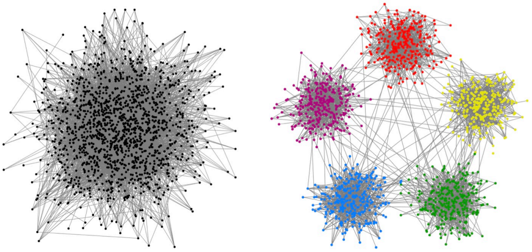
Community detection from a Random Graphs Perspective

Most basic unsupervised learning task on networks.

- Communities are usually densely connected parts of the network.



Given an unlabelled observation of the graph (on the left) our task is to produce the communities (on the right).

Numerous applications in

- Recommender systems
- Webpage sorting
- functionality of human brain
- Social networks.

Questions :-

① Can we always extract communities if they are present?

② How to asses performance of community detection algorithms?

These questions have led to the field of community detection as we know it today. We will try to answer these questions for a probabilistic model called "Stochastic block Model" which is also known as the "mother model" for community detection.

Stochastic Block Model :-

Model parameters :-

- K (number of communities)
- $p = (p_1 \dots p_K)$ a probability vector (dictating sizes of communities)
- Q a $K \times K$ symmetric matrix (connection probabilities) $\left[Q_{ij} > 0 \quad p_i > 0 \right]$
- ρ_n a sparsity parameter.

Model :- $SBM(n, p, \rho_n Q)$

- Generate $\underline{\sigma(u) \overset{iid}{\sim} p}$ for each vertex $u = 1 \dots n$
- Add an edge $\{u, v\}$ with probability $\underline{\rho_n Q_{\sigma(u)\sigma(v)}}$.

Note :- Let $\Sigma_i = \{ u : \sigma(u) = i \}$ (ie. set of vertices in community i)

Then $\dfrac{|\Sigma_i|}{n} \xrightarrow{P} p_i$ (Law of large numbers)

Recovery problems :- We will formulate community detection as a statistical question where $(G, \sigma) \sim SBM(n, p, Q)$.

$G = g$ is observed, but σ is unknown.

- want to find a good estimator $\hat{\sigma}$ of σ.

Agreement :- $A(\sigma, \hat{\sigma}) = \max\limits_{\pi \in S_K} \dfrac{1}{n} \sum\limits_{i=1}^{n} \mathbb{1}\{ \sigma(i) = \pi(\hat{\sigma}(i)) \}$

$\nwarrow$ set of permutations of $(1, \dots K)$.

Partition associated with $\sigma, \hat{\sigma}$:-

Let $\Sigma_i = \{ u : \sigma(u) = i \}$ and $\Sigma = \{ \Sigma_1, \dots, \Sigma_K \}$

Similarly $\hat{\Sigma}_i = \{u : \sigma(u) = i\}$ and $\hat{\Sigma} = \{\hat{\Sigma}_1, \dots, \hat{\Sigma}_k\}$

Exact recovery :- Find estimator $\hat{\sigma}$ s.t.

$$\lim_{n\to\infty} P\left(A(\sigma,\hat{\sigma}) = 1 \right) = 1.$$

Equivalently $\lim_{n\to\infty} P\left(\Sigma = \hat{\Sigma} \right) = 1$.

When $\rho_n = \frac{\log n}{n}$, ie. average degrees scale as $\log n$, then one sees a sharp phase transition from information-theoretic impossibility and achievability.

Almost exact recovery :- Find estimator $\hat{\sigma}$ s.t.

$$\lim_{n\to\infty} P\left(A(\sigma,\hat{\sigma}) \geq 1-\varepsilon \right) = 1 \quad \forall \ \varepsilon > 0.$$

Possibility depends on whether $n\rho_n \to \infty$ or not. No sharp phase transition here.

Partial recovery :- Defining this is tricky. For example, if $\hat{\sigma}$ is an independent copy of σ, then

$$\mathbb{E}\left[A(\sigma, \hat{\sigma}) \right] \geq \frac{1}{n} \sum_u P\left(\hat{\sigma}(u) = \sigma(u) \right)$$

$$= \sum_i p_i^2 .$$

Thus one can get a good overlap by just random guessing, which is obviously not indicative of community recovery.

Another example is $\hat{\sigma}(u) = \underset{i \in [K]}{\arg\max}\, p_i \quad \forall u$.

Then $\mathbb{E}[A(\sigma, \hat{\sigma})] \geq \max p_i$

Due to such scenarios, one defines partial recovery in terms of

$$\tilde{A}(\sigma, \hat{\sigma}) = \max_{\pi \in S_K} \frac{1}{K} \sum_{i=1}^{K} \frac{\#\{u : \sigma(u) = i,\ \sigma(u) = \pi(\hat{\sigma}(u))\}}{\#\{u : \sigma(u) = i\}}$$

Then partial recovery is possible if

$$\lim_{n \to \infty} \mathbb{P}\left(\tilde{A}(\sigma, \hat{\sigma}) \geq \alpha\right) = 1 \quad \text{for } \alpha > \frac{1}{K}.$$

For the symmetric SBM, with $Q_{ij} = \begin{cases} q_{in} & \text{for } i = j \\ q_{out} & \text{for } i \neq j \end{cases}$ and $p_i = 1/K$, we can define it as

$$\boxed{\lim_{n \to \infty} P\left(A(\sigma, \hat{\sigma}) \geq \alpha\right) = 1}$$

There is a whole literature on partial recovery showing sharp information-theoretic transition at $\delta_n = \frac{1}{n}$ and even that no computationally efficient algorithm can achieve partial recovery upto information theoretic threshold if $K \geq 4$. We will not go into this interesting literature in this course.

Exact recovery

Maximum A Posteriori (MAP) Estimator :-

Defn :- MAP estimator partition, denoted by $\hat{\Sigma}_{MAP}$, is computed by solving the following:

maximize $P(\Sigma = S \mid G = g)$ over all partitions $S = \{S_1, \dots, S_K\}$

If there are multiple maximizers, we pick one uniformly among them.

Lemma :- Let $\hat{\Sigma}_{MAP}$ be the MAP estimator partition, and $\hat{\Sigma}$ be the partition from any other estimator.

Then $P(\hat{\Sigma}_{MAP} \neq \Sigma) \leq P(\hat{\Sigma} \neq \Sigma)$

Pf :-

$$P(\Sigma \neq \hat{\Sigma}) = \sum_{g} \underbrace{P(\Sigma \neq \hat{\Sigma} \mid G = g)}_{\text{MAP minimizes this for any given } g.} P(G = g)$$

Therefore $P(\Sigma \neq \hat{\Sigma}) \geq P(\Sigma \neq \hat{\Sigma}_{MAP})$ for any $\hat{\Sigma}$ ■

Conclusion :- If $\hat{\Sigma}_{MAP}$ fails with probability bounded away from zero, then exact recovery is impossible.

If $\hat{\Sigma}_{MAP}$ succeeds whp, then obviously exact recovery is possible.

(Throughout whp means with probability → 1 as $n \to \infty$)

Let's try to compute $\hat{\Sigma}_{MAP}$ in a simple scenario ....

- σ is chosen uniformly from $\mathcal{X} = \{ s : \#\{u : s(u) = 1\} = n/2 \}$

  i.e. σ has two communities of equal size.

- Let $\rho_n Q = \begin{pmatrix} q_{in} & q_{out} \\ q_{out} & q_{in} \end{pmatrix}$ and $q_{in} > q_{out}$

- Thus $\Sigma = (\Sigma_1, \Sigma_2)$ is a uniform partition with $|\Sigma_1| = |\Sigma_2| = n/2$.

Now, by Bayes theorem

$$\mathbb{P}(\Sigma = S \mid G = g) \propto \mathbb{P}(G = g \mid \Sigma = S) \; \mathbb{P}(\Sigma = S)$$

$$= \mathbb{P}(G = g \mid \Sigma = S)$$

So computing MAP is equivalent to

$$\underset{S \in \Sigma}{\text{maximize}} \; P(G = g \mid \Sigma = S) \longrightarrow \textcircled{*}$$

Let

$N_{out}(g, S) :=$ number of between community edges in g w.r.t. S.

$N(g) :=$ number of edges in g.

$$\mathbb{P}(G = g \mid \Sigma = S)$$

$$= q_{out}^{N_{out}(g,S)} (1 - q_{out})^{\frac{n}{2} \times \frac{n}{2} - N_{out}(g,S)} \times$$

$$q_{in}^{N(g) - N_{out}(g,S)} \qquad (1 - q_{in})^{\binom{\frac{n}{2}}{2} + \binom{\frac{n}{2}}{2} - (N(g) - N_{out}(g,S))}$$

$$\propto \left(\frac{q_{out}/(1-q_{out})}{q_{in}/(1-q_{in})} \right)^{N_{out}(g,S)}$$

(the propotionality symbol hides terms that do not depend on S)

Now, $f(x) = \frac{x}{1-x}$ is a strictly increasing f^n and

therefore $$\frac{q_{out}/(1-q_{out})}{q_{in}/(1-q_{in})} < 1$$ whenever $q_{out} < q_{in}$

By $\circledast$,

$$\hat{\Sigma}_{MAP} = \operatorname{argmin} N_{out}(g,S), \text{ over all } (S_1,S_2) \text{ s.t. } |S_1| = |S_2| = n/2 .$$

(This is the minimum bisection problem)

Challenges to analyze $\hat{\Sigma}_{MAP}$:-

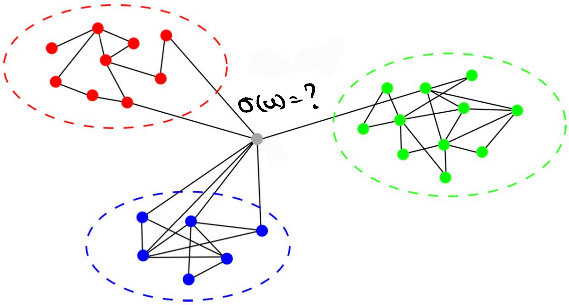
- Computing $\hat{\Sigma}_{MAP}$ is NP-hard
- Distribution of min-bisection is very hard to obtain which makes analysis of $\mathbb{P}(\hat{\Sigma}_{MAP} \neq \Sigma)$ extremely difficult.

To this end, let us look a concept called "genie-based estimator" that will be useful for us to prove the impossibility result as well as finding an efficient estimator later.

Genie-based estimator :-

Consider the hypothetical scenario, where in addition to the graph $G=g$, there is a "genie" who gives us access to $\sigma_{-u} := \{\sigma(v)\}_{v \neq u}$.

Our goal now is to estimate $\sigma(u)$.



We can again consider the MAP estimator, but now the given data is $G=g$ and σ_{-u}.

The resulting estimator is the genie-based estimator.

More precisely,

$$\boxed{\hat{\sigma}_{genie}(u) := \underset{s=1,\dots,K}{\arg\max}\ \mathbb{P}\left(\sigma(u)=s \mid G=g,\ \sigma_{-u}\right)} \longrightarrow (*)$$

Now,

$$\hat{\sigma}_{genie}(u) := \underset{j=1,\dots,K}{\arg\max}\ \mathbb{P}\left(\sigma(u)=j \mid G=g,\ \sigma_{-u}\right)$$

$$= \underset{j=1,\dots,K}{\arg\max}\ \mathbb{P}\left(G=g \mid \sigma(u)=j,\ \sigma_{-u}\right) p_j \quad \text{(By Bayes theorem)}$$

$$= \underset{j=1,\dots,K}{\arg\max}\ \mathbb{P}\left(D(u)=d(u) \mid \sigma(u)=j,\ \sigma_{-u}\right) p_j \longrightarrow (**)$$

where $D(u) = (D_1(u), D_2(u), \dots D_K(u))$ is the degree profile of u

and $D_i(u) = \#$ vertices in community i from u (Note that degree profiles depend on G, σ_{-u})

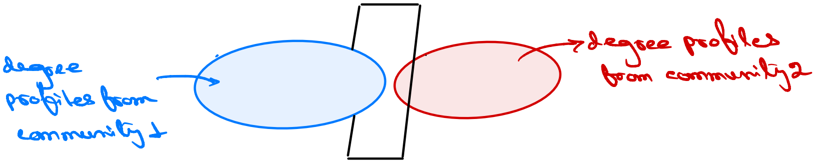
So $\hat{\sigma}_{genie}(u)$ maximizes the (weighted) likelihood of observing the given degree profile.

Exercise:- Consider the simple example discussed on page 5.

Prove that $\hat{\sigma}_{genie}(u) = \begin{cases} 1 & \text{if } D_1(u) \geq D_2(u) \\ 2 & \text{if } D_1(u) < D_2(u) \end{cases}$

Another way to look at it is as follows:

if we plot the degree profiles $D(u) = (D_1(u), D_2(u))$ in $\mathbb{R}^2$, then $\hat{\sigma}_{genie}$ tries to seperate them using a hyperplane


degree profiles from community 1
degree profiles from community 2

This idea of seperating degree profiles via hyperplanes is true in the general K-community case as well.

Coming back to the general case,

we have a pretty good handle on the distributions of $D(u) \mid \sigma(u) = j$. The distribution is multivariate binomial which is asymptotically close to a multivariate Poisson distribution.

Next, lets analyze the error probability of $\hat{\sigma}_{genie}(u)$...

$$P_e(u) = \mathbb{P}\left(\hat{\sigma}_{genie}(u) \neq \sigma(u) \mid \sigma_{-u} \right)$$

$$= \sum_{i=1}^{k} \mathbb{P}\left(D(u) \in Bad(i) \mid \sigma(u) = i, \sigma_{-u} \right) p_i$$

($Bad(i)$ is the region in $\mathbb{Z}_+^k$ where i is not maximizing ⊛⊛)

$$\leq \sum_{i=1}^{k} \sum_{j \neq i} \mathbb{P}\left(D(u) \in Bad_j(i) \mid \sigma(u) = i, \sigma_{-u} \right) p_i$$

($Bad_j(i)$ is the region in $\mathbb{Z}_+^k$ where j beats i while maximizing ⊛⊛.
Hence, $d \in Bad_j(i)$ if
$\mathbb{P}(D(u) = d \mid \sigma(u) = i, \sigma_{-u}) p_i$
$< \mathbb{P}(D(u) = d \mid \sigma(u) = j, \sigma_{-u}) p_j$)

$$= \sum_{i<j} \sum_{d \in \mathbb{Z}_+^k} \min\left\{ \mathbb{P}(D(u) = d \mid \sigma(u) = i, \sigma_{-u}) p_i , \mathbb{P}(D(u) = d \mid \sigma(u) = j, \sigma_{-u}) p_j \right\}$$

$$=: \tilde{P}_e$$

Exercise :- Prove that $P_e(\omega) \geq \frac{1}{K-1} \tilde{P}_e$

Hint : $\sum_{j \neq i} \mathbb{P}\left(\sigma(u) \in \mathrm{Bad}_j(i) \mid \sigma(u) = i, \sigma_{-u}\right)$

$$\leq (K-1)\, \mathbb{P}\left(\sigma(u) \in \mathrm{Bad}(i) \mid \sigma(u) = i, \sigma_{-u}\right)$$

Lemma 1 (Abbe, Sandon '15) :- If $p_n = \log n / n$, then

$$\tilde{P}_e = n^{-I(P,Q) + o\left(\frac{\log\log n}{\log n}\right)} \quad \text{whp}$$

where,

$$I(P,Q) = \min_{i<j} CH\left((P_\ell Q_{i\ell})_{\ell=1}^{K}, (P_\ell Q_{j\ell})_{\ell=1}^{K}\right)$$

$$CH(\mu, \nu) = \max_{t \in [0,1]} \sum_x \nu(x) f_t\left(\frac{\mu(x)}{\nu(x)}\right)$$

$$f_t(y) = 1 - t + ty - y^t$$

$CH(\cdot,\cdot)$ is called Chernoff-Hellinger divergence, a notion of distance between two probability distributions, much like KL divergence.

Remark :- Lemma 1 is our crucial estimate. We will show that

$I(P,Q) < 1$, ie. $n P_e(u) \to \infty$ ⇒ impossibility

$I(P,Q) > 1$, ie. $n P_e(u) \to 0$ ⇒ achievability.

Impossibility :-

Th^{m} 1:- Suppose $(G,\sigma) \sim SBM(n, k, p, \frac{\log n}{n} Q)$ and $I(P,Q) < 1$. Then, for any estimator $\hat{\sigma}$,

$$\operatorname*{Lt}_{n\to\infty} \mathbb{P}(\hat{\sigma} \neq \sigma) = 1,$$ i.e. exact recovery is impossible.

Proof :- Enough to prove that

$$\mathbb{P}(\hat{\Sigma}_{MAP} \neq \Sigma) \to 1.$$

Now, $\mathbb{P}(\hat{\Sigma}_{MAP} \neq \Sigma)$

$$\geq \mathbb{P}(\exists u : \hat{\sigma}_{genie}(u) \neq \sigma(u))$$

[Since $\hat{\sigma}_{genie}(u)$ is a more powerful estimator as it minimize error with additional information σ_{-u}] .

Now, by Lemma above,

$$\mathbb{E}\left[\#\{u : \hat{\sigma}_{genie}(u) \neq \sigma(u)\}\right] = n^{1 - I(P,Q) + o(1)}$$

$$\to \infty \quad \text{if } I(P,Q) < 1$$

A basic fact in probability theory states that if $\{E_u\}_u$ are pairwise independent, and $\sum_u \mathbb{P}(E_u) \to \infty$, then

$$\mathbb{P}\left(\bigcup_u E_u\right) \to 1.$$

(Exercise: Prove this basic fact)

However, $\{\hat{\sigma}_{genie}(u) \neq \sigma(u)\}$ and $\{\hat{\sigma}_{genie}(v) \neq \sigma(v)\}$ are not really independent — but the dependence is weak.

In fact, one can proceed as follows:

Let $Z_u = \mathbb{1}\{\hat{\sigma}_{genie}(u) \neq \sigma(u)\}$

and $Z = \sum_u Z_u$

Then

① $\mathbb{E} Z \to \infty$ (by Lemma as discussed above)

② $\dfrac{\mathbb{P}(Z_v=1 \mid Z_u=1)}{\mathbb{P}(Z_v=1)} \to 1$. (Left as an exercise, note that the dependence is only caused by one edge $\{u,v\}$).

③ Above ② implies $\dfrac{\mathrm{Var}(Z)}{(\mathbb{E} Z)^2} \to 0$ (Exercise).

This completes the proof of Th^{m} 1.

∎

Finding estimators achieving information-theoretic threshold :-

We want to find a good estimator $\hat{\sigma}$ s.t. it achieves exact recovery whenever $I(P,Q) > 1$.

We will again seek help of the "genie-based" estimator.

The genie-based estimator $\hat{\sigma}_{genie}(u)$ depends on the graph G and $\sigma_{-u} := \{\sigma(v)\}_{v \neq u}$.

i.e. $\hat{\sigma}_{genie}(u) = \hat{\sigma}_{genie}(u, G, \sigma_{-u})$

Since σ_{-u} is unknown, $\hat{\sigma}_{genie}(u)$ is not computable.

Idea for two-step estimator (Algorithm 1):

(S1) Good guess: Take an initial estimator that is good enough, i.e. $\hat{\sigma}_1$ s.t. $A(\sigma, \hat{\sigma}) \approx 1$.

(S2) Clean up: compute $\hat{\sigma}_2(u) = \hat{\sigma}_{genie}(u, G, \hat{\sigma}_{1,-u})$ for all u. $(\hat{\sigma}_{1,-u} = \{\hat{\sigma}_1(v)\}_{v \neq u}$

- $\hat{\sigma}_1$ can be viewed as a "noisy version" of the information provided by the genie. As long as the estimator $\hat{\sigma}_{genie}$ is robust to small noise, we can expect it to perform our desired task.

Advantages :- ① If $\hat{\sigma}_1$ is good enough then $\hat{\sigma}_2 \approx \hat{\sigma}_{genie}$, and therefore $\hat{\sigma}_2$ recovers communities upto the information-theoretic threshold.

② If $\hat{\sigma}_1$ is efficient, then $\hat{\sigma}_2$ is overall efficient.

Challenges :- ① How to find good $\hat{\sigma}_1$?

② $\hat{\sigma}_2(u) = \hat{\sigma}_{genie}(u, G, \hat{\sigma}_{1,-u}(G))$ is difficult to analyze

both dependence on G makes the distribution of degree profiles hard to analyze.

We will deal with challenge ① later.

- For challenge ② we will use an idea called "<u>graph splitting</u>".

Basically, we want to split the data G in two roughly independent parts (G_1, G_2). Then we compute $\hat{\sigma}_1$ on G_1 and then use G_2 in clean-up step. In otherwords, we compute

$$\hat{\sigma}_2(u) := \hat{\sigma}_{genie}(u ; G_2 , \hat{\sigma}_{1,-u}(G_1))$$

If (G_1, G_2) are independent, then the degree profiles would become poisson again and $\hat{\sigma}_{genie}$ becomes easier again.

Def^{n} (Graph splitting):- (G_1, G_2) is constructed from G as:

① G_1 is the graph by picking each edge of G with prob γ independently.

② $G_2 = G \setminus G_1$ contains rest of the edges.

Note :- If $G \sim SBM(n, K, p, \frac{\log n}{n} Q)$,

then $G_1 \sim SBM(n, K, p, \gamma \frac{\log n}{n} Q)$

and $G_2 | G_1 \sim SBM(n, K, p, (1-\gamma) \frac{\log n}{n} Q)$ with edges of G_1 forbidden.

However, (G_1, G_2) are still dependent.

Next lemma shows that replacing G_2 by an independent copy $\tilde{G}_2$

Lemma 2:- Let $G \sim SBM(n, K, p, \frac{\log n}{n} Q)$, and (G_1, G_2) be a graph splitting as above with $\gamma = o\left(\frac{\log \log n}{\log n}\right)$

Let $\hat{\sigma}_1 = \hat{\sigma}_1(G_1)$ be s.t. it achieves almost exact recovery,

i.e. $\lim_{n \to \infty} P(A(\hat{\sigma}_1, \sigma) \geq 1 - \varepsilon_n) = 1$ for some $\varepsilon_n \to 0$

Take $\tilde{G}_2 \sim SBM(n, K, p, (1-\gamma) \frac{\log n}{n} Q)$ (independent of G)

Then for any u, and $d \in \mathbb{Z}_+^K$, (D_u = degree profile of u).

$$P(D_u(\hat{\sigma}_1, G_2) = d \mid G_1, \sigma(u) = i, \hat{\sigma}_{1,-u})$$
$$\leq (1 + o(1)) P(D_u(\hat{\sigma}_1, \tilde{G}_2) = d \mid \sigma(u) = i, \hat{\sigma}_{1,-u}) + n^{-\omega(1)}$$

with high probability,

where $\omega(1) \to \infty$.

The proof of this lemma is skipped here. The idea is that the difference in degree profiles can only occur due to edges that are in G_1 and also in $\tilde{G}_2$. The expected number of such edges is $O\left(\binom{n}{2}\left(\frac{\log n}{n}\right)^2\right) = O(\log^2 n)$ which is much smaller than the expected number of edges in G_2 or $\tilde{G}_2$.

The next corollary shows usefulness of this Lemma:

Corollary 1:-

$$\mathbb{P}\left(\hat{\sigma}_{genie}\left(u, G_2, \hat{\sigma}_{1,-u}(G_1)\right) \text{ makes error}\right)$$
$$\leq \mathbb{P}\left(\hat{\sigma}_{genie}\left(u, \tilde{G}_2, \hat{\sigma}_{1,-u}(G_1)\right) \text{ makes error}\right)$$

Pf (Hint):- Recall $Bad(i) := \{ d : \left(\mathbb{P}(D(u)=d \mid \sigma = j)\, p_j\right)_{j=1}^{k}$ is not maximized at $i \}$

use these events, and use Lemma 2 to conclude.

Th^m 3 :- Let $G \sim SBM\left(n, k, p, \frac{\log n}{n} Q\right)$, and (G_1, G_2) be a graph splitting as above with $\gamma = o\left(\frac{\log \log n}{\log n}\right)$

Let $\hat{\sigma}_1 = \hat{\sigma}_1(G_1)$ be s.t. it achieves almost exact recovery, i.e. $\lim_{n \to \infty} \mathbb{P}\left(A(\hat{\sigma}_1, \sigma) \geq 1 - \varepsilon_n\right) = 1$ for some $\varepsilon_n \to 0$.

Then, $\hat{\sigma}_2 := \left(\hat{\sigma}_{genie}\left(u, G_2, \sigma_{1,-u}(G_1)\right)\right)_{u=1}^{n}$ achieves exact recovery whenever $I(P, Q) > 1$.

Pf :-

$\mathbb{P}(\hat{\sigma}_2 \text{ makes error})$

($x \lesssim y$ means $x \le (1+o(1))y + o(1)$)

$$\lesssim \mathbb{P}\left(\hat{\sigma}_2 \text{ makes error} \mid A(\hat{\sigma}_1, \sigma) \geq 1-\varepsilon_n\right)$$

$$\leq n\,\mathbb{P}\left(\hat{\sigma}_{\text{genie}}(u; G_2, \hat{\sigma}_{1,-u}) \text{ makes error} \mid A(\hat{\sigma}_1, \sigma) \geq 1-\varepsilon_n\right)$$

$$\lesssim n\,\mathbb{P}\left(\hat{\sigma}_{\text{genie}}(u; \tilde{G}_2, \hat{\sigma}_1) \text{ makes error} \mid A(\hat{\sigma}_1, \sigma) \geq 1-\varepsilon_n\right)$$

(By Lemma 2)

$$\leq n^{1+o(1)}\,\mathbb{P}\left(\hat{\sigma}_{\text{genie}}(u; \tilde{G}_2, \sigma_{-u}) \text{ makes error} \mid A(\hat{\sigma}_1, \sigma) \geq 1-\varepsilon\right) + o(1)$$

($\hat{\sigma}_{\text{genie}}(u; \tilde{G}_2, \hat{\sigma}_1)$ computes estimator of $\sigma(u)$ based on a potentially noisy degree profile $\tilde{D}$ when the real degree profile is D

Using that degree profile has Poisson distribution,

one can prove: $\dfrac{\mathbb{P}(\tilde{D}(u) = d \mid \sigma(u) = i)}{\mathbb{P}(D(u) = d \mid \sigma(u) = i)} \leq n^{o(1)}$)

$$= n^{1+o(1)}\,\underbrace{\mathbb{P}\left(\hat{\sigma}_{\text{genie}}(u; \tilde{G}_2, \sigma_{-u}) \neq \sigma(u)\right)}_{\text{using that } \tilde{G}_2 \text{ is independent of } \hat{\sigma}_1.} + o(1)$$

$$= n^{1+o(1)} \cdot n^{-I(P,Q) + o(1)} + o(1)$$

(Using Lemma 1 from Genie-based estimator error).

$\to 0$.

Corollary :- For two community SBM with $p = (1/2, \frac{1}{2})$

$$Q = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \quad \rho_n = \frac{\log n}{n}$$

$\sqrt{a} - \sqrt{b} < \sqrt{2} \implies$ impossibility

$\sqrt{a} - \sqrt{b} > \sqrt{2} \implies$ achievability.

How to produce a good $\hat{\sigma}$, ?

Our task now reduces to the following:

Consider $(G, \sigma) \sim SBM(n, K, p, \rho_n Q)$ where $n\rho_n \to \infty$

find estimator $\hat{\sigma}$, s.t.

$$\lim_{n\to\infty} P(A(\hat{\sigma}, \sigma) \geq 1 - \varepsilon_n) = 1$$

- whether such an estimator exists or not depends on the matrix Q.

Approach 1 (Sphere comparison algorithm) :-

we will just do heuristics.

Take two vertices v, v' and we want to decide whether they are in the same community or not. Let $N_r(v)$ be the number of vertices at exact distance r from v.

The idea is that the limiting behavior of $|N_r(v) \cap N_r(v')|$ (after suitable scaling) depends the community assignment for large enough r.

However $|N_r(v) \cap N_r(v')|$ is usually difficult to analyze since the events $\{u \in N_r(v)\}$ and $\{u \in N_r(v')\}$ are usally dependent. This is again resolved using the graph splitting idea.

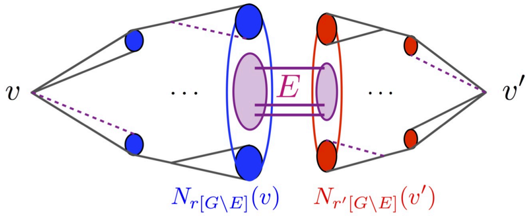
Choose a subset of edges E by sampling from all edges with fixed probability c.

Let $N_{r,r'}(v, v', E)$ be the number of pairs of vertices (v_1, v_2) s.t.

① v_1 is at distance r from v in $G \setminus E$.

② v_2 is at distance r from v' in $G \setminus E$.

③ $(v_1, v_2) \in E$.



[Abbe, Sandon '15] considered the following statistic:

$$I_{r,r'}(v, v', E) = N_{r+2,r'}(v, v', E)\, N_{r,r'}(v, v', E) - N^2_{r+1,r'}(v, v', E)$$

They showed that if Q is an irreducible matrix such that no two of its rows are identical, then for $r = r' = \frac{3}{4} \frac{\log n}{\log d}$ and c fixed $I_{r,r'}(v, v', \mathbb{B})$ is able to distinguish communitie between any two fixed vertices v, v' with high probability.

This allows them to find an estimator $\hat{\sigma}$, that achieves almost exact recovery.

Approach 2 (Spectral Algorithms) :- We will investigate this in detail.

Let's stick to the special case with two communities each having size $n/2$ for simplicity.

Think of σ to be chosen uniformly from the set

$$\Sigma = \left\{ s : \{1, \ldots, n\} \to \{1, 2\} \; : \; \#\{u : s(u) = 1\} = n/2 \right\}$$

Also, let $Q = p_n \begin{pmatrix} a & b \\ b & a \end{pmatrix}$ where $n p_n = \log \log n$.

Further, suppose there is a self-loop at each vertex with probability $p_n a$.

Lets look at the spectrum of the matrix $\bar{A} = \mathbb{E} A$.

$$\bar{A}(u,v) = \begin{cases} a\gamma_n & \text{if } \sigma(u) = \sigma(v) \\ b\gamma_n & \text{if } \sigma(u) \neq \sigma(v) \end{cases}$$

(Note that, due to self-loops, we have $\bar{A}(u,u) = \frac{a \log n}{n}$)

$\bar{A}$ is a rank 2 matrix.

- The top eigenvalue is $\bar{\lambda}_1 = \left(\frac{a+b}{2}\right) n\gamma_n$ with eigenvector $\bar{\phi}_1 = \frac{1}{\sqrt{n}}(1, \ldots 1)$.

- The second eigenvalue is $\bar{\lambda}_2 = \left(\frac{a-b}{2}\right) n\gamma_n$ with eigenvector $\bar{\phi}_2$, where

$$\bar{\phi}_2(u) = \begin{cases} 1 & \text{if } \sigma(u) = 1 \\ -1 & \text{if } \sigma(u) = 2 \end{cases}$$

Now, we can view $A = \underbrace{\bar{A}}_{\text{signal}} + \underbrace{(A - \bar{A})}_{\text{noise}}$

If (λ_1, ϕ_1), (λ_2, ϕ_2) are the top two eigenpairs, we can expect

$$\boxed{\begin{array}{lll} \lambda_1 \approx \bar{\lambda}_1, & \phi_1 \approx \bar{\phi}_1 & \text{when } \|A - \bar{A}\| \text{ is small} \\ \lambda_2 \simeq \bar{\lambda}_2, & \phi_2 \simeq \bar{\phi}_2 & \end{array}}$$

Thus we expect ϕ_2 to be able to perform community recovery.

To make this precise, we need help from the theory of matrix perturbation.

Some Matrix perturbation theory :-

Suppose B, B_0 are any two symmetric matrices.

Let $\lambda_i(X), \phi_i(X)$ be the largest eigenvalue and eigenvector of the matrix X.

Let $\|X\|_{op} := \sup_{y \neq 0} \frac{\|Xy\|_2}{\|y\|_2}$ be the operator norm of X.

Thm. 1 (Weyl's inequality) :- $|\lambda_i(B) - \lambda_i(B_0)| \leq \|B - B_0\|_{op}$

Thm. 2 (Davis-Kahan theorem) :-

Suppose $B_0 = \sum_{j=1}^{K} \lambda_j(B_0)\, \phi_j(B_0)\, \phi_j(B_0)^T$ and $\operatorname{rank}(B_0) = K$.

Then

$$\min_{s \in \{\pm 1\}} \|s\phi_i(B) - \phi_i(B_0)\|_2$$

$$\leq \frac{C\|B - B_0\|_{op}}{\min\{\lambda_{i-1}(B_0) - \lambda_i(B_0),\ \lambda_i(B_0) - \lambda_{i+1}(B_0)\}}$$

where C is an absolute constant.

Conclusion :- Since $\bar{\lambda}_i \asymp n\rho_n$, we can show that

$\min_{s \in \{\pm 1\}} \|s\phi_i - \bar{\phi}_i\|_2 \xrightarrow{P} 0$ as long as $\|A - \bar{A}\|_{op} = o(n\rho_n)$ whp.

Existing random matrix theory shows that

$$\|A - \bar{A}\|_{op} \lesssim \sqrt{n\rho_n} \quad \text{as long as } n\rho_n \geq c\log n$$

Unfortunately, this does not hold for $n\rho_n \to \infty$.

As the graph gets sparser, the top eigenvalues and vectors of A tend to get dictated by high-degree vertices. For instance, for any graph,

$$\left\{\frac{1}{n}\sum d_i, \sqrt{d_{max}}\right\} \leq \lambda_1(A) \leq d_{max}$$

which shows that the top eigenvalue is always dictated by the max degree on a sparse graph.

To fix this issue, one can look at the 'trimmed' matrix $\tilde{A}$ defined by:

$$\tilde{A}_{ij} = \begin{cases} A_{ij} & \text{if } \quad d_i \text{ or } d_j < 20\|Q\|_\infty n\rho_n \\ 0 & \text{ow} \end{cases}$$

Lemma (Chin, Rao, Vu '15):- $\|\tilde{A} - \bar{A}\|_{op} \leq c_0\sqrt{n\rho_n}$ whp. for some constant $c_0 > 0$

Remark :- This lemma holds for arbitrary Q and $\rho_n \geq \frac{c_1}{n}$.

Corollary:- If $\text{rank}(\text{diag}(P)\, Q) = K$, then

$$\min_{s \in \{\pm 1\}} \| s\phi_j - \bar{\phi}_j \|_2 \xrightarrow{P} 0 \quad \forall j = 1, \dots, K.$$

For 2 communities, this suggests the following algo:

Algorithm:-

(S1) Construct $\tilde{A}$ as

$$\tilde{A}_{ij} = \begin{cases} A_{ij} & \text{if } \; d_i \text{ or } d_j < 20 \|Q\|_\infty n \rho_n \\ 0 & \text{ow} \end{cases}$$

(S2) Output $\text{sign}(\phi_2(\tilde{A}))$.

Remark:- For the K community case, one can compute the $n \times K$ matrix U that with the i-th column being u_i (the i-th largest eigenvalue).

Now to each vertex i, we can assign the i-th row of U as its "spectral embedding".

Then we can compute community assignments using a K-means clustering. One of the main problems with this method is that u_i's approximate $\bar{u}_i$'s only up to a sign. To that end, given any $s \in \{\pm 1\}^K$, we can compute $\hat{\sigma}(s)$ using the spectral embedding from $u \, \mathrm{diag}(s)$. Now we can choose the estimator maximizing weighted likelihood:

$$\hat{\sigma}_{spee} = \underset{s \in \{\pm 1\}^K}{\arg\max} \; P(\sigma = \hat{\sigma}(s) \,|\, G = g) \, P(\sigma = \hat{\sigma}(s))$$

How well does direct spectral algorithms perform?:

Consider again the simple case with two communities of equal sizes

$$Q = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \qquad \rho_n = \frac{\log n}{n}.$$

Define $\hat{\sigma}_{spec}(u) = \begin{cases} 1 & \text{if } \phi_2(u) > 0 \\ 2 & \text{if } \phi_2(u) \le 0 \end{cases}$

Does $\hat{\sigma}_{spec}$ achieve exact recovery whenever $\sqrt{a} - \sqrt{b} > \sqrt{2}$?

The Davis-Kahan theorem can never give us exact recovery guarantee.

To answer the above question, we need a stronger perturbation bound on the eigenvector ϕ_2.

More precisely, we need a perturbation bound on the ℓ_∞ norm.

Is $\|\phi_2 - \bar{\phi}_2\|_\infty = o\left(\frac{1}{\sqrt{n}}\right)$ whp in the $\rho_n = \frac{\log n}{n}$ regime?

This remained as a challenging question until [Abbe, Fan, Wang, Zhong '20] proved instead

$$\left\|\phi_2 - \frac{A\bar{\phi}_2}{\bar{\lambda}_2}\right\|_\infty = o\left(\frac{1}{\sqrt{n}}\right) \quad \text{whp}.$$

The result was proved for more general random matrices (not just for SBM) which we discuss below:-

Assumptions:-

Let A be a random matrix with $(a_{ij})_{1\le i\le j\le n}$ being independent and $a_{ij}=a_{ji}$. Let $\bar{A}=\mathbb{E}A$

Let $(\lambda_i,\varphi_i)_{i=1}^{K}$, $(\bar\lambda_i,\bar\varphi_i)_{i=1}^{K}$ be the top-K eigenpairs of $A,\bar{A}$.

① K is fixed and $\bar{A}$ has K distinct non-zero eigenvalues $\bar\lambda_1>\bar\lambda_2>\dots>\bar\lambda_K$ and $\bar\lambda_1=\Theta(\bar\lambda_K)$

Let $\Delta=\min_{i\in[K]}\{\bar\lambda_{i-1}-\bar\lambda_i\}$ $\quad \bar\lambda_0=+\infty$ (called eigengap)

② $\exists\ \gamma=\gamma_n\to 0$ s.t. $\|A-\bar{A}\|_{op}\le\gamma\Delta$ whp.

and $\|\bar{A}\|_{2\to\infty}\le\gamma\Delta$

③ $\exists\ \varphi:\mathbb{R}_+\to\mathbb{R}_+$ (continuous, non-decreasing, possibly depend on n) such that $\varphi(0)=0$, $\varphi(1)=O(1)$, $\frac{\varphi(x)}{x}$ is non increasing and for any $m\in[n]$, $w\in\mathbb{R}^n$,

$$|\langle A-\bar{A},w\rangle|\le\Delta\|w\|_\infty\,\varphi\left(\frac{\|w\|_2}{\sqrt{n}\,\|w\|_\infty}\right)$$

(When $A_{ij}\sim$ Normal$(\bar{A}_{ij},\sigma^2)$, then $\varphi(x)\propto x$

When $A_{ij}\sim$ Bernoulli$(\bar{A}_{ij})$, then $\varphi(x)\propto\frac{1}{(1\vee\log(1/x))}$)

Theorem (Abbe, Fan, Wang, Zhang '20) :- Under above conditions

$$\min_{s\in\{\pm1\}}\left\|\varphi_i-s\frac{A\bar\varphi_i}{\bar\lambda_i}\right\|_\infty=o(\|\varphi_i\|_\infty)\quad\text{whp.}$$

$$\forall\ i\in[K]$$

Pf :- will do the proof sketch in simple case $K=1$.

Suppose $\gamma < \frac{1}{2}$.

Then $|\lambda_1 - \bar{\lambda}_1| < \|A - \bar{A}\|_{op}$ (Weyl's inequality)

$\leq \gamma \bar{\lambda}_1$ (By Assumption 2)

$\Rightarrow \lambda_1 \geq \frac{\bar{\lambda}_1}{2}$ and $\left| \frac{1}{\lambda_1} - \frac{1}{\bar{\lambda}_1} \right| = \frac{|\lambda_1 - \bar{\lambda}_1|}{|\lambda_1 \bar{\lambda}_1|} \leq \frac{2\gamma}{\bar{\lambda}_1}$

Thus,

$$\left\| \phi_1 - \frac{A\bar{\phi}_1}{\lambda_1} \right\|_\infty$$

$$\leq \left| \frac{1}{\lambda_1} - \frac{1}{\bar{\lambda}_1} \right| \|A\bar{\phi}_1\|_\infty + \frac{1}{\lambda_1} \|A(\phi_1 - \bar{\phi}_1)\|_\infty$$

$$\leq \underbrace{\frac{2}{\bar{\lambda}_1} \gamma \|A\bar{\phi}_1\|_\infty}_{\textcircled{I}} + \underbrace{\frac{1}{\lambda_1} \|A(\phi_1 - \bar{\phi}_1)\|_\infty}_{\textcircled{II}}$$

$\textcircled{I} \leq \frac{2\gamma}{\bar{\lambda}_1} \left(\bar{\lambda}_1 \|\bar{\phi}_1\|_\infty + \|(A - \bar{A})\bar{\phi}_1\|_\infty \right)$ (using $\bar{A}\bar{\phi}_1 = \bar{\lambda}\bar{\phi}_1$)

$\leq \frac{2\gamma}{\bar{\lambda}_1} \left(\bar{\lambda}_1 \|\bar{\phi}_1\|_\infty + \bar{\lambda}_1 \|\bar{\phi}_1\|_\infty \psi(1) \right)$ (By Assumption 3)

$= o\left(\|\bar{\phi}_1\|_\infty \right)$.

Analysis of $\textcircled{II}$ is a bit delicate but we can do a trick called '<u>leave-one-out</u>'.

Let $\phi_1^{(m)}$ be the eigenvector $A^{(m)}$

where $A^{(m)}_{ij} = A_{ij} \mathbb{1}\{i \neq m, j \neq m\}$

ie. $A^{(m)}$ is obtained from A by deleting m-th row, column.

If A_m denotes the m-th row of A, then

$$\textcircled{\text{II}} = \frac{1}{\bar{\lambda}_1} \max_m \left| \langle A_m, \phi_1 - \bar{\phi}_1 \rangle \right|$$

Now $\left| \langle A_m, \phi_1 - \bar{\phi}_1 \rangle \right|$

$$\leq \left| \langle A_m, \phi_1 - \phi_1^{(m)} \rangle \right| + \left| \langle A_m, \phi_1^{(m)} - \bar{\phi}_1 \rangle \right|$$

$$\leq \underbrace{\|A\|_{2\to\infty}}_{\substack{\leq \|\bar{A}\|_{2\to\infty} \\ + \|A-\bar{A}\|_{op} \\ \leq 2\gamma \|u\|_\infty \\ \text{whp.}}} \underbrace{\|\phi_1 - \phi_1^{(m)}\|_2}_{\substack{\text{Can be bounded} \\ \text{by Davis-Kahan} \\ \text{theorem}}} + \underbrace{\left| \langle A_m, \phi_1^{(m)} - \bar{\phi}_1 \rangle \right|}_{\substack{\text{Can be bounded by} \\ \text{Assumption 3 since} \\ \phi_1^{(m)} \text{ and } A_m \text{ are independent}}}$$

We skip the rest of the proof.