Spectral thresholds in the Bipartite Stochastic Block Model

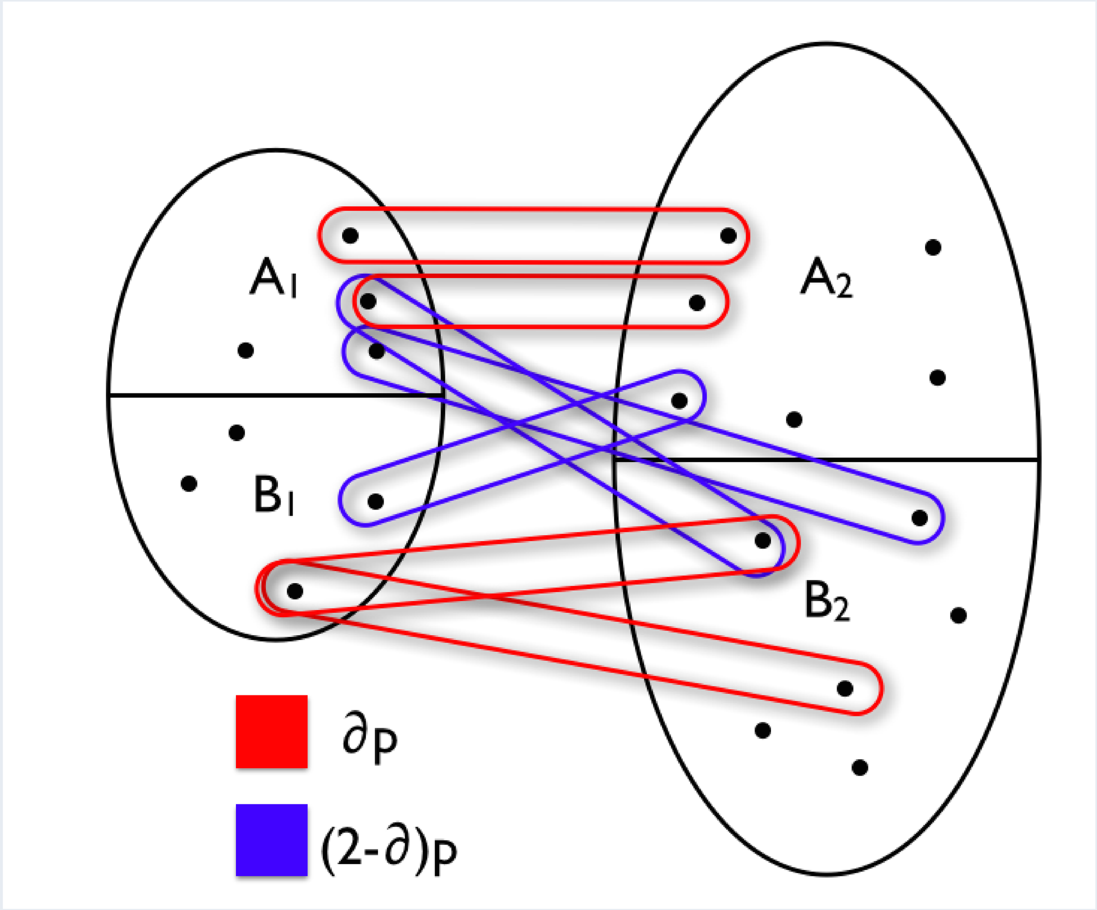
Laura Florescu and Will Perkins

NYU and University of Birmingham

Bipartite Stochastic Block Model

Definition

Let $G = (V_1, V_2, E)$ graph with two vertex sets V_1, V_2, with $|V_1| = n_1$, $|V_2| = n_2$, $\mathbf{n_2 \gg n_1}$.



Questions

- For what $p = p(n_1, n_2)$ can we detect the partition?
- For what $p = p(n_1, n_2)$ do SVD and modifications of it work?

Motivation

- **Planted random k-SAT**
- **inverting Goldreich's PRG**
- **finding planted partition in a random hypergraph**

In [1], Feldman, Perkins, and Vempala reduced above to a BSBM.

Their algorithm used power iteration with a sequence of independent matrices, recovering the partition with $p = \tilde{O}((n_1 n_2)^{-1/2})$.

They asked the question: would the plain SVD recover the partition at the same density?

Detection vs Non-detection

Theorem

δ fixed, $n_2 = \omega(n_1)$. Can efficiently detect V_1 whp if

$$p > \frac{1+\epsilon}{(\delta - 1)^2 \sqrt{n_1 n_2}},$$

$\forall$ fixed ϵ.

Theorem

If $n_2 \geq n_1$ and

$$p \leq \frac{1}{(\delta - 1)^2 \sqrt{n_1 n_2}},$$

no algorithm can detect the partition whp.

Algorithms

Let M be the $n_1 \times n_2$ adjacency matrix of the random bipartite graph.

- **Vanilla SVD**: Compute the second left singular vector of M and round to a vector $z \in \{\pm 1\}^{n_1}$.
- **Diagonal deletion SVD**: Set the diagonal entries of MM^T to 0. Compute the second eigenvector of this matrix and round to $z \in \{\pm 1\}^{n_1}$.
- DD succeeds at a polynomially sparser density than vanilla!

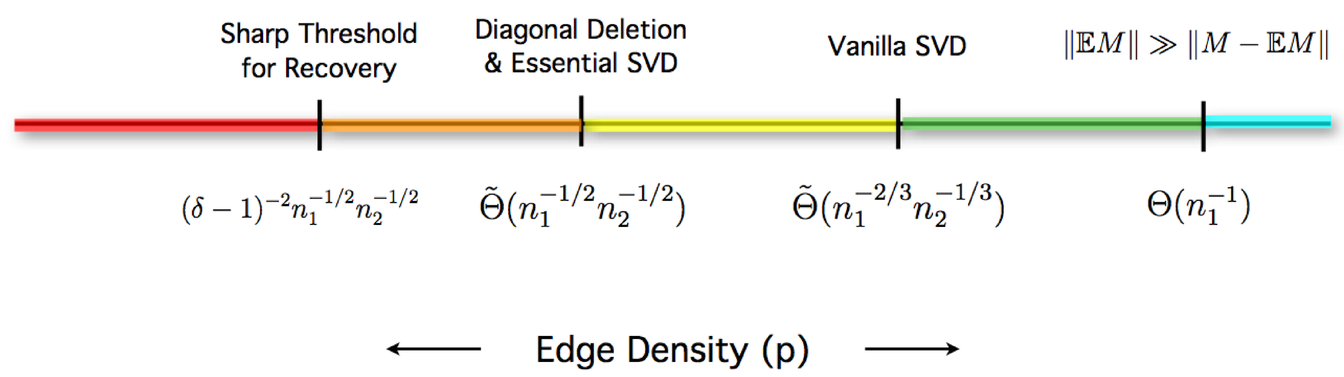
Spectral thresholds

Theorem

Let $n_1 \leq n_2$, with $n_1 \to \infty$. Let $\delta \in [0, 2] \setminus \{1\}$ be fixed with respect to n_1, n_2. Then

1. *If $p \geq (n_1 n_2)^{-1/2} \log n_1$, then whp the diagonal deletion SVD algorithms recovers the partition $V_1 = A_1 \cup B_1$.*
2. *If $p \geq n_1^{-2/3} n_2^{-1/3} \log n_1$, then whp the vanilla SVD algorithm recovers the partition.*

SBM reduction: Construct G_2 on V_1 by joining u and w if $u\tilde{v}$, $w\tilde{v}$ for $v \in V_2$ and $deg(v) = 2$. Sparsify G_2 by randomly removing some edges. Then apply a SBM algorithm (Massoulié [3] or Mossel, Neeman, Sly [4]) to G_2.



Conclusions

- We found multiple thresholds for spectral algorithms in this model.
- Spectral algorithms can succeed in recovering the planted partition even when the spectral norm of the noise matrix dominates than that of the signal matrix.
- The usual SVD fails at densities at which subsampling, diagonal deletion, and vertex deletion succeed, answering the question in [1].
- Diagonal deletion is an extremely simple but effective algorithmic technique.

References

[1] Vitaly Feldman, Will Perkins, Santosh Vempala. Subsampled power iteration: a unified algorithm for block models and planted CSP's, NIPS 2015.

[2] Laura Florescu, Will Perkins. Spectral thresholds in the bipartite stochastic block model, COLT 2016.

[3] Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. *STOC 2014.*

[4] Elchanan Mossel, Joe Neeman, Allan Sly. A proof of the block model threshold conjecture.