### 9.3.1 Three of more communities

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 sparse regime, $p = \frac{a}{n}$ and $q = \frac{b}{n}$, the threshold at which it is possible to make an estimate that correlates with the original partition is open.

**Open Problem 9.1** *Consider the balanced Stochastic Block Model for $k > 3$ (constant) communities with inner probability $p = \frac{a}{n}$ and outer probability $q = \frac{b}{n}$, what is the threshold at which it becomes possible to make an estimate that correlates with the original partition is open (known as the partial recovery or detection threshold). We refer the reader to [DKMZ11, ZMZ14, GZC+15] for more information on this and many other interesting conjectures often motivated from statistical physics.*

## References


[DKMZ11] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84, December 2011.

[GZC+15] Amir Ghasemian, Pan Zhang, Aaron Clauset, Cristopher Moore, and Leto Peel. Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Available online at arXiv:1506.06179 [stat.ML]*, 2015.

[ZMZ14] Pan Zhang, Cristopher Moore, and Lenka Zdeborova. Phase transitions in semisupervised clustering of sparse networks. *Phys. Rev. E*, 90, 2014.

MIT OpenCourseWare
http://ocw.mit.edu

18.S096 Topics in Mathematics of Data Science
Fall 2015

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.