

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

Abstract

The COVID-19 pandemic severely changed the way of life in the United States (US). From early scattered regional outbreaks to current countrywise 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 exploited confirmed case counts to unveil spatio-temporal contagion patterns of COVID-19 incidence for each US county. Overall, results here support the promising prospect of using GSP tools for epidemiology knowledge discovery on graphs.

Motivation and context

- Huge interest in understanding the spread patterns of the virus
- Previous work mostly focused on
 - \Rightarrow Pathology analysis from biological perspectives
 - \Rightarrow Contagion within specific locations
 - \Rightarrow Case prediction through forecasting



- Macroscopic view of the contagion within the nation
 - \Rightarrow Spatio-temporal analysis of the contagion
 - \Rightarrow Graph frequency analysis
 - \Rightarrow Study of transformation between contagion patterns
- Investigate COVID-19 spread pattern across all counties in the US

Graph signal processing - 101

- Network as graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$: encode pairwise relationships
- \blacktriangleright Interest in both \mathcal{G} , and in data associated with nodes in \mathcal{V} \Rightarrow The object of study is a graph signal
- **Ex:** Opinion profile, buffer congestion levels, neural activity, epidemic
- Graph SP: need to broaden classical SP results to graph signals \Rightarrow Our view: GSP well suited for spatio-temporal analysis

Yang Li and Gonzalo Mateos E-mail: yli131@ur.rochester.edu

Dept. of Electrical and Computer Engineering, University of Rochester

Graph signal fundamentals	COVII
► Weighted, undirected, and connected graph G(V, E, W) ⇒ Set of N nodes V = {1,, N}	► C
 ⇒ Set of edges E ⊆ V × V Consider the signal x = [x₁,,x_N]^T ∈ ℝ^N ⇒ x_i denotes the signal value at node i ∈ V W : V × V → ℝ₊: set of unordered pairs of nodes ⇒ W_{ij} ≥ 0, W_{ij} = W_{ji} Large W_{ij}: signal values x_i and x_j tend to be similar Small W_{ij}: signal values x_i and x_j not directly related 	► C
Graph Fourier transform and smoothness	
 Eigendecomposition of graph Laplacian L := diag(W1) - W Decompose the symmetric graph Laplacian as L = VΛV^T ⇒ Diagonal eigenvalue matrix Λ := diag(λ₁,, λ_N) ⇒ Orthonormal matrix of eigenvectors V := [v₁,, v_N] Graph Fourier transform (GFT) ⇒ x̃ = [x̃₁,, x̃_N]^T = V^Tx ⇒ iGFT (inverse GFT) of x̃: x = Vx̃ 	 H C Graph
$\mathbf{v} = \mathbf{V}\tilde{\mathbf{v}} = \sum_{i=1}^{N} \tilde{\mathbf{v}}_{i} \mathbf{v}_{i}$	► C
Synthesize x as a sum of orthogonal frequency components \mathbf{v}_k GFT coefficient \tilde{x}_k : contribution of \mathbf{v}_k to signal x \Rightarrow Notion of signal variability over the graph Total variation of the graph signal x $TV(\mathbf{x}) := \mathbf{x}^T \mathbf{L} \mathbf{x} = \sum_{i \neq j} W_{ij}(x_i - x_j)^2$ Total variation of the eigenvectors \mathbf{v}_k $TV(\mathbf{v}_k) = \mathbf{v}_k^T \mathbf{L} \mathbf{v}_k = \lambda_k, 0 = \lambda_1 < \lambda_2 \le \ldots \le \lambda_N$ Indicate how eigenvectors (frequency components) vary over graph \mathcal{G} GFT/iGFT enables decomposition of graph signal x \Rightarrow Into spectral components, and \Rightarrow Characterize different levels of variability	
Graph filtering	► S
 Graph signal x with GFT coefficients x̃ ⇒ Eigenvalues of the Laplacian correspond to graph frequencies ⇒ Eigenvectors serve as frequency basis Isolate lowest N_L eigenvalues and corresponding eigenvectors ⇒ Low-pass filter executed as x̃_L = H̃_Lx̃ ⇒ H̃_L = diag(H̃_L), ñ̃_{L,n} = I {n < N_L} x_L = Vx̃_L = VH̃_Lx̃ = VH̃_LV^Tx = H_Lx 	Freque C X Ta
 ► Graph band-pass filter H_M and high-pass filter H_H ⇒ Mutually exclusive and span all graph frequencies ► Decomposes the original graph signal into x = x_L + x_M + x_H ⇒ Increases the resolution of the signal ⇒ Low, medium and high variability w.r.t underlying graph 	

D-19 data as graph signals

Cumulative number of confirmed COVID-19 cases per 100k residents \Rightarrow For each of the N = 3142 counties in US \Rightarrow From Jan 22 to August 31, $\mathbf{x} \in \mathbb{R}^{3142 \times 223}$ Compute daily confirmed cases per 100k residents per county



-lidden relationships between the signals of each county Consider commute flow between counties

n construction

Commute flow data from year 2011 to 2015 \Rightarrow Population mobility between US counties

- Neighted undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$ $\Rightarrow N = 3142$ counties as nodes, $\binom{3142}{2}$ edges
- $\Rightarrow W_{ii} :=$ average population mobility flow between two counties



_arge portion of the commute flows happen within states Significant amount of connections between states

ency decomposition of graph signals

Construct low/high-pass filter

 \Rightarrow Take the lowest/highest one fifth of the eigenvalues is now decomposed into $\mathbf{X}_L, \mathbf{X}_M, \mathbf{X}_H$ Take row-wise average of the absolute values in $\mathbf{X}_L, \mathbf{X}_H$



P regions: similar number of daily cases with nearby counties \Rightarrow Aligned with the commute flows between counties **HP regions**: distinct and abnormal signals against nearby counties \Rightarrow Against the population mobility between them



Temporal analysis of signals







► Transformation of spread pattern in NYC area \Rightarrow Localized outbreak in one county (row 1), to \Rightarrow Wide spread across counties (row 2)

Discussion and road ahead

References



Frequency analysis w.r.t contagion patterns

Partition signals of 223 days into windows of 7 days \Rightarrow Take row-wise average of each window Localized outbreak identified as a high-pass signal first \Rightarrow Westchester county at week 6

Magnitudes of high-pass signals decreased

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

► G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," IEEE Signal Process. Mag., vol. 36, no. 3, pp. 16-43, 2019.

A. Ortega, P. Frossard, et al. "Graph signal processing: Overview, challenges, and applications," Proc. IEEE, vol. 106, no. 5, pp. 808-828, 2018.

> Y. Li and G. Mateos, "Graph frequency analysis of COVID-19 prevalence in the United States," KDD Workshop on Mining and Learning with Graphs, San Diego, CA, Aug. 24, 2020.