Problem Set 9: Project

Date: 01.05.2014 Due by 29.05.2014

Instructions : You are allowed to work in groups of up to 2 people. In the first page, write your name, Sciper number, and the name of the other (eventual) person which is in your group. Hand in a written report, one per group suffices. We will not make a distinction between the members of your group. Every group member gets the same grade. The report does not need to be long, but it should explain what it is that you did and give answers to the posed questions. Send everything by Thursday May 29th to our email addresses `{marco.mondelli, ruediger.urbanke}@epfl.ch`

Rules : As for your graded homework, if you used any resources (books, Wikipedia, etc.), list these sources at the beginning of your report. This does not influence your grade. Nevertheless, if we discover outside influences which are not listed, we will consider it as cheating. Please note that EPFL has a VERY strict policy on cheating and you might be in BIG trouble. It is simply not worth it.

You are at a cocktail party in which there are n people. You know that each of these people drinks either *red Rivella* or *green Rivella*, and none of them drinks both. You would like to discover the identities of the two groups, i.e. who drinks the red Rivella and who drinks the green one. The people at the party will not tell you directly what type of Rivella they drink. However, by by observing their pairwise interactions, you will be able to deduce if a couple drinks the same type of Rivella or not. Suppose that you observe each pairwise interaction independently with probability p.

1. *[5points]* Phrase your task as a graph problem. Let $p = 1$. Are you able to tell who drinks the red Rivella and who drinks the green Rivella?

2. *[5points]* Suppose that among the n people there is your friend Jean. You know for sure that he is a heavy drinker of red Rivella! Can you solve the problem when $p = 1$? Take a specific set of pairwise interaction and give a necessary and sufficient condition on them s.t. you are able to solve the problem of distinguishing the two groups.

3. *[10points]* Suppose that n is large. Your aim is to find the range of values of p for which the problem can be solved.

   To do so, you need to perform some simulations. Pick $n = 10000$ and let $M = 50$ be the number of trials that you are going to perform. Let $c \in \{0, 0.1, 0.2, \cdots, 1.9, 2\}$.

   a) Take $p = \dfrac{c}{n}$ and compute the probability that you can solve the problem as a function of c.

   b) Take $p = \dfrac{c \ln n}{n}$ and compute the probability that you can solve the problem as a function of c.

   c) Take $p = \dfrac{c}{\sqrt{n}}$ and compute the probability that you can solve the problem as a function of c.

   Plot your results. What can you conclude?

4. Now, suppose that there is a group of k people that took the course *Graph Theory Applications* last year. They will talk all the time about graphs and other weird stuff, so you will be able to detect each of the possible $\binom{k}{2}$ pairwise interactions between people of this group. However, the members of this group can still interact with the rest of the party (!) and you will see such pairwise interactions independently with probability p as usual. Take $p = \frac{1}{2}$. Your aim is to find the people of this group.

   a) *[10points]* First of all, phrase this task as a graph problem. Take $k = 4\sqrt{n \log_2 n}$. Describe a polynomial (in n) algorithm which solves this problem. Implement the solution with your favorite programming language as follows.

   The input to the program is the value of n. Select at random the group of k people which are in the group and write in the file `weird_people.txt` their names. Generate the remaining interactions independently with probability $p = \frac{1}{2}$. Then, run your algorithm which looks for this group of k people and write their names in the file `weird_people_found.txt`. Compare the two files and return success if they are equal and error otherwise.

   Explicitly check that for large values of n the probability of success is close to 1.

   Hint 1. Consider the random variable which counts the number of pairwise interactions of a person. Pick a person that did not take the course *Graph Theory Applications* last year. What is the mean and the variance of this random variable? How does the answer change for the people that took *Graph Theory Applications*?

   Hint 2. In case you are ambitious and you want to write down a formal argument why your algorithm works for large values of n, the following classical result can also be helpful. Let X be the sum of n i.i.d. Bernoulli(1/2) random variables, namely, $X = \sum_{i=1}^{n} X_i$ where the X_i are i.i.d. and s.t. $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = 1/2$. Then, for any real number $t \in (0, n/2)$,

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

$$\mathbb{P}\left(X \leq \frac{n}{2} - t\right) \leq \exp\left(-\frac{2t^2}{n}\right).$$

   This result is known under the name of *Chernoff bound*.

   b) *[10points]* Take $k = 10\sqrt{n}$. Run some experiments to determine if the algorithm of point a) yields a probability of success close to 1 for large values of n. If it does not, consider the following algorithm based on spectral techniques.

   Find the second eigenvector v_2 of the adjacency matrix associated to the graph. Sort the vertices by decreasing order of the absolute values of their coordinates in v_2 (you can break the ties arbitrarily). Let W be the first k vertices in this order. Let Q_0 be the set of all vertices with at least $\frac{3k}{4}$ neighbors in W. Output Q_0, which is the desired set of vertices with high probability.

   Implement the solution with your favorite programming language as in point a) and explicitly check that for large values of n the probability of success is close to 1.