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 \( \bar{x} \) and \( \bar{y} \) 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, \( \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \) and \( \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \). The Pearson Correlation Coefficient (PCC) \( r \) for the sample is given by

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

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 \( \Delta_1, \Delta_2, \Delta_3, \Delta_4 \) and \( \Delta_5 \) 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.

<table>
<thead>
<tr>
<th></th>
<th>#Nodes</th>
<th>#Edges</th>
<th>Avg. degree $\Delta$</th>
<th>#Triangles $T$</th>
<th>Diameter $D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_1$</td>
<td>10,729</td>
<td>11,000</td>
<td>2.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_2$</td>
<td>6927</td>
<td>11,850</td>
<td>3.421</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_3$</td>
<td>10,670</td>
<td>22,002</td>
<td>4.124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_4$</td>
<td>10,900</td>
<td>31,180</td>
<td>5.721</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_5$</td>
<td>33,696</td>
<td>180,811</td>
<td>10.732</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(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 $\{(\Delta_1, T_1), (\Delta_2, T_2), \ldots, (\Delta_5, T_5)\}$ and the other showing the pairs $\{(\Delta_1, D_1), (\Delta_2, D_2), \ldots, (\Delta_5, D_5)\}$. In each case, please show the $\Delta$ values along the $X$ axis and the other value along the $Y$ axis.