Model: Hidden partition

\[ g = \begin{cases} \frac{n}{2} \text{ when } \sigma_i = \sigma_j + 1 \\ \sigma_i \neq \sigma_j + 1 \end{cases} \]

Key theme: Spectral algns might work
Threshold: If \((a-b)^2 < 2(a+b)\) into theoretically
"can't do better then random."

Else, think we can.

One idea: Use expected adj. matrix:

\[
\mathbb{E}[A_{uv}] = \frac{1}{n} \left[ \frac{a+b}{2} + \sigma_u \sigma_v \frac{a-b}{2} \right]
\]

\[
= \begin{cases} \frac{a}{n} & \text{if } \sigma_u = \sigma_v \\ \frac{a}{n} & \text{if } \sigma_u \neq \sigma_v \end{cases}
\]

\[ \Rightarrow n \mathbb{E}[A] = \frac{a+b}{2} \cdot \overrightarrow{1} \cdot \overrightarrow{1}^T + \frac{a-b}{2} \cdot \overrightarrow{\sigma} \cdot \overrightarrow{\sigma}^T \]

Rank 2, high eigenvalues are \(1 \& \overline{\sigma} \)

R 1. a have A not \(\mathbb{E}[A] \)
But we have $A$, not $\mathbb{E}[A]$

If $a, b = \Theta(\log n)$, then the graph looks locally tree-like (every small enough subset looks like a tree).

Note: Expected degree is $\frac{a+b}{2}$.

Expected "bias" (difference in neighbors vs. same vs. different spin) is $\frac{a-b}{2} \to$ call this $M$

Claim: At distance $r$, have eigenvector $f^{(r)}$, s.t.

$$f^{(r)}_v = \frac{1}{M^r} \sum_{u : d(u,v) = r} \sigma_u$$

i.e.

$$f^r_v = \frac{1}{M^r} \text{ bias of the } 3^r \text{ vertices.}$$

Why is $f^{(r)}$ an eigenvector?

$$(A f^{(r)})_v = M^{-r} \left( \sum_{u : d(u,v) = r+1} \sigma_u + (d_v - 1) \sum_{u : d(u,v) = r-1} \sigma_u \right)$$

Since $(A f^{(r)})_v$ sums over all neighbors $u$ of $v$, the quantity $f^r_v$.

However, some vertices that are dist $r$ from $v$'s neighbors are dist $r-1$ from $v$, giving 2nd term.
\[(Af^{(c)})_v = \sum_{u \in V} f_v^{(r+1)} \text{ for these } u\]

(Note, using locally tree-like, with this property have to consider distances besides \(r+1\) \& \(r-1\))

Then:

\[(Af^{(c)})_v = M f_v^{(r+1)} + (d_v - 1) \frac{1}{m} f_v^{(r-1)}\]

Now, \[\mathbb{E} \left[ \sum_{u \in V} \sigma_u \right] = m \sigma_v \text{ \quad \text{linearity of exp.}}\]

\[\mathbb{E} [f_v^{(r+2)} - f_v^{(r)}] = 0\]

(can also show \(|f_v^{(r+1)} - f_v^{(r)}| = O\left(\left(\frac{\alpha + b}{2}\right)^{r/2} \frac{1}{m}\right)\)

i.e., \(f_v^{(r+2)} \approx f_v^{(r)}\) are basically the same.

\[\Rightarrow (Af^{(c)})_v = M f_v^{(r)} + (d_v - 1) \frac{1}{m} f_v^{(r)}\]

\[= (M + (d_v - 1) \frac{1}{m}) f_v^{(r)}\]
Which if $d_v = 1$ for all $v$, would show $f_v$ is an eigenvector w/ eigenvalue $\lambda = m + \frac{d-1}{m}$

(Even if $d_v$ varies,

$Af = mIf + \frac{1}{m}(D-I)f$ for $D$ the diagonal degree vector.

$\Rightarrow f$ is an eigenvector of $(A - \frac{1}{m}D) \Rightarrow$ eigenvalue $m - \frac{1}{m}$)

If $a, b = O(\log n)$, all degrees concentrate, so $f$ actually an eigenvector approximately.

A random $d$-regular graph has top eigenvector $2\sqrt{d}$

So if $m + \frac{1}{m}(d-1) \geq 2\sqrt{d}$, can extract $f$ which correlates to the truth.

When $a, b$ constant, get some large degree vertex $v$

(degree $\Omega \left( \frac{\log n}{\log \log n} \right)$) w.h.p

Then, looking at adj matrix, for $x = 1$ at $v$, $0$ elsewhere

$y = \sqrt{\frac{\log n}{\log \log n}}$ at $v$'s neighbors, 0 else

$x^T Ay = \sqrt{\frac{\log n}{\log \log n}} = \omega(d)$, so the spectrum of $A$ gets dominated by eigenvectors with mass concentrated on $v$ & $v$'s neighbors; eigenvectors correlated to communities are not top eigenvectors.
correlated to communities are not top eigenvalues.

Instead, use the non-backtracking walk matrix:

Matrix indexed on directed edges:

\[ B_{(u \to v), (x \to y)} = \begin{cases} 
1 & \text{if } v = x, u \neq y \\
0 & \text{else}
\end{cases} \]

I.e.,

\[ \text{or} \]

Then, consider

\[ g^{(r)}_{(u \to v)} = \frac{1}{m^r} \sum_{\substack{w_{(u \to v), (v \to w)} = r}} g_w \]

means

\[ \text{sum over spins here} \]

Can show, again by "locally realizability":

\[ (B g^{(r)})_{(u \to v)} = \frac{1}{m^r} \sum_{\substack{v_{(u \to v), (v \to w)} = r+1}} g_w \]

(No backtracking term anymore, because no...
(No backtracking term anymore, because no backtracking walk!)

So implies:

$$ (B g^{(r)}) = M g^{(r+1)} $$

Now if we can show $M$ is a top eigenvector, then we can recover $g$ which is correlated w/ the true communities.

Note that $B^{(r)}$'s elements are # of non-backtracking paths of two directed edges of length $r$. But, since the graph is locally tree-like, this is either 0 or 1 for any pair of edges!

Now $\text{tr}(B^{r}(B^{r})^\top) \geq \sum_i |\lambda_i|^2 \geq \lambda^2$ for # eigenvalues.

$$ \implies \mathbb{E}[|\lambda|^2] \leq \mathbb{E}[\text{tr}(B^{r}(B^{r})^\top)] $$

$$ \leq \left( \frac{a+b}{2} \right)^r $$

$$ \implies \text{whp } |\lambda| \leq \sqrt{\frac{a+b}{2}} \text{ by Markov's} $$

So $g$ has eigenvalue $\frac{a+b}{2}$ but most eigenvectors have eigenvalue $\leq \sqrt{\frac{a+b}{2}}$. 