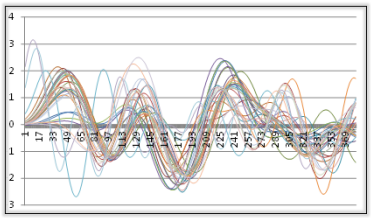
Investigating graph clustering methods for the power grid

Gabriela Radu, Emilie Hogan

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Overview:

Given time series data from multiple power generators containing measurements of phase angle taken at many points in time. The goal is to cluster together generators which have similar phase angle behavior.



Graph of phase angle behavior of the multiple generators over a give time series

In order to achieve this goal, Fast Fourier Transforms and Euclidean distances were used to quantify similarities between generators.
A graph was created in which two generators were connected if they were sufficiently similar (known as a nearest neighbor graph). Using different nearest neighbor values, matrices were generated based on similarities between generators. Specific time intervals were taken from the large data set to assess the time dependency of the methods.

Main Question:

Are there different procedures that can be applied to the original generated matrices that can produce even better clustering of the given objects?

Markov Clustering Algorithm:

The first alternative procedure was applying the Markov Clustering Algorithm (MCL). Which consisted of column normalizing the original matrix to generate the probability matrix. Then the matrix was expanded four times to simulate the random walks of the probability matrix (Markov Chains) .This was followed with an inflation of the matrix which consisted of raising each column to a non-negative power and then re-normalizing.

Normalizing/Squaring:

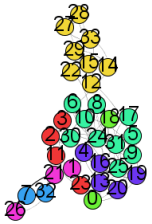
The second alternative procedure was to use the original matrix and produce the probability matrix followed by squaring the entries to strengthen and weaken the edge weights.

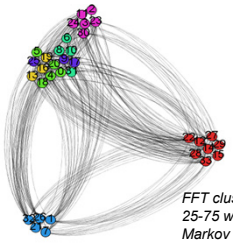
Markov Chains with Pruning:

The third alternative procedure was to obtain the probability matrix followed by expansion of the matrix (four times). The matrix was then pruned (any entry with edge weight <0.0001 was replaced by 0).

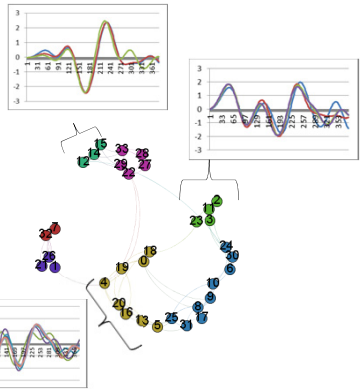
Fast Fourier Transforms

FFT cluster based on time series 25-75 with 3 nearest neighbor. Original generated matrix. Modularity: 0.162





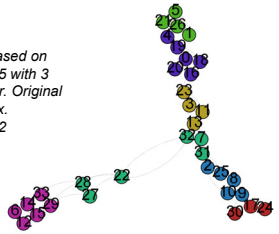
FFT cluster based on time series 25-75 with 3 nearest neighbor. Markov Cluster Algorithm. Modularity: 0.498

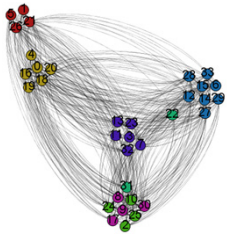


FFT cluster based on time series 25-75 with 3 nearest neighbor. Normalizing and Squaring. Modularity: 0.755

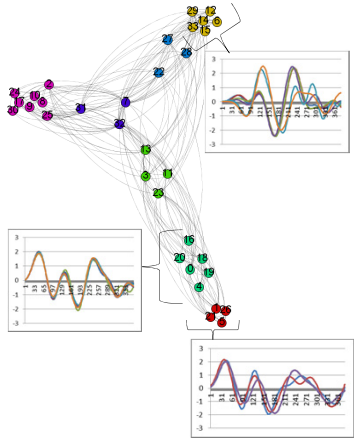
Euclidean Distances

EUCD cluster based on time series 25-75 with 3 nearest neighbor. Original generated matrix. Modularity: 0.472





EUCD cluster based on time series 25-75 with 3 nearest neighbor. Markov Cluster Algorithm. Modularity: 0.575



EUCD cluster based on time series 25-75 with 3 nearest neighbor. Markov Chains with Pruning. Modularity: 0.539

Results:

Applying the MCL to both FFT and EUCD produces somewhat viable clusters but majority are not consistent with the visual representations based on the times series of the generators.

Both Normalizing/Squaring and Markov chains with pruning were applied to FFT and EUCD. While both produced better clusters then the MCL, Normalizing/Squaring saw better performance in FFT and Markov Chains with pruning performed better under EUCD. Visual representations were consistent with the clusters that were generated.

Applications:

The main scope of this project was for the direct application to Power Grid Model Reduction. The procedures can also be applied towards:

- Network Maps
- Bioinformatics

References:

Bavaud, F.: Euclidean Distances, soft and spectral Clustering on Weighted Graphs 1-4 (2012)
Houry, J.K. Application to Markov Chains. Retrieved 07/08/2013 from http://aix1.uottawa.ca/~jkhoury/markov.htm
Macropol,K. Clustering on Graphs: The Markov Cluster Algorithm (MCL). Retrieved 06/19/2013 from http://www.cs.ucsb.edu/~xyan/classes/CS595D-2009winter/MCL_Presentation2.pdf