CS-E5740 Complex Networks, Answers to exercise set 8

Alex Herrero Pons, Student number: 918697

November 16, 2020

Problem 1

a) In the Table 1 we can observe the obtained results for this section.

samp.	triangles	two-stars	transit.	triang.frac.	two-st.frac.
node	3192	4503	0.7089	0.0608	0.0617
edge	3216	11488	0.2800	0.0612	0.1574
star	18300	28958	0.6319	0.3483	0.3968
orig.	52536	72976	0.7199	1.0000	1.0000

Table 1: Empirical estimates of the number of triangles, the number of two-stars, and transitivity of the different sampling schemes.

If we compare the fraction of preserved triangles we can see that with the node and edge samplings we obtain similar values while with the star sampling is way higher. In the other hand the fraction of preserved two-stars we can see that the value with the star sampling is twice as big as the one with edge sampling, and this last one is twice as big as the value with the node sampling.

Observing the results we can conclude that the node and edge sampling schemes affect in a similar way the number of triangles, but not the number of two-stars. And in both cases the effect is different with the star sampling scheme.

Comparing to the original network we can observe that the transitivity via node sampling is quite similar to the real value, this happens because sampling all nodes with the same probability will tend to preserve the same fraction of triangles and two-stars. We can check in next section how the probability of preserving a triangle $p^n_{\triangle}$ is the same as preserving a two-star $p^n_{\angle}$.

b) We know that for the Bernoulli sampling of edges $p^e_{\angle} = p^2$ and $p^e_{\triangle} = p^3$.

For the other samplings we have that:

i) Bernoulli sampling of nodes: $p^n_\angle = p^3$ and $p^n_\triangle = p^3$ since we need to observe the three nodes in both cases.

ii) Star sampling: $p^s_\angle = p$ since for every directly observed node we will have all its two-stars and $p^s_\triangle = 3 * p^2 * (1-p) + p^3$ since we need to observe at least 2 directed nodes, so we have 3 ways of observing 2 directed nodes plus the probability of observing 3 directed nodes.

c) Knowing that,

$$\hat{\tau}^{HT} = \frac{1}{p_\tau}\hat{\tau}$$

We can obtain:

$$\hat{\tau}_\triangle^{HT,n} = \frac{1}{p_{\tau_\triangle}}\hat{\tau}^n_\triangle = \frac{1}{p_{\tau_\triangle}}\hat{\tau}^n_\triangle = \frac{1}{p^3}\hat{\tau}^n_\triangle$$

$$\hat{\tau}_\angle^{HT,n} = \frac{1}{p_{\tau_\angle}}\hat{\tau}^n_\angle = \frac{1}{p_{\tau_\angle}}\hat{\tau}^n_\angle = \frac{1}{p^2}\hat{\tau}^n_\angle$$

$$\hat{\tau}_C^{HT,n} = \frac{\hat{\tau}_\triangle^{HT,n}}{\hat{\tau}_\angle^{HT,n}}$$

$$\hat{\tau}_\triangle^{HT,e} = \frac{1}{p_{\tau_\triangle}}\hat{\tau}^e_\triangle = \frac{1}{p_{\tau_\triangle}}\hat{\tau}^e_\triangle = \frac{1}{p^3}\hat{\tau}^e_\triangle$$

$$\hat{\tau}_\angle^{HT,e} = \frac{1}{p_{\tau_\angle}}\hat{\tau}^e_\angle = \frac{1}{p_{\tau_\angle}}\hat{\tau}^e_\angle = \frac{1}{p^3}\hat{\tau}^e_\angle$$

$$\hat{\tau}_C^{HT,e} = \frac{\hat{\tau}_\triangle^{HT,e}}{\hat{\tau}_\angle^{HT,e}}$$

$$\hat{\tau}_\triangle^{HT,s} = \frac{1}{p_{\tau_\triangle}}\hat{\tau}^s_\triangle = \frac{1}{p_{\tau_\triangle}}\hat{\tau}^s_\triangle = \frac{1}{p^3}\hat{\tau}^s_\triangle$$

$$\hat{\tau}_\angle^{HT,s} = \frac{1}{p_{\tau_\angle}}\hat{\tau}^s_\angle = \frac{1}{p_{\tau_\angle}}\hat{\tau}^s_\angle = \frac{1}{3 * p^2 * (1-p) + p^3}\hat{\tau}^s_\angle$$

$$\hat{\tau}_C^{HT,s} = \frac{\hat{\tau}_\triangle^{HT,s}}{\hat{\tau}_\angle^{HT,s}}$$

d) We can observe the obtained plots in Figures 1-3.

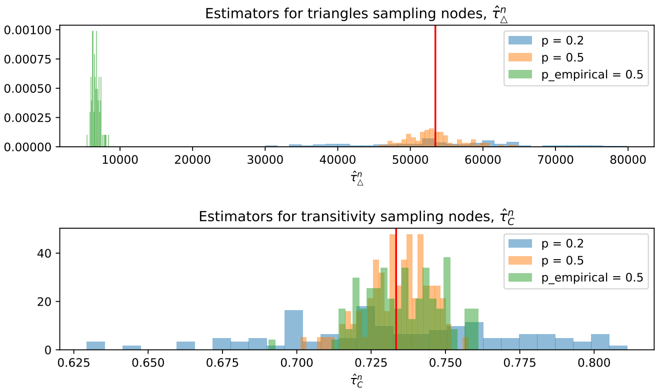

Figure 1: HT estimator of the node sampling.

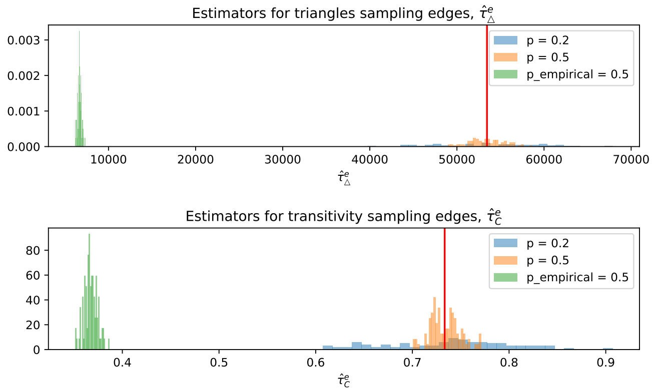

Figure 2: HT estimator of the edge sampling.

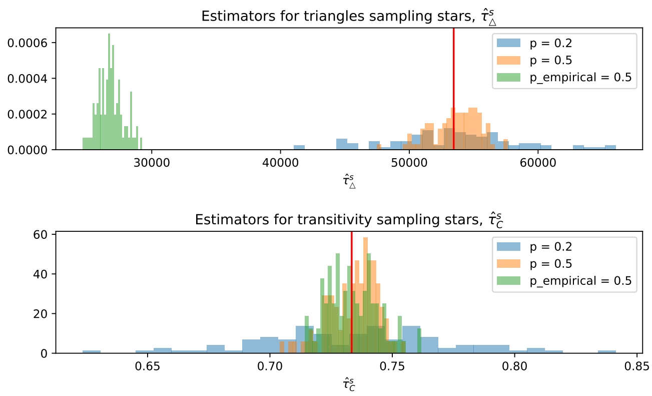

Figure 3: HT estimator of the star sampling.

Problem 2

a)

$$Q_i = \frac{6}{13} - \left(\frac{13}{13 \times 2}\right)^2 = 0.423$$

$$Q_{ii} = 1 - \left(\frac{26}{13 \times 2}\right)^2 = 0$$

b)

$$\Delta Q = \frac{l_{ab}}{L} - \left(\frac{d_{ab}}{2L}\right)^2 - \frac{l_a}{L} + \left(\frac{d_b}{2L}\right)^2 - \frac{l_b}{L} + \left(\frac{d_b}{2L}\right)^2$$

c)

$$d_{ab} =$$

$$l_{ab} =$$

$$\Delta Q = \frac{l_{ab}}{L} - \left(\frac{d_{ab}}{2L}\right)^2 - \frac{l_a}{L} + \left(\frac{d_b}{2L}\right)^2 - \frac{l_b}{L} + \left(\frac{d_b}{2L}\right)^2$$

Seeing this, we can conclude that we will always merge two clusters (a and b) if $L > \frac{d_a d_b}{2}$ without minding the shape of a and b.

d) I don't know.