Analyzing Multivariate Time Series Data Using a Network-Based Approach

Nasir Ahmad
Problem Statement

• Analyzing complex systems.

• Myriads of interacting components.

• Evolving over time.
Rationale

• Transform a dataset to a network.

• Use the structural properties of a network.
Data Source

• Data from the World Bank Data Catalogue is used for this analysis.

• Four different indicators are used, which are 1) Energy Use 2) CO₂ Emission 3) GDP and 4) Unemployment.
Methodology

• The main concept of the methodology is to compare values of a distribution with one another by measuring whether they are similar or not (i.e., whether they lie within a certain range of one another). In other words, while traditional histograms have fixed ranges, this methodology adapts the ranges around individual values.

• This process is close to the formation of a network, where two values become “connected” if they are within a certain range of one another.

• A network is analytically represented by its adjacency matrix $A_{ij}$ in which each cell takes the value of 1 if nodes $i$ and $j$ are connected and 0 otherwise; more formally in this case:

$$A_{ij} = \begin{cases} 
1 & \text{if } (x_i - \zeta) \leq x_j \leq (x_i + \zeta) \\
0 & \text{otherwise}
\end{cases}$$

where, $x_i$ represents the value of node $i$, $x_j$ represents the value of node $j$, and $\zeta$ is called the cutoff percentage.

$x_i = 100$

$\zeta = 5$
Selection of $\zeta$

• The main objective is to select a $\zeta$ for which the network properties have become stable.

• As we slowly increase the value of $\zeta$, clusters (akin to communities) form and a larger / giant cluster rapidly emerges containing $V_g$ nodes.

• Proportional number of vertices in the giant cluster defined as $p_g = \frac{V_g}{V}$.

• When $p_g$ becomes stable and a certain proportion of nodes (e.g. $p_g \geq 0.68$) are present in the giant cluster, that $\zeta$ is selected as the optimum cutoff.

Orbital Position

• Then, using the optimal network, a technique is developed to track the discrete relative position of any node compared to the mode of the distribution (e.g. how far or close that node is from the mode).

• This discrete relative position of a node $i$ from the mode is defined as orbital position $O_i$ and expressed mathematically as:

$$O_i = \frac{x_i - M}{\zeta} + C$$
Orbital Diagram

• Orbital positions of every node are plotted over time in polar format to generate the orbital diagram for each country, where the timeline proceeds in the counter clockwise direction.
Travelled distance and Orbital Speed

• The change in orbital positions from the previous year is also computed by taking the difference between orbital positions in two consecutive years. We can then define a total ‘Travelled Distance’, $D_i$, for each country by taking the summation of those differences in orbital positions for all the timelines. For a timeline from 1 to $t$, $D_i$ is expressed as:

$$D_i = \sum_{2}^{t} O_i - O_{i-1}$$

• The speed at which the countries orbital positions are changing is calculated by dividing the $D_i$ of any country by its number of available data points.
Thank You!

Nasir Ahmad
sahmad38@uic.edu

Sybil Derrible
derrible@uic.edu
Questions?