Final Project

Carlos Pinzón, Camilo Rocha

Árboles y Grafos

2019-1

The project consists of three parts, each involving the computation and analysis of some properties of randomly generated graphs.

1. Random directed graphs based on the Erdős-Renyi model.
2. Random undirected graphs based on the Erdős-Renyi model.
3. Random block-chain trees based on the model explained in class.

Part 1 - Preamble

Consider the Erdős-Renyi model for generating random directed graphs, and consider the following metrics of interest.

1. Number of strongly connected components.
2. Average size of the strongly connected components.
3. Number of edges that connect two distinct SCCs.

Notice that each metric can be seen as a function $f : \mathbb{D} \to \mathbb{R}^{\geq 0}$, where $\mathbb{D}$ represents the set of all finite directed graphs.

Exercises

Implement a random graph generator based on the Erdős-Renyi model, and use it to solve the exercises.

Assume the following lists:

- `p_small = [0, 0.005, 0.010,..., 0.095]`, and
- `p_large = [0.1, 0.15, 0.2,..., 0.95, 1]`.

Notice that $p_{\text{small}} \cup p_{\text{large}}$ is simply a set of samples from the interval $[0, 1]$ with a higher granularity in the range $[0, 0.1)$ than in $[0.1, 1]$.

Exercise 1.1

Generate one directed random graph G with parameters $n = 15$, $p = 0.2$. Use a seed so that your results are repeatable.

1. Draw it or print it using any software you wish, e.g. graphviz + dot, networkx + iGraph, or, csv + spreadsheet software.
2. Compute $f(G)$ for each directed graph metric f mentioned in the preamble, and corroborate your results with the picture.

Exercise 1.2

Consider for each metric f, the function $F(n, p)$ that computes the **expected value** of $f(G)$ where G is generated randomly using your algorithm with parameters n, p.

We would like to estimate $F(n, p)$ for $n = 100$, for each value of $p \in P_{\text{small}}$, and for each undirected graph metric f in the preamble.

To do that, you must generate for each p, at least 200 random graphs, and compute the mean $\hat{F}(n, p)$ of the outputs of $f(G)$ for each f.

Save your results into a text file (csv or npy recommended), and plot the curves of each metric using the following format:

- x-axis: p.
- y-axis: $\hat{F}(100, p)$.

Exercise 1.3

Using at least 1000 runs and a divide and conquer strategy, estimate

- the value p_1 that maximizes $\hat{F}(100, p)$, where f is the metric that computes the number of edges connecting two SCCs.
- the inflection point p_2 of the curve $\hat{F}(100, p)$, where f is the metric that computes the average size of SCCs.

Part 2 - Preamble

Consider the Erdős-Renyi model for generating random undirected graphs, and consider the following metrics of interest.

1. Number of connected components.
2. Average size of the connected components.
3. Number of articulation points.
4. Number of bridges.
5. Average node degree.
6. Number of triplets ($u \leftrightarrow v \leftrightarrow w$).
7. Number of triangles ($u \leftrightarrow v \leftrightarrow w$ with $u \leftrightarrow w$).

Notice that each metric can be seen as a function $g : \mathbb{G} \to \mathbb{R}^{\geq 0}$, where $\mathbb{G} \subseteq \mathbb{D}$ represents the set of all finite undirected graphs.

Exercise 2.1

Repeat the exercise 1.1 on an undirected graph using the same parameters and computing the respective metrics.

Exercise 2.2

Repeat the previous exercise using undirected graphs and their respective metrics.

For the number of CCs, triplets and triangles, use the full range $p_{\text{small}} \cup p_{\text{large}}$.

Plot also the quotient $\frac{\text{\# triangles}}{\text{\# triplets}}$. Is there any relation between p and this quotient?

Exercise 2.3

Using at least 1000 runs and a divide and conquer strategy, estimate

- the value p_1 that maximizes $\hat{F}(100, p)$, where f is the metric that computes the number of bridges.
- the value p_2 that maximizes $\hat{F}(100, p)$, where f is the metric that computes the number of articulation points.
- the inflection point p_3 of the curve $\hat{F}(100, p)$, where f is the metric that computes the average size of CCs.

Part 3 - Preamble

Recall the blockchain model explained in class in which network delay is taken into account.

Let p be the network lag relative to the average mining time, i.e. $p = \frac{d}{\tau}$ where d is the average time (in seconds) that it takes for a message to be sent between a pair of randomly chosen nodes in the network, and τ is the average time between subsequent block discoveries in the network.

Suppose a simulation until n blocks are found, and consider the following metrics once the simulation stops:

1. Number of lost blocks, in other words, blocks that .
2. Maximal length of published blocks.

Exercise 3.1

Estimate the expected value of the two metrics with at least 200 simulations.

Fix $n = 1000$ and vary p in the range

- `[0, 0.05, 0.1, ..., 1.95, 2]`

Plot your results as done in the exercises of the previous parts.

Exercise 3.2

1. What happens with the two curves as $p \to \infty$?
2. Can you support the hypothesis that as n becomes larger, the expected error of the two metric estimates with respect to their true values decreases?