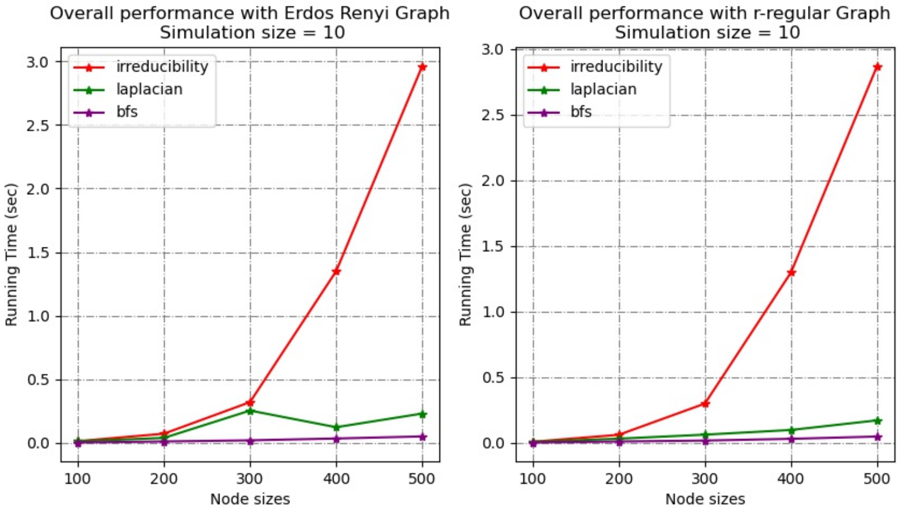
Networking for Big Data - Challenge #1

Group 18 - Rivest

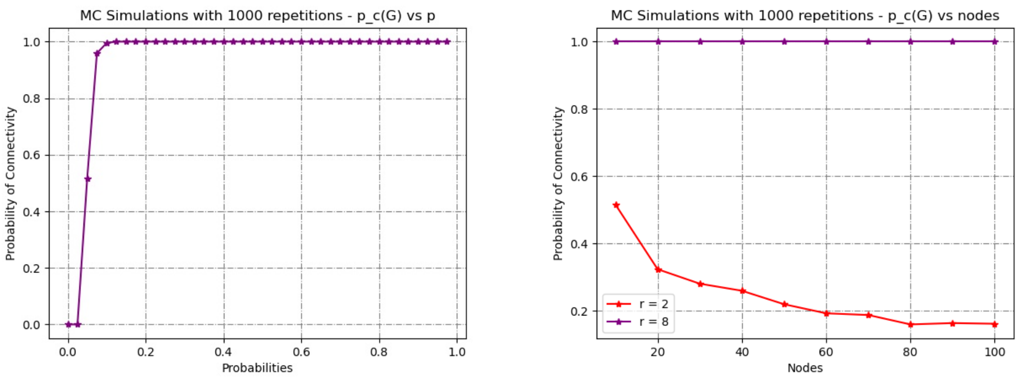
Alessandra della Fazia 1970722
Jeremy Sapienza 1960498
Simone Fiorellino 1960415
Stefano D'Arrigo 1960500

April 20 2021

1 Part 1



Question 1 As we can observe in the plot above, we consider as a measure of performance the running time (expressed in seconds). We perform 10 simulations, in which, for each size of the graph (100, 200, 300, 400, 500 nodes), we compute the average running time for each of the three methods. The `irreducibility` method tends to have a running time higher than the other methods. Instead, The `BFS` method has a better performance. The same behaviours appeared between the Erdos-Renyi and r-regular random graphs.



Question 2 As we can observe in the left plot, with the probability $p \geq 0.1$, the Erdos-Renyi graph returns (with 100 nodes) a probability of connectivity equal to 1.

Question 3 Consider the r-regular random graph with degrees equal to 2 and 8. As we can observe in the right plot, with a degree equal to 8, the probability of connectivity is equal to 1. Instead, with degree equal to 2, increasing the number of nodes the probability of connectivity decreases to 0.

2 Part 2

Question 1 We aim to find an expression for r as a function of n such that N, S, L are the same in both the Fat-Tree and the Jellyfish topologies. Hence, we solve the following system:

$$\begin{cases} N_{fat} = N_{jelly} \\ S_{fat} = S_{jelly} \\ L_{fat} = L_{jelly} \end{cases} \begin{cases} \frac{n^3}{4} = S_{jelly} \cdot (n-r) \\ S_{jelly} = \frac{5}{4}n^2 \\ L_{fat} = L_{jelly} \end{cases} \implies \begin{array}{l} \dfrac{n^3}{4} = \dfrac{5}{4}n^2(n-r) \\ \mathbf{r} = \dfrac{\mathbf{4}}{\mathbf{5}}\mathbf{n} \end{array} \tag{1}$$

The equality $L_{fat} = L_{jelly}$ can be proven.

Question 2 We know that the application-oblivious throughput bound is $TH \le \frac{l}{\bar{h} \cdot \nu_f}$. In a r-regular random graph l is equal to $l = S_{jelly} \cdot r$. We also know that each server is connected to one switch and each switch is connected to other r switches. Thus, the number of flows among the servers in the topology can be derived as $\nu_f = N_{jelly} \cdot (N_{jelly} - 1)$. Ultimately, solving the algebraic steps, the throughput bound as function of $\bar{h}$ and n is:

$$TH \le \frac{l}{\bar{h} \cdot \nu_f} = \frac{S_{jelly} \cdot r}{\bar{h} \cdot N_{jelly} \cdot (N_{jelly} - 1)} = \cdots = \frac{4}{\bar{h} \cdot [\frac{1}{4}n^3 - 1]} \tag{2}$$

Question 3 Finally, we derive the expression for the average shortest path in a Fat-Tree topology. We know that the length of the shortest paths between servers takes values in the set $\{2, 4, 6\}$; thus, for each length, we count how many shortest paths are present:

$$\begin{aligned} \bar{h}_{fat} &= \frac{n\left(\frac{n^3}{4} - \frac{n^2}{4}\right)\left(\left(\frac{n}{2}\right)^2 \cdot 6\right) + \frac{n^2}{2}\left(\left(\frac{n}{2}\right)^2 - \frac{n}{2}\right)\left(\frac{n}{2} \cdot 4\right) + n\left(\frac{n}{2}\right)^2\left(\frac{n}{2} - 1\right)(1 \cdot 2)}{n\left(\frac{n^3}{4} - \frac{n^2}{4}\right)\left(\frac{n}{2}\right)^2 + n\frac{n}{2}\left(\left(\frac{n}{2}\right)^2 - \frac{n}{2}\right)\left(\frac{n}{2}\right) + n\left(\frac{n}{2}\right)^2\left(\frac{n}{2} - 1\right)} \\ &= \frac{\left(\frac{n}{2}\right)^4(n-1) \cdot 6 + \left(\frac{n}{2}\right)^3\left(\frac{n}{2} - 1\right) \cdot 4 + \left(\frac{n}{2}\right)^2\left(\frac{n}{2} - 1\right) \cdot 2}{\left(\frac{n}{2}\right)^4(n-1) + \left(\frac{n}{2}\right)^3\left(\frac{n}{2} - 1\right) + \left(\frac{n}{2}\right)^2\left(\frac{n}{2} - 1\right)} \end{aligned} \tag{3}$$

The value of $\bar{h}$ in a Jellyfish topology has been proven to be lower-bounded by:

$$\bar{h}_{jelly} \ge \frac{\sum_{j=1}^{k-1} j \cdot r \cdot (r-1)^{j-1} + k \cdot R}{S_{jelly} - 1} = \frac{\sum_{j=1}^{k-1} j \cdot \frac{4}{5}n \cdot (\frac{4}{5}n - 1)^{j-1} + k \cdot R}{\frac{5}{4}n^2 - 1} \tag{4}$$

having substituted to the variables R and k the analytic expressions dependent on the number of ports n.

In the end, the numerical values of n, N, S, L and of the throughputs are shown in the Table.

l	n	N	S	L	$TH_{Fat-Tree}$	$TH_{Jellyfish}$
1	10	250	125	1000	2.7×10^{-3}	6.6×10^{-3}
2	20	2000	500	8000	3.3×10^{-4}	8.1×10^{-4}
3	30	6750	1125	27000	9.9×10^{-5}	2.4×10^{-4}
4	40	16000	2000	64000	4.2×10^{-5}	1.0×10^{-4}
5	50	31250	3125	125000	2.1×10^{-5}	5.1×10^{-5}