Computational trade-offs in graph models

Quentin Berthet

Google Research

Simons Institute - 2021

Modern data landscape

- Data-driven approaches prevalent in most scientific fields
- Important aspect: Data collected without discernment
  - Most of the data not relevant to the problem at hand.
  - Data can be complex: heterogeneity, privacy, errors.
- **Intrinsic** computational costs to large datasets in complex models?

Libraries of Babel, past and present

Computational aspects of high-dimensional statistics

- Flood of data, high dimensional problems:
  - Higgs boson: 800 trillion events/year. Genome: 3,000 megabase pairs.
  - **Data:** $X_1, \ldots, X_n$ **High-dimensional parameter:** $X_i \sim \mathbf{P}_v$, $v \in \mathbf{R}^d$.
- Structure, sparsity $\rightarrow$ combinatorial problems in likelihood methods.
  - **Estimation** $\rightarrow$ **Optimization**

$$\hat{v} \in \operatorname*{argmax}_{v \in \mathcal{S}} \ell_{\mathbf{X},n}(v) \, .$$

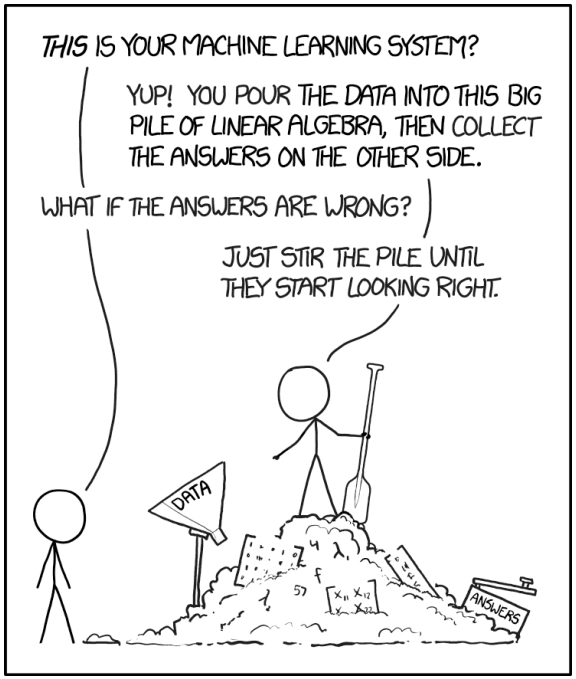
  - **Hypothesis testing** $\rightarrow$ **Averages**

$$\Psi(\mathbf{X}, n) : \quad \frac{1}{|\mathcal{S}_1|} \sum_{v \in \mathcal{S}_1} L_{\mathbf{X},n}(v) > \frac{1}{|\mathcal{S}_0|} \sum_{v \in \mathcal{S}_0} L_{\mathbf{X},n}(v) \, .$$

- Objective: computationally efficient statistical methods.

Recent progress in Machine Learning

- In practice: using heuristics to solve optimization problems
- Very good empirical results, not a **complete** theoretical understanding



- Are all "learning problems" computationally easy?
- Mathematical analysis reveals finer understanding

Framework - Toy models

- Detection or estimation of structured signal of intensity
  - **Absence of signal** Some high-dimensional distribution $\mathbf{P}_0$.
  - **Structured signal** For some $v \in \mathcal{S}$, $\mathbf{P}_{\theta,v}$, for some $\theta > 0$.
- **Examples:**
  - $\mathbf{P}_0 = \mathcal{N}(0, I_d)$ and $\mathbf{P}_{\theta,v}$, for k-sparse $v = \mathbf{1}_S$; $\mathbf{P}_{\theta,v} = \mathcal{N}(\theta v, I_d)$.
  - $\mathbf{P}_0 = \mathcal{N}(0, I_d)$ and $\mathbf{P}_{\theta,v}$, for k-sparse $v = \mathbf{1}_S$; $\mathbf{P}_{\theta,v} = \mathcal{N}(0, I_d + \theta v v^\top)$.
  - $\mathbf{P}_0 = \mathrm{Ber}(1/2, G)$ and $\mathbf{P}_{\theta,H} = \mathrm{Ber}(1/2, G) \cup \mathrm{Ber}(1/2 + \theta, H)$.
- Active line of research in statistics, initially removed from algorithmic concerns.

  Ingster (1982), Donoho & Jin (2004) Addario-Berry et al (2010),

  Arias-Castro & Verzelen (2013), B. & Rigollet (2013),...
- Formulated as hypothesis testing problem based on independent $X_1 \dots, X_n$

$$H_0 : X_i \sim \mathbf{P}_0 \, ; \qquad H_1 : X_i \sim \mathbf{P}_{\theta,v} \, , v \in \mathcal{S}$$

Approach - Toy models

- A test $\psi : \mathcal{X} \to \{0, 1\}$ has a performance measured by probability of error

$$\mathbf{P}_0(\psi = 1) + \max_{v \in \mathcal{S}} \mathbf{P}_{\theta,v}(\psi = 0)\,, \quad \text{or} \quad \mathbf{P}_0(\psi = 1) + \frac{1}{|\mathcal{S}|} \sum_{v \in \mathcal{S}} \mathbf{P}_{\theta,v}(\psi = 0)$$

- How large should $\theta > 0$ be to have a error $\leq \delta$, for $\delta \in (0, 1)$?
- Related to the notion of distance between $\mathbf{P}_0$ and $\mathbf{P}_{\theta,\mathcal{S}} = \frac{1}{|\mathcal{S}|} \sum_{v \in \mathcal{S}} \mathbf{P}_{\theta,v}$.

Information-theoretic tools

- Maximum likelihood ratio test as a combinatorial optimization problem

$$\max_{v \in \mathcal{S}} \frac{\mathrm{d}\mathbf{P}_{\theta,v}}{\mathrm{d}\mathbf{P}_0}(X) > k\,.$$

Computational aspects of high-dimensional statistics

- What if likelihood-based method is NP-hard?

  Sparse linear regression, stochastic block model for clustering

- Computational limits in statistics:
  - Worst-case hardness: 0th order information.
  - Algorithm for frequent instances of NP-hard problem:

    Clustering, Nonconvex regression, Alternating minimization.
  - Proxy functional and optimization problem.

    Lasso, Convex relaxations.

- Average-case hypotheses:

  Some task is hard to achieve consistently and efficiently.

Computational hardness in statistics

- Hardness of a statistical problem (information theory)

$$|\mathbf{P}_0(\Psi = 0) - \mathbf{P}_1(\Psi = 0)| \leq \varepsilon\,.$$

  For all tests $\Psi \to \{0, 1\}$, **statistical difficulty**.

- Hardness of a statistical problem (complexity theory)
  - For some test Ψ,

$$|\mathbf{P}_0(\Psi = 0) - \mathbf{P}_1(\Psi = 0)| \approx 1\,.$$

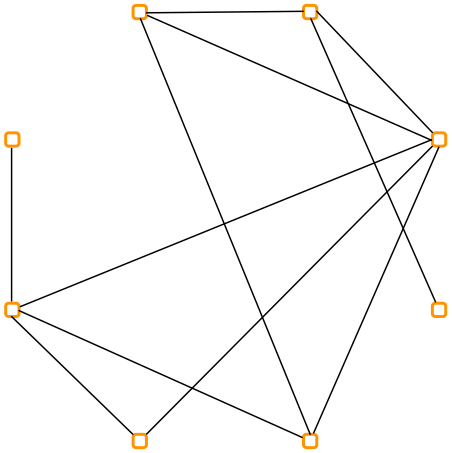
  - For all "tractable" tests Φ,

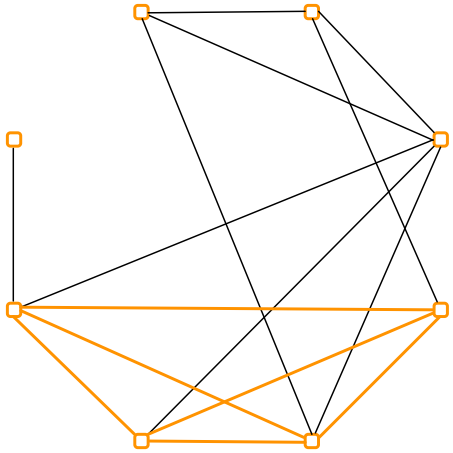
$$|\mathbf{P}_0(\Phi = 0) - \mathbf{P}_1(\Phi = 0)| \leq \varepsilon\,.$$

  Statistical difficulty for tractable methods.

Planted clique

- **Assumption:** No polynomial-time algorithm detects the presence of a planted clique in regime $k = n^c$, $c < 1/2$.





- Wide range of evidence for this assumption Jerrum (92), Feige and Krauthgamer (92), Alon et al. (98), Juels and Peinado (00), Ames and Vavasis (11), Dekel et al. (10), Feige et al. (00),Feldman et al. (12) Alon et al. (07), Hazan et al. (11)
- **Consequences:** Reductions based on this problem to establish hardness of other statistical problems (Sparse PCA, Planted dense subgraph, Sparse Gaussian mixtures, etc.)

Link prediction with logistic model

Nicolai Baldin

- **Statistical and computational rates in graph logistic regression**

  Q. B., N. Baldin

  AISTATS 2019

Statistical study of graphs

- Observed graph $G = ([n], E)$.
- Link between presence of edge $(i, j) \in E$ and properties of vertices i and j.
- **Stochastic block model:** unsupervised problem, for $\tau : [n] \to \{1, 2\}$

$$\mathbf{P}\big((i, j) \in E\big) = P_{\tau(i)\tau(j)}\,,$$

recover $\tau(i)$ for all $i \in [n]$. Graph G generated by model.

Holland et al. (83), Mossel et al. (14), Javanmard et al. (16), Abbe et al. (17)

- **Total variation denoising:** supervised problem, for $i \in [n]$

$$y_i = \theta_i^* + \varepsilon_i \qquad \sum_{(i,j)\in E} |\theta_i^* - \theta_j^*|^q \leq R\,,$$

estimate θ^*, using the graph as structure.

Tibshirani et al (16-17), Hütter and Rigollet (16)

Link prediction: problem description

- **Data:**
  - Covariates $X_i \in \mathbf{R}^d$ for $i \in [n]$, information about each vertex, deterministic.
  - Partial observation of G, $Y_{(i,j)} \in \{0,1\}$ for $(i,j) \in \Omega$, subset of size N

- **Generative model:** For some symmetric $\Theta^* \in \mathbf{R}^{d \times d}$, and $(i,j) \in \Omega$

$$\mathbf{P}\big(\underbrace{(i,j) \in E}_{Y_{(i,j)}=1}\big) = \sigma(X_i^\top \Theta^* X_j) = \frac{1}{1+\exp(-X_i^\top \Theta^* X_j)}\,, \quad \text{indep.}$$

- **Objective:** Estimate Θ^* under a block-sparse and low-rank assumption

$$\Theta^* \in \mathcal{P}_{k,r}(M) = \big\{\Theta \in \mathbf{S}_d : \|\Theta\|_{0,0} \le k\,,\ \ \mathsf{rank}(\Theta) \le r\,,\ \ \|\Theta\|_{1,1} \le M\big\}\,.$$

- **Challenges:** High-dimensional problem $d^2 \gg N$, combinatorial constraints.

Motivation

- Generating model

$$\mathbf{P}\big(\underbrace{(i,j)\in E}_{Y_{(i,j)}=1}\big) = \sigma(X_i^\top \Theta^* X_j) = \frac{1}{1+\exp(-X_i^\top \Theta^* X_j)}$$

- Probability of edge (i,j) depends on an **affinity**.

$$X_i^\top \Theta^* X_j = \sum_{s,t\in[d]} X_i^{(s)} X_j^{(t)} \Theta^*_{s,t}$$
$$= \sum_{\ell=1}^{r} \lambda_\ell (X_i^\top u_\ell)(X_j^\top u_\ell)\,.$$

- Affinity depends on $(X^\top u_\ell)$, weighted by λ_ℓ with
  - Few vectors u_ℓ **(low rank)**.
  - Few relevant coefficients of X_i, X_j **(block sparsity)**.
- Goes beyond geometric description: $\lambda_\ell \leq 0$ is possible.

Relationship with other models

- **Metric learning:** n objects, each with representation $V_i \in \mathbf{R}^r$

  Queries of **similarity** S_{ij} between object i and j,

$$\mathbf{P}\big(S_{ij} = 1\big) = \sigma(V_i^\top V_j) = \frac{1}{1 + \exp(-V_i^\top V_j)}\,.$$

  Recovered in this model by $X_i = e_i$, $n = p$, $\Theta^* = VV^\top$, $X_i^\top \Theta^* X_j = V_i^\top V_j$.

  Special case of our model for parameter space $\Theta^* \succeq 0$ and $\mathrm{rank}(\Theta^*) = r$.

- **Graph prediction** Biau and Bleakley (06), joint distribution on (X_i, X_j, Y_{ij})

$$\mathbf{P}(Y_{ij} = 1 \,|\, X_i, X_j) = \eta(X_i, X_j)$$

  Classification setting, through empirical risk minimization, objective

$$\min_g \mathbf{P}_{X,X',Y}(g(X, X') \neq Y)$$

Generalized linear models

- Matrix logistic regression: can be rewritten

$$X_i^\top \Theta^* X_j = \langle X_j X_i^\top, \Theta^* \rangle = \mathbb{D}_\Omega^{(i,j)} \cdot \mathsf{vec}(\Theta^*)$$

For $\mathbb{D}_\Omega \in \mathbf{R}^{N \times d^2}$ with rows $\mathbb{D}_\Omega^{(i,j)} = \mathsf{vec}(X_j X_i^\top)$ and $\mathsf{vec}(\Theta^*) \in \mathbf{R}^{d^2}$,

$$\mathbf{P}\big(y_{(i,j)} = 1\big) = \frac{1}{1 + \exp(-\mathbb{D}_\Omega^{(i,j)} \cdot \mathsf{vec}(\Theta^*))}$$

with $y \in \mathbf{R}^N$ indexed by $(i,j) \in \Omega$. van de Geer (08)

- Structure assumption are deeply connected to matrix formulation.
- Logistic version of trace regression models, for covariates $A_{(i,j)} = X_j X_i^\top$

$$X_i^\top \Theta^* X_j = \mathbf{Tr}(A_{(i,j)}^\top \Theta^*)\,.$$

Rohde and Tsybakov (11), Koltchinskii et al. (16), Fan et al. (17)

Assumptions

- Problem ill-posed for $N \leq d^2$, structural assumptions (block-sparsity and rank).

- Probabilities bounded-away from 0 and 1, classical in logistic estimation.

$$\varepsilon < \mathbf{P}(Y_{ij} = 1) < 1 - \varepsilon \Rightarrow |X_i^\top \Theta^* X_j| \leq M$$

  Can be guaranteed by taking $\|X_j X_i^\top\|_\infty \cdot \|\Theta^*\|_{1,1} \leq M$.

  Is satisfied for $\|X_j X_i^\top\|_\infty \leq 1$ and $\|\Theta^*\|_{1,1} \leq M$.

  Required for identifiability, large values impossible to estimate.

  van de Geer (08), Abramovich and Grinshtein (16), Bach (10)

- Assumption on the covariates, akin to restricted isometry assumption, required.

  Candès and Tao (05)

Block isometry property

- Conditioning on covariates $\mathbf{X} \in \mathbf{R}^{d \times N}$: exists $\Delta_{\Omega,s}(\mathbf{X}) \in (0,1)$ such that

$$N\big(1 - \Delta_{\Omega,s}(\mathbf{X})\big)\|B\|_F^2 \leq \|\mathbf{X}^\top B\,\mathbf{X}\|_{F,\Omega}^2 \leq N\big(1 + \Delta_{\Omega,s}(\mathbf{X})\big)\|B\|_F^2\,,$$

  for all s-block-sparse matrices.

- We have $(\mathbf{X}^\top B\,\mathbf{X})_{i,j} = X_i^\top B X_j = \mathbb{D}_\Omega^{(i,j)}\,\mathrm{vec}(B)$ so

$$\mathrm{vec}_\Omega(\mathbf{X}^\top B\,\mathbf{X}) = \mathbb{D}_\Omega \cdot \mathrm{vec}(B)\,.$$

  Conditioning is equivalent to an RIP-type condition on $\mathbb{D}_\Omega/\sqrt{N}$, restricted to vector forms of s-block-sparse matrices. Traonmilin, Gribonval (15)

- Satisfied with high-probability under various randomness models.

  e.g. independent $X_i^{(t)} \in [-1,1]$, random centered, or Ω uniformly random.

Likelihood-based approach

- Log-likelihood

$$\ell_{\mathbf{X},Y}(\Theta) = -\sum_{(i,j)\in\Omega} \log\left(1+\exp\left((2Y_{(i,j)}-1)X_i^\top \Theta X_j\right)\right).$$

- Classically we have

$$\ell_{\mathbf{X},Y}(\Theta) = \ell_{\mathbf{X},\mathsf{pop}}(\Theta) + \langle \xi, \Theta\rangle\,, \quad \xi = \sum_{(i,j)\in\Omega} \big(Y_{ij} - \mathbf{E}_{i,j}[Y_{ij}]\big) X_j X_i^\top\,.$$

$$\ell_{\mathbf{X},\mathsf{pop}}(\Theta) = \ell_{\mathbf{X},\mathsf{pop}}(\Theta^*) - \mathsf{KL}(\mathbf{P}_{\Omega,\Theta^*}, \mathbf{P}_{\Omega,\Theta})\,.$$

- With identifiability assumption and noise model
  - Understanding of the geometry of $\ell_{\mathbf{X},\mathsf{pop}}(\Theta)$ around Θ^*
  - Control of the deviation terms: sum of i.i.d. sub-Gaussian variables.

Results - Statistical complexity

- Under the assumptions on covariates $\mathbf{X}$ and Θ^*, for combinatorial MLEs

$$\mathbf{E}[\|\hat{\Theta} - \Theta^*\|_F^2] \lesssim \frac{1}{1 - \Delta_{\Omega,2k}(\mathbf{X})} \left(\frac{kr}{N} + \frac{k}{N} \log \left(\frac{de}{k} \right) \right)$$

- Terms correspond to statistical complexities of
  - Estimating a $k \times k$ symmetric matrix of rank r.
  - Finding the support of size k among d variables.
- Matching lower bounds up to constants.
- Both estimators are not computationally tractable.
- Can we get both? Optimal statistical rates and polynomial-time estimator.

Results - Convex relaxation

- Relaxation of the block-sparsity penalty in an ℓ_1, coefficient-wise penalty

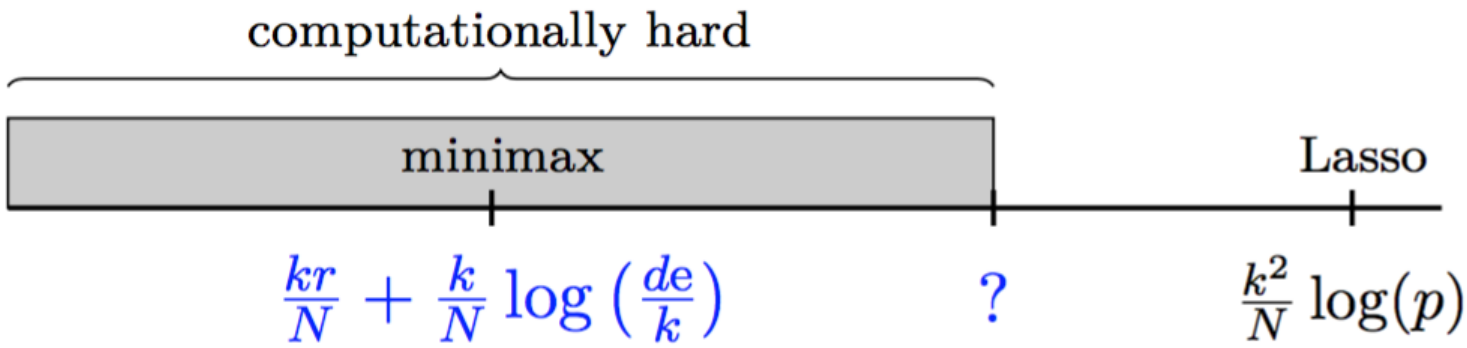
$$\hat{\Theta}_{1,1} \in \operatorname{argmin} -\ell_{\mathbf{X},Y}(\Theta) + \lambda \|\Theta\|_{1,1} \, .$$

- Written in vector form, equivalent to logistic LASSO. van de Geer (08)
- For $\lambda = C\sqrt{\log(d)}$ we recover that

$$\mathbf{E}[\|\hat{\Theta}_{1,1} - \Theta^*\|_F^2] \lesssim \frac{1}{1 - \Delta_{\Omega,2k}(\mathbf{X})} \frac{k^2}{N} \log(d) \, .$$

- Computationally tractable estimator but suboptimal statistical performance.
- Corresponds to the statistical complexity over k^2-sparse matrices: ℓ_1 norm is invariant to permutation of the coefficients, block structure is unseen.

Rates comparaison



Rates of estimation with computational aspects

- All rates scale as expected in effective sample size N and ambient dimension d
- Behavior in other parameters of the problem, big gap for $r \ll k$.
- For fixed constant rank $r = 1, 2, 3$, gap from k to k^2 observed.

Reduction - Informal

Looking for Θ_0 and collection of Θ_S such that for given $\mathbf{X} \in \mathbf{R}^{d \times N}$, with $m = d = n$, $k = n^{1/2-\varepsilon}$, and $q \in (1/2, 1)$

- $\Pi_0 = \sigma(\mathbf{X}^\top \Theta_0 \mathbf{X}) = \mathbf{1}\mathbf{1}^\top/2$, with $\Theta_0 = \mathbf{0}$ for any $\mathbf{X}$.
- $\Pi_S = \sigma(\mathbf{X}^\top \Theta_S \mathbf{X}) = \mathbf{1}\mathbf{1}^\top/2 + (q - 1/2)\mathbf{1}_S\mathbf{1}_S^\top$.
  - Taking $X_i = N^{1/4} e_i = n^{1/2} e_i$, $X_j X_i^\top = N^{1/2} E_{i,j}$, so $\mathbb{D}_\Omega/\sqrt{N} = I_N$
  - For $\alpha > 0$ such that $q = 1/(1 + e^{-\alpha})$
  - Taking $\Theta_S = \alpha \mathbf{1}_S\mathbf{1}_S^\top/\sqrt{N}$, so $\Theta_{S,(i,j)} = \alpha/\sqrt{N} = \alpha/n > 0$ if $(i,j) \in S^2$.
- We have

$$\|\Theta_S - \Theta_0\|_F^2 = \frac{\alpha k^2}{N}\,.$$

Reduction - Informal

Looking for Θ_0 and collection of Θ_S such that for given $\mathbf{X} \in \mathbf{R}^{d \times N}$, with $m = d = n$, $k = n^{1/2-\varepsilon}$, and $q \in (1/2, 1)$

- $\Pi_0 = \sigma(\mathbf{X}^\top \Theta_0 \mathbf{X}) = \mathbf{1}\mathbf{1}^\top/2$, with $\Theta_0 = \mathbf{0}$ for any $\mathbf{X}$.
- $\Pi_S = \sigma(\mathbf{X}^\top \Theta_S \mathbf{X}) = \mathbf{1}\mathbf{1}^\top/2 + (q - 1/2)\mathbf{1}_S\mathbf{1}_S^\top$, for $\mathbf{X}$ rescaled canonical basis.
- We have

$$\|\Theta_S - \Theta_0\|_F^2 = \frac{\alpha k^2}{N}\,.$$

- If there is an estimator $\hat{\Theta}_{\text{poly}}$, computable in polynomial time and for which

$$\mathbf{E}_{\Theta^*}[\|\hat{\Theta}_{\text{poly}} - \Theta^*\|_F^2] \leq \frac{c\alpha k^2}{N}$$

We can run this estimator on known $\mathbf{X}$ and $Y = \text{adj}(G)$, solve planted clique.

Conclusion

- **Contributions**
  - Logistic regression, supervised model for graphs, non i.i.d. edges
  - Optimal statistical rates for combinatorial estimators.
  - Suboptimal performance for LASSO-type estimator, convex relaxation.
  - Computational lower bounds in a regression problem.

- **Open questions**
  - More general reduction scheme, for any covariate design.
  - More general link functions, nonconvex optimization problems.

Ising blockmodel

P. Rigollet (MIT)

P. Srivastava (TIFR)

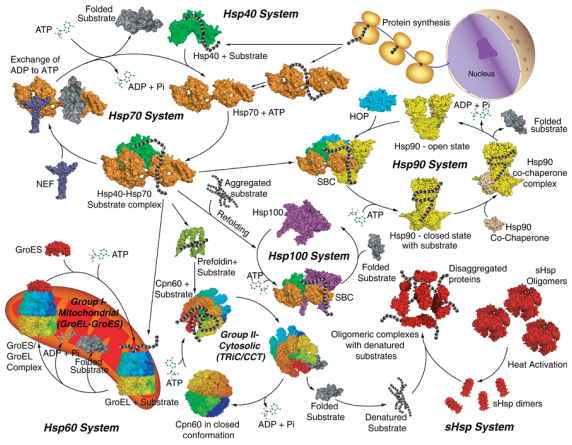
- **Exact recovery in the Ising blockmodel**

  Q. B., P.Rigollet, and P. Srivastava

  Ann. Statis. 2019

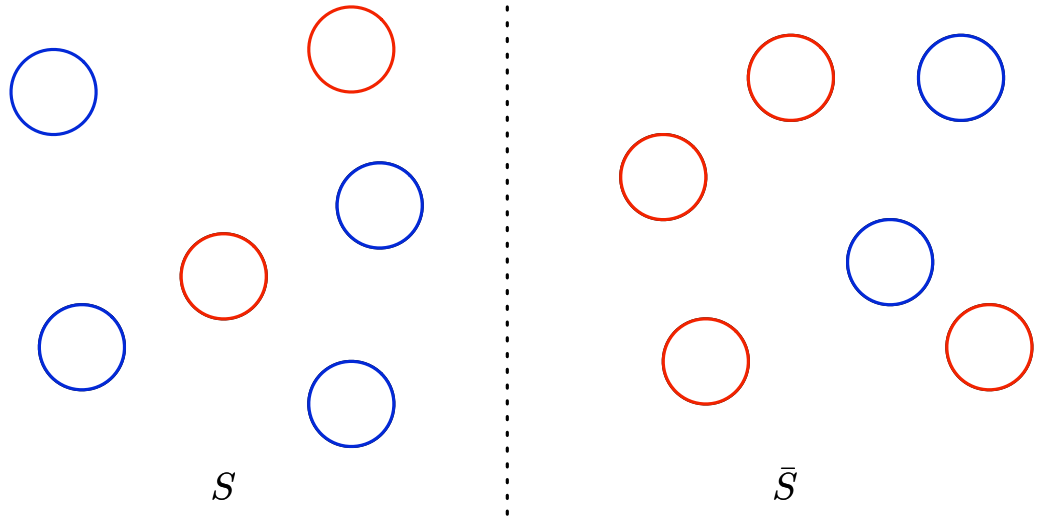
Motivation

- Finding communities in populations, based on similar **behavior** and **influence**.
- One of the justifications for **stochastic blockmodels**
- What if we observe the **behavior**, not the graph?



Motivation

Model with p individuals, observation $\sigma \in \{-1, 1\}^p$, balanced communities $(S, \bar{S})$.



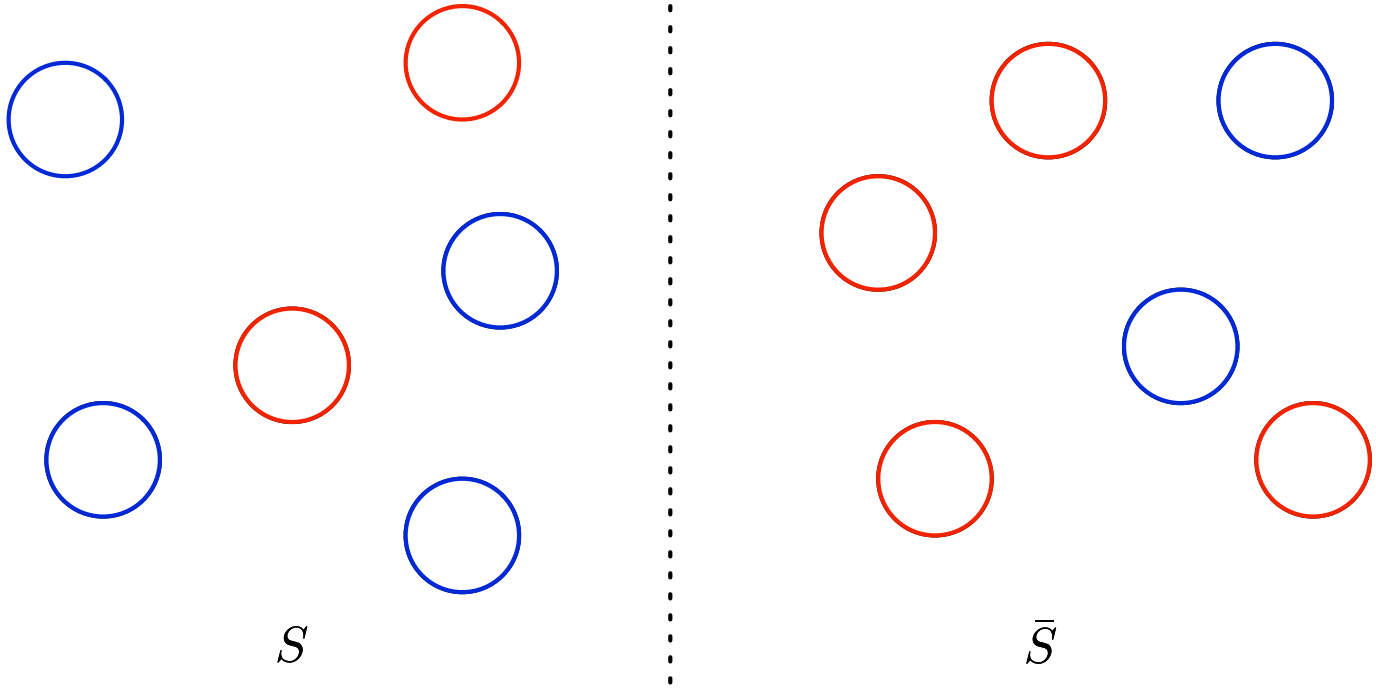
For a given question: members of the community influence an individual's answer.

Example: A metal sheet has temperature at point x, y given by $T(x, y) = T_0 + k(x^2 + y^2)$. What is $T(r, \theta)$?

$+1$) $T(r, \theta) = T_0 + kr^2$

-1) $T(r, \theta) = T_0 + k(r^2 + \theta^2)$

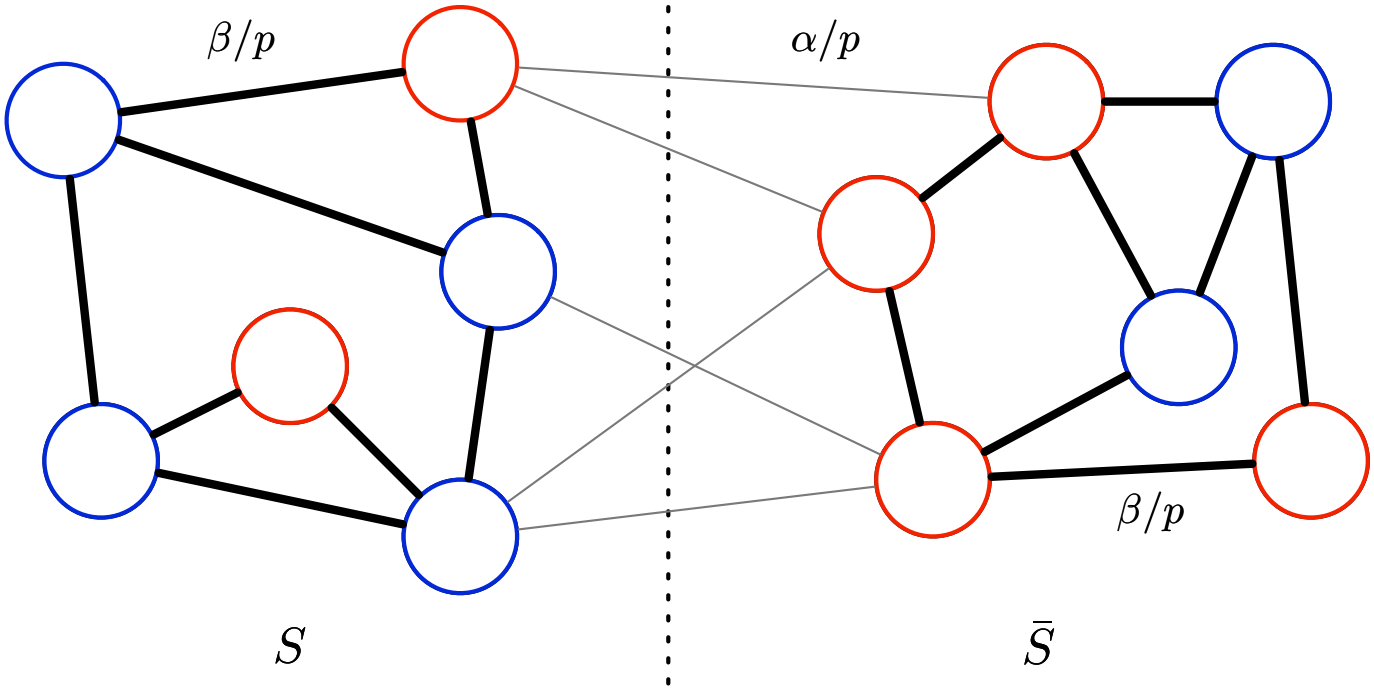
Motivation



Model with p individuals, $\sigma \in \{-1, 1\}^p$ and balanced communities $(S, \bar{S})$.

$$\mathbf{P}_S(\sigma) = \underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{?}.$$

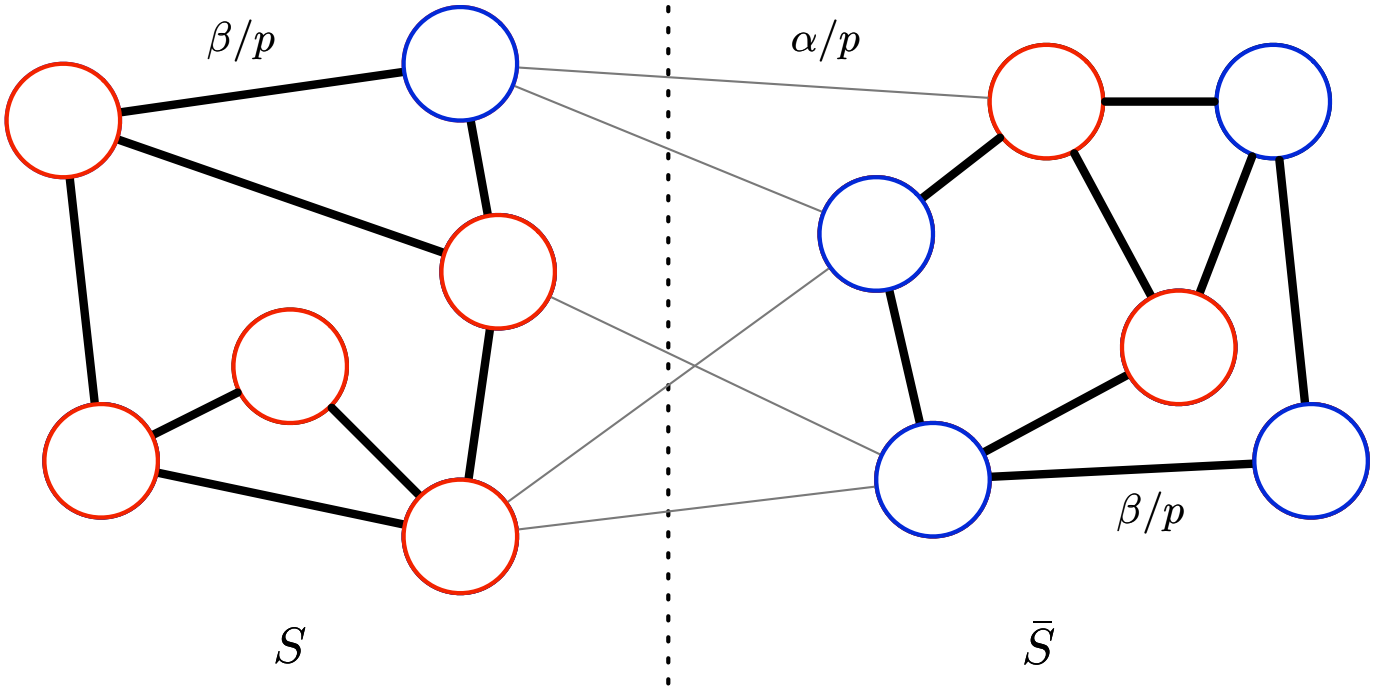
Motivation



Model with p individuals, $\sigma \in \{-1, 1\}^p$ and balanced communities $(S, \bar{S})$.

$$\mathbf{P}_S(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\left[\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\right].$$

Motivation



Model with p individuals, $\sigma \in \{-1, 1\}^p$ and balanced communities $(S, \bar{S})$.

$$\mathbf{P}_S(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\left[\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\right].$$

Problem description

Ising blockmodel:

$$\mathbf{P}_S(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\left[\frac{\beta}{2p}\sum_{i\sim j}\sigma_i\sigma_j + \frac{\alpha}{2p}\sum_{i\nsim j}\sigma_i\sigma_j\right] = \frac{1}{Z_{\alpha,\beta}} \exp\left(-\mathcal{H}_{S,\alpha,\beta}(\sigma)\right).$$

Energy decreases (probability increases) with more agreement inside each block.

- Blockmodel: $\mathbf{P}_S(\sigma_i = \sigma_j) = \begin{cases} b & \text{for all } i \sim j \\ a & \text{for all } i \nsim j \end{cases}$
- Balance: $|S| = |\bar{S}| = p/2$,
- Homophily: $\beta > 0 \Leftrightarrow b > 1/2$,
- Assortativity: $\beta > \alpha \Leftrightarrow b > a$.
- Relationship with Hopfield model, and posterior for the SBM.

Observations: $\sigma^{(1)}, \ldots, \sigma^{(n)} \in \{-1, 1\}^p$ i.i.d. from $\mathbf{P}_S$.

Objective: recover the *balanced* partition $(S, \bar{S})$ from observations.

Stochastic blockmodels

- **one** observation of random graph on p vertices

$$\mathbf{P}(i \leftrightarrow j) = \begin{cases} b & \text{for all } i \sim j \\ a & \text{for all } i \nsim j \end{cases}$$

- Exact recovery using SDP iff

$$a = \mathsf{a}\frac{\log p}{p}, b = \mathsf{b}\frac{\log p}{p}$$

and $(\mathsf{a}+\mathsf{b})/2 > 1 + \sqrt{\mathsf{ab}}$

Abbe, Bandeira, Hall '14
Mossel, Neeman, Sly '14
Xu, Lelarge, Massoulieé '14
Hajek, Wu '16

Graphical models / MRF

- n observations $\sigma^{(1)}, \dots, \sigma^{(n)}$ i.i.d.

$$\mathbf{P}(\sigma) \propto \exp\left[\frac{\beta}{2p}\sum_{i,j} J_{ij}\sigma_i\sigma_j\right]$$

- Goal estimating sparse $J = \{J_{ij}\}$ (max degree d)
- Sample complexity $n \gg 2^d \log p$

Chow-Liu '68
Bresler, Mossel, Sly '08
Santhanam, Wainwright '12
Bresler '15
Vuffray, Misra, Lokhov, Chertkov '16

Wigner matrices

Wishart matrices

Problem overview

- By symmetry $\mathbf{E}[\sigma_i] = 0$, what of the second moment?
- Structure of the problem visible in the **covariance matrix** Σ

$$\Sigma = \mathbf{E}[\sigma\sigma^\top] = \left(\begin{array}{c|c} \Delta & \Omega \\ \hline \Omega & \Delta \end{array}\right) + (1-\Delta) I_p \, .$$

- Difficulty of the problem related with the values of quantities $\Delta, \Omega \in (-1, 1)$

$$\Delta = 2b - 1 \, , \quad \Omega = 2a - 1 \, .$$

- Parallel with the **stochastic block model** on graphs with independent edges
- **Summary - Analysis**
  - **Deviations:** Behavior of $\hat{\Sigma}$ around Σ, sample size guarantees.
  - **Population:** Results function of $\Delta - \Omega$, scaling in p and α, β.

Maximum likelihood estimation

- Log-likelihood $\mathcal{L}_n(S) = -n \log Z_{\alpha,\beta} + \frac{n}{2}\mathbf{Tr}[\hat{\Sigma} Q_S]$
- Maximum likelihood estimator:

$$\hat{V} \in \operatorname*{argmax}_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Sigma} V]\,, \quad \text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}\,.$$

- Define $\Gamma = P\Sigma P$ and $\hat{\Gamma} = P\hat{\Sigma}P$, for a projector P on the orthogonal of $\mathbf{1}$:

$$\Gamma = (1-\Delta)P + p\frac{\Delta - \Omega}{2} u_S u_S^\top\,, \qquad u_S = \frac{1}{\sqrt{p}}(\mathbf{1}_S - \mathbf{1}_{\bar{S}})$$

- For all $V \in \mathcal{P}$, $\mathbf{Tr}[\hat{\Gamma} V] = \mathbf{Tr}[\hat{\Sigma} V]$, so equivalently

$$\hat{V} \in \operatorname*{argmax}_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Gamma} V]$$

SDP relaxation

$$\hat{V} \in \operatorname*{argmax}_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Gamma} V]\,, \quad \text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}\,.$$

NP-Hard (Min bisection)

- Semidefinite convex relaxation of $\mathcal{P}$

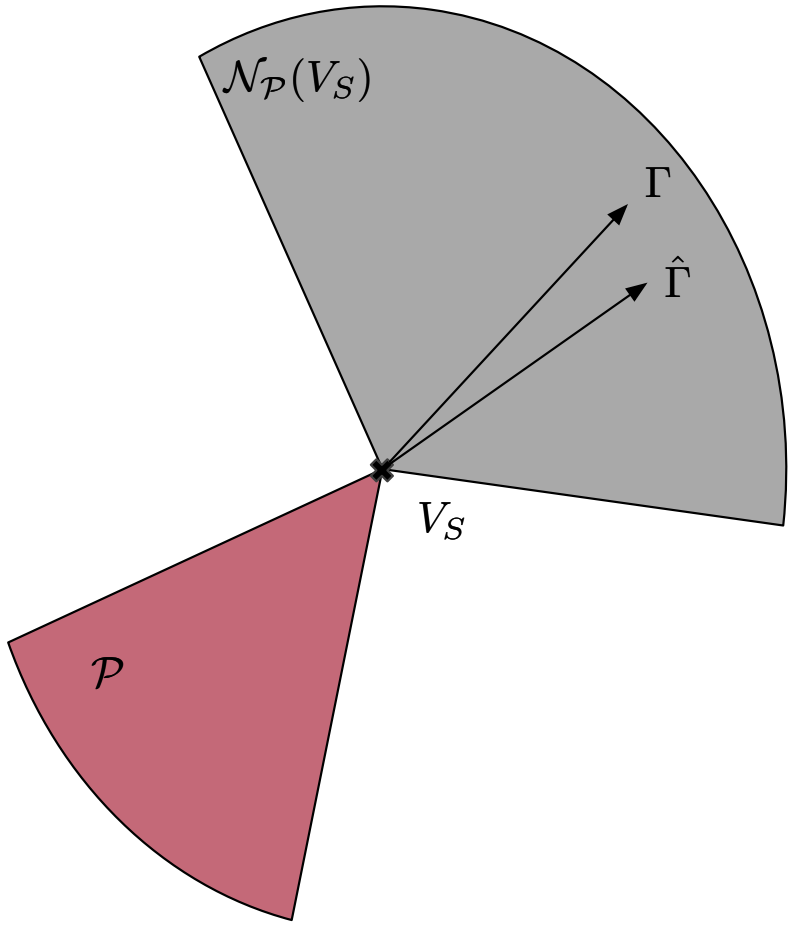
$$\mathcal{E} = \{V : \mathrm{diag}(V) = \mathbf{1}, V \succeq 0\}\,.$$

Change of variable $V = vv^\top$

MAXCUT Goemans-Williamson (95)

- Point V solution of $\max_{V \in \mathcal{C}} \mathbf{Tr}[\hat{\Gamma} V]$ equivalent to $\hat{\Gamma} \in \mathcal{N}_{\mathcal{C}}(V)$

- Relaxation is tight for population matrix Γ: $\hat{V} = V_S$ if $n = \infty$.



SDP relaxation

$$\hat{V} \in \operatorname*{argmax}_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Gamma} V]\,, \quad \text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}\,.$$

NP-Hard (Min bisection)

- Semidefinite convex relaxation of $\mathcal{P}$

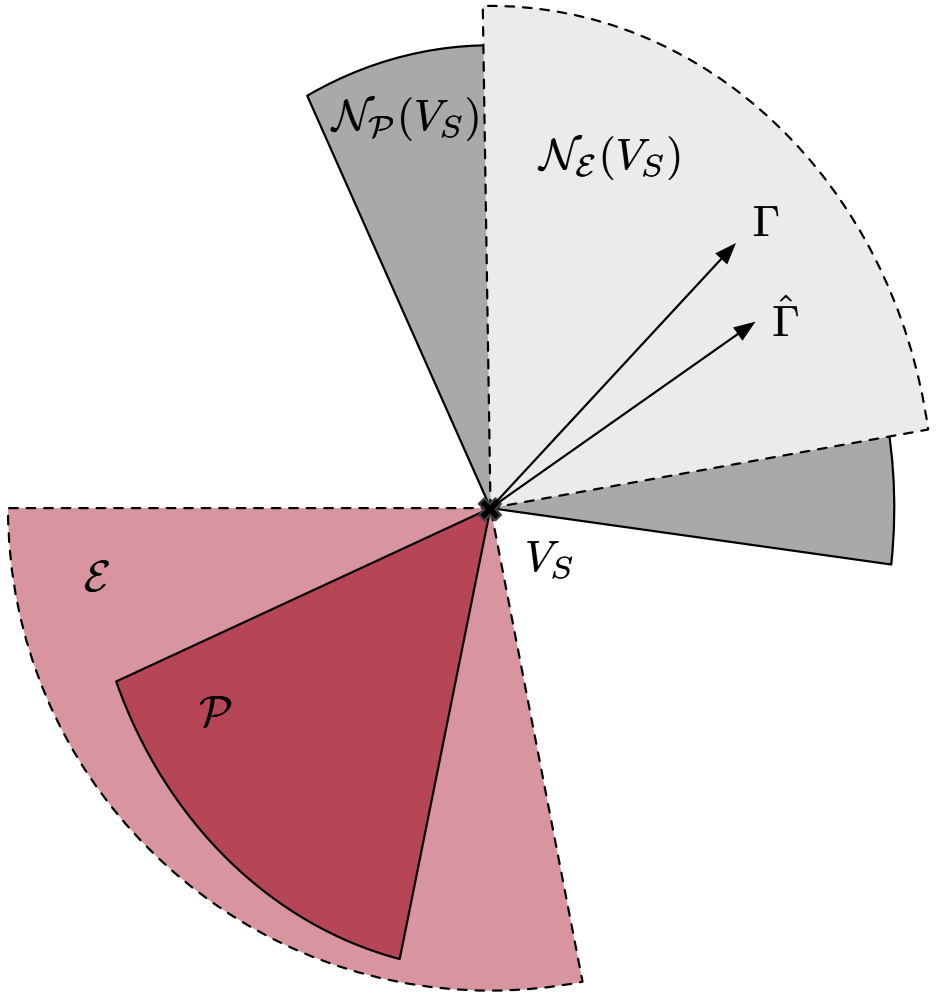
$$\mathcal{E} = \{V : \mathrm{diag}(V) = \mathbf{1}, V \succeq 0\}\,.$$

Change of variable $V = vv^\top$

MAXCUT Goemans-Williamson (95)

- Point V solution of $\max_{V \in \mathcal{C}} \mathbf{Tr}[\hat{\Gamma} V]$ equivalent to $\hat{\Gamma} \in \mathcal{N}_{\mathcal{C}}(V)$

- Relaxation is tight for population matrix Γ: $\hat{V} = V_S$ if $n = \infty$.



Exact recovery

- **Upper bound**: we have $\hat{V} = V_S$ with probability $1 - \delta$ for

$$n \gtrsim \frac{1}{C_{\alpha,\beta}} \frac{\log(p/\delta)}{\Delta - \Omega},$$

  by bounding function of $\Gamma - \hat{\Gamma}$, a sum of independent matrices. Tropp (12)

- **Matching lower bound**: Information theoretic argument yields

$$n \leq \frac{\gamma}{\beta - \alpha} \frac{\log(p/4)}{\Delta - \Omega} \implies \mathbf{P}(\text{recovery}) \lesssim \gamma$$

- Full understanding of the scaling of $\Delta - \Omega$ needed.

The Ising blockmodel model ($\alpha < \beta$)

$$\Sigma = \left(\begin{array}{c|c} \Delta & \Omega \\ \hline \Omega & \Delta \end{array}\right) + (1-\Delta) I_p$$

- **Block magnetizations**: $\mu_S = \frac{\mathbf{1}_S^\top \sigma}{p/2}, \mu_{\bar{S}} = \frac{\mathbf{1}_{\bar{S}}^\top \sigma}{p/2} \in [-1,1]$. Observe that

$$\Delta \approx \frac{2}{p^2} \sum_{i \sim j} \mathbf{E}[\sigma_i \sigma_j] \approx \frac{1}{2} \mathbf{E}[\mu_S^2 + \mu_{\bar{S}}^2] \quad \text{and} \quad \Omega = \frac{2}{p^2} \sum_{i \nsim j} \mathbf{E}[\sigma_i \sigma_j] = \mathbf{E}[\mu_S \, \mu_{\bar{S}}]$$

- **Free energy**: $(\mu_S, \mu_{\bar{S}}) \in [-1,1]^2$ is a sufficient statistic

$$\mathbf{P}_S(\mu_S, \mu_{\bar{S}}) \approx \frac{1}{Z_{\alpha,\beta}} \exp\Big(-\frac{p}{8} \, g_{\alpha,\beta}(\mu_S, \mu_{\bar{S}}) \Big)$$

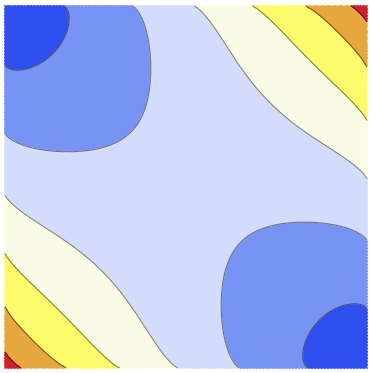
- **Ground states**: Minimizers $G \subset [-1,1]^2$ of $g^{\mathsf{CW}}_{\alpha,\beta}(\mu_S, \mu_{\bar{S}})$.

- **Concentration**: $(\mu_S, \mu_{\bar{S}}) \approx$ ground states with exp. large probability so

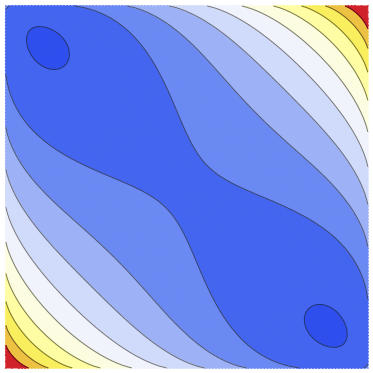
$$\Delta - \Omega \approx \frac{1}{2} \mathbf{E}[(\mu_S - \mu_{\bar{S}})^2] \approx \frac{1}{|G|} \sum_{\mathsf{s} \in G} (\mathsf{s}_1 - \mathsf{s}_2)^2 .$$

Ground states for the Ising blockmodel

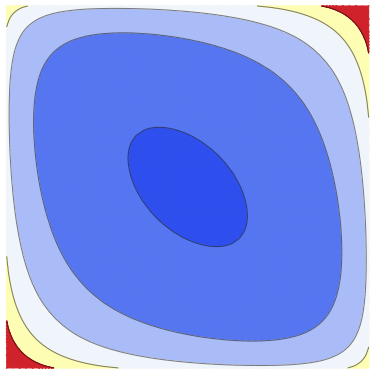
$$g_{\alpha,\beta}(\mu_S,\mu_{\bar{S}}) = -\beta\mu_S^2 - \beta\mu_{\bar{S}}^2 - 2\alpha\,\mu_S\,\mu_{\bar{S}} + 4h\Big(\frac{1+\mu_S}{2}\Big) + 4h\Big(\frac{1+\mu_{\bar{S}}}{2}\Big)$$



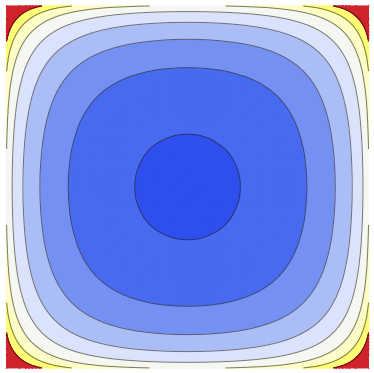
$\alpha = -6$



$\alpha = -2.5$

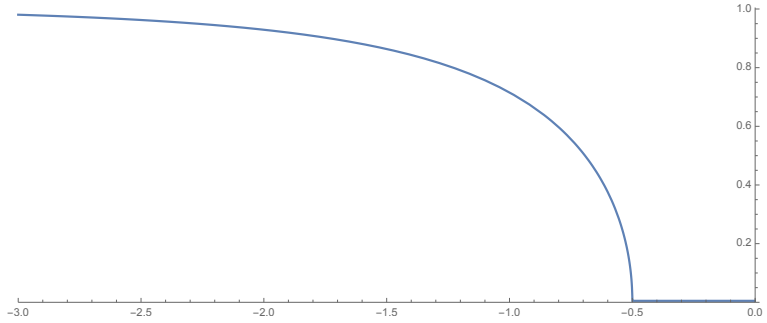


$\alpha = -0.5$



$\alpha = 0$

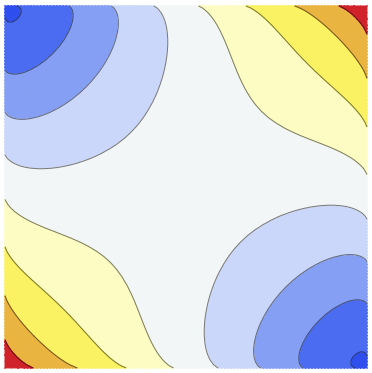
Ground states on the skew-diagonal $(\tilde{\mu}_S = -\tilde{\mu}_{\bar{S}})$ for $\alpha \leq 0$ and fixed $\beta = 1.5 < 2$



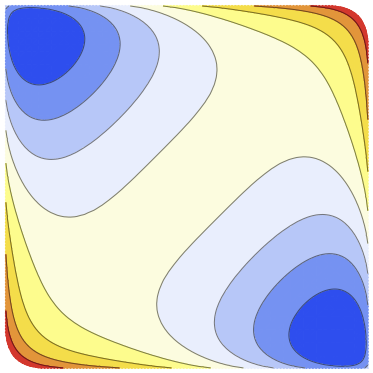
$\alpha \mapsto \tilde{\mu}_S(\alpha, \beta = 1.5)$

Ground states for the Ising blockmodel

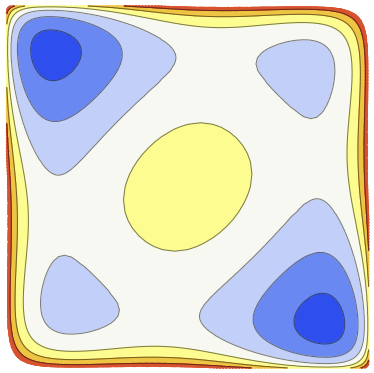
$$g_{\alpha,\beta}(\mu_S,\mu_{\bar{S}}) = -\beta\mu_S^2 - \beta\mu_{\bar{S}}^2 - 2\alpha\,\mu_S\,\mu_{\bar{S}} + 4h\Big(\frac{1+\mu_S}{2}\Big) + 4h\Big(\frac{1+\mu_{\bar{S}}}{2}\Big)$$



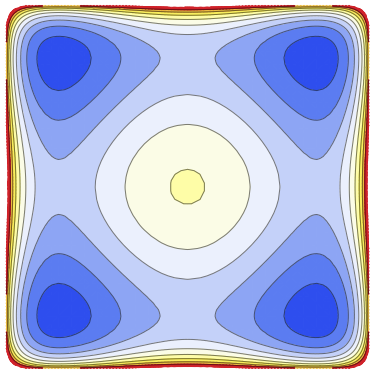
$\alpha = -4$



$\alpha = -0.9$

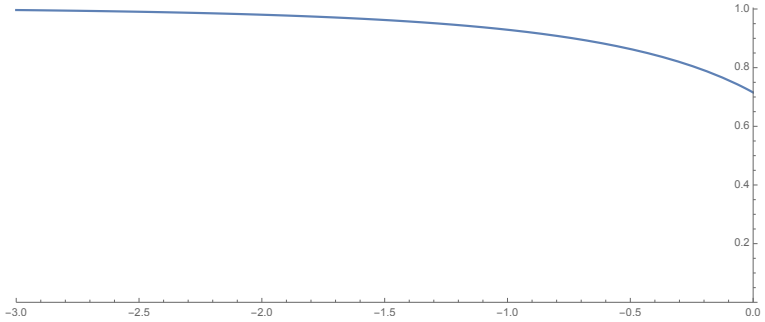


$\alpha = -0.2$



$\alpha = 0$

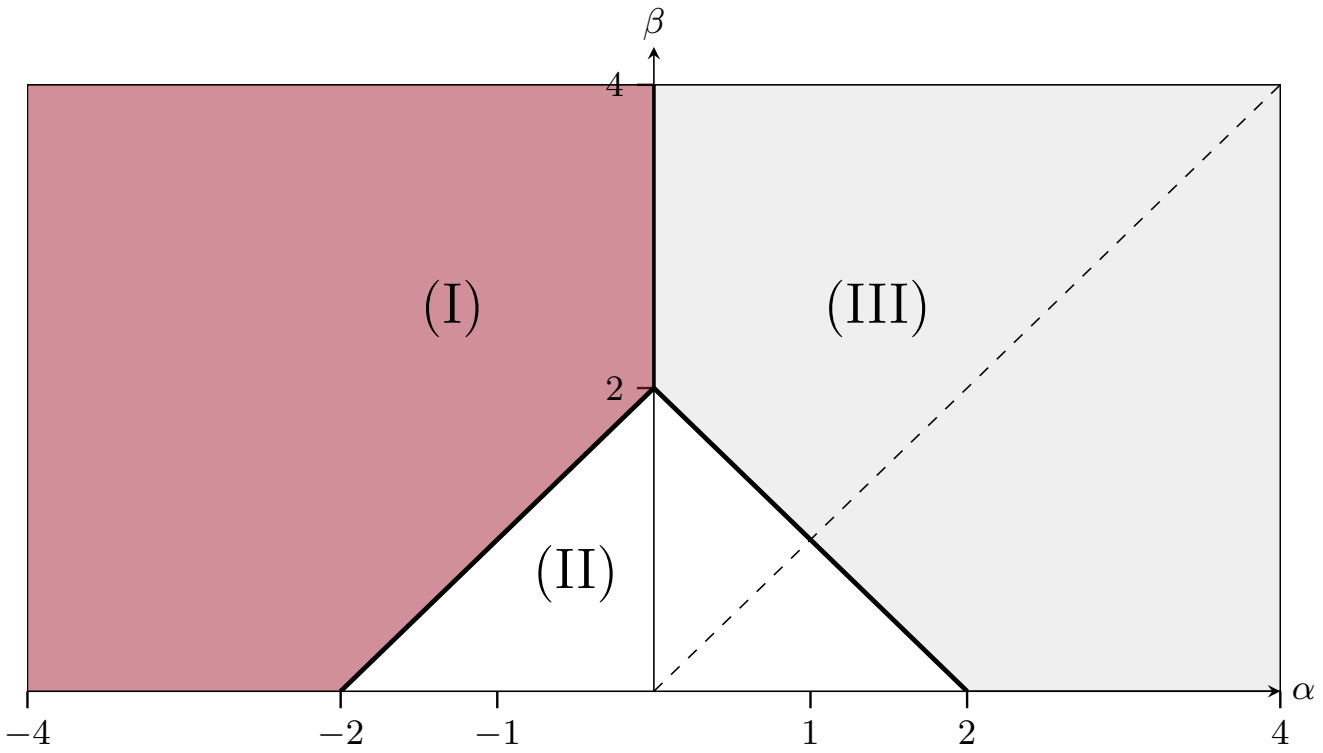
Ground states on the skew-diagonal $(\mu_S = -\mu_{\bar{S}})$ for $\alpha \leq 0$ and fixed $\beta = 2.5 > 2$



$\alpha \mapsto \tilde{\mu}_S(\alpha,\beta) \qquad \beta = 2.5$

Phase diagram

Full understanding of the position of the ground states for $\beta > 0$, $\alpha < \beta$



- Phase diagram for all the parameter regions
  - Region (I): Two ground states $(\tilde{\mu}_S, \tilde{\mu}_{\bar{S}}) = \pm(\tilde{x}, -\tilde{x})$.
  - Region (II): One ground state at $(0, 0)$.
  - Region (III): Two ground states $(\tilde{\mu}_S, \tilde{\mu}_{\bar{S}}) = \pm(\tilde{x}, \tilde{x})$.

Concentration

- Quantities of interest as expectations of the mean block magnetizations

$$\Delta \approx \frac{1}{2}\mathbf{E}[\mu_S^2 + \mu_{\bar{S}}^2] \quad , \quad \Omega \approx \mathbf{E}[\mu_S \mu_{\bar{S}}] \quad \text{and} \quad \Delta - \Omega \approx \frac{1}{2}\mathbf{E}[(\mu_S - \mu_{\bar{S}})^2]\,.$$

- Gaussian approximation of the discrete distribution with $Z \sim \mathcal{N}(0, I_2)$.

$$\mathbf{E}_{\alpha,\beta}[\varphi(\mu)] \simeq_p \frac{1}{|G|}\sum_{\tilde{s}\in G}\mathbf{E}\big[\varphi(\tilde{s} + 2\sqrt{\frac{2}{p}}H^{-1/2}Z)\big] \qquad \forall\, \varphi\,.$$

- Approximation of the gap $\Delta - \Omega$:

$$\Delta - \Omega \simeq_p \begin{cases} 2\tilde{x}^2 & \text{in region (I)} \\ \dfrac{C_{\alpha,\beta}}{p} & \text{in region (II)} \\ \dfrac{C'_{\alpha,\beta}}{p} & \text{in region (III)} \end{cases}$$

Naive estimation

- Covariance matrix:

$$\Sigma = \mathbf{E}[\sigma\sigma^\top] = \left(\begin{array}{c|c} \Delta & \Omega \\ \hline \Omega & \Delta \end{array}\right) + (1-\Delta)I_p\,.$$

- Empirical covariance matrix:

$$\hat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n} \sigma^{(t)}\sigma^{(t)\top} = \Sigma \pm \sqrt{\frac{\log p}{n}} \quad \text{entrywise}$$

- Threshold off-diagonal entries of $\hat{\Sigma}$ at $(\Delta+\Omega)/2$

- Exact recovery if

$$n \gtrsim \begin{cases} \log p & \text{in region (I)} \\ p^2 \log p & \text{in region (II)} \\ p^2 \log p & \text{in region (III)} \end{cases}$$

Exact recovery

- **Upper bound**: we have $\hat{V} = V_S$ with probability $1 - \delta$ for

$$n \gtrsim \frac{1}{C_{\alpha,\beta}} \frac{\log(p/\delta)}{\Delta - \Omega},$$

by bounding function of $\Gamma - \hat{\Gamma}$, a sum of independent matrices. Tropp 12

- **Matching lower bound**: Fano's inequality yields

$$n \leq \frac{\gamma}{\beta - \alpha} \frac{\log(p/4)}{\Delta - \Omega} \implies \mathbf{P}(\text{recovery}) \lesssim \gamma$$

- Full understanding of the scaling of $\Delta - \Omega$ gives optimal rates.

$$n \gtrsim \begin{cases} \log p & \text{in region (I)} \\ p \log p & \text{in regions (II) and (III)} \end{cases}$$

with constant factors illustrating further these transitions.

Conclusion

- **Contributions**
  - Model for interactions between individuals in different communities.
  - Analysis from statistical physics to understand parameters of the problem.
  - Study of convex relaxations with an analysis on normal cones.

- **Open questions**
  - Exact recovery threshold, conjecture that $n^* = \frac{C^* \log(p)}{(\beta - \alpha)(\Delta - \Omega)}$.
  - Rates for partial recovery in Hamming distance.
  - Generalization to multiple blocks, more complex structures.

THANK YOU