MAP554 – Networks: PC 5

28 September 2016

Emergence of small sub-graphs in Erdős-Rényi random graphs

Consider a $G(n,p)$ random graph. The aim of the exercise is to determine for what regime of p as a function of n specific sub-graphs appear in $G(n,p)$. We focus specifically on the presence of triangles, i.e. unordered sets (i,j,k) of vertices such that all three edges (i,j), (j,k), (k,i) occur in the graph.

1.1 Use the first moment method to show that with high probability, there are no triangles in $G(n,p)$ when $np = o(1)$.

1.2 Use the second moment method to show that with high probability, there is at least one triangle in $G(n,p)$ when $np = \omega(1)$.

1.3 Fix some constant $\lambda > 0$. Use Poisson approximation and the Stein-Chen method to determine as a function of λ the limiting probability as $n \to \infty$ that there is at least one triangle in $G(n, \lambda/n)$.

1.4 Can you extend this analysis to the emergence of K-cliques, i.e. complete sub-graphs on K vertices, for some fixed $K \geq 3$?

A spectral approach to jointly rate workers and jobs in crowd-sourcing platforms

A crowdsourcing platform has n workers, and $m = \alpha n$ jobs need to be performed by the platform ($\alpha \in]0,1[$ is fixed, and n large). Each job j consists of tagging a corresponding object by -1 or 1. Assume that item $j \in [m]$ has a "true tag" $y_j \in \{-1, 1\}$, while worker $i \in [n]$ has a "reliability factor" $p_i \in [0,1]$, so that when tagging an object it will give the correct tag with probability p_i and the wrong one with probability $1 - p_i$. The tags produced by workers are assumed independent across workers and jobs.

We consider the following assignment strategy: for each (worker,object) pair $(i,j) \in [n] \times [m]$, with probability d/n worker i is asked to treat job j.

2.1 Let $C_{ij} \in \{-1, 1\}$ be the tag provided by worker i on job j if it did treat job j, and $C_{ij} = 0$ otherwise. Express the expectation of matrix C as a function of the p_i and y_j.

2.2 Show how to transform the data matrix C into an $(n+m) \times (n+m)$ symmetric matrix A whose expectation has rank at most 2, and whose entries are independent up to symmetry.

2.3 Describe the eigenstructure of matrix $\bar{A} = \mathbb{E}(A)$ as a function of vectors $Y := \{y_j\}_{j \in [m]}$, $P := \{2p_i - 1\}_{i \in [n]}$.

2.4 We now assume that the vector P has an L^2 norm of $\beta\sqrt{n}$ for some fixed $\beta > 0$. Using the first bound on the spectral radius of a noise matrix given in the analysis of the Stochastic Block Model, relate the eigenstructures of A and $\bar{A}$ when $d = n^\delta$ for some fixed $\delta \in]0, 1[$. Argue using the sharper result of Feige and Ofek that the relation between the spectral structures still holds when $d >> \sqrt{\log(n)}$.

2.5 Let $\{(2r_i - 1)_{i \in [n]}, (z_j)_{j \in [m]}\}$ be the leading eigenvector of A associated with a positive eigenvalue, normalised so that the L^2 norm of $(z_j)_{j \in [m]}$ equals $\sqrt{m}$. Deduce that there exists a sign $\sigma \in \{-1, 1\}$ such that:

$$\sum_{i \in [n]} \left(\frac{||y||}{||P||} \sigma(2p_i - 1) - (2r_i - 1) \right)^2 = o(n), \quad \sum_{j \in [m]} (\sigma y_j - z_j)^2 = o(m).$$

2.6 Further assume that for some randomly chosen job j_0, we have obtained its true tag y_{j_0} by other means. Let

$$\hat{y}_j = y_{j_0} \text{sign}(z_{j_0}) \text{sign}(z_j), \quad j \in [m].$$

Show that with high probability we have

$$\sum_{j \in [m]} \mathbf{1}_{\hat{y}_j \neq y_j} = o(m).$$

2.7 Show that the estimates $\hat{p}_i$ defined as

$$(2\hat{p}_i - 1) = y_{j_0} \text{sign}(z_{j_0})(2r_i - 1)\frac{\lambda n}{md},$$

where λ is the leading eigenvalue of A, are such that $\sum_{i \in [n]} (\hat{p}_i - p_i)^2 = o(n)$.