Regionalization via Network-Constrained Clustering

David Sparks
Duke University, d.sparks@duke.edu

Follow this and additional works at: http://opensiuc.lib.siu.edu/pnconfs_2010

Recommended Citation
http://opensiuc.lib.siu.edu/pnconfs_2010/47

This Article is brought to you for free and open access by the Conference Proceedings at OpenSIUC. It has been accepted for inclusion in 2010 by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.
Regionalization via network-constrained clustering

David Sparks
d.sparks@duke.edu
Duke University
Political Science

This network graph depicts adjacencies among the 3,117 counties in the continental United States. This represents borders between neighboring counties, while nodes are colored according to the 2008 presidential election results, and sized according to total votes cast. Nodes are positioned in the graph according to the Kamada-Kawai force-directed algorithm.

The visualization supports observation due to the prevalence of a number of states that are located in the central United States, but severely under-represented are small states and large metropolitan areas, which are clearly visible in this large and detailed force-directed graph and automatic learning voting classifications.

Using the percentage of votes cast in each county, clustering algorithms generate distinct or similar clusters of electoral data, illustrating political preference within the continental United States.

Constrained clustering, as seen in the above graph, provides a clearer picture of partisan geographic tie than traditional clustering techniques that generalize familiar electoral data to identify regions of similar political preference within the continental United States.

Constrained clustering is a family of classification techniques that generate hard clustering algorithms. A constrained clustering algorithm ensures a clustering result contains a specified set of observations (clusters). Here, I applied network-constrained clustering in electoral data to identify regions of political preference within the continental United States.

Network-constrained clustering proceeds by generating a dendrogram, which can be used to determine the number of clusters in the data. The dendrogram represents a hierarchical clustering of the data, with each node representing a cluster of observations.

Each of the other maps visualizes contiguous 3, 4, and 12 clusters, giving a clearer picture of partisan geographic tie than traditional constrained clustering techniques that generalize familiar electoral data to identify regions of political preference within the continental United States.

The left advantages of network-constrained clustering is that it allows consideration of both measured variables and network position in clustering; improving clusters or communities within the network.

Furnishing counties into 12 clusters begins to show broader and politically meaningful regions.

References

This clustering allows a geographically constrained plot of voting for the 3, 4, and 12 cluster results, with all elections and geographic data.