# Graph-Based Model Election Forecasting

Matthew Goodman
Advisors: Prof. Alejandro Ribeiro & Fernando Gama
Dept. of Electrical and Systems Engineering
University of Pennsylvania
goodmanm@seas.upenn.edu

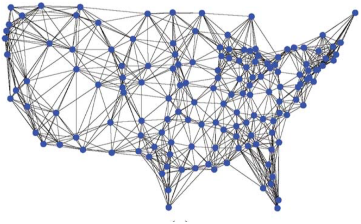
April 26, 2016

# Polling Shortcomings

- Elections are fundamental to democratic society
  $\Rightarrow$ Draw widespread attention from the public $\Rightarrow$ Forecasting
- Standard way to forecast election is through polling
- Polling is costly and resource intensive
  $\Rightarrow$ Classify population $\Rightarrow$ Sample $\Rightarrow$ Predict.
- Counties are the electoral units $\Rightarrow$ Report accurate results
- Relationship between counties is non-linear $\Rightarrow$ Model as a network

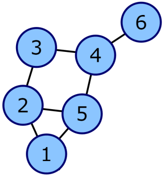
- Election results are reported in terms of counties
- Counties as nodes on a graph $\Rightarrow$ Election results as a graph signal
- Every pair of counties has a weighted connection
- Connection weight derived from historical voting patterns
- Determine the subset of counties to poll

  $\Rightarrow$ Exploit theory of sampling graph signals



# Election Network

- Election network as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{W})$
  $\Rightarrow$ $\mathcal{N}$: counties; $(\mathcal{E}, \mathcal{W})$: historical voting patterns
- Election results $x_i$ for each county $\Rightarrow$ Signal $\mathbf{x} \in \mathbb{R}^n$
- Graph shift operator $\mathbf{S}$: captures the strength of the connection
- $\mathbf{S}$ is a normal matrix: $\mathbf{S}\mathbf{S}^H = \mathbf{S}^H\mathbf{S}$ $\Rightarrow$ Diagonalizable $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$
- Define a graph Fourier transform $\hat{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$
  $\Rightarrow$ If $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_k \ \mathbf{0}_{n-k}]^T$ $\Rightarrow$ $\mathbf{x}$ is $k$-bandlimited




$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & 0 & 0 & S_{15} & 0 \\ S_{21} & S_{22} & S_{23} & 0 & S_{25} & 0 \\ 0 & S_{23} & S_{33} & S_{34} & 0 & 0 \\ 0 & 0 & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & 0 & S_{54} & S_{55} & 0 \\ 0 & 0 & 0 & S_{64} & 0 & S_{66} \end{pmatrix}$$

# Sampling of Graph Signals

- $\mathbf{x}$ is *$k$-bandlimited* $\Rightarrow$ $\hat{x}_k$ freq. components; $\mathbf{V}_k$ eigenvectors
  $\Rightarrow$ Perform sampling and reconstruction
- Determine which $p \geq k$ nodes to sample $\Rightarrow$ Sampling matrix $\mathbf{C}$
- Find $\mathbf{C}$ such that $(\mathbf{C}\mathbf{V}_k)$ is invertible
  $\Rightarrow$ Reconstruct $\tilde{\mathbf{x}} = \mathbf{V}_k(\mathbf{C}\mathbf{V}_k)^{-1}(\mathbf{C}\mathbf{x})$
- In the presence of noise $\mathbf{C}^* = \text{argmax}\{\sigma_{\text{min}}(\mathbf{C}\mathbf{V}_k)\}$
  $\Rightarrow$ Select the nodes that let the least noise contaminate the signal
  $\Rightarrow$ Solvable by computationally tractable algorithms
- Determining nodes to sample $\Rightarrow$ Selecting counties to poll

# Results

- Built the network using Presidential election results from 1984-2008
- Added senatorial results for cycles between those same years
- Attempted to predict 2012 election results based on previous years

  $\Rightarrow$ Tested different ways of relating the data $\Rightarrow$ Use covariance

- 23 counties to poll from states:

  $\Rightarrow$ Florida, California, Montana, Virginia, Massachusetts, etc

- Prediction relative error: 4.99% of aggregate votes over all counties
- Counties with worst prediction:

  $\Rightarrow$ Kenton (KY), Red River (TX), Windham (VT)

- Counties with best prediction:

  $\Rightarrow$ Morrill (NE), Washington (PA), Macon (TN)

- Prediction matched actual result: Democrat Win

# Conclusions & Future Directions

- Used historic election results to identify voting patterns
- Built an election network with counties (electoral units) as nodes
- Defined election results as a signal defined on this graph
- Used sampling theory to determine which counties to poll
- Contrasted predicted results with latest 2012 elections

- Adjust the election model to incorporate verifiable census data
  $\Rightarrow$ Includes median income data, poverty index data, etc.
- Explore newly proposed methods to generate graph shift operator.
  $\Rightarrow$ Aggregate all available data into single shift operator.
- Apply the model to forecast the 2016 election.