9.10 More communities

A natural question is to understand what is the exact recovery threshold for the Stochastic Block Model on $k \geq 2$ communities. Recall the definition: The stochastic block model can be similarly defined for any $k \geq 2$ communities: $G$ is a graph on $n = km$ nodes divided on $k$ groups of $m$ nodes each. Similarly to the $k = 2$ case, for each pair $(i, j)$ of nodes, $(i, j)$ is an edge of $G$ with probability $p$ if $i$ and $j$ are in the same set, and with probability $q$ if they are in different sets. Each edge is drawn independently and $p > q$. In the logarithmic degree regime, we’ll define the parameters in a slightly different way: $p = \frac{\alpha\log m}{m}$ and $q = \frac{\beta\log m}{m}$. Note that, for $k = 2$, we roughly have $\alpha = 2\alpha'$ and $\beta = 2\beta'$, which means that the exact recovery threshold, for $k = 2$, reads as: for

$$\sqrt{\alpha'} - \sqrt{\beta'} > 1$$

recovery is possible (and with the SDP algorithm), and for $\sqrt{\alpha'} - \sqrt{\beta'} < 1$ exact recovery is impossible.

Clearly, for any $k > 2$, if $\sqrt{\alpha'} - \sqrt{\beta'} < 1$ then exact recovery will also be impossible (simply imagine that $n$ oracle tells us all of the community memberships except for those of two of the clusters, then the problem reduces to the $k = 2$ case). The remarkable fact is that, for $k = o(\log m)$ this is enough, not only for exact recovery to be possible, but also for an SDP based algorithm (very similar to the one above) to achieve exact recovery (see [AS15, ABKK15, HWX15, PW15]). However, for $k \approx \log n$, the situation is not understood.

Open Problem 9.2 What is the threshold for exact recovery on the balanced symmetric Stochastic Block Model in $k \approx \log n$ communities and at what threshold does the SDP succeed at exactly determining the communities? (see [ABKK15]).

References


