Networks and their Structure Assignment

Network Science Topic 1

Note that the networks in this exercise are all directed.

1. [15 marks] Obtain the code and dataset (under Topic 1 on Learn Ultra) and load the `citation_graph`. Two vertices u and v are *connected* in this graph if there is a path from u to v or from v to u (or both). A *connected component* of the graph is a maximal set of vertices such that each pair of vertices is connected. How many vertices are there in the largest connected component of the `citation_graph`? Let G be the graph formed by the largest connected component of the `citation_graph` (that is, obtain G by removing all vertices not in the largest connected component). Create two plots showing the normalized distributions of the in-degree and out-degree of G.
2. [15 marks] Recall the PA graph model that constructed graphs one vertex at a time. In this model the out-degrees were all (almost) the same. Define a version of the model where the out-degree varies in a way that is similar to the distribution found for G in Question 1. Construct instances of the model and plot the normalized distributions of the in-degree and out-degree and compare them to those of G. (Your model might turn out to be a poor model for G. This does not matter as long as you can motivate your definition and implement it correctly.)