CSE 494: Artificial Intelligence for Cyber Security: Spring 2019

Lab 8: Social Network Analysis - an introduction

Objectives of the lab:

- Scrape and analyze the dataset from https://arxiv.org – this will be the network data for the lab
- Plot the Degree distribution of the authorship network in last 5 years and in last 15 years
- Visualize the authorship network

Task 1: Scraping the Arxiv data and creating an adjacency graph.

Install the python arxiv scraping module from https://github.com/Mahdisadjadi/arxivscraper. You can choose any category of research from arxiv. The website also provides a clear documentation on how to scrape and store the data using the popular pandas dataframe object. Pandas is a powerful tool to store and visualize data, but you are free to choose your database of choice.

Task2:

After fetching the data, create an authorship graph. There are several ways to create a graph. The general techniques include adjacency list and adjacency graph. We will use networkx for this task.

Task3:

Obtain the degree of each author in the network and plot the degree distribution