AMRITA VISHWA VIDYAPEETHAM

DEEMED TO BE UNIVERSITY

19CSE337 Social Networking Security

Lecture 13

Topics to Discuss

- Some network analysis terminologies
- Modularity
- Computing modularity of a network

Network Analysis Terminologies

Size

- The size of a network is characterized by the numbers of nodes and edges in it.

Density

- The density of a network is the fraction between 0 and 1 that tells what portion of all possible edges are actually realized in the network. (i.e; ratio between observed edges and maximum possible edges).
- For a network G made of n nodes and m edges, the density $\rho(G)$ is given by

  $\rho(G)=m/(n(n-1)/2)=2m/n(n-1)$ for an undirected network.

  $\rho(G)=m/(n-1)$, for a directed network

Network Analysis Terminologies

Characteristic path length

- It is the average length of shortest paths for all possible node pairs in the network, giving an expected distance between two randomly chosen nodes.
- This is an intuitive characterization of how big (or small) the world represented by the network is.
- It is calculated as

$$L=\sum_{i,j} d(i \rightarrow j)/n(n-1)$$ where n is the number of nodes.

- This formula works for both undirected and directed networks.

Network Analysis Terminologies

Eccentricity

- This metric is defined for each node and gives the maximal shortest path length a node can have with any other node in the network.
- This tells how far the node is to the farthest point in the network.
- It is calculated as

$$\epsilon(i)=\max_j d(i \rightarrow j)$$

Network Analysis Terminologies

Diameter

- This metric gives the maximal eccentricity in the network.
- Intuitively, it tells us how far any two nodes can get from one another within the network.
- Nodes whose eccentricity is D are called peripheries.
- D is calculated as $D=\max_i \epsilon(i)$

Network Analysis Terminologies

Radius

- This metric gives the minimal eccentricity in the network.
- Intuitively, it gives the smallest number of steps required to reach every node from an optimal node as a starting point.
- Nodes whose eccentricity is R are called centers.
- R is calculated as $R=\min_i \epsilon(i)$

Network Analysis Terminologies

Degree Distribution

- Degree of each nodes in a network can be represented as a probability distribution.
- A degree distribution of a network is a probability distribution given by

$$P(k)=\{i|deg(i)=k\}/n$$

i.e., the probability for a node to have degree k.

Network Analysis Terminologies

Assortativity

- Degrees are a metric measured on individual nodes.
- But when we focus on the edges, there are always two degrees associated with each edge, one for the node where the edge originates and the other for the node to where the edge points.
- So, if we take the former for x and the latter for y from all the edges in the network, we can produce a scatter plot that visualizes a possible degree correlation between the nodes across the edges.
- Such correlations of node properties across edges can be generally described with the concept of assortativity.
  - Assortativity (positive assortativity): The tendency for nodes to connect to other nodes with similar properties within a network.
  - Disassortativity (negative assortativity): The tendency for nodes to connect to other nodes with dissimilar properties within a network.

Network Analysis Terminologies

Assortativity coefficient

- The assortativity coefficient is a Pearson correlation coefficient of some node property f between pairs of connected nodes.
- Positive coefficients imply assortativity, while negative ones imply disassortativity.
- If the measured property is a node degree (i.e; f=degree), this is called the degree assortativity coefficient.

Network Analysis Terminologies

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r	=	correlation coefficient
x_i	=	values of the x-variable in a sample
$\bar{x}$	=	mean of the values of the x-variable
y_i	=	values of the y-variable in a sample
$\bar{y}$	=	mean of the values of the y-variable

Network Analysis Terminologies

- It can be measured as

$$r=\sum_{(i,j)\epsilon E}(fi-f1)(fj-f2)/((\sum_{(i,j)\epsilon E}(fj-f2))^{1/2})^{*}\ ((\sum_{(i,j)\epsilon E}(fj-f2)^{2})^{1/2}$$

- Where E is the set of directed edges (undirected edges should appear twice in E in two directions), and

$$f1=\sum_{i,j\epsilon E}f(i)/|E|,\quad f2=\sum_{i,j\epsilon E}f(j)/|E|$$

- For directed networks, each of f1 and f2 can be either in-degree or out-degree, so there are four different degree assortativities can measure: in-in, in-out, out-in, and out-out.

Network Analysis Terminologies

- The properties of networks for the analysis is classified into three as
  - mesoscopic properties (eg; modularity)
  - microscopic (e.g; degrees or clustering coefficients)
  - macroscopic (e.g; density, characteristic path length)
- Among modularity is very important as it can tell us how a network is organized at spatial scales intermediate between those two extremes, and therefore, these concepts are highly relevant to the modeling and understanding of complex systems too.

Modularity

- **Modularity**: The extent to which a network is organized into multiple communities.
- It is a measure of the structure of networks or graphs which measures the strength of division of a network into modules (also called groups, clusters or communities).
- Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.
- Modularity is often used in optimization methods for detecting community structure in networks.

Modularity

- The modularity of a given set of communities in a network is defined as follows

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

where

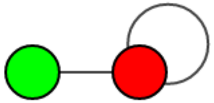
Aij is the weight of the edge between i and j.
ki is the sum of weights of the vertex attached to the vertex I, also called as degree of the node
ci is the community to which vertex i is assigned
δ(x,y) is 1 if x = y and 0 otherwise
m = (1/2)Σ_{ij} Aij   i.e number of links

Modularity

- Modularity is a number for a graph and a clustering of its nodes, it says how "good" is your clustering.
- It takes values from -1 to 1 , where 1 means super-good clusters found, 0 - clustering is not better than just random, high negative number would correspond to something like anti-clusters.
- Modularity in the range 0.3-0.7 means good clusters found: that numbers people want to see in practice.

Modularity Computation

- Let us consider the simplest undirected graph as shown in figure.

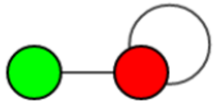


- It has two nodes: red and green connected with an edge and one additional self-loop edge from the red node to itself.
- Calculate modularity for it manually and compare it with the value obtained by using built-in functions available in NetworkX or in similar packages.

Modularity Computation

- Since it is undirected graph edge weights are considered as 1 and 2 for self edges.
- The adjacency matrix for the graph is

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}$$



- The weighted degrees $K_i = \sum A_{i,j}$

  k_{green}=A00+A10=0+1=1

  k_{red}=A01+A11=1+2=3

Modularity Computation

Calculating modularity

- The modularity is the sum of 4 summands (all pairs of i,j)
  - Case 1: i=j=green
  - Case 2: i=green, j=red
  - Case 3: i=red, j=green
  - Case 4: i=j=red

Modularity Computation

$Q=1/2m\sum_{ij}[Aij-(kikj/2m)]\delta(Ci,Cj)$

- Case 1: i=j=Green

  $=1/2*2[A_{green,green}-(k_{green}k_{green}/2*2)]\delta(C_{green},C_{green})$

  $=¼[A_{00}-k_{green}k_{green}/4]\ \delta(C_{green},C_{green})$

  $=¼[0-1/4]*1= -1/16$

Modularity Computation

- Case 2: i=green, j=red

  $=¼[A_{green,red}-k_{green}k_{red}/4]*\delta(C_{green},C_{red})$

  $=¼[1-(1*3/4)]*0=0$

- Case 3: i=red, j=green - does not contribute, since i,j in different groups. Value=0

Modularity Computation

- Case 4: i=j=red

  $=\frac{1}{4}[A_{red,red} - k_{red}\, k_{red}/4]\ \delta(C_{red}, C_{red})$

  $=\frac{1}{4}[2-(3*3/4)]*1$

  $=\frac{1}{4}[2-(9/4)]$

  $= -1/16$

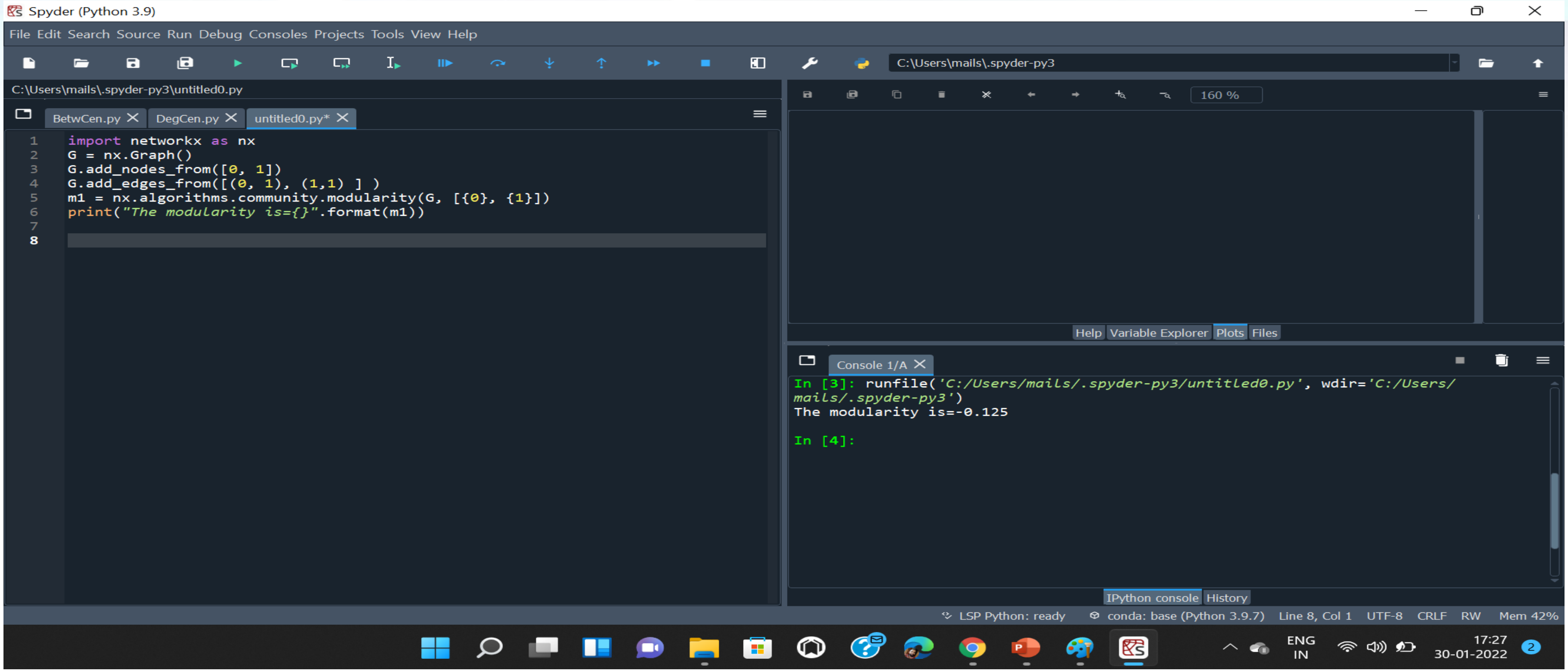
- Since modularity is the sum of all the four terms.

  $Q= -1/16+0+0+(-1/16)$

  $=-2/16=-1/8$

  $= -0.125$

Modularity using NetworkX



Thanks…………