Community Detection in Colorful Graphs

December 17, 2015

1 Problem Set-up

Let $P = (P_0, P_1, ..., P_L)$ and $Q = (Q_0, Q_1, ..., Q_L)$ be discrete distributions over L colors and 0. Let $\sigma_0 : [n] \to K$ be the true clustering. For a pair (i, j) in the same cluster $(\sigma_0(i) = \sigma_0(j))$, suppose $A_{ij} \sim P$ and for a pair (i, j) in different clusters, we suppose $A_{ij} \sim Q$.

We observe the colorful matrix A but not P and Q and the goal is to recover the clusters. We refer to the setting where P, Q are known as the oracle setting.

Note that under the oracle setting, the log-likelihood of a clustering σ is

$$\begin{aligned} l(A \,|\, \sigma) &= \sum_{i<j} \mathbf{1}_{\sigma(i)=\sigma(j)} \sum_{l=0}^{L} \log P_l \mathbf{1}_{A_{ij}=l} + \sum_{i<j} \mathbf{1}_{\sigma(i)\neq\sigma(j)} \sum_{l=0}^{L} \log Q_l \mathbf{1}_{A_{ij}=l} \\ &= \sum_{i<j} \mathbf{1}_{\sigma(i)=\sigma(j)} \sum_{l=0}^{L} \log \frac{P_l}{Q_l} \mathbf{1}_{A_{ij}=l} + \sum_{i<j} \sum_{l=0}^{L} \log Q_l \mathbf{1}_{A_{ij}=l} \end{aligned}$$

Let $I_{tot} = -2 \log \sum_l \sqrt{P_l Q_l}$ be the 1/2 Renyi-divergence between the distributions P, Q.

2 Weak Recovery Under the Oracle Setting

[1] characterizes the minimax rate of weak recovery for Bernoulli P, Q. Their results and proofs can be extended in a straightforward manner to general discrete P, Q under the oracle setting.

Proposition 2.1. *(Upper bound) Assume $\frac{nI_{tot}}{K \log K} \to \infty$. The maximum likelihood estimator $\widehat{\sigma}$ in the oracle setting achieves:*

$$\sup_{\Theta(n,K,\beta,P,Q)} \mathbb{E} r(\widehat{\sigma}, \sigma) \leq \begin{cases} \exp\left(-(1+o(1))\frac{nI_{tot}}{2}\right), & K = 2, \\ \exp\left(-(1+o(1))\frac{nI_{tot}}{\beta K}\right), & K \geq 3 \end{cases}$$

Proof. Proof of this proposition follows that of Theorem 3.2 in [1]. We describe only the parts that need to be modified.

Because we have a different likelihood function, our $T(\sigma)$ takes on a different form from that of [1] at the bottom of page 8:

$$T(\sigma) = \sum_{i<j} \mathbf{1}_{\sigma(i)=\sigma(j)} \sum_{l=0}^{L} \log \frac{P_l}{Q_l} \mathbf{1}_{A_{ij}=l}$$

Let σ_0 denote the true community assignment. We make a mistake if for some other community assignment σ, we get $T(\sigma) > T(\sigma_0)$. The key part of the proof is to bound

$$P_m \equiv P\left(\exists \sigma : T(\sigma) > T(\sigma_0),\, d_H(\sigma, \sigma_0) = m\right)$$

To that end, we bound the probability of error of a fixed σ m-distant from σ_0 in Hamming distance. We will prove an analogue of Proposition 5.1 in [1]:

Let σ be an arbitrary assignment satisfying $d(\sigma, \sigma_0) = m$. Let X_i, Y_i be random variables such that

$$X_i = \log \frac{P_l}{Q_l} \text{ w.p. } P_l \qquad Y_i = \log \frac{P_l}{Q_l} \text{ w.p. } Q_l$$

and α, γ be integers where

$$\alpha = |\{(i,j) : \sigma_0(i) = \sigma_0(j) \wedge \sigma(i) \neq \sigma(j)\}| \quad \gamma = |\{(i,j) : \sigma_0(i) \neq \sigma_0(j) \wedge \sigma(i) = \sigma(j)\}|$$

Then

$$P(T(\sigma) \geq T(\sigma_0)) \leq P\left(\sum_{i=1}^{\alpha} X_i - \sum_{i=1}^{\gamma} Y_i < 0\right) \leq \exp(-\frac{\gamma+\alpha}{2} I) \tag{2.1}$$

Lemma 5.3 from [1] bounds α, γ and Proposition 5.2 bounds the number of σ's (up to equivalent classes) such that $d_H(\sigma, \sigma_0) = m$. These pieces together bounds P_m. The rest of the proof follows [1] exactly starting from Page 16.

We devote the rest of the proof toward proving equation 2.1.

$$\begin{aligned} T(\sigma_0) - T(\sigma') &= \sum_{i<j} \mathbf{1}_{\sigma_0(i)=\sigma_0(j) \wedge \sigma'(i) \neq \sigma'(j)} \sum_{l=1}^{L} \mathbf{1}_{A_{ij}=l} \log \frac{P_l}{Q_l} \\ &\quad - \sum_{i<j} \mathbf{1}_{\sigma_0(i) \neq \sigma_0(j) \wedge \sigma'(i)=\sigma'(j)} \sum_{l=1}^{L} \mathbf{1}_{A_{ij}=l} \log \frac{P_l}{Q_l} \\ &= \sum_{i=1}^{\alpha} X_i - \sum_{i=1}^{\gamma} Y_i \end{aligned}$$

$$
\begin{aligned}
P(\sum_{i=1}^{\gamma} Y_i - \sum_{i=1}^{\alpha} X_i > 0) &\leq \mathbb{E}\left(e^{-t\sum_{i=1}^{\alpha} X_i} e^{t\sum_{i=1}^{\gamma} Y_i}\right) \\
&\leq \mathbb{E}\left(e^{-tX_1\alpha} e^{tY_1\gamma}\right) \\
&\leq \left(\mathbb{E}e^{-tX_i}\mathbb{E}e^{tY_1}\right)^{(1-w)\alpha + w\gamma} \frac{\left(\mathbb{E}e^{tY_1}\right)^{(1-w)(\gamma-\alpha)}}{\left(\mathbb{E}e^{-tX_1}\right)^{w(\gamma-\alpha)}}
\end{aligned}
$$

We will show that when $t = 1/2$ and $w = 1/2$, the fraction term equals 1 and the first term equals $\exp\left(-(1/2\alpha + 1/2\gamma)I\right)$.

Note that

$$
\begin{aligned}
\mathbb{E}e^{-tX_1} &= \sum_l P_l e^{-t\log\frac{P_l}{Q_l}} \\
&= \sum_l P_l \left(\frac{Q_l}{P_l}\right)^t \\
&= \sum_l \sqrt{P_l Q_l} \quad \text{(if } t = 1/2\text{)}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}e^{tY_1} &= \sum_l Q_l e^{t\log\frac{P_l}{Q_l}} \\
&= \sum_l Q_l \left(\frac{P_l}{Q_l}\right)^t \\
&= \sum_l \sqrt{P_l Q_l} \quad \text{if } t = 1/2
\end{aligned}
$$

$$
\begin{aligned}
P(\sum_{i=1}^{\gamma} Y_i - \sum_{i=1}^{\alpha} X_i > 0) &\leq \left(\sum_l \sqrt{P_l Q_l}\right)^{\alpha+\gamma} \\
&\leq \exp(-\frac{1}{2}I)^{\alpha+\gamma} \\
&\leq \exp\left(-\frac{(\alpha+\gamma)}{2}I\right)
\end{aligned}
$$

$\square$

3 Fixed L Case

Suppose L is fixed. Suppose also that $\lim_{n\to\infty} P_0 \wedge Q_0 \to 1$ so that $P_l, Q_l \to 0$ for $l \neq 0$.

Proposition 3.1. *Suppose $\frac{nI}{K} \to \infty$. If weak recovery (consistency) is possible under the oracle setting, then it is possible when P, Q are unknown.*

Proof. It is clear that weak recovery is possible iff $\frac{nI_{tot}}{K \log K} \to \infty$.

$$\begin{aligned} I_{tot} &= -2 \log \left(\sum_{l=1}^{L} \sqrt{P_l Q_l} + \sqrt{(1 - \sum_{l=1}^{L} P_l)(1 - \sum_{l=1}^{L} Q_l)} \right) \\ &= -2 \log \left(1 - \frac{1}{2} \left(\sum_{l=0}^{L} (\sqrt{P_l} - \sqrt{Q_l})^2 \right) \right) \\ &= (1 + o(1)) \sum_{l=0}^{L} (\sqrt{P_l} - \sqrt{Q_l})^2 \end{aligned}$$

Since $I_{tot} = \omega(\frac{K \log K}{n})$ by hypothesis, it must be that, for some l, $(\sqrt{P_l} - \sqrt{Q_l})^2 = \omega(\frac{K \log K}{n})$.

We choose such an l and consider an estimator $\widehat{\sigma}_l$ that uses only the information $\mathbf{1}_{A_{ij}=l}$.

Since the Renyi-divergence I_l of $Ber(P_l)$ and $Ber(Q_l)$ is

$$I_l = (1 + o(1)) \left((\sqrt{P_l} - \sqrt{Q_l})^2 + (\sqrt{1 - P_l} - \sqrt{1 - Q_l})^2 \right)$$

We have that $\frac{nI_l}{K \log K} \to \infty$ and weak consistency is thus achievable with $\widehat{\sigma}_l$. $\square$

Although weak consistency is achievable with the estimator $\widehat{\sigma}_l$ that considers only $\mathbf{1}_{A_{ij}=l}$, the estimator converges at $\exp(-\frac{nI_l}{\beta K})$ and therefore does not converge at the same rate. It is easy to see that $I_l \geq I_{tot} \geq \frac{1}{L} I_l (1 - o(1))$.

We propose a procedure to remove the $\frac{1}{L}$ factor and get oracle rate of convergence:

1. Split data into two halves.
2. For each color $l = 1, ..., L$ and for $l = 0$, cluster the first half of the data based on $\mathbf{1}_{A_{ij}=l}$ and get $\widehat{\sigma}_l$. Use the second half to verify the accuracy, either by likelihood evaluation or by clustering the second half as well and testing consistency.
3. Cluster all data with the best l and use the rough clustering to estimate $\widehat{P}_l$ and $\widehat{Q}_l$.
4. Plug in the estimated $\widehat{P}_l$ and $\widehat{Q}_l$ into the maximum likelihood estimator.

As a first step toward analyzing such a procedure, we ask the following question: suppose σ^* is an imperfect clustering of points $(2, ..., n)$ such that $d_H(\sigma_0, \sigma^*) = \gamma$. We want to assign a cluster to node 1 based on σ^*. We will pretend we can see P, Q exactly and assign node 1 by maximum likelihood. We will for simplicity also assume two clusters each of size $n/2$ for now. Assume without loss of generality that $\sigma_0(1) = 1$.

The assignment is

$$\operatorname*{argmax}_{k=1,2} \sum_{j\,:\,\sigma^*(j)=k} \sum_{l=0}^{L} \log \frac{P_l}{Q_l} \mathbf{1}_{A_{1j}=l}$$

We can think of this as assigning a weight of $\log \frac{P_l}{Q_l}$ to edges with color l and then assigning node 1 to the cluster that maximizes the weights of the edges $(1, j)$ for j in that cluster.

What is the probability that point 1 will be assigned the wrong cluster using σ^*? There are four categories of points:

1. j such that $\sigma_0(j) = 1$ and $\sigma^*(j) = 1$. There are $n/2 - \gamma/2$ of these. $A_{1j} \sim P$.
2. j such that $\sigma_0(j) = 2$ and $\sigma^*(j) = 1$. There are $\gamma/2$ of these. $A_{1j} \sim Q$.
3. j such that $\sigma_0(j) = 1$ and $\sigma^*(j) = 2$. There are $\gamma/2$ of these. $A_{1j} \sim P$.
4. j such that $\sigma_0(j) = 2$ and $\sigma^*(j) = 2$. There are $n/2 - \gamma/2$ of these. $A_{1j} \sim Q$.

Thus,

$$\sum_{j\,:\,\sigma^*(j)=1} \sum_{l=0}^{L} \log \frac{P_l}{Q_l} \mathbf{1}_{A_{1j}=l} \leq \sum_{j\,:\,\sigma^*(j)=2} \sum_{l=0}^{L} \log \frac{P_l}{Q_l} \mathbf{1}_{A_{1j}=l}$$

if and only if

$$\sum_{j=1}^{n/2-\gamma/2} X_i + \sum_{j=1}^{\gamma/2} Y_i \leq \sum_{j=1}^{n/2-\gamma/2} Y_j + \sum_{j=1}^{\gamma/2} X_j$$
$$\sum_{j=1}^{n/2-\gamma} X_i \leq \sum_{j=1}^{n/2-\gamma} Y_j$$

where $X_j = \log \frac{P_l}{Q_l}$ with probability P_l and $Y_j = \log \frac{P_l}{Q_l}$ iwth probability Q_l. Note that $\mathbb{E}[X_j] > \mathbb{E}[Y_j]$.

Standard Chernoff argument yield that the probability of error is at most $\exp(-(1 - 2\frac{\gamma}{n})\frac{n}{2} I)$. This seems to indicate that refinement based on σ^* will achieve that oracle rate so long as σ^* is consistent, i.e., $\gamma = o(n)$.

4 $L \to \infty$ Case.

In this case, suppose that

$$\frac{n}{K \log K} \sum_{l=0}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 \to \infty$$

Again, we want to know when we can reduce the multi-color problem to a Bernoulli problem. More precisely, we would like to say that there exists a subset of colors $C \subset \{0, ..., l\}$ such that $\frac{n}{K \log K} I_{tot} \to \infty$ implies that $\frac{n}{K \log K} I_C \to \infty$. This would mean that clustering based on the labels

$$\widetilde{A}_{ij} = I_{A_{ij} \in C}$$

is consistent whenever consistency is possible under the oracle setting.

The current results in this section are:

1. Choosing C as a singleton is inconsistent.
2. Choosing $C = \{l : P_l > Q_l\}$ is inconsistent in general. It is consistent under further assumptions.

Let us prove the first bullet point.

Proposition 4.1. *There exist distributions P, Q such that, for any l, we have that $\frac{n}{K \log K} I_{tot} \to \infty$ but $\frac{n}{K \log K} I_C \to c < \infty$ for $C = \{l\}$.*

Proof. We define P, Q as such:

$$P_l = \begin{cases} \frac{1}{n} & \text{for } l = 1, ..., \log n \\ 0 & \text{for } l > \log n \end{cases} \qquad Q_l = \begin{cases} \frac{1}{2n} & \text{for } l = 1, ..., 2 \log n \\ 0 & \text{for } l > 2 \log n \end{cases}$$

with $P_0 = 1 - \frac{\log n}{n}$ and $Q_0 = 1 - \frac{\log n}{n}$.

$$\begin{aligned} I_{tot} &= \left(\sum_{l=1}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 \right) (1 + o(1)) \quad \text{(by lemma 5.1)} \\ &\geq \left(\log n \left(\sqrt{\frac{1}{n}} - \sqrt{\frac{1}{2n}} \right)^2 \right) (1 + o(1)) \\ &= \frac{\log n}{n} (1 - \sqrt{1/2})^2 (1 + o(1)) \end{aligned}$$

Therefore, for fixed K, $\frac{n}{K \log K} I_{tot} \to \infty$.

Let us now upper bound I_C. Clearly, if $C = \{0\}$, then $I_C = 0$. Therefore, let $C = \{l\}$ for some $l > 0$. First, suppose $l \in \{1, ..., \log n\}$.

$$\begin{aligned} I_C &= (\sqrt{P_l} - \sqrt{Q_l})^2 (1 + o(1)) \\ &= (\sqrt{1/n} - \sqrt{1/2n})^2 (1 + o(1)) \\ &\leq \frac{1}{n} (1 + o(1)) \end{aligned}$$

If $l \in \{\log n + 1, ..., 2\log n\}$, then it is clear that $I_C = \frac{1}{2n}(1 + o(1))$.

Therefore, $\frac{nI_C}{K \log K} \to c < \infty$

$\square$

Now we move on to the second point.

We first state a useful lemma:

Lemma 4.1. *Let $C \subset \{0, ..., l\}$ be a subset of the colors. Let $P_C = \sum_{l \in C} P_l$ and Q_C be similarly defined.*

Let $d_C = P_C - Q_C$ and assume w.l.o.g. $d_C > 0$. Suppose

$$\frac{d_C}{P_C} \to 0 \qquad P_C \to 0$$

Then, we have that

$$I_C = \Theta\left(\frac{d_C^2}{P_C}\right)$$

Proof.

$$\begin{aligned}
I_C &= (\sqrt{P_C} - \sqrt{Q_C})^2(1 + o(1)) \\
&= \left(P_C(1 - \sqrt{\frac{Q_C}{P_C}})^2\right)(1 + o(1)) \\
&= \left(P_C\left(1 - \sqrt{1 - \frac{P_C - Q_C}{P_C}}\right)^2\right)(1 + o(1)) \\
&= \left(P_C\left(1 - (1 - \frac{1}{2}\frac{d_C}{P_C}(1 + o(1)))\right)^2\right)(1 + o(1)) \\
&= P_C\left(\frac{1}{2}\frac{d_C}{P_C}\right)^2(1 + o(1)) \\
&= \frac{1}{4}\frac{d_C^2}{P_C}(1 + o(1))
\end{aligned}$$

$\square$

Now, we have the following unachievability proposition.

Proposition 4.2. *There exist distributions P, Q such that $\frac{n}{K \log K} I_{tot} \to \infty$ but $\frac{n}{K \log K} I_C \to c < \infty$ for $C = \{l : P_l \geq Q_l\}$.*

Proof. Let $x, d > 0$ such that $x, d \to 0$. Let P, Q be defined as the following:

$$P_l = \begin{cases} d & \text{if } l = 1 \\ (x - d)2^{-l+1} & \text{if } l > 1 \end{cases} \qquad Q_l = \begin{cases} 0 & \text{if } l = 1 \\ (x - d)2^{-l+1} & \text{if } l > 1 \end{cases}$$

We therefore have $P_0 = 1 - x$ and $Q_0 = 1 - (x - d)$.

We have that

$$\begin{aligned} I_{tot} &= \sum_{l=1}^{\infty}(\sqrt{P_l} - \sqrt{Q_l})^2(1+o(1)) \\ &= d(1+o(1)) \end{aligned}$$

We then set $d = \frac{\log n}{n}$ and $x = \frac{\log^2 n}{n}$ so that $I_C = \Theta(\frac{d^2}{x})$ by Lemma 4.1.

$$I_{tot} = \Theta\left(\frac{\log n}{n}\right) \qquad I_C = \Theta\left(\frac{1}{n}\right)$$

□

Finally, we have a condition under which we have achievability (ignoring computational difficulties). The intuition is that we need P_l and Q_l to be on the same scale.

Proposition 4.3. *Let $f(n)$ be some sequence such that $f(n) \to 0$. Let $a_1, a_2, \ldots$ be constants such that $\sum_{l=1}^{\infty} a_l < \infty$ and $b_1, b_2, \ldots$ be constants such that $\sum_{l=1}^{\infty} b_l < \infty$.*

Let $P_l = a_l f(n)$ and $Q_l = b_l f(n)$ for $l > 0$. Then, we have, so long as $\frac{n}{K \log K} I_{tot} \to \infty$, that

$$\frac{n}{K \log K} I_C \to \infty$$

for $C = \{l : P_l \geq Q_l\}$.

Proof. First, it is clear that the total variation distance

$$\begin{aligned} \|P - Q\|_{TV} &= \frac{1}{2}\sum_{l=0}^{\infty}|P_l - Q_l| \\ &= \frac{1}{2}\left(|P_0 - Q_0| + \sum_{l=1}^{\infty}|a_l f(n) - b_l f(n)|\right) \\ &= \frac{1}{2}\left(|1 - \sum_{l=1}^{\infty} a_l f(n) - (1 - \sum_{l=1}^{\infty} b_l f(n))| + \|a - b\|_1 f(n)\right) \\ &= \frac{1}{2} f(n)\left(|\sum_{l=1}^{\infty}(a_l - b_l)| + \|a - b\|_1\right) \end{aligned}$$

Thus, the total variation, which we denote by d, is of the order $d = \Theta(f(n))$. Since $I_{tot} \leq 2d(1+o(1))$, we have that $I_{tot} = O(f(n))$ as well.

We claim that $I_C = \Theta(f(n))$. The proof would be complete if we can validate this claim. Without loss of generality, let's suppose that $P_0 < Q_0$.

$$I_C = P_C\left(1 - \sqrt{\frac{Q_C}{P_C}}\right)^2(1+o(1))$$

Since $\frac{Q_C}{P_C} = \frac{\sum_{l \in C} b_l}{\sum_{l \in C} a_l} = c < 1$, we have that $I_C = \Theta(P_C) = \Theta(f(n))$.

$\square$

5 Technical Lemmas

Lemma 5.1. *Let $P = \{P_l\}_{l=0,\ldots,\infty}$ and $Q = \{Q_l\}_{l=0,\ldots,\infty}$ be two discrete distributions and suppose $P_0, Q_0 \to 1$.*

Then, we have that $I \to 0$ and

$$I = (1 + o(1)) \sum_{l=1}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2$$

Proof. First, it is clear that if $P_0, Q_0 \to 1$, then

$$\begin{aligned}
\sum_{l=0}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 &= (\sqrt{P_0} - \sqrt{Q_0})^2 + \sum_{l=1}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 \\
&= (\sqrt{P_0} - \sqrt{Q_0})^2 + \sum_{l=1}^{\infty} P_l + \sum_{l=1}^{\infty} Q_l - 2 \sum_{l=1}^{\infty} \sqrt{P_l Q_l} \\
&\leq (\sqrt{P_0} - \sqrt{Q_0})^2 + \sum_{l=1}^{\infty} P_l + \sum_{l=1}^{\infty} Q_l
\end{aligned}$$

Therefore, $\lim_{n \to \infty} \sum_{l=0}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 = 0$.

$$\begin{aligned}
I &= -2 \log \sum_{l=0}^{\infty} \sqrt{P_l Q_l} \\
&= -2 \log \left(1 - \frac{1}{2} \sum_{l=0}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 \right) \\
&= (1 + o(1)) \sum_{l=0}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 \quad \text{(since the sum tends to 0)}
\end{aligned}$$

We will show that $(\sqrt{P_0} - \sqrt{Q_0})^2 = o\left(\sum_{l=1}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2 \right)$ and the result follows immediately.

Let $P' = 1 - P_0$ and $Q' = 1 - Q_0$.

$$
\begin{aligned}
(\sqrt{P_0} - \sqrt{Q_0})^2 &= (\sqrt{1-P'} + \sqrt{1-Q'})^2 \\
&= (1-P')\left(1 - \sqrt{\frac{1-Q'}{1-P'}}\right)^2 \\
&= (1-P')\left(1 - \sqrt{1 - \frac{Q'-P'}{1-P'}}\right)^2 \\
&\le (1-P')\left(1 - (1 - \frac{1}{2}\left(\frac{Q'-P'}{1-P'}\right)(1+o(1)))\right)^2 \\
&\le (1-P')\left(\frac{1}{2}\left(\frac{Q'-P'}{1-P'}\right)(1+o(1))\right)^2 \\
&\le \frac{1}{4}\left(\frac{Q'-P'}{1-P'}\right)^2(1+o(1)) \le \frac{1}{4}(Q'-P')^2(1+o(1))
\end{aligned}
$$

$$
\begin{aligned}
\sum_{l=1}^{\infty}(\sqrt{P_l} - \sqrt{Q_l})^2 &= \sum_{l=1}^{\infty} P_l + Q_l - 2\sqrt{P_l Q_l} \\
&\ge P' + Q' - 2\sqrt{\left(\sum_{l=1}^{\infty} P_l\right)\left(\sum_{l=1}^{\infty} Q_l\right)} \\
&= P' + Q' - 2\sqrt{P'Q'} \\
&= (\sqrt{P'} - \sqrt{Q'})^2 \\
&= P'\left(1 - \sqrt{\frac{Q'}{P'}}\right)^2 \\
&= P'\left(1 - \sqrt{1 - \frac{P'-Q'}{P'}}\right)^2 \\
&\ge P'\left(1 - (1 - \frac{1}{2}\frac{P'-Q'}{P'}(1+o(1)))\right)^2 \\
&\ge P'\left(\frac{1}{2}\frac{P'-Q'}{P'}(1+o(1))\right)^2 \\
&\ge \frac{1}{4}\left(\frac{(P'-Q')^2}{P'}\right)(1+o(1))
\end{aligned}
$$

Thus, we have shown that

$$
\begin{aligned}
(\sqrt{P_0} - \sqrt{Q_0})^2 &\le \frac{1}{4}(Q'-P')^2(1+o(1)) \\
\sum_{l=1}^{\infty}(\sqrt{P_l} - \sqrt{Q_l})^2 &\ge \frac{1}{4}\frac{(P'-Q')^2}{P'}(1+o(1))
\end{aligned}
$$

Since $P' \to 0$, the proof is complete.

$\square$

6 Continuous Distributions

In this section, we suppose that the weights of within-cluster edges are drawn from a density p and that of between-cluster edges are drawn from a density q.

In the following analysis, we will suppose that p, q are mixtures. The first components of both p, q are identical, and the second component of q is a mean-shift of that of p.

$$\begin{aligned} p(x) &= \lambda f(x) + (1-\lambda) g(x) \\ q(x) &= \lambda f(x-\alpha) + (1-\lambda) g(x) \end{aligned}$$

for some constant $\lambda \in (0,1)$, some density $g(x)$ supported on the interval $[0,1]$, and some density $f(x)$ supported on the interval $[0, 1-\alpha]$.

For simplicity, we will suppose that $g(x) = 1$.

Proposition 6.1. *Let $I = -2\log \int \sqrt{p(x)q(x)}dx$ be the continuous Renyi divergence.*

Suppose α is sufficiently small, then, for some constants c, C, we have that

$$c\alpha^2 \le I \le C\alpha^2$$

Proof. First, denote $H = \int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx$ as the continuous Hellinger distance and we will first show that $c\alpha^2 \le H \le C\alpha^2$.

Note that $f(x-\alpha) = f(x) + f'(c_x)\alpha$ for some c_x in between $x - \alpha, x$. Therefore, $q(x) = p(x) - \lambda f'(c_x)\alpha$.

$$\begin{aligned} &\int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx \\ =&\int (\sqrt{p(x)} - \sqrt{p(x) - \lambda f'(c_x)\alpha})^2 dx \\ =&\int p(x) \left(1 - \sqrt{1 - \frac{\lambda f'(c_x)\alpha}{p(x)}}\right)^2 dx \end{aligned}$$

Since f' is bounded and $p(x)$ is bounded away from zero, we have that, for some η close to 0,

$$\begin{aligned} =&\int p(x) \left(1 - (1 - \frac{1}{2}\frac{\lambda f'(c_x)\alpha}{p(x)}(1-\eta))\right)^2 dx \\ =&\int p(x) \left(\frac{1}{2}\frac{\lambda f'(c_x)\alpha}{p(x)}\right)^2 (1-\eta) dx \\ =&\alpha^2(1-\eta) \int p(x) \left(\frac{1}{2}\frac{\lambda f'(c_x)}{p(x)}\right)^2 dx \end{aligned}$$

Now, it remains to bound I in terms of H. Since

$$\begin{aligned}
I =& -2\log \int \sqrt{p(x)q(x)}dx \\
=& -2\log\left(1 - \frac{1}{2}\int(\sqrt{p(x)} - \sqrt{q(x)})^2 dx\right) \\
=& -2\log(1 - \frac{1}{2}H)
\end{aligned}$$

Thus, for α small enough such that $H \leq 1/2$, we have that $2H \geq I \geq H$. $\square$

Let the interval $[0,1]$ be divided into L equally spaced sub-intervals. Let l index each of these sub-intervals and let $B = 1/L$ be the length of each sub-interval.

Let $P_l = \int_{\mathrm{Bin}_l} p(x)dx$ and $Q_l = \int_{\mathrm{Bin}_l} q(x)dx$. We define the *discretized Renyi divergence* as $\widetilde{I} = -2\log\sum_{l=1}^{L}\sqrt{P_l Q_l}$ and $\widetilde{H} = \sum_{l=1}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2$.

We will show that $\widetilde{H} \geq c\alpha^2$ for some constant c and for small enough α. We will also show that, for any α, $\lim_{L\to\infty}\widetilde{I} = I$.

These facts combined show that $\widetilde{I} \geq \widetilde{H} \geq c\alpha^2$ and therefore, $\widetilde{I} - I = \eta\alpha^2$ where $\lim_{L\to\infty}\eta = 0$.

Proposition 6.2. *Let $P_l, Q_l, \widetilde{I}$ be defined as above. Suppose $|p'(x)|, |q'(x)| \leq M$ for some absolute constant M.*

Then,

$$\lim_{L\to\infty}\widetilde{I} = I$$

Proof. Let $\mathrm{Bin}_l = [a_l, b_l]$.

$$\begin{aligned}
P_l &= \int_{\mathrm{Bin}_l} p(x)dx \\
&= \int_{a_l}^{b_l} p(x)dx \\
&= \int_{a_l}^{b_l} p(a_l) + p'(c_x)(x - a_l)dx \quad \text{for } c_x \in [a_l, b_l] \\
&= Bp(a_l) + B^2\xi_l \quad \text{where } |\xi_l| \leq M/2
\end{aligned}$$

Likewise, we have that $Q_l = Bq(a_l) + B^2\xi_l'$.

$$\begin{aligned}
\widetilde{I} &= -2\log\sum_{l=1}^{L}\sqrt{P_l Q_l} \\
&= -2\log\sum_{l=1}^{L} B\sqrt{(p(a_l) + B\xi_l)(q(a_l) + B\xi_l')} \\
&= -2\log\sum_{l=1}^{L} B\sqrt{p(a_l)q(a_l)} + \left(B^2\frac{\xi_l}{p(a_l)} + B^2\frac{\xi_l'}{q(a_l)}\right)(1+\eta)
\end{aligned}$$

where $\eta \to 0$ as $B \to 0$. $p(a_l), q(a_l)$ are lower bounded and ξ_l, ξ_l' are bounded. Thus, we have that

$$\begin{aligned}\lim_{B\to 0} \widetilde{I} &= \lim_{B\to 0} -2\log\left(\sum_{l=1}^{L} B\sqrt{p(a_l)q(a_l)} + \sum_{l=1}^{L} B^2 c_l(1+\eta)\right) \\ &= -2\log\left(\int \sqrt{p(x)q(x)}dx\right)\end{aligned}$$

□

Proposition 6.3. *Let $\widetilde{I} = -2\log\sum_{l=1}^{L}\sqrt{P_l Q_l}$. Then, for sufficiently small α, $\widetilde{I} \geq c\alpha^2$ for some absolute constant c.*

Proof.

$$\begin{aligned}\widetilde{H} &= \sum_{l=1}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2 \\ &= \sum_{l=1}^{L} P_l\left(1 - \sqrt{Q_l}P_l\right)^2 \\ &= \sum_{l=1}^{L} P_l\left(1 - \sqrt{1 - \frac{P_l - Q_l}{P_l}}\right)^2\end{aligned}$$

We bound the terms $P_l - Q_l$ and P_l separately.

$$\begin{aligned}P_l - Q_l &= \int_{\mathrm{Bin}_l} p(x) - q(x)dx \\ &= \int_{a_l}^{b_l} \lambda f'(c_x)\alpha dx \\ &= \lambda\xi_l\alpha B\end{aligned}$$

where $|\xi_l| \leq M$.

$$P_l = \int_{\mathrm{Bin}_l} p(x)dx \geq B(1-\lambda)$$

Thus, for small enough α, we have that

$$\begin{aligned}\widetilde{H} &= \sum_{l=1}^{L} P_l\left(c_l\alpha\right)^2 \\ &\geq c\alpha^2\end{aligned}$$

for some constant c. The claim follows since $\widetilde{I} \geq \widetilde{H}$.

□

7 A More General Condition

References

[1] Anderson Y Zhang and Harrison H Zhou. Minimax rates of community detection in stochastic block model.