Graph Theory Applications EPFL, Spring 2014

Solution to Problem Set 9: Project

Date: 01.05.2014 Due by 29.05.2014

1. Consider a graph $G(n,p)$ with n vertices s.t. for any pair of nodes the edge connecting them is included in the graph with probability p independently from every other edge. Recall that you saw a similar model in Problem Set 4. Associate to the i-th vertex a value $v_i \in \{0,1\}$, which represents the color of Rivella which is drunk by that specific person. In addition, associate to the edge connecting the i-th vertex with the j-th vertex the value $v_i \oplus v_j$, where $\oplus$ denotes the XOR operation. The aim is to find the vector $v = (v_1, \cdots, v_n)$, given the values of the edges of the graph.

   Let $p = 1$. Then, all the edges are known, i.e., all the pairwise XORs are known. Suppose that you find a specific solution $v^* = (v_1^*, \cdots, v_n^*)$, which fulfills all the conditions coming from the edges. Consider the vector $\bar{v}^* = (\bar{v}_1^*, \cdots, \bar{v}_n^*) = (1 \oplus v_1^*, \cdots, 1 \oplus v_n^*)$, which is obtained by negating each component of v^*. It is easy to check that also $\bar{v}^*$ fulfills all the conditions coming from the edges. Consequently, the solution to the problem is not unique and you are not able to discriminate who drinks red Rivella and who drinks green Rivella.

2. If you know the value of a specific vertex, then you can distinguish the two groups if and only if $G(n,p)$ is connected. The proof of this statement goes as follows.

   Denote by v_1 the vertex whose value is known (i.e., the vertex associated to your friend Jean). Suppose that $G(n,p)$ is connected. Then, for any $i \in \{2, \cdots, n\}$, there exists a walk from v_1 to v_i, say $(e_1^{v_1,v_i}, \cdots, e_m^{v_1,v_i})$. Consequently, $v_i = v_1 \oplus e_1^{v_1,v_i} \oplus \cdots \oplus e_m^{v_1,v_i}$. Therefore, you can distinguish the two groups.

   Suppose that $G(n,p)$ is not connected. Then, there are at least two connected component inside the graph and one of those connected components does not contain v_1. For the same reason seen at point 1, the values of the vertices of this connected component cannot be found unequivocally. Hence, the problem does not have a unique solution and the two groups cannot be distinguished.

   When $p = 1$, the $G(n,p)$ is the complete graph with n vertices, which is clearly connected. Hence, the problem can be solved.

3. Given the previous discussion, the aim is to find the range of values of p s.t. $G(n,p)$ is connected with high probability when n is large.

   As can be seen in Figure 1, if p scales as $\frac{1}{n}$, then with high probability $G(n,p)$ is not connected. If p scales as $\frac{1}{\sqrt{n}}$, then with high probability $G(n,p)$ is connected. No threshold behavior can be observed in these two cases. On the other hand, if p scales as $\frac{\ln n}{n}$, then there is a sharp threshold for $c = 1$, i.e. for $p < \frac{\ln n}{n}$ with high probability $G(n,p)$ is not connected and for $p > \frac{\ln n}{n}$ with high probability $G(n,p)$ is connected.

4. The problem consists in recovering a hidden clique (complete graph) of k vertices inside a bigger graph of n vertices. More specifically, consider the random graph $G(n, 1/2)$ and pick at random a subset of k vertices. Pick the subgraph associated to this subset of vertices and make it the complete graph, i.e., add all the missing edges. Let us denote this new graph as $G(n, 1/2, k)$.

Take a graph $G(n, 1/2, k)$. Now, the task consists of finding the hidden complete subgraph of size k in $G(n, 1/2, k)$. If k is large enough, the first idea is to check the degree of the vertices. Indeed, intuitively the vertices in the hidden clique are more connected than the other vertices.

More specifically, let us denote by S the set of vertices in the clique and by d_i the degree of vertex i. Suppose that $i \notin S$. Then, d_i is the sum of $n-1$ i.i.d. $\mathrm{Bernoulli}(1/2)$ random variables. Therefore, by Chernoff bound,

$$\mathbb{P}\left(d_i \geq \frac{n-1}{2} + t\right) \leq \exp\left(-\frac{2t^2}{n}\right),$$

which, by taking $t = \sqrt{n \log n} + \frac{1}{2}$, yields

$$\mathbb{P}\left(d_i \geq \frac{n}{2} + \sqrt{n \log n}\right) \leq \exp\left(-2\frac{(\sqrt{n \log n} + 1/2)^2}{n}\right) \leq \frac{1}{n^2}.$$

Consequently, as the random variables d_i for $i \notin S$ are i.i.d.,

$$\begin{aligned} \mathbb{P}\left(\max_{i \notin S} d_i \geq \frac{n}{2} + \sqrt{n \log n}\right) &= 1 - \mathbb{P}\left(\max_{i \notin S} d_i < \frac{n}{2} + \sqrt{n \log n}\right) \\ &= 1 - \prod_{i \notin S} \mathbb{P}\left(d_i < \frac{n}{2} + \sqrt{n \log n}\right) \leq 1 - \left(1 - \frac{1}{n^2}\right)^n \leq \frac{1}{n}. \end{aligned} \tag{1}$$

Since the last term goes to 0 as n grows large, we can conclude that with high probability $\max_{i \notin S} d_i \leq \frac{n}{2} + \sqrt{n \log n}$.

Consider now a vertex which is in the clique. Then, d_i is equal to $k-1$ plus the sum of $n-k$ i.i.d. $\mathrm{Bernoulli}(1/2)$ random variables. Therefore, by Chernoff bound,

$$\mathbb{P}\left(d_i \leq k - 1 + \frac{n-k}{2} - t\right) \leq \exp\left(-\frac{2t^2}{n-k}\right),$$

which, by taking $t = \sqrt{n \log n} - 1$ and $k = 4\sqrt{n \log n}$ yields

$$\mathbb{P}\left(d_i \leq \frac{n}{2} + \sqrt{n \log n}\right) \leq \exp\left(-2\frac{(\sqrt{n \log n} - 1)^2}{n - 4\sqrt{n \log n}}\right).$$

The RHS is clearly $O\left(\frac{1}{n^2}\right)$. Hence, since the random variables d_i for $i \in S$ are also i.i.d., by using similar passages to (1), we have that

$$\mathbb{P}\left(\min_{i \in S} d_i \geq \frac{n}{2} + \sqrt{n \log n}\right) = O\left(\frac{1}{n}\right),$$

which implies that with high probability $\min_{i \in S} d_i \geq \frac{n}{2} + \sqrt{n \log n}$.

This suffices to conclude that with high probability the k vertices in the hidden clique have the highest degrees.

Suppose now that $k = 10\sqrt{n}$. Then this method does not work simply because the size of the clique is comparable to the variance of d_i for $i \notin S$. In other words, the (deterministic) increase in the degree due to the fact that a vertex is in the clique is comparable to the random variations of the vertex degree itself. Therefore, we need to resort to more sophisticated techniques, such as, for instance, the algorithm which is suggested in the text.

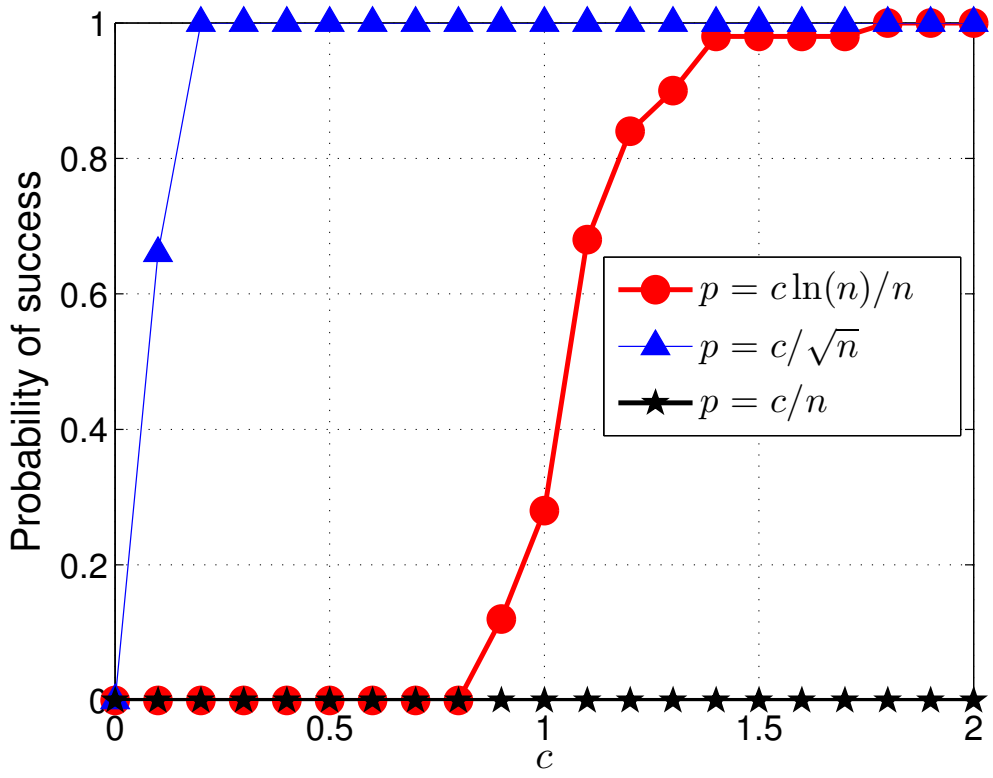

Figure 1: Probability of success as a function of $c \in \{0, 0.1, \cdots, 1.9, 2\}$ when p scales as $\frac{\ln n}{n}$, as $\frac{1}{\sqrt{n}}$, and as $\frac{1}{n}$.