

APPM 5515: High Dimensional Probability– Fall 2020 —

Homework 7

Brandon Finley

1. Compute a bound for the second eigenvalue.

if $v = n(p - q) > \sqrt{2n(p + q)}$, then λ_2 is no longer in the “semi-circle.” And we know $\lambda_2 = \frac{n(p-q)}{2} + \frac{p+q}{p-q}$. So the question boils down to the following

$$\text{is } \lambda_2 = \frac{n(p-q)}{2} + \frac{p+q}{p-q} > \sqrt{2n(p+q)} ? \quad (1)$$

We can extend this to say

$$\begin{aligned} \frac{n(p-q)}{2} + \frac{p+q}{p-q} &> \frac{\sqrt{2n(p+q)}}{2} + \frac{p+q}{p-q} \\ \implies \frac{p+q}{p-q} &\geq \frac{\sqrt{2n(p+q)}}{2} \\ &= \sqrt{\frac{n(p+q)}{2}} \end{aligned}$$

which we can verify is true experimentally.

2. If $p = \alpha \frac{\log n}{n}$ and $q = \beta \frac{\log n}{n}$.

$$\begin{aligned} \implies n \left(\alpha \frac{\log n}{n} - \beta \frac{\log n}{n} \right) &> \sqrt{2n \left(\alpha \frac{\log n}{n} + \beta \frac{\log n}{n} \right)} \\ \implies \log(n)(\alpha - \beta) &> \sqrt{2 \log(n) (\alpha + \beta)} \\ \implies \alpha - \beta &> \frac{\sqrt{2(\alpha + \beta)}}{\sqrt{\log(n)}} = \frac{2}{\sqrt{\log(n)}} \sqrt{\frac{\alpha + \beta}{2}} \end{aligned}$$

3. In practice $p \ll \frac{\log(n)}{n}$ so it is better to write $p = a/n$ and $q = b/n$. Thus we can say the following

$$\begin{aligned} \implies n \left(\frac{a}{n} - \frac{b}{n} \right) &> \sqrt{2n \left(\frac{a}{n} + \frac{b}{n} \right)} \\ \implies a - b &> \sqrt{2(a + b)} \\ \implies \frac{a - b}{2} &> \frac{\sqrt{2(a + b)}}{2} = \sqrt{\frac{a + b}{2}} \end{aligned}$$

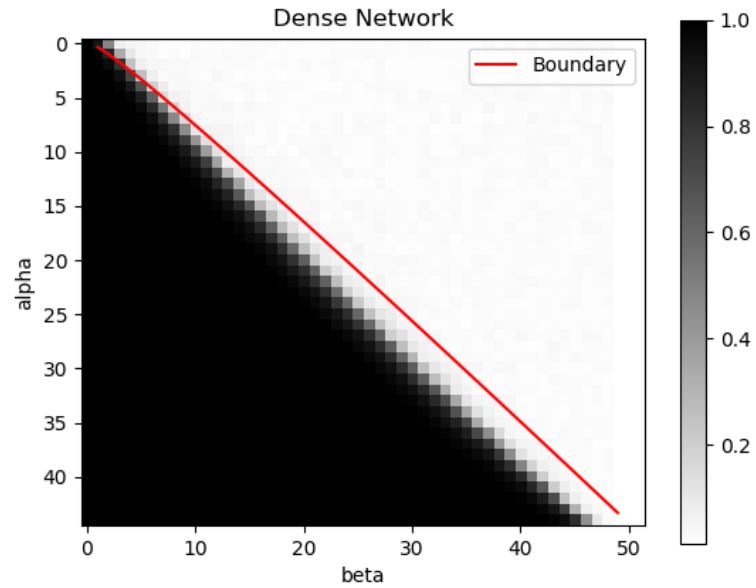
4. If $w = w$, then $\delta_{w_i, w_i} = 1$ and $\delta_{-w_i, w_i} = 0$. Thus,

$$\begin{aligned}
 \text{rawoverlap} &= \max \left(\sum_{i=1}^n \delta_{w_i, w_i}, \sum_{i=1}^n \delta_{-w_i, w_i} \right) \\
 &= \max \left(\sum_{i=1}^n 1, 0 \right) \\
 &= \max(n, 0) = n \\
 \implies \text{overlap} &= \frac{2}{n} \cdot \text{rawoverlap} - 1 = \frac{2n}{n} - 1 = 1
 \end{aligned}$$

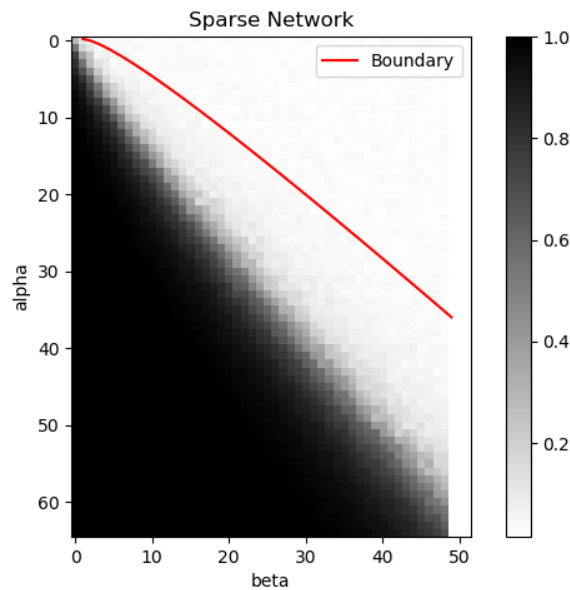
We also note that if we randomly partition our nodes, that is, we have that $w_i = -1, 1$ with $p_i = 0.5$, we can show that then the expected value of $\delta_{w_i, w_i} = \delta_{-w_i, w_i} = \frac{1}{2}$ and so we have that

$$\begin{aligned}
 \text{rawoverlap} &= \max \left(\sum_{i=1}^n \delta_{w_i, w_i}, \sum_{i=1}^n \delta_{-w_i, w_i} \right) \\
 &= \max \left(\sum_{i=1}^n \frac{1}{2}, \sum_{i=1}^n \frac{1}{2} \right) \\
 &= \max \left(\frac{n}{2}, \frac{n}{2} \right) \\
 &= \frac{n}{2} \\
 \implies \text{overlap} &= \frac{2}{n} \cdot \text{rawoverlap} - 1 = \frac{2}{n} \cdot \frac{n}{2} - 1 = 0
 \end{aligned}$$

5. Now, onto the code we have written to validate our findings in this and the last Homework. *Note: The axis correspond to the indices of the matrix, so when $\alpha = 0$ on the graph, $\alpha = 5$. Likewise, when $\beta = 0$ on the graph, $\beta = 1$.* Also note that the values I got for the network where it was near 0, was not completely zero. This was different from what was seen in the problem set. Thus, it caused the decision boundary curve to be slightly higher than what seems it should be. I wonder why the values were close to zero but not exactly zero, especially for the dense network. Anyways, we have the dense network:



Along with the sparse network:



And finally, using the same method on Zachary's karate club, we have that score = 1 with the following prediction, where a community value of 1 corresponds to community 1 and -1 corresponds to community 2. Although, since we just care about the partitions, they have arbitrary meanings.

| Node | Community |
|------|-----------|
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| 9 | -1 |
| 10 | -1 |
| 11 | 1 |
| 12 | 1 |
| 13 | 1 |
| 14 | 1 |
| 15 | -1 |
| 16 | -1 |
| 17 | 1 |
| 18 | 1 |
| 19 | -1 |
| 20 | 1 |
| 21 | -1 |
| 22 | 1 |
| 23 | -1 |
| 24 | -1 |
| 25 | -1 |
| 26 | -1 |
| 27 | -1 |
| 28 | -1 |
| 29 | -1 |
| 30 | -1 |
| 31 | -1 |
| 32 | -1 |
| 33 | -1 |
| 34 | -1 |

— END —