}\mathbb{E}\left[\log \mu_{\beta;t}\Big(\exp(\frac{2\beta}{\sqrt{\gamma}}\langle\sigma_i,\sigma_j\rangle)\Big)\right] + o(1)\,.$$

Semidefinite programming

Basic remark

$$\left.\begin{array}{r}\mathsf{mcut}\\ \mathsf{MCUT}\\ \mathsf{MaxCut}\end{array}\right\}\text{ are NP-hard}$$

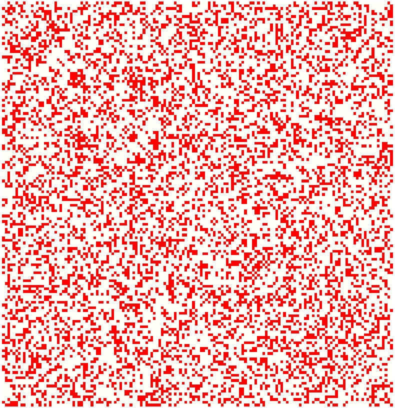
(and hard to approximate)

(. . . very rich theory)

What about random instances?

(relevant for machine learning, statistics, signal processing,. . .)

Adjacency matrix



$$(A_G)_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Consider min bisection

$$
\begin{aligned}
\text{maximize} \quad & \langle \sigma, A_G \sigma \rangle \,, \\
\text{subject to} \quad & \sum_{i=1}^{n} \sigma_i = 0 \,, \\
& \sigma_i \in \{+1, -1\} \,.
\end{aligned}
$$

Lagrangian

$$\begin{aligned} \text{maximize} \quad & \langle \sigma, A_G \sigma \rangle - \lambda \Big(\sum_{i=1}^{n} \sigma_i \Big)^2 , \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} . \end{aligned}$$

$\lambda =$ regularization parameter

Lagrangian

$$\begin{aligned} \text{maximize} \quad & \langle \sigma, (A_G - \lambda \mathbf{1}\mathbf{1}^{\mathsf{T}})\sigma \rangle \,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \,. \end{aligned}$$

$\lambda =$ regularization parameter

Sparse random graphs

- $G \sim \mathcal{G}(n, \gamma/n)$
- 'Right' regularization $\lambda = \gamma/n$

Lagrangian

$$\begin{aligned} \text{maximize} \quad & \langle \sigma, \left(A_G - \frac{\gamma}{n} \mathbf{1}\,\mathbf{1}^{\mathsf{T}} \right) \sigma \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\}\,. \end{aligned}$$

Lagrangian

$$\begin{aligned} \text{maximize} \quad & \langle \sigma, (A_G - \mathbb{E}\{A_G\})\sigma \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\}\,. \end{aligned}$$

Value (first part of the talk)

$$\begin{aligned} \mathsf{OPT}(G) &= 2n\mathsf{P}_*\sqrt{\gamma} + n\, o_\gamma(\sqrt{\gamma}) \\ &\approx 1.5264\, n\sqrt{\gamma} \end{aligned}$$

Lagrangian

$$\begin{aligned} \text{maximize} \quad & \langle \sigma, (A_G - \mathbb{E}\{A_G\})\sigma \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\}\,. \end{aligned}$$

Value (first part of the talk)

$$\begin{aligned} \mathsf{OPT}(G) &= 2n\mathsf{P}_*\sqrt{\gamma} + n\, o_\gamma(\sqrt{\gamma}) \\ &\approx 1.5264\, n\sqrt{\gamma} \end{aligned}$$

Idea: Convex relaxations

Spectral relaxation

$$\begin{aligned} \text{maximize} \quad & \langle \sigma, (A_G - \mathbb{E}\{A_G\})\sigma \rangle \,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \quad . \end{aligned}$$

Spectral relaxation

$$
\begin{aligned}
\text{maximize} \quad & \langle \sigma, (A_G - \mathbb{E}\{A_G\})\sigma \rangle\,, \\
\text{subject to} \quad & \sout{\sigma_i \in \{+1, -1\}} \;\; \|\sigma\|_2 = n\,.
\end{aligned}
$$

Spectral relaxation

$$\begin{aligned} \text{maximize} \quad & \langle \sigma, (A_G - \mathbb{E}\{A_G\})\sigma \rangle \,, \\ \text{subject to} \quad & \cancel{\sigma_i \in \{+1, -1\}} \;\; \|\sigma\|_2 = n \,. \end{aligned}$$

Value

$$\lambda_{\max}(A_G - \mathbb{E}A_G)\, n$$

Spectral relaxation very bad in the sparse regime!

Theorem (Krivelevich, Sudakov 2003+Vu 2005)

If $G \sim \mathcal{G}(n, \gamma/n)$*, then, with high probability,*

$$\lambda_{\max}(A_G - \mathbb{E}A_G) = \begin{cases} 2\sqrt{\gamma}\,(1 + o(1)) & \text{if } \gamma \gg (\log n)^4, \\ \sqrt{\log n/(\log\log n)}(1 + o(1)) & \text{if } \gamma = O(1). \end{cases}$$

Compare with $\mathrm{OPT}(G)/n \approx 1.5264\,\sqrt{\gamma}$

Spectral relaxation very bad in the sparse regime!

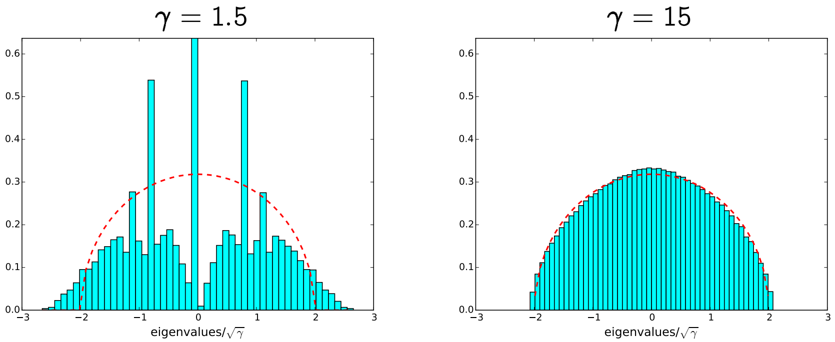
Theorem (Krivelevich, Sudakov 2003+Vu 2005)

If $G \sim \mathcal{G}(n, \gamma/n)$, then, with high probability,

$$\lambda_{\max}(A_G - \mathbb{E}A_G) = \begin{cases} 2\sqrt{\gamma}\,(1 + o(1)) & \text{if } \gamma \gg (\log n)^4, \\ \sqrt{\log n/(\log\log n)}(1 + o(1)) & \text{if } \gamma = O(1). \end{cases}$$

Compare with $\mathrm{OPT}(G)/n \approx 1.5264\,\sqrt{\gamma}$

Illustration: $n = 10,000$



Semidefinite Programming (SDP)

$$\begin{aligned} \text{maximize} \quad & \langle (A_G - \mathbb{E}\{A_G\}), \sigma\sigma^{\mathsf{T}} \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\}\,. \end{aligned}$$

SDP(G)

$$\begin{aligned} \text{maximize} \quad & \langle (A_G - \mathbb{E}\{A_G\}), X \rangle\,, \\ \text{subject to} \quad & X_{ii} = 1\,, \\ & X \succeq 0\,. \end{aligned}$$

Semidefinite Programming (SDP)

$$\begin{aligned} \text{maximize} \quad & \langle (A_G - \mathbb{E}\{A_G\}), \sigma\sigma^{\mathsf{T}} \rangle \,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \,. \end{aligned}$$

SDP(G)

$$\begin{aligned} \text{maximize} \quad & \langle (A_G - \mathbb{E}\{A_G\}), X \rangle \,, \\ \text{subject to} \quad & X_{ii} = 1 \,, \\ & X \succeq 0 \,. \end{aligned}$$

SDP is continuous sparse–dense

Theorem (Montanari, Sen 2015)

If $G \sim \mathcal{G}(n; \gamma/n)$, $\gamma = O(1)$, then, with high probability,

$$\frac{1}{n}\mathsf{SDP}(G) = 2\sqrt{\gamma} + o(\sqrt{\gamma})$$

- $\mathsf{SDP}(G)$ behaves much better than the principal eigenvalue

SDP is continuous sparse–dense

Theorem (Montanari, Sen 2015)

If $G \sim \mathcal{G}(n; \gamma/n)$, $\gamma = O(1)$, then, with high probability,

$$\frac{1}{n}\mathsf{SDP}(G) = 2\sqrt{\gamma} + o(\sqrt{\gamma})$$

- $\mathsf{SDP}(G)$ behaves much better than the principal eigenvalue

Proof strategy

Define $\mathsf{OPT}_k(G)$

$$\begin{aligned} \text{maximize} \quad & \langle (A_G - \mathbb{E}\{A_G\}), X \rangle\,, \\ \text{subject to} \quad & X_{ii} = 1\,, \\ & X \succeq 0\,, \\ & \text{rank}(X) = k \end{aligned}$$

- $k = 1$: Minimum bisection
- $k = n$: SDP

Proof strategy

Define $\mathsf{OPT}_k(G)$

$$\begin{aligned} \text{maximize} \quad & \langle (A_G - \mathbb{E}\{A_G\}), X \rangle\,, \\ \text{subject to} \quad & X_{ii} = 1\,, \\ & X \succeq 0\,, \\ & \text{rank}(X) = k \end{aligned}$$

- $k = 1$: Minimum bisection
- $k = n$: SDP

Proof strategy

$\mathsf{OPT}_k(G)$, equivalent definition

$$\begin{aligned} \text{maximize} \quad & \langle (A_G - \mathbb{E}\{A_G\}), \boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}} \rangle \,, \\ \text{subject to} \quad & \boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n)^{\mathsf{T}} \in \mathbb{R}^{n\times k} \,, \\ & \boldsymbol{\sigma} \in \mathbb{R}^k \,, \quad \|\boldsymbol{\sigma}_i\|_2 = 1 \,. \end{aligned}$$

Proof strategy

$\mathsf{OPT}_k(G)$, equivalent definition

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j=1}^{n} \big(A_G - \mathbb{E}\{A_G\}\big)_{i,j} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \,, \\ \text{subject to} \quad & \boldsymbol{\sigma} \in \mathbb{R}^k \,, \quad \|\boldsymbol{\sigma}_i\|_2 = 1 \,. \end{aligned}$$

Spin model!

Proof strategy

$\mathsf{OPT}_k(G)$, equivalent definition

$$\begin{aligned} &\text{maximize} \quad \sum_{i,j=1}^{n} \big(A_G - \mathbb{E}\{A_G\}\big)_{i,j} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \,, \\ &\text{subject to} \quad \boldsymbol{\sigma} \in \mathbb{R}^k \,, \quad \|\boldsymbol{\sigma}_i\|_2 = 1 \,. \end{aligned}$$

Spin model!

Proof strategy

- Higher-rank Grothendieck inequality

$$\mathsf{SDP}(G) - \frac{\text{const.}}{k} \le \mathsf{OPT}_k(G) \le \mathsf{SDP}(G)\,.$$

- Interpolation for $\mathsf{OPT}_k(G)$ (free energy)

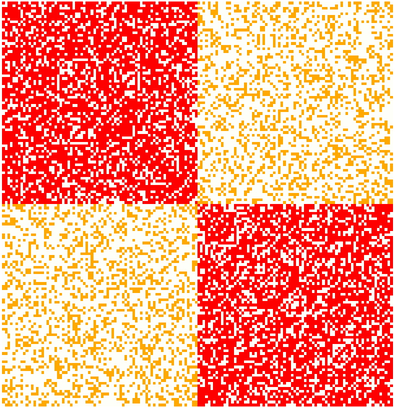
$$(A_G - \mathbb{E}\{A_G\}) \longleftrightarrow J \sim \text{GOE}(n)$$

- Analyze SDP for $J \sim \text{GOE}(n)$.

One application

The hidden partition problem

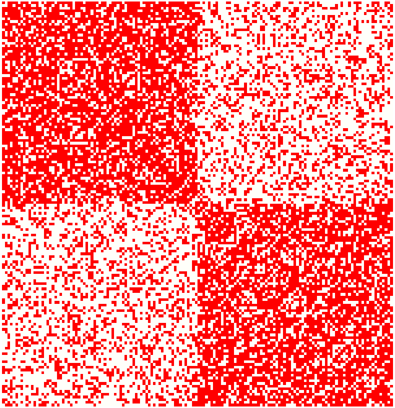
(a.k.a. community detection)



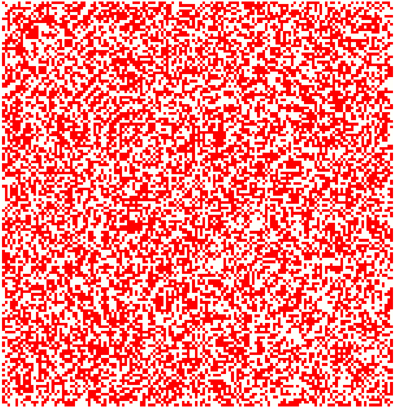
Vertices V, $|V| = n$, $V = S_1 \cup S_2$, $|S_1| = |S_2| = n/2$

$$\mathbb{P}\{(i,j) \in E\} = \begin{cases} p & \text{if } \{i,j\} \subseteq S_1 \text{ or } \{i,j\} \subseteq S_2, \\ q < p & \text{otherwise.} \end{cases}$$

Of course entries are not colored. . .



. . . and rows/columns are not ordered



Background: Hypothesis testing (sparse graph $p = a/n$, $q = b/n$)

Hypothesis H_0:

$$\mathbb{P}\{(i,j) \in E\} = \frac{a+b}{2n}$$

Hypothesis H_1: $V = S_1 \cup S_2$, $|S_1| = |S_2| = n/2$

$$\mathbb{P}\{(i,j) \in E\} = \begin{cases} a/n & \text{if } \{i,j\} \subseteq S_1 \text{ or } \{i,j\} \subseteq S_2, \\ b/n & \text{otherwise.} \end{cases}$$

Detection thresholds

Classical Statistics / Information Theory:
Is there *any algorithm* that detects the hidden structure?

Computational:
Is there *a poly-time algorithm* that detects the hidden structure?

Detection thresholds

Classical Statistics / Information Theory:
Is there *any algorithm* that detects the hidden structure?

Computational:
Is there *a poly-time algorithm* that detects the hidden structure?

Information theory threshold

Theorem (Mossel, Neeman, Sly, 2012)

There is a test that succeed with high probability if and only if $a + b > 2$ *and*

$$\frac{a - b}{\sqrt{2(a + b)}} > 1\,.$$

[Proves conjecture by Decelle, Krzakala, Moore, Zdeborova, 2011]

Information theory threshold

Theorem (Mossel, Neeman, Sly, 2012)

There is a test that succeed with high probability if and only if $a + b > 2$ *and*

$$\frac{a-b}{\sqrt{2(a+b)}} > 1\,.$$

[Proves conjecture by Decelle, Krzakala, Moore, Zdeborova, 2011]

Computational threshold

- Dyer, Frieze 1989 — $p = na > q = nb$ fixed.
- Condon, Karp 2001 — $a - b \gg n^{1/2}$
- McSherry 2001 — $a - b \gg \sqrt{b \log n}$
- Coja-Oghlan 2010 — $a - b \gg \sqrt{b}$
- Massoulie 2013 and Mossel, Neeman, Sly, 2013

$$\frac{a-b}{\sqrt{2(a+b)}} > 1$$

Very ingenious algorithms!

Computational threshold

- Dyer, Frieze 1989 — $p = na > q = nb$ fixed.
- Condon, Karp 2001 — $a - b \gg n^{1/2}$
- McSherry 2001 — $a - b \gg \sqrt{b \log n}$
- Coja-Oghlan 2010 — $a - b \gg \sqrt{b}$
- Massoulie 2013 and Mossel, Neeman, Sly, 2013

$$\frac{a-b}{\sqrt{2(a+b)}} > 1$$

Very ingenious algorithms!

Computational threshold

- Dyer, Frieze 1989 $p = na > q = nb$ fixed.
- Condon, Karp 2001 $a - b \gg n^{1/2}$
- McSherry 2001 $a - b \gg \sqrt{b \log n}$
- Coja-Oghlan 2010 $a - b \gg \sqrt{b}$
- Massoulie 2013 and Mossel, Neeman, Sly, 2013

$$\frac{a-b}{\sqrt{2(a+b)}} > 1$$

Very ingenious algorithms!

What if I am not ingenious?

$$T_{\mathsf{SDP}}(G) = \begin{cases} 1 & \text{if } \mathsf{SDP}(G) \geq \theta_*, \\ 0 & \text{otherwise.} \end{cases}$$

- This is really **off-the-shelf**
- How well does it work?

It works well

Theorem (Montanari, Sen 2015)

The test based on $\mathsf{SDP}(G)$ *works with high probability if*

$$\frac{a-b}{\sqrt{2(a+b)}} > 2 + o_{a+b}(1)\,.$$

- At most suboptimal by factor $2 + o_{a+b}(1)$
- Under a radom matrix theory conjecture: $2 \to 1$
- Earlier result by Guedon, Vershynin 2015: $(a-b) \geq 10^4\sqrt{a+b}$.
- (studied also in Abbe, Bandeira, Hall 2014+ Hajek, Wu, Xu 2015)

It works well

Theorem (Montanari, Sen 2015)

The test based on SDP(G) *works with high probability if*

$$\frac{a-b}{\sqrt{2(a+b)}} > 2 + o_{a+b}(1)\,.$$

- At most suboptimal by factor $2 + o_{a+b}(1)$
- Under a radom matrix theory conjecture: $2 \to 1$
- Earlier result by Guedon, Vershynin 2015: $(a-b) \geq 10^4\sqrt{a+b}$.
- (studied also in Abbe, Bandeira, Hall 2014+ Hajek, Wu, Xu 2015)

It works well

Theorem (Montanari, Sen 2015)

The test based on SDP(G) *works with high probability if*

$$\frac{a-b}{\sqrt{2(a+b)}} > 2 + o_{a+b}(1)\,.$$

- At most suboptimal by factor $2 + o_{a+b}(1)$
- Under a radom matrix theory conjecture: $2 \to 1$
- Earlier result by Guedon, Vershynin 2015: $(a-b) \geq 10^4\sqrt{a+b}$.
- (studied also in Abbe, Bandeira, Hall 2014+ Hajek, Wu, Xu 2015)

Conclusion

Conclusion

Two families of spin models on sparse random graphs

- **Ising spins** $\sigma_i \in \{+1, -1\}$
  - Extremal cuts in random graphs
  - Much remains to understand
- **Vector spins** $\boldsymbol{\sigma}_i \in \mathsf{S}^{k-1}$
  - Semidefinite relaxations
  - Even more remains to understand!

Thanks!

Conclusion

Two families of spin models on sparse random graphs

- **Ising spins $\sigma_i \in \{+1, -1\}$**
  - Extremal cuts in random graphs
  - Much remains to understand

- **Vector spins $\boldsymbol{\sigma}_i \in \mathsf{S}^{k-1}$**
  - Semidefinite relaxations
  - Even more remains to understand!

Thanks!

Conclusion

Two families of spin models on sparse random graphs

- **Ising spins** $\sigma_i \in \{+1, -1\}$
  - Extremal cuts in random graphs
  - Much remains to understand

- **Vector spins** $\boldsymbol{\sigma}_i \in \mathsf{S}^{k-1}$
  - Semidefinite relaxations
  - Even more remains to understand!

Thanks!

Conclusion

Two families of spin models on sparse random graphs

- **Ising spins** $\sigma_i \in \{+1, -1\}$
  - Extremal cuts in random graphs
  - Much remains to understand

- **Vector spins** $\boldsymbol{\sigma}_i \in \mathsf{S}^{k-1}$
  - Semidefinite relaxations
  - Even more remains to understand!

Thanks!