Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

Clustering on the Power Grid

EDUARDO COTILLA-SANCHEZ†, MAHANTESH HALAPPANAVAR*, EMILIE HOGAN* (presenter)

*Pacific Northwest National Laboratory, †Oregon State University

Joint Mathematics Meetings, January 17, 2014
MAA Session on Mathematics Experiences in Business, Industry, and Government, I

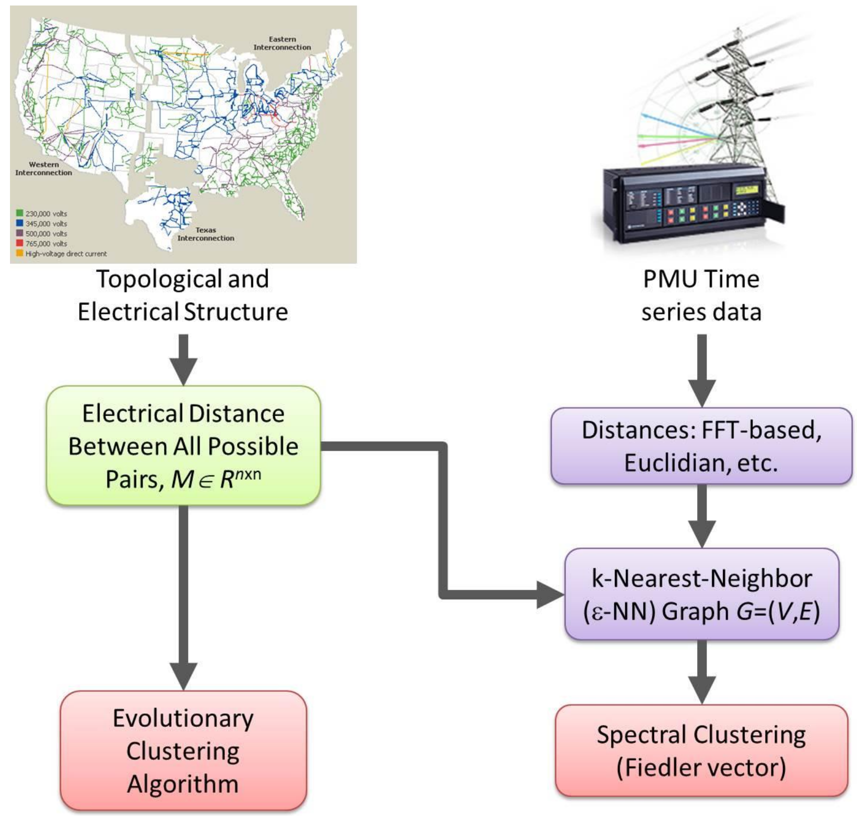
PNNL-SA-100103

Outline

- Overview
- Clustering Approaches
  - Fiedler vector
  - Evolutionary algorithm
- Case Studies
  - IEEE 50 generator system

Overview

- Two types of clustering
  - Evolutionary
  - Spectral
- Two types of data
  - Topological – uses information about connections in the grid
  - Derived – reporting from within the grid
- Two data sources
  - Electrical distance → topological
  - PMU data → derived



Overview (cont.)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- **Goal**: group generators which have similar behavior after a disturbance
- Clustering from topological data groups generators together based on their proximity and connectivity
  - Predicts which generators should have similar behavior
- Clustering from derived data uses the observed behavior of the generators to cluster
  - Should be more reliable, but relies on data from after the disturbance
- We compare the results of one type of clustering on topological data with another type of clustering on derived data

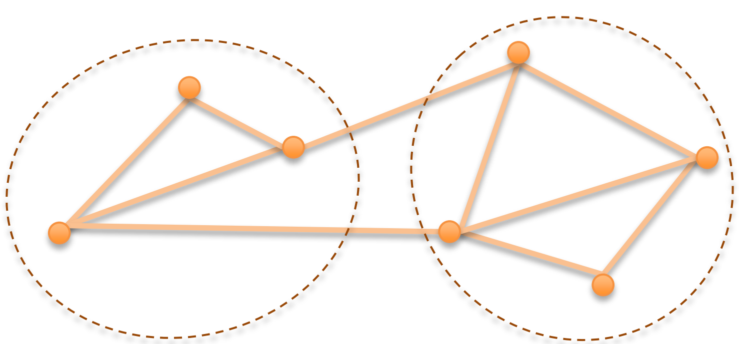
Outline

- Overview
- Clustering Approaches
  - Fiedler vector
  - Evolutionary algorithm
- Case Studies
  - IEEE 50 generator system

Fiedler Clustering

Introduction

- Input:
  - Graph – $G = (V, E, w)$ where V is a set of entities, E is a set of relationships between elements of V, w is a weight function on the edges indicating a level of similarity between connected entities
  - Integer k indicating number of clusters to break V into
- Output:
  - Partition of V into k groups, $\langle V_1, V_2, \dots, V_k \rangle$, so that their union is V and $V_i \cap V_j = \emptyset$ for all $i \neq j$.
    - Two elements within the same partition should be more similar than two elements in different partitions



Fiedler Clustering

Background

- Given a weighted graph we calculate the weighted Laplacian,
  - Let $d_i = \sum_{(i,j)\in E} w(i,j)$, then the Laplacian is defined as

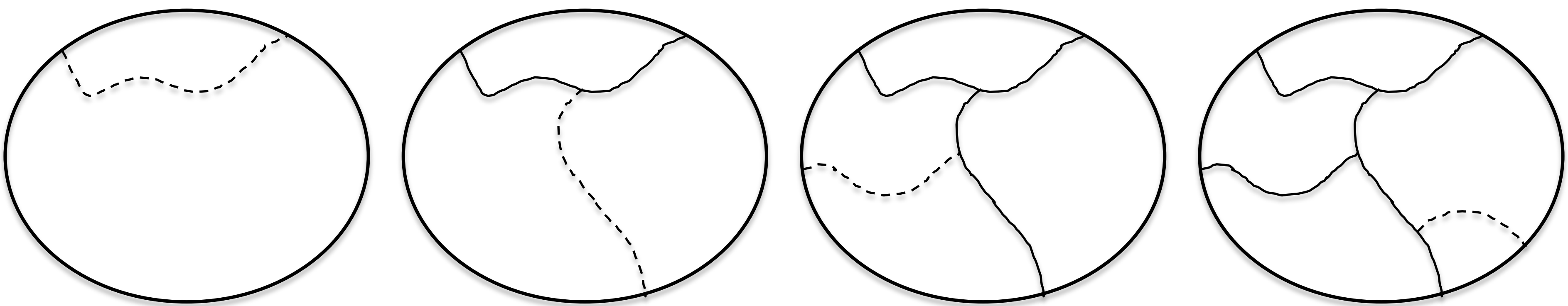
$$L(G) = \left(L_{i,j}\right)_{i,j=1..n} \text{ where } L_{i,j} = \begin{cases} d_i & i = j \\ -w(i,j) & (i,j) \in E \\ 0 & (i,j) \notin E \end{cases}$$

- Smallest eigenvalue is always zero (rows sum to zero)
  - Number of zero eigenvalues indicates number of connected components in the graph
- Eigenvector associated to smallest non-zero eigenvalue is called the *Fiedler vector*
  - Vertices make a continuous "choice" between -1 and +1, $-1 \leq f_i \leq 1$
  - Partition vertices into two groups – negative and positive
  - Zero values split between the two groups arbitrarily

Fiedler Clustering

Algorithm

- Repeatedly bipartition sets of vertices
  - Split every group to always yield 2^n clusters for some n, or
  - Split groups one at a time based on some measure
- Some choices:
  - How to split the vertices based on the Fiedler vector?
    - Positive vs. negative, or less than midpoint vs. greater than midpoint?
  - How to pick the next element to split?



Evolutionary Clustering

Electrically Cohesive Clusters

- Electrical distance
  - Relationship between electric power transaction and voltage phase angle difference.
  - Derive electrical distance from conductance matrix G or power flow matrix
    - Vertices are generators or buses
    - Edge weights are based on conductance, impedance, voltage
- Evolutionary clustering algorithm needs fitness function

Evolutionary Clustering

Measures of clustering quality

Aggregate clustering fitness, f

$$f = ECI^{\alpha} \cdot BCCI^{\beta} \cdot CCI^{\gamma} \cdot CSI^{\zeta} \cdot CC$$

Electrical Cohesiveness Index, ECI (i.e., tightness index)

$$ECI = 1 - \frac{sum\ of\ weights\ within\ clusters}{sum\ of\ all\ weights}$$

Between-cluster Connectedness Index, BCCI

$$BCCI = 1 - \frac{sum\ of\ one\ over\ weights\ for\ edges\ between\ clusters}{sum\ of\ one\ over\ all\ weights}$$

Cluster Count Index, CCI

Function which decreases as more clusters emerge

Cluster Size Index, CSI

Function which decreases as average size increases

Connectedness Index, CC

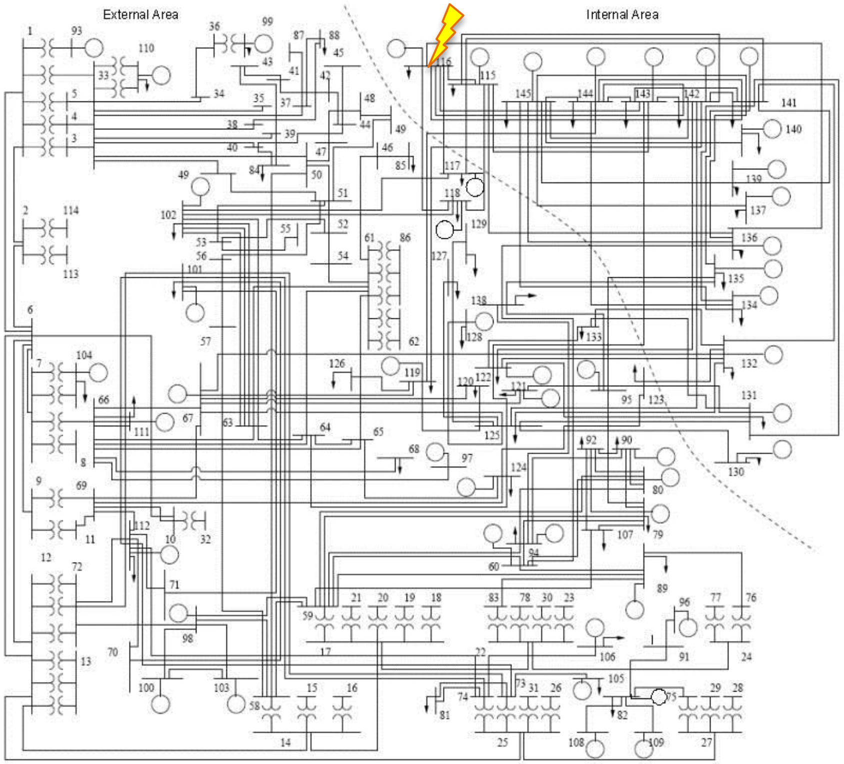
$$CC = \begin{cases} 1 & clusters\ are\ connected \\ 0 & clusters\ are\ not\ connected \end{cases}$$

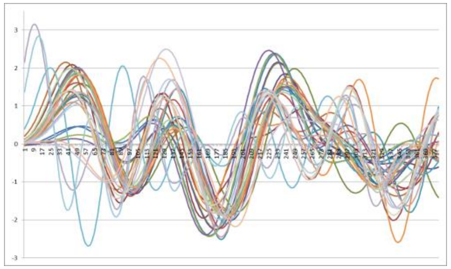
Outline

- Overview
- Clustering Approaches
  - Fiedler vector
  - Evolutionary algorithm
- Case Studies
  - IEEE 50 generator system

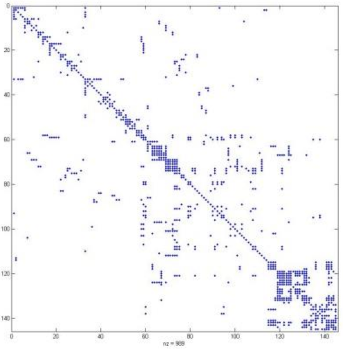
IEEE 50 generator system

- 145-buses, 50-generators
- Internal area has 16 generators, external has 34
- Generators modeled using classical machine dynamics with 2nd order swing equation
- Short-circuit fault at Bus 116 on line 116–136 for 60ms, then line tripped
- Data:
  - Oscillations of rotor angles in external system generators
  - Y_{bus} matrix → electrical distance





Rotor angle time series for each generator



Nonzeros in Y_{bus} matrix

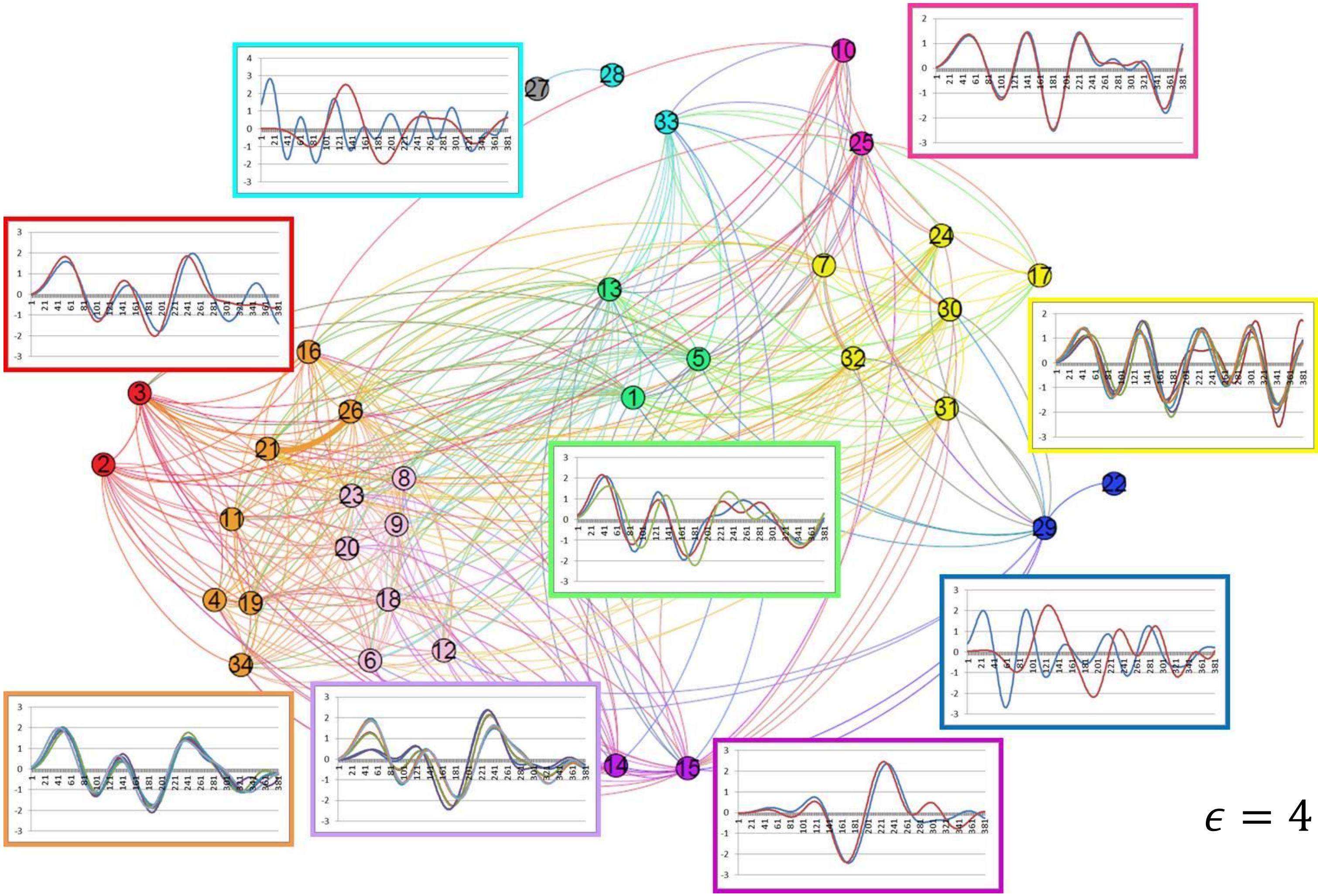
Fiedler cluster, IEEE 50

Setup

- Algorithm requires a graph: vertices are generators, what are edges?
  - Create ϵ-nearest-neighbor graph from rotor angle data
  - Need notion of distance between generators based on their rotor angle time series
    - Euclidean distance between time series vectors
    - Euclidean pseudo-distance between phase vectors of the DFT, weighted by absolute value of difference between amplitude vectors
  - $(v_i, v_j) \in E$ if the distance is less than some specified ϵ
    - Larger or smaller ϵ will generate more or less sparse graph
  - Weights must be similarities, not distances, so
    - $w(v_i, v_j) = \frac{1}{d(v_i, v_j)}$ if there is an edge between v_i and v_j
- Set number of clusters, $k = 10$

Fiedler cluster, IEEE 50

Results



$$\epsilon = 4$$

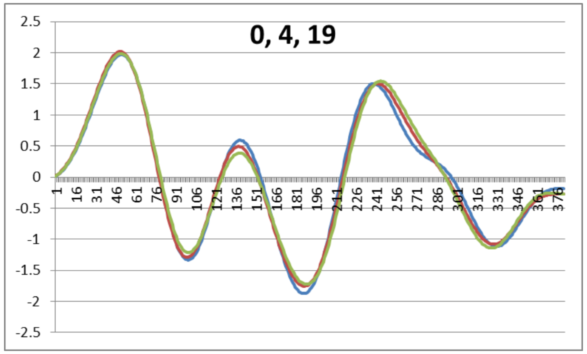
Evolutionary clustering, IEEE 50

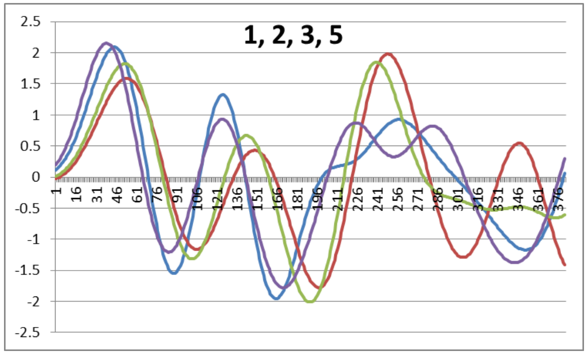
Setup

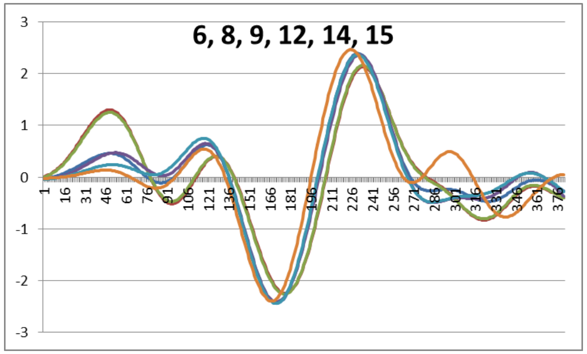
- Algorithm is designed for clustering buses rather than generators
- System has 145 buses – generators connect to 50 of those buses
- We cluster the generators based on which cluster its bus is in
- Use the electrical distance matrix, derived from the Y_{bus}
  - Ybus is nodal admittance matrix – rows and columns correspond to buses, entries are non-zero if there is a power line connecting buses, and the value of the entry is the admittance on that line

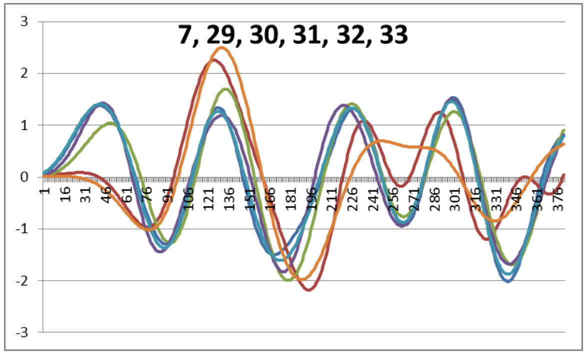
Evolutionary clustering, IEEE 50

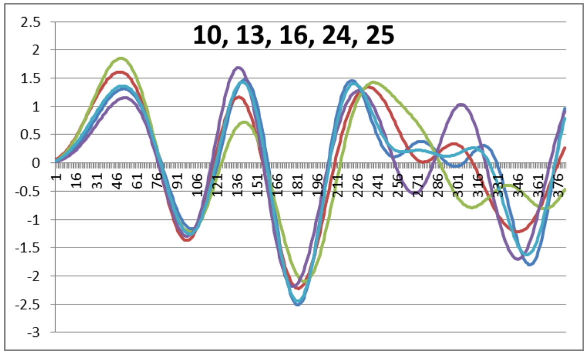
Results

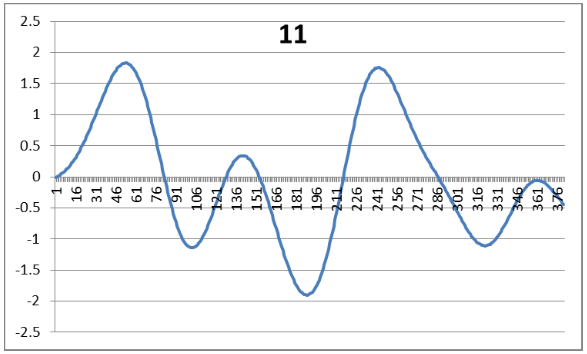


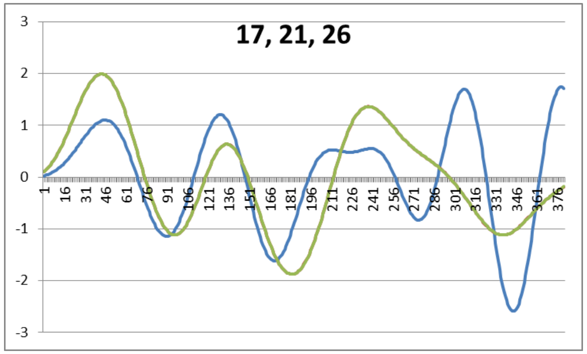


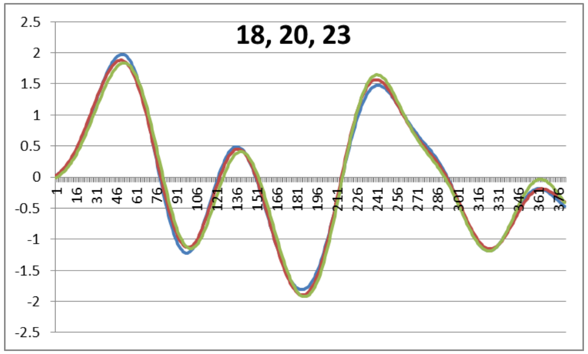


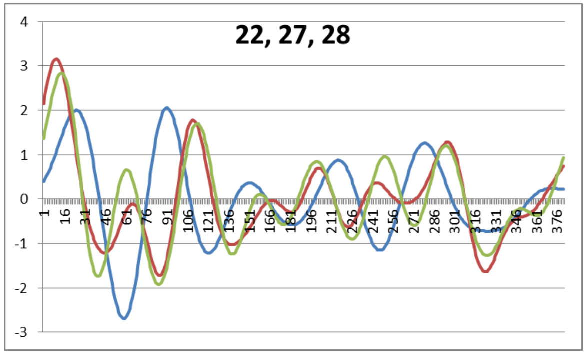












Comparison

- Visual inspection of the time series for each clustering method shows that both do reasonably well
  - Both have some clusters that don't appear to be good fits
  - Both have many clusters that appear to behave very similarly
- Clusters from Fiedler method generally intersect a small number of clusters from evolutionary method
  - Generally taking a large portion of one cluster and smaller pieces of other clusters.

Applications

- Model reduction
  - Grouping generators that swing together and selecting a characteristic generator from each group to represent that group's behavior
- Defensive Islanding
  - Controlled partitioning as response to instabilities
  - Isolate self-sufficient portions (islands)
  - Incorporate electrical characteristics to identify portions that could be self-sufficient
- Optimal placement of PMUs
  - PMU installation is expensive
  - Use clustering methods to identify groups that can share a PMU

Future Work

- Exploring other types of spectral clustering
  - K-means on eigenvectors
- Further comparison with other clustering methods
  - e.g., SVD methods, DBSCAN
- Implementation of some of the applications mentioned
  - Particularly model reduction

References

- E. Hogan, E. Cotilla-Sanchez, M. Halappanavar, S. Wang, P. Mackey, P. Hines, Z. Huang, "Towards Effective Clustering Techniques for the Analysis of Electric Power Grids", Proc. 3rd Int'l Workshop on High Performance Computing, Networking and Analytics for the Power Grid, 2013
- E. Cotilla-Sanchez, P.D.H. Hines, C. Barrows, S. Blumsack, and M. Patel. Multi-attribute partitioning of power networks based on electrical distance. IEEE Transactions on Power Systems, In press (special section on `Analysis and simulation of very large power systems'), 2013.

Acknowledgements

- The authors acknowledge the contributions of Shaobu Wang, Patrick Mackey, Paul Hines, Zhenyu Huang, Shuai Lu, Guang Lin, and Mihai Anitescu in this work.
- This work was supported in part by the Applied Mathematics Program of the Office of Advance Scientific Computing Research within the Office of Science of the U.S. Department of Energy (DOE).
- Pacific Northwest National Laboratory (PNNL) is operated by Battelle for the DOE under Contract DE-AC05-76RL01830.
- Additional funding for this work was provided by PJM Applied Solutions, NSF ECCS award #0848247, and US DOE award #DE-OE0000447.

BACKUP SLIDES

Fiedler Clustering

Algorithm

- Input: $G = (V, E, w)$, $k \in \mathbb{N}$
- $P \leftarrow \{V\}$ Partition of vertices
- for each p in P until $|P| = k$
  - $L \leftarrow L(G[p])$ Laplacian of graph restricted to vertices in p
  - $F \leftarrow fielder(L)$
  - $p_- \leftarrow$ vertices with negative Fiedlers in F (and some zeros)
  - $p_+ \leftarrow$ vertices with positive Fiedlers in F (and some zeros)
  - $P \leftarrow P \setminus \{p\} \cup \{p_-, p_+\}$ Replace p with the partition of p into two sets
- RETURN P

- Some options:
  - How to split the vertices based on the Fiedler vector? Positive and negative as we have here, or less than midpoint vs. greater than midpoint?
  - How to pick the next element of P to split?

Fiedler clustering on IEEE 50 system

- Algorithm requires a graph: vertices are generators, what are edges?
  - Create ϵ-nearest-neighbor graph from rotor angle data
  - Need notion of distance between generators based on their rotor angle time series – based on discrete Fourier transform of time series vector

$$DFT(v) = \langle v^{ph}, v^{amp} \rangle$$

$$d_{DFT}(v,w) = \sqrt{\sum_{i=1}^{n} \left(v_i^{ph} - w_i^{ph}\right)^2 \cdot \left|v_i^{amp} - w_i^{amp}\right|}$$

Evolutionary Clustering

Measures of clustering quality

Electrical Cohesiveness Index, ECI (i.e., tightness index)

$$ECI = 1 - \frac{\hat{e}(C)}{\hat{e}_{max}} = 1 - \frac{\sum_{i=1}^{n} \sum_{j \in M_i} e_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}}$$

Between-cluster Connectedness Index, BCCI

$$BCCI = 1 - \frac{h(C)}{h_{max}} = 1 - \frac{\sum_{i=1}^{n} \sum_{j \notin M_i} 1/e_{ij}}{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} 1/e_{ij}}$$

Cluster Count Index, CCI

$$CCI = e^{\frac{-(\ln p - \ln p_\star)^2}{2\sigma^2}}$$

Cluster Size Index, CSI

$$CSI = e^{\frac{-(\ln \bar{s} - \ln s_\star)^2}{2\sigma^2}}$$

Connectedness Index, CC

$$CC = \begin{cases} 1 & \textit{clusters are connected} \\ 0 & \textit{clusters are not connected} \end{cases}$$

Aggregate clustering fitness, f

$$f = ECI^{\alpha} \cdot BCCI^{\beta} \cdot CCI^{\gamma} \cdot CSI^{\zeta} \cdot CC$$

Evolutionary Clustering

GA configuration and parameters

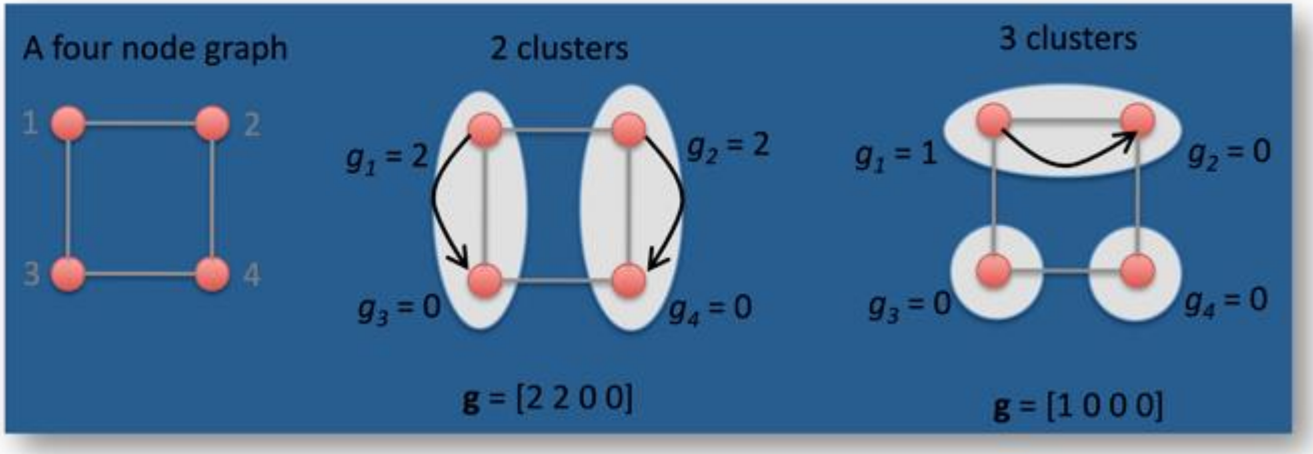
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Parameter	Value
Representation	String of integers indicating connections
Fitness function	$f = ECI \times BCCI \times CSI \times CCI$
Recombination method	1-point crossover
Recombination probability	80%
Mutation method	Random resetting with the feasible set $\{0,1,\ldots,m_i\}\ \forall i$
Mutation probability, p_m	$1/n$
Parent selection	Tournament selection without replacement
Survival selection	Roulette wheel with elitism
Population size	400
Initialization	Randomly generate clusters
Termination condition	Run until sufficient quality is achieved

Evolutionary Clustering

GA configuration and parameters

Parameter	Value
Representation	String of integers indicating connections
Fitness function	$f = ECI \times RCCI \times CSI \times CCI$
Recomb	
Recomb	
Mutatic	$m_i\} \ \forall i$
Mutatic	
Parent s	
Surviva	
Population size	400
Initialization	Randomly generate clusters
Termination condition	Run until sufficient quality is achieved



Polish System

- 110kV, 220kV and 400kV networks for a snapshot of the 1999-2000 winter peak conditions
- Model the boundary conditions with artificial buses that represent inter-tie power transactions
  - Resulting network is comprised of 2383 buses, 2896 transmission lines and 327 generators
- Collapse the network in preprocessing step: assume busses with degree one belong to the same cluster as their neighbor
  - Now system has 1733 busses
- Looking for 5 clusters since Polish grid was split into 5 different areas

Clustering, Polish system

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Fiedler
  - Don’t have PMU data for Polish system
  - Need different way of defining distance between generators
  - Use electrical distance calculated from Y_{bus}
  - Two clusters of size 217, three of size 433
- Evolutionary
  - Initial population a combination of random and k-means solutions
  - 6000 generations
  - Cluster sizes: 283, 285, 287, 297, and 581
- Comparison
  - Each evolutionary cluster intersects with each Fiedler cluster
  - Usually one large intersection and four small ones