Graph frequency analysis of COVID-19 incidence to identify contagion patterns in different counties of the United States

Yang Li and Gonzalo Mateos
E-mail: yli131@ur.rochester.edu
Dept. of Electrical and Computer Engineering, University of Rochester

Abstract

The COVID-19 pandemic severely changed the way of life in the United States (US). From early scattered regional outbreaks to current country-wide spread, and from rural areas to highly populated cities, the contagion exhibits different patterns at various timescales and locations. We conducted a graph frequency analysis to investigate the spread patterns of COVID-19 in different US counties. The commute flows between US counties were used to construct the graph that measures the population mobility between counties. The numbers of daily confirmed COVID-19 cases per county were collected and represented as graph signals, which were then mapped into the frequency domain via the Graph Fourier Transform. The concept of graph frequency in Graph Signal Processing (GSP) enables the decomposition of graph signals (i.e., daily confirmed cases) into modes with smooth or rapid variations with respect to the underlying commute graph. Follow-up analysis revealed the relationship between graph frequency components and the COVID-19 spread pattern within and across counties. Transformation between different spread patterns within the same region was also revealed by graph frequency analysis on finer temporal scales. Specifically, our preliminary graph frequency analysis of COVID-19 data as graph signals exhibits different patterns at various timescales and locations.

Motivation and context

- Huge interest in understanding the spread patterns of the virus
- Previous work mostly focused on

Pathology analysis from biological perspectives
- Contagion within specific locations
- Case prediction through forecasting
- Macroscopic view of the contagion within the nation
- Temporal analysis of signals
- Local outbreaks in highly-concentrated places
- Localized outbreak in one county (row 1), to Wide spread across counties (row 2)

Graph signal fundamentals

- Weighted, undirected, and connected graph \((G(V, E; W))\)
  - Set of nodes \(V = \{1, \ldots, N\}\)
  - Consider the signal \(x = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^N\)
  - \(x_i\) denotes the signal value at node \(i \in V\)
- Set of edges \(E = \{(i,j) \in V \times V\}\)
  - \(W_{ij}\) are weights on the connected pairs of nodes
  - Large \(W_{ij}\) signal values \(x_i, x_j\) tend to be similar
  - Small \(W_{ij}\) signal values \(x_i, x_j\) tend not to be related

Graph Fourier transform and smoothness

- Eigendecomposition of graph Laplacian \(L = \text{diag}(\lambda_i)L = W \Lambda W^T\)
  - Diagonal eigenvalue matrix \(\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)\)
  - Orthornormal matrix of eigenvectors \(W = [v_1, \ldots, v_N]\)

Graph Fourier transform (GFT)

- \(\tilde{x} = x - \bar{x}\) where \(\bar{x} = \frac{1}{N} \sum x_i\)
- Synthesize \(x\) as a sum of orthogonal frequency components \(v_k\)
  - \(\text{GFT coefficient} \: \lambda_k \text{ contribution of} \: v_k\text{ to signal} \: x\)
  - Notion of signal variability over the graph

- Total variation of the graph signal \(x\)
  - \(\sum \text{degree}(x_i) = \sum W_{ij}x_i\)
- Total variation of the eigenvectors \(v_k\)
  - \(\sum v_k^2 = \sum \lambda_k = 0\)
- Indicate how eigenvectors (frequency components) vary over graph \(G\)
  - \(\text{GFT/IGFT enables decomposition of graph signal} \: x\)
  - Into spectral components, and
  - Characterize different levels of variability

Graph filtering

- Graph signal \(x\) with GFT coefficients \(\lambda_k\)
  - Eigenvalues of the Laplacian correspond to graph frequencies
  - Eigenvectors serve as frequency basis
  - Isolate lowest \(N_k\) eigenvalues and corresponding eigenvectors
  - \(\text{Low-pass filter} \: x_k = \text{proj}_{\Lambda_k}\)
- \(L = \text{diag}(\lambda_i)\), \(L = (\lambda_i < \lambda_j)\)
  - \(L_{ij} = x_i - x_j\)
  - \(\Lambda_{ij} = \text{val}, \: V_{ij} = \text{val}, \: \Lambda_{ii} = \text{val}, \: \Lambda_{ji} = \text{val}\)

- Graph band-pass filter \(H_{Bk}\) and High-pass filter \(H_{Ak}\)
  - Mutually exclusive and span all graph frequencies
  - \(\text{Decomposes the original graph signal into} \: x = x_k + x_{Ak}\)
  - \(\text{Increases the resolution of the signal} \: x_{Ak}\)
  - \(\text{Low, medium and high-variability w.r.t underlying graph}\)

COVID-19 data as graph signals

- Cumulative number of confirmed COVID-19 cases per 100k residents
  - For each of the \(N = 3142\) counties in US
  - From Jan 22 to August 31, \(x_{2021}\)
  - \(W_{ij}\) average population mobility flow between two counties

Graph construction

- Commute flow data from year 2011 to 2015
- Population mobility between US counties
- Weighted undirected graph \((G(V, E; W))\)
  - \(N = 3142\) counties as nodes, \((i,j) \in E\)
  - \(W_{ij}\) average population mobility flow between two counties

Frequency decomposition of graph signals

- \(\text{Construct low-pass filter} \: x_{LP}\)
  - Take the lowest/highest one fifth of the eigenvalues
- \(\text{X is now decomposed into} \: x_{LP}, x_{MP}, x_{HP}\)
  - Take row-wise average of the absolute values in \(x_{LP}, x_{HP}\)

Frequency analysis w.r.t contagion patterns

- LP regions: more daily cases and take longer to reach peak
- HP regions: Local outbreaks in highly-concentrated places

Discussion and road ahead

- Graph Signal Processing for COVID-19 contagion patterns
- Information extracted from graph frequency domain
- \(GSP\) vs across/within county contagion
- Future work shall be devoted to
  - Graph capturing dynamic population mobility
  - Spatial-temporal analysis on finer scales

References