Network Dynamics and Simulation Sciences Laboratory
Virginia Bioinformatics Institute
Virginia Tech, Blacksburg, VA

Computer Science Department
College of Engineering & Applied Sciences
University at Albany – SUNY
Albany, NY

Workshop on CINET – August 11, 2015

CINET/GRANITE Exercise – I(b)

Notes: Please take a look at Exercise I(a) if you need information about using CINET to compute measures for a network. Some knowledge of programming will be useful in doing this exercise.

Overview of the Exercise: The goal of this exercise is to check whether certain pairs of graph measures are correlated. We will use **Pearson Correlation Coefficient** (PCC) as the measure of correlation. The definition of PCC and its significance is briefly discussed in the boxed text below.

Definition: Suppose we are given a data sample consisting of $n \geq 1$ pairs of numbers $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. Let $\overline{x}$ and $\overline{y}$ denote respectively the mean values of the sets $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$; that is, $\overline{x} = (\sum_{i=1}^{n} x_i)/n$ and $\overline{y} = (\sum_{i=1}^{n} y_i)/n$. The **Pearson Correlation Coefficient** (PCC) r for the sample is given by

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{[\sum_{i=1}^{n}(x_i - \overline{x})^2]}\ \sqrt{[\sum_{i=1}^{n}(y_i - \overline{y})^2]}}$$

where positive square roots are used for both the terms in the denominator. The PCC value r defined above satisfies the condition $-1 \leq r \leq 1$. The value $r = 1$ indicates that a linear equation describes the relationship between the two sets X and Y. Similarly, $r = -1$ indicates a linear relationship between the two sets, with Y values decreasing as the X values increase. The value $r = 0$ indicates that X and Y are not correlated.

Procedure: This exercise has six steps which are shown below. (A table to record your results appears on the next page.)

1. For this exercise, please use the following five networks available in CINET:

   G_1 : Autonomous systems - Oregon-1-010407
   G_2 : Erdos Collaboration Network
   G_3 : Autonomous systems - Oregon-1-010331
   G_4 : Autonomous systems - Oregon-2-010331
   G_5 : Enron Giant Component

2. For each of the above networks, the **average node degree** can be computed as follows: If the network has n nodes and e edges, the average node degree is given by the formula $2e/n$. Let the average node degree values of G_1, G_2, G_3, G_4 and G_5 be Δ_1, Δ_2 , Δ_3, Δ_4 and Δ_5 respectively. (The table on the next page shows the average node degree values.)

3. For each of the five networks, find the **number of triangles**. (CINET provides a measure called "Compute the Number of Triangles".) Let the number of triangles in G_1, G_2, G_3, G_4 and G_5 be T_1, T_2, T_3, T_4 and T_5 respectively.

4. For each of the five networks, find the **diameter**. (CINET provides a measure called "Find Diameter of a Graph".) Let the diameters of G_1, G_2, G_3, G_4 and G_5 be D_1, D_2 , D_3, D_4 and D_5 respectively. (Since each of the five networks is connected, all the five diameter values will be finite.)

5. Compute the PCC value r_1 for the sample $\{(\Delta_1, T_1), (\Delta_2, T_2), \ldots, (\Delta_5, T_5)\}$.

6. Compute the PCC value r_2 for the sample $\{(\Delta_1, D_1), (\Delta_2, D_2), \ldots, (\Delta_5, D_5)\}$.

Note: You can compute the PCC values r_1 and r_2 using a calculator or a simple program.

	#Nodes	**#Edges**	**Avg. degree** (Δ)	**#Triangles** (T)	**Diameter** (D)
G_1	10,729	11,000	2.051		
G_2	6927	11,850	3.421		
G_3	10,670	22,002	4.124		
G_4	10,900	31,180	5.721		
G_5	33,696	180,811	10.732		

(a) PCC value r_1 (from Step 5) = ____________________

(b) PCC value r_2 (from Step 6) = ____________________

Note: Time permitting, you can also try to prepare two **scatter plots**, one showing the pairs $\{(\Delta_1, T_1)$, (Δ_2, T_2), …, $(\Delta_5, T_5)\}$ and the other showing the pairs $\{(\Delta_1, D_1)$, (Δ_2, D_2), …, $(\Delta_5, D_5)\}$. In each case, please show the Δ values along the X axis and the other value along the Y axis.