# Optimal rates for community estimation in the weighted stochastic block model

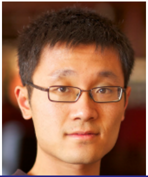
Po-Ling Loh

University of Cambridge
Department of Pure Mathematics and Mathematical Statistics

BIRS workshop on random graphs and statistical inference

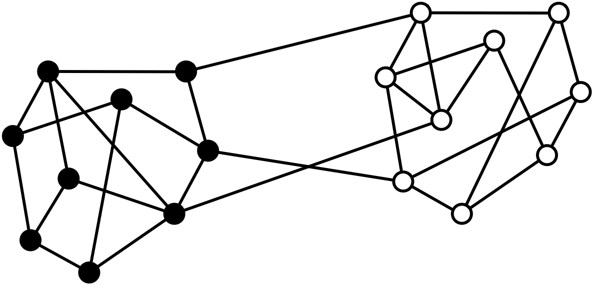
August 10, 2021

Joint work with Min Xu (Rutgers) and Varun Jog (Cambridge)



# Community recovery/estimation

- **Given:** Undirected graph $G = (V, E)$ on $n$ nodes
- **Goal:** Partition nodes into communities based on relative connectivity



- Network structure may be assortative (homophilic), disassortative (heterophilic), etc.

# Stochastic block model

- Probabilistic model introduced by Holland et al. '83 for generating community-structured data

# Stochastic block model

- Probabilistic model introduced by Holland et al. '83 for generating community-structured data
- $K \geq 2$ communities
- For node $i$, let $\sigma(i) \in \{1, 2, \ldots, K\}$ denote community assignment
- Presence of edge depends only on communities of incident vertices:

$$W_{i,j} \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(p_{\sigma(i)\sigma(j)})$$

# Stochastic block model

- Probabilistic model introduced by Holland et al. '83 for generating community-structured data
- $K \geq 2$ communities
- For node $i$, let $\sigma(i) \in \{1, 2, \dots, K\}$ denote community assignment
- Presence of edge depends only on communities of incident vertices:

$$W_{i,j} \overset{\text{i.i.d.}}{\sim} \text{Ber}(p_{\sigma(i)\sigma(j)})$$

- Gives rise to "block" structure:

$$W = \begin{array}{c c} & \begin{array}{cccc:cccc} & & A & & & & B & \end{array} \\ \begin{array}{c} A \\ \\ B \end{array} & \left(\begin{array}{cccc:cccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ \hdashline 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right) \end{array}$$

# Questions

- What are good algorithms for community estimation?
- Can "optimal" accuracy be obtained with computationally tractable algorithms?

# Measure of accuracy

- For clustering algorithm $\widehat{\sigma}$, define *misclassification error*

$$\ell(\widehat{\sigma}(W), \sigma_0) := \min_{\tau \in S_K} \frac{1}{n} d_H(\widehat{\sigma}(W), \tau \circ \sigma_0)$$

# Measure of accuracy

- For clustering algorithm $\widehat{\sigma}$, define *misclassification error*

$$\ell(\widehat{\sigma}(W), \sigma_0) := \min_{\tau \in S_K} \frac{1}{n} d_H(\widehat{\sigma}(W), \tau \circ \sigma_0)$$

- Define *risk*

$$R(\widehat{\sigma}, \sigma_0) := \mathbb{E}\left[\ell(\widehat{\sigma}(W), \sigma_0)\right]$$

# Measure of accuracy

- For clustering algorithm $\widehat{\sigma}$, define *misclassification error*

$$\ell(\widehat{\sigma}(W), \sigma_0) := \min_{\tau \in S_K} \frac{1}{n} d_H(\widehat{\sigma}(W), \tau \circ \sigma_0)$$

- Define *risk*

$$R(\widehat{\sigma}, \sigma_0) := \mathbb{E}\left[\ell(\widehat{\sigma}(W), \sigma_0)\right]$$

- Characterize minimax risk

$$\inf_{\widehat{\sigma}} \sup_{\sigma_0, P \in \Theta} R(\widehat{\sigma}, \sigma_0)$$

# Sharp thresholds (Zhang & Zhou '15)

- Minimax bounds for misclassification error in (roughly) equal-sized communities:

$$\inf_{\widehat{\sigma}} \sup_{\sigma_0, P \in \Theta} R(\widehat{\sigma}, \sigma_0) = \exp\left(-(1+o(1))\frac{nI_n}{K}\right)$$

where $\Theta$ is parameter space such that within-community edges have probability $\geq \frac{a}{n}$, between-community edges have probability $\leq \frac{b}{n}$, and

$$\begin{aligned} I_n &= -2\log\left(\sqrt{\frac{a}{n}}\sqrt{\frac{b}{n}} + \sqrt{1-\frac{a}{n}}\sqrt{1-\frac{b}{n}}\right) \\ &= D_{1/2}\left(\mathrm{Ber}\left(\frac{a}{n}\right) \Big\| \mathrm{Ber}\left(\frac{b}{n}\right)\right) \end{aligned}$$

is *Renyi divergence* of order $\frac{1}{2}$

# Sharp thresholds (Zhang & Zhou '15)

- Implies weak consistency is governed by behavior of $I_n$: If $nI_n \to \infty$,

$$\inf_{\widehat{\sigma}} \sup_{\sigma_0, P \in \Theta} R(\widehat{\sigma}, \sigma_0) \to 0$$

# Sharp thresholds (Zhang & Zhou '15)

- Implies weak consistency is governed by behavior of $I_n$: If $nI_n \to \infty$,

$$\inf_{\widehat{\sigma}} \sup_{\sigma_0, P \in \Theta} R(\widehat{\sigma}, \sigma_0) \to 0$$

- **Rough intuition:** $I_n = D_{1/2}\left(\text{Ber}\left(\frac{a}{n}\right) \middle\| \text{Ber}\left(\frac{b}{n}\right)\right)$ quantifies *distinguishability* of within-community and between-community edges

# Sharp thresholds (Zhang & Zhou '15)

- Implies weak consistency is governed by behavior of $I_n$: If $nI_n \to \infty$,

$$\inf_{\widehat{\sigma}} \sup_{\sigma_0, P \in \Theta} R(\widehat{\sigma}, \sigma_0) \to 0$$

- **Rough intuition:** $I_n = D_{1/2}\left(\text{Ber}\left(\frac{a}{n}\right) \middle\| \text{Ber}\left(\frac{b}{n}\right)\right)$ quantifies *distinguishability* of within-community and between-community edges

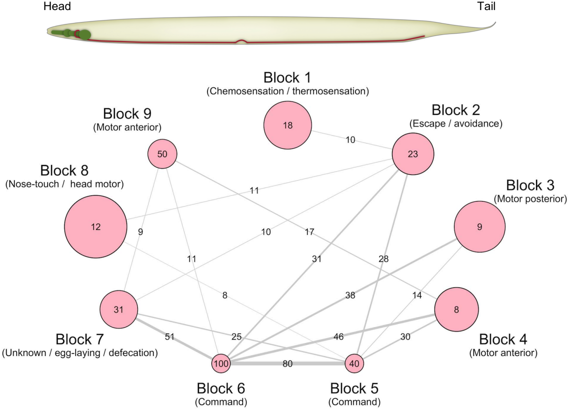
- **Question:** What if edge distributions are not Bernoulli?

# Motivating examples

- Weighted/labeled graphs occur naturally in social networks, airline networks, neural networks, etc.

**C. elegans nervous system:**

- Chemical synapses
- Gap junctions
- Neuromuscular junctions

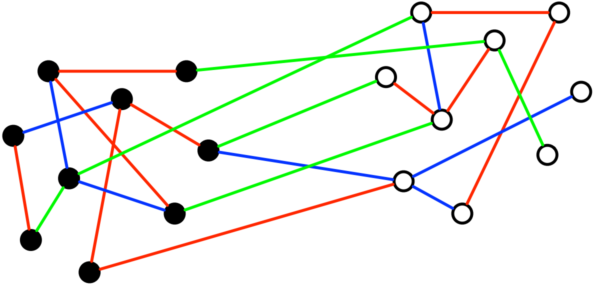



"Stochastic blockmodeling of the modules and core of the Caenorhabditis elegans connectome"

Pavlovic et al. (2014)

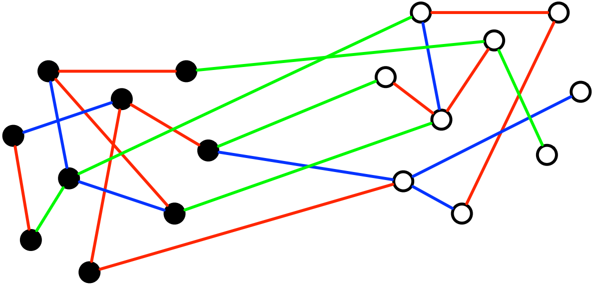
# Weighted stochastic block model (discrete case)

- Entries of adjacency matrix drawn from general distributions



# Weighted stochastic block model (discrete case)

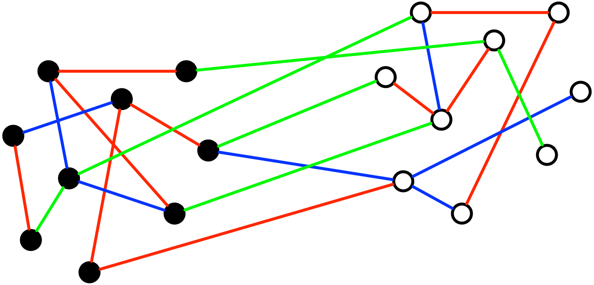
- Entries of adjacency matrix drawn from general distributions



- $p_n, q_n$ denote within-community, between-community distributions

# Weighted stochastic block model (discrete case)

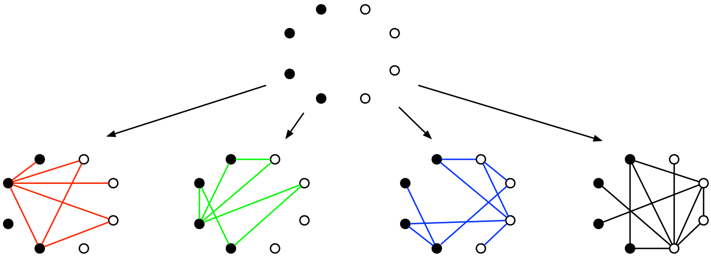
- Entries of adjacency matrix drawn from general distributions



- $p_n, q_n$ denote within-community, between-community distributions
- Edge labels may contain valuable information for community recovery

# Example: Multilayer networks

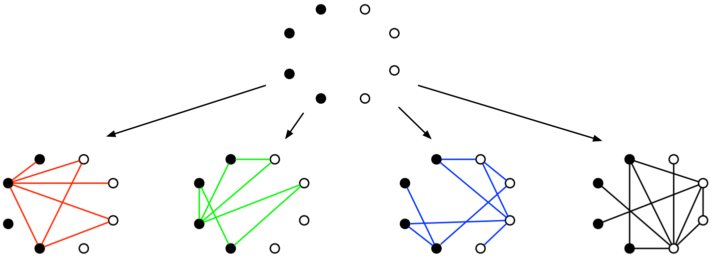
- Nodes in graph give rise to $m$ different edge sets/ "views"



- **Goal:** Combine all views to recover underlying community structure

# Example: Multilayer networks

- Nodes in graph give rise to $m$ different edge sets/ "views"



- **Goal:** Combine all views to recover underlying community structure
- Can be viewed as single weight distribution taking $2^m$ possible values

# Algorithms for unweighted SBMs

- MLE formulation has good theoretical properties, but computationally infeasible

# Algorithms for unweighted SBMs

- MLE formulation has good theoretical properties, but computationally infeasible
- Popular approach based on **spectral clustering**: $W$ is close to $\mathbb{E}[W]$, which (viewed columnwise) has well-separated cluster structure

# Algorithms for unweighted SBMs

- MLE formulation has good theoretical properties, but computationally infeasible
- Popular approach based on **spectral clustering**: $W$ is close to $\mathbb{E}[W]$, which (viewed columnwise) has well-separated cluster structure
  - Performing spectral clustering on $W$ yields weak recovery under appropriate problem scaling (Lei & Rinaldo '15)

# Algorithms for unweighted SBMs

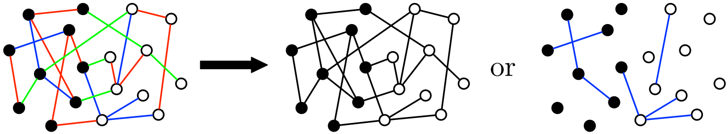
- MLE formulation has good theoretical properties, but computationally infeasible
- Popular approach based on **spectral clustering**: $W$ is close to $\mathbb{E}[W]$, which (viewed columnwise) has well-separated cluster structure
  - Performing spectral clustering on $W$ yields weak recovery under appropriate problem scaling (Lei & Rinaldo '15)
  - To achieve *optimal misclassification error*, require additional refinement steps involving local MLE calculations (Gao et al. '15)

# Algorithms for unweighted SBMs

- MLE formulation has good theoretical properties, but computationally infeasible
- Popular approach based on **spectral clustering**: $W$ is close to $\mathbb{E}[W]$, which (viewed columnwise) has well-separated cluster structure
  - Performing spectral clustering on $W$ yields weak recovery under appropriate problem scaling (Lei & Rinaldo '15)
  - To achieve *optimal misclassification error*, require additional refinement steps involving local MLE calculations (Gao et al. '15)
- Naively applying spectral clustering to weighted adjacency matrix *cannot* be optimal, since numerical labels may be arbitrary

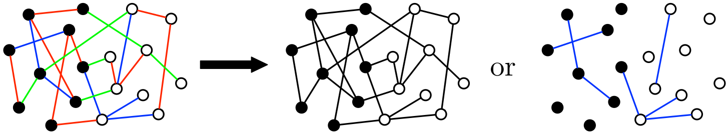
# Reducing to unweighted case

- **Idea:** Map weighted graph to unweighted graph, then perform recovery algorithm for unweighted SBM



# Reducing to unweighted case

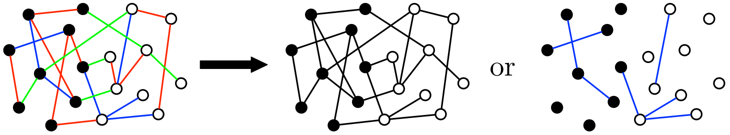
- **Idea:** Map weighted graph to unweighted graph, then perform recovery algorithm for unweighted SBM




- **Questions:** Could this possibly lead to optimal estimation rates...? How to perform mapping?

# Reducing to unweighted case

- **Idea:** Map weighted graph to unweighted graph, then perform recovery algorithm for unweighted SBM



- **Questions:** Could this possibly lead to optimal estimation rates...? How to perform mapping?
- Modifying Zhang & Zhou '15 analysis yields lower bound

$$\inf_{\widehat{\sigma}} \sup_{\sigma_0,(p_n,q_n)\in(\mathcal{P},\mathcal{Q})} R(\widehat{\sigma},\sigma_0) \geq \exp\left(-(1+o(1))\frac{nI_n}{K}\right),$$

but this rate cannot be achieved simply by dropping edge weights

# Characterization via Hellinger distance

- In setting where $I_n \to 0$,

$$
\begin{aligned}
I_n &= -2 \log \left( \sum_{\ell=0}^{L} \sqrt{p_n(\ell) q_n(\ell)} \right) \\
&= \left( \sum_{\ell=0}^{L} \left( \sqrt{p_n(\ell)} - \sqrt{q_n(\ell)} \right)^2 \right) (1 + o(1))
\end{aligned}
$$

- Latter expression known as *Hellinger distance* between $p_n$ and $q_n$

# Initial color identification

- **Key insight:** If $nI_n \to \infty$ (regime where weak recovery is possible), $\exists \ell^* \in \{1, \ldots, L\}$ such that

$$n\left(\sqrt{p_n(\ell^*)} - \sqrt{q_n(\ell^*)}\right)^2 \to \infty,$$

so community recovery based on $\ell^*$ alone achieves weak recovery

# Initial color identification

- **Key insight:** If $nI_n \to \infty$ (regime where weak recovery is possible), $\exists \ell^* \in \{1, \dots, L\}$ such that

$$n \left( \sqrt{p_n(\ell^*)} - \sqrt{q_n(\ell^*)} \right)^2 \to \infty,$$

  so community recovery based on $\ell^*$ alone achieves weak recovery

- **Problem:** How to identify $\ell^*$ based on observing graph?

# Initial color identification

- **Key insight:** If $nI_n \to \infty$ (regime where weak recovery is possible), $\exists \ell^* \in \{1, \dots, L\}$ such that

$$n\left(\sqrt{p_n(\ell^*)} - \sqrt{q_n(\ell^*)}\right)^2 \to \infty,$$

so community recovery based on $\ell^*$ alone achieves weak recovery

- **Problem:** How to identify $\ell^*$ based on observing graph?
- How to refine initial clustering based on $\ell^*$ to obtain optimal error rate?

# Choosing $\ell^*$

- For each $1 \leq \ell \leq L$, perform spectral clustering on unweighted adjacency matrix $A_\ell$ to obtain community assignments $\widehat{\sigma}_\ell$

# Choosing $\ell^*$

- For each $1 \leq \ell \leq L$, perform spectral clustering on unweighted adjacency matrix $A_\ell$ to obtain community assignments $\widehat{\sigma}_\ell$
- Estimate within- and between-community edge probabilities:

$$\widehat{P}_\ell = \frac{\sum_{u \neq v : \widehat{\sigma}_\ell(u) = \widehat{\sigma}_\ell(v)} (A_\ell)_{uv}}{|\{u \neq v : \widehat{\sigma}_\ell(u) = \widehat{\sigma}_\ell(v)\}|}, \quad \widehat{Q}_\ell = \frac{\sum_{u \neq v : \widehat{\sigma}_\ell(u) \neq \widehat{\sigma}_\ell(v)} (A_\ell)_{uv}}{|\{u \neq v : \widehat{\sigma}_\ell(u) \neq \widehat{\sigma}_\ell(v)\}|}$$

# Choosing $\ell^*$

- For each $1 \le \ell \le L$, perform spectral clustering on unweighted adjacency matrix $A_\ell$ to obtain community assignments $\widehat{\sigma}_\ell$
- Estimate within- and between-community edge probabilities:

$$\widehat{P}_\ell = \frac{\sum_{u \neq v : \widehat{\sigma}_\ell(u) = \widehat{\sigma}_\ell(v)} (A_\ell)_{uv}}{|\{u \neq v : \widehat{\sigma}_\ell(u) = \widehat{\sigma}_\ell(v)\}|}, \quad \widehat{Q}_\ell = \frac{\sum_{u \neq v : \widehat{\sigma}_\ell(u) \neq \widehat{\sigma}_\ell(v)} (A_\ell)_{uv}}{|\{u \neq v : \widehat{\sigma}_\ell(u) \neq \widehat{\sigma}_\ell(v)\}|}$$

- Select

$$\ell^* := \arg\max_\ell \left\{ \frac{(\widehat{P}_\ell - \widehat{Q}_\ell)^2}{\widehat{P}_\ell \vee \widehat{Q}_\ell} \right\}$$

to be color achieving "maximal separation"

# Achieving optimal rate

- However, recovery algorithm $\widehat{\sigma}_{\ell^*}$ yields *suboptimal* error rate

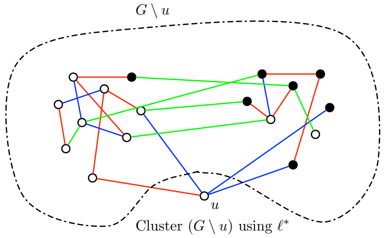
# Achieving optimal rate

- However, recovery algorithm $\widehat{\sigma}_{\ell*}$ yields *suboptimal* error rate
- Additional steps (based on Gao et al. '15) can bootstrap $\widehat{\sigma}_{\ell*}$ to optimal recovery rates:

# Achieving optimal rate

- However, recovery algorithm $\widehat{\sigma}_{\ell^*}$ yields *suboptimal* error rate
- Additional steps (based on Gao et al. '15) can bootstrap $\widehat{\sigma}_{\ell^*}$ to optimal recovery rates:
  1. For each node $u$,
     - Cluster $G \backslash u$ via spectral method with color $\ell^*$ to obtain $\widehat{\sigma}_u$
     - Use local MLE to estimate assignment of $u$:

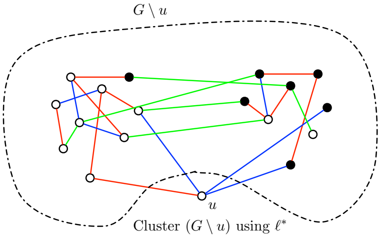
$$\widehat{\sigma}_u(u) = \arg\max_k \sum_{v:\widehat{\sigma}_u(v)=k, v\neq u} \sum_{\ell=0}^{L} \log \frac{\widehat{P}_\ell^*}{\widehat{Q}_\ell^*}(A_\ell)_{uv}$$



# Achieving optimal rate

- However, recovery algorithm $\widehat{\sigma}_{\ell^*}$ yields *suboptimal* error rate
- Additional steps (based on Gao et al. '15) can bootstrap $\widehat{\sigma}_{\ell^*}$ to optimal recovery rates:
  1. For each node $u$,
     - Cluster $G \backslash u$ via spectral method with color $\ell^*$ to obtain $\widehat{\sigma}_u$
     - Use local MLE to estimate assignment of $u$:

$$\widehat{\sigma}_u(u) = \arg\max_k \sum_{v:\widehat{\sigma}_u(v)=k, v\neq u} \sum_{\ell=0}^{L} \log \frac{\widehat{P}_\ell^*}{\widehat{Q}_\ell^*}(A_\ell)_{uv}$$




  2. Combine assignments $\{\widehat{\sigma}_u\}_{u \in V}$ to obtain overall clustering

# Theorem: Optimal recovery (discrete)

**Theorem**

*Suppose* $I_n \to 0$ *and* $nI_n \to \infty$. *Then*

$$\lim_{n\to\infty} \mathbb{P}\left(\ell(\widehat{\sigma}(W), \sigma_0) \leq \exp\left(-\frac{nI_n}{K}(1+o(1))\right)\right) = 1.$$

- Conditions $I_n \to 0$ and $nI_n \to \infty$ imply weight distributions converge together, but not *too* quickly

# Theorem: Optimal recovery (discrete)

**Theorem**

*Suppose* $I_n \to 0$ *and* $nI_n \to \infty$. *Then*

$$\lim_{n\to\infty} \mathbb{P}\left(\ell(\widehat{\sigma}(W), \sigma_0) \leq \exp\left(-\frac{nI_n}{K}(1 + o(1))\right)\right) = 1.$$

- Conditions $I_n \to 0$ and $nI_n \to \infty$ imply weight distributions converge together, but not *too* quickly
- Matches lower bound stated above: Misclassification error governed by Renyi divergence

# Continuous setting

- Also desirable to perform community recovery in settings where edge weights are continuous
- **Examples:** Gaussian distributions, exponential distributions, etc.

# Continuous setting

- Also desirable to perform community recovery in settings where edge weights are continuous
- **Examples:** Gaussian distributions, exponential distributions, etc.
- Previous work relatively scarce, depends on *parametric* assumptions

# Continuous setting

- Also desirable to perform community recovery in settings where edge weights are continuous
- **Examples:** Gaussian distributions, exponential distributions, etc.
- Previous work relatively scarce, depends on *parametric* assumptions

- **Our contribution:** Nonparametric, computationally feasible recovery algorithm achieving optimal error rates

# Mathematical formulation

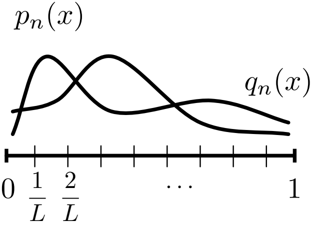
- Suppose weight distribution is **mixture** of point mass at 0 with probability $P_{0,n}$, continuous distribution with density $p_n(x)$ with probability $1 - P_{0,n}$ (similarly for $Q_{0,n}$ and $q_n(x)$)

# Mathematical formulation

- Suppose weight distribution is **mixture** of point mass at 0 with probability $P_{0,n}$, continuous distribution with density $p_n(x)$ with probability $1 - P_{0,n}$ (similarly for $Q_{0,n}$ and $q_n(x)$)
- In order to obtain optimal error rate, **discretize** $p_n(x)$ and $q_n(x)$ more and more finely and reduce to previous algorithm

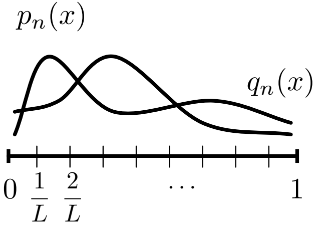
# Mathematical formulation

- Suppose weight distribution is **mixture** of point mass at 0 with probability $P_{0,n}$, continuous distribution with density $p_n(x)$ with probability $1 - P_{0,n}$ (similarly for $Q_{0,n}$ and $q_n(x)$)
- In order to obtain optimal error rate, **discretize** $p_n(x)$ and $q_n(x)$ more and more finely and reduce to previous algorithm
  - If $p_n(x)$ and $q_n(x)$ have bounded support $S = [0, 1]$, partition into $L_n$ equal-length intervals



# Mathematical formulation

- Suppose weight distribution is **mixture** of point mass at 0 with probability $P_{0,n}$, continuous distribution with density $p_n(x)$ with probability $1 - P_{0,n}$ (similarly for $Q_{0,n}$ and $q_n(x)$)
- In order to obtain optimal error rate, **discretize** $p_n(x)$ and $q_n(x)$ more and more finely and reduce to previous algorithm
  - If $p_n(x)$ and $q_n(x)$ have bounded support $S = [0, 1]$, partition into $L_n$ equal-length intervals




  - Otherwise, first apply transformation $\Phi : S \to [0, 1]$ and then partition into $L_n$ equal-length intervals

# Theoretical guarantees

**Theorem**

*Suppose* $I_n \to 0$ *and* $nI_n \to \infty$ *and* $p_n$ *and* $q_n$ *satisfy appropriate regularity conditions with respect to* $\Phi$*. If* $L_n \to \infty$ *is chosen such that* $\frac{nI_n}{L_n \exp(L_n^{1/r})} \to \infty$*, then*

$$\lim_{n\to\infty} \mathbb{P}\left(\ell(\widehat{\sigma}(W), \sigma_0) \leq \exp\left(-\frac{nI_n}{K}(1 + o(1))\right)\right) = 1.$$

# Theoretical guarantees

## Theorem

*Suppose $I_n \to 0$ and $nI_n \to \infty$ and $p_n$ and $q_n$ satisfy appropriate regularity conditions with respect to $\Phi$. If $L_n \to \infty$ is chosen such that $\frac{nI_n}{L_n \exp(L_n^{1/r})} \to \infty$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(\ell(\widehat{\sigma}(W), \sigma_0) \leq \exp\left(-\frac{nI_n}{K}(1 + o(1))\right)\right) = 1.$$

- Note that if $nI_n \to \infty$, appropriate discretization level $L_n$ always exists

# Theoretical guarantees

## Theorem

*Suppose $I_n \to 0$ and $nI_n \to \infty$ and $p_n$ and $q_n$ satisfy appropriate regularity conditions with respect to $\Phi$. If $L_n \to \infty$ is chosen such that $\frac{nI_n}{L_n \exp(L_n^{1/r})} \to \infty$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(\ell(\widehat{\sigma}(W), \sigma_0) \le \exp\left(-\frac{nI_n}{K}(1+o(1))\right)\right) = 1.$$

- Note that if $nI_n \to \infty$, appropriate discretization level $L_n$ always exists
- Under additional condition $\limsup_n \frac{nI_n}{\log n} \le 1$, can obtain

$$\mathbb{E}\left[\ell(\widehat{\sigma}(W), \sigma_0)\right] \le \exp\left(-\frac{nI_n}{K}(1+o(1))\right),$$

agreeing with lower bound on risk

# Examples

- When $S = [0, 1]$, simplest case satisfying conditions is when $p_n, q_n, |p_n'|$, and $|q_n'|$ are uniformly bounded away from 0 and $\infty$

# Examples

- When $S = [0, 1]$, simplest case satisfying conditions is when $p_n, q_n, |p_n'|$, and $|q_n'|$ are uniformly bounded away from 0 and $\infty$

- When $S = \mathbb{R}$, consider *location-scale family*:

$$f_{\mu,\sigma}(x) = f\left(\frac{x-\mu}{\sigma}\right) - \log \sigma,$$

with parameter space $\mu \in [-C_\mu, C_\mu]$ and $\sigma \in \left[\frac{1}{c_\sigma}, c_\sigma\right]$

# Examples

- When $S = [0, 1]$, simplest case satisfying conditions is when $p_n, q_n, |p_n'|$, and $|q_n'|$ are uniformly bounded away from 0 and $\infty$

- When $S = \mathbb{R}$, consider *location-scale family*:

$$f_{\mu,\sigma}(x) = f\left(\frac{x-\mu}{\sigma}\right) - \log\sigma,$$

  with parameter space $\mu \in [-C_\mu, C_\mu]$ and $\sigma \in \left[\frac{1}{c_\sigma}, c_\sigma\right]$
- Densities given by

$$p_n(x) = \exp\left(f_{\mu_{1,n},\sigma_{1,n}}(x)\right), \quad q_n(x) = \exp\left(f_{\mu_{2,n},\sigma_{2,n}}(x)\right)$$

# Examples

- When $S = [0, 1]$, simplest case satisfying conditions is when $p_n, q_n, |p_n'|$, and $|q_n'|$ are uniformly bounded away from 0 and $\infty$

- When $S = \mathbb{R}$, consider *location-scale family*:

$$f_{\mu,\sigma}(x) = f\left(\frac{x-\mu}{\sigma}\right) - \log \sigma,$$

  with parameter space $\mu \in [-C_\mu, C_\mu]$ and $\sigma \in \left[\frac{1}{c_\sigma}, c_\sigma\right]$
- Densities given by

$$p_n(x) = \exp\left(f_{\mu_{1,n},\sigma_{1,n}}(x)\right), \quad q_n(x) = \exp\left(f_{\mu_{2,n},\sigma_{2,n}}(x)\right)$$

- Requirements:
  - $|f^{(k)}(x)|$ bounded for some $k \geq 2$,
  - $\exists c, M$ such that $f'(x) > M$ for $x < -c$ and $f'(x) < -M$ for $x > c$

  (satisfied for Gaussian and Laplace families)

# Main contributions

- Misclassification error rate for weighted SBMs sharply characterized by Renyi divergence

$$I_n = -2 \log \int \sqrt{p_n(x) q_n(x)} dx$$

# Main contributions

- Misclassification error rate for weighted SBMs sharply characterized by Renyi divergence

$$I_n = -2 \log \int \sqrt{p_n(x) q_n(x)} dx$$

- Computationally feasible algorithm for optimal recovery:
  1. (If necessary) discretize continuous weight distribution
  2. Cluster according to each individual color, find $\ell^*$ with greatest separation
  3. Refine weakly consistent clustering based on $\ell^*$ to obtain optimal error rate

# Reference

- Xu, Jog & Loh (2020). Optimal rates for community estimation in the weighted stochastic block model. *Annals of Statistics*.

**Thank you!**