Engineering Probability and Statistics Course Project
Flow of information in social networks
Part I:Percolation

Seyed Iman Hosseini Zavaraki

March 15, 2018

1 Terms and conditions

By submitting a report on this course project you accept the following terms:
1) You waive any right to object to the grading procedure and final grade.
2) The grading system is not final and can be subject to change without notice.
3) Your submission may be forfeited in case of objection to the grading system.
4) Any submission after the deadline, WILL BE DISCARDED.
5) It is the duty of the student to submit a detailed and readable report including the supplementary materials (e.g. source code of programs he has made) to illustrate his/her work.
6) What you hand in should be your own work. The consequences of cheating are severe and you are strongly advised that you do not cheat.

2 Introduction

Since the late 1890s, both mile Durkheim and Ferdinand Tnnies heralded the concept of social networks in their theories and research of social groups. And since then mathematical treatment of social actors, interacting in a medium gained widespread traction.

Mathematical models, both analytical and numerical have been used in a wide range of problems such as voter models (to forecast results of elections), evaluating financial markets and influence maximization in Instagram and other social networks to name a few. The tools used in these studies, can also be used in problems like congestion control in designing computer networks, or reliability problems for robust network coverage.

In this project we are going to consider the problem of how news flow throw a network. We will be using different models for our network and our actors(the users in that network) and derive results on thresholds on the initial state of the system, such that a certain news traverses the whole network.

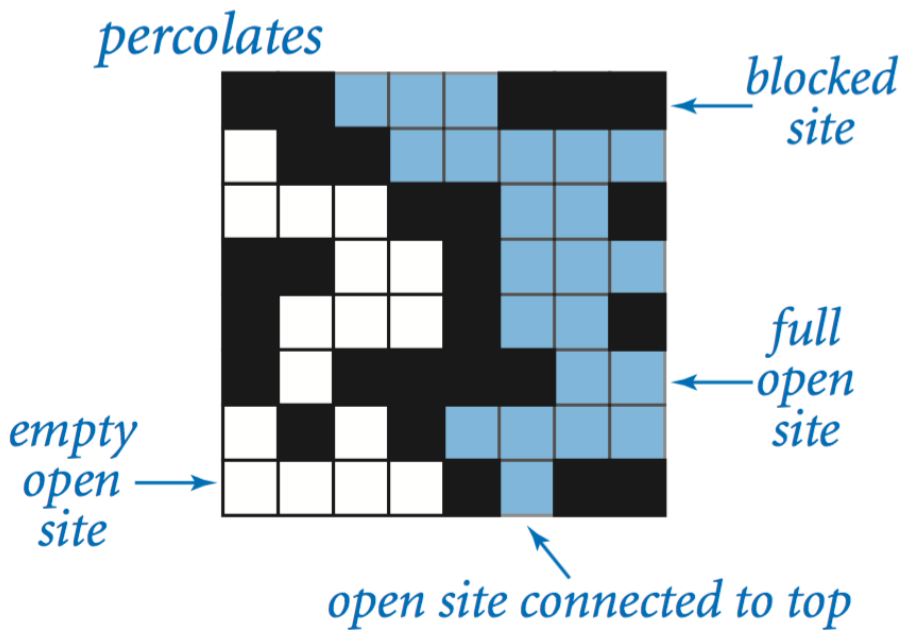

Figure 1: Percolation in 2D

The goal of this project is to illustrate how probability and statistics, are the language of formulating and approaching diverse problems and how they occur both in theoretical, and practical contexts.

3 Coffee, Kolmogorov and Whitney's theorem

Coffee is being made by hot water flowing throw ground coffee-beans, and due to chemical extraction from the coffee, the water comes away the other side as "Coffee". The medium in which hot water flows through, for example in a moka pot is filled with coffee, but between the packed grains of coffee, there is empty space for the water to stream and drizzle downwards.

Think of it as a 3D grid with integer cells, in which some cells are occupied with grains of coffee and some are free cells, through which water can move.

Let's consider a simpler 2D configuration, our medium is now a 2D grid with dimensions NxN and each cell is either free or occupied. To decide the state of cells(empty/occupied) we fill each cell with probability p. So our model now has two parameters, p and N, where N illustrates the size of our medium (the top container in the moka pot) and p is the probability that a cell is open, so it shows how "packed" the coffee is. For example in the extreme case of applying high pressure and packing the coffee, the density of the grains would be high and all cells would be occupied by grains, and this identifies with $p = 0$ in

Figure 2: Making coffee via moka pot

our model.

Evidently filling our 2D grid, is a random process so each time we populate our grid we get a different configuration. Imagine with some specific value for parameters p and N,we make a grid 1000 times, and each time we check whether there exist a free (unblocked) path from the upper row, to the lower row. Now what happens if we do it for 100000 times? Intuitively we would expect the ratio of number of times there is a path, to the total number of times, to converge to a certain value depending only on p and N.

Task 1.1 (5 points)

Define $\theta(p)$ as the probability that a configuration with p,(the probability of a cell being open) percolates, i.e. there is an open path from the upper edge to the lower edge.

For different values of N plot $\theta(p)$. ($N \in 20, 50, 100, 400, 1000$)

What do you observe? What happens at $N \to \infty$?

As seen in Task 1.1, it seems that $\theta(p)$ is monotonic as it is intuitively expected. Also it seems that for large N there is a threshold p (let's call this p_c) that for $p > p_c$ $\theta = 1$ and for any p less, $\theta = 0$. So let us follow through with this intuition and attempt a rigorous proof via the formalism we have learned.

Task 1.2 (15 points)

Lemma 3.1 *Prove the following lemma:*

The Borel-Cantelli Lemma

For a probability triple of Ω*, the sample space,* F *a* σ*-algebra of* Ω *and a probability measure* P*,*

let $A_1, A_2, .. \in F$ *then*

I) If $\sum_n P(A_n) < \infty$ *then* $P(A_n \ infinitely \ often) = 0$

II) If $\sum_n P(A_n) = \infty$, $\{A_n\}_{n=1}^{\infty}$ *independent, then* $P(A_n \ infinitely \ often) = 0.$

What is the gist of this lemma? Lets see an application of this lemma:

Consider an infinite heavily weighted coin tossing. Let our independent events be $H_1, H_2, H_3, ...$, where H_i is the event that the i thcoin is heads. Suppose also that our coins are heavily weighted against flipping heads, with $P(H_n) = 1/n$ (e.g. the millionth coin has only a one in a million chance at being heads). Using the above lemma prove that we will still flip infinitely many heads. (hint:use the divergence of the harmonic series)

For an even less obvious, and more counter-intuitive result, consider the above problem but this time with $P(H_n) = (\frac{99}{100})^n$ In other words, there is a 99% chance the first coin is heads, a 98.01 % chance that the second one is heads, etc Now prove that in this scenario, we cannot have infinitely many heads.

To expand on the above lemma, we need one more definition

Given a sequence of events,

$$A_1, A_2, ... \in F$$

we define their *tail field* as

$$\tau = \cap_{n=1}^{\infty}(A_n, A_{n+1}, ...)$$

The tail field thus, is a -algebra whose members we call *tail events.*

Now we can state the Kolmogorov's Zero-One Law:

Theorem 3.2 *Kolmogorov's Zero-One Law*

Given a probability triple (Ω*,*F*,*P*) and a sequence of independent events* $A_1, A_2, .. \in F$ *with tail field* τ *, if* $T \in \tau$ *then* $P(T) \in 0, 1.$

This theorem can be used to prove that in fact, for the percolation problem stated above, and in fact in a more general setting (not restricted to the grid graph described above) in the limit that the graph grows large, there will be a threshold p_c which above that threshold percolation will happen. Now from experiment (Task 1.1) you probably found out the value of p_c for a 2-D grid, but analytical proof that p_c is actually (spoiler alert!) 1/2 is far from trivial and took more than 20 years to surface. Despite this, we are going to prove bounds on the value of p_c.

> Task 1.3 (10 points)
>
> Prove that $p_c > 1/3$. Or in other words we're proving that if $p < 1/3$ then $\theta(p) = 0$.
>
> Hint: you can use (without proof) that this is equivalent to:
>
> Fix a certain point as centre, and imagine you're playing Snake, beginning from that point move through the sites such that you don't cross your path [we call this a self-avoiding path]. Now what we want to investigate is having a self-avoiding path which only involves open sites, with infinite size. (if it happens, then percolation has occured) Now Imagine we're given a certain path of length N, what is the probability that it is a valid path (all it's sites are open), now starting from our centre point, how many different, self-avoiding path of length N do we have?
>
> Now this second question is hard to answer, but we can come up with an upper-bound on the number of these paths. Afterwards, call this approximate probability (which is an upper-bound on the exact probabilty) $P(n)$ and investigate what will happen as $n -> \infty$ and what requirement on p we should have, such that $P_\infty = 0$.

Now to prove our upper-bound on p_c we need a theorem by Whitney, this theorem is pure graph theory and we do not state the proof here, but you can view the following figure and see why it works.
Note: Each connected component made out of open sites, is called a 'cluster'.

Theorem 3.3 *There isn't any infinite cluster involving the centre, if there exists a cycle made out of closed sites surrounding the centre.*

The intuition is clear: if you have an infinite cluster, you cannot enclose it in a fence of closed sites.

> Task 1.4 (25 points)
>
> Prove that $p_c < 2/3$. This is very much similar to the former result. Use the Whitney's theorem and investigate the probability of having such 'enclosing' closed sites of size N, and take the limit of large N.

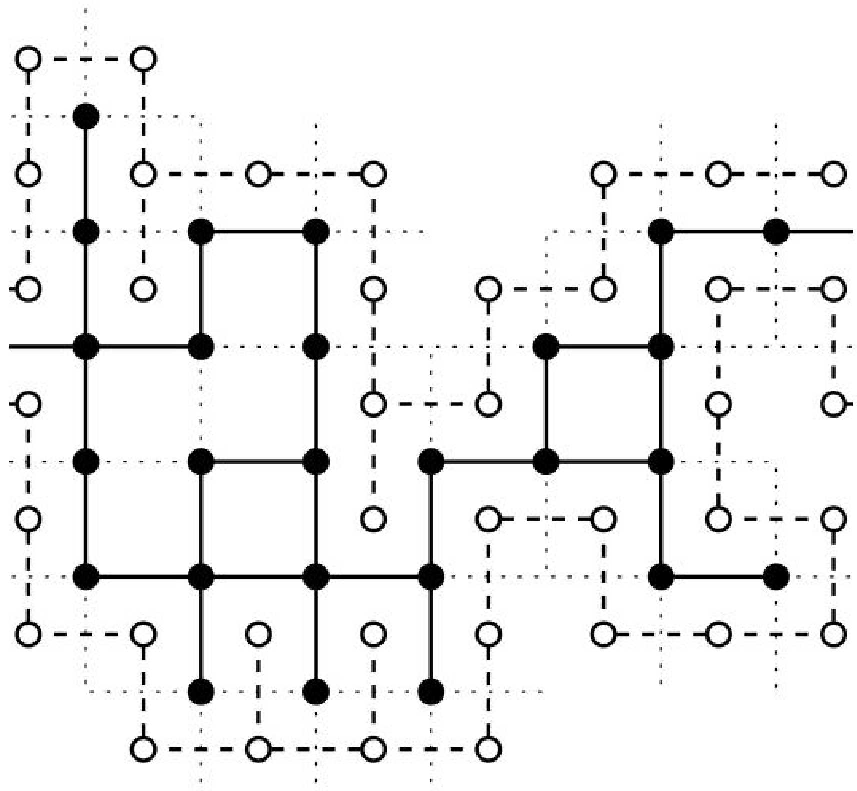

Figure 3: Whitney's theorem