NON NEGATIVE KERNEL GRAPHS FOR TIME-VARYING SIGNALS USING VISIBILITY GRAPHS

Ecem Bozkurt, Antonio Ortega

Department of Electrical and Computer Engineering
University of Southern California

ABSTRACT

In this work, we present a novel framework to represent time-varying signals as dynamic graphs using the Non Negative Kernel (NNK) graph construction algorithm. Visibility Graphs (VG) are utilized to transform the time-series signals from time domain to graph domain. Degree similarity and clustering coefficients of these temporal graphs are used to characterize the similarity between different nodes. Using this similarity metric, a spatiotemporal graph is obtained by NNK algorithm. In addition, by introducing the delayed signals of the same node in the dictionary, we explored the relation between the delayed signals, which helped us discovering the synchronization between time-series signals. This technique is applied on a temperature dataset which consists of daily temperatures that are collected by 160 US weather stations in 2020. We compared the results of our proposed method to an alternative graph construction technique. Finally, we have shown that, for a highly correlated dataset, proposed NNK algorithm can still achieve a sparse graph that is compliant with the geographical information.

***Index Terms*—** Time Varying Graphs, Temporal Graphs, Dynamic Graphs, Visibility Graphs, Graph Signal Processing

1. INTRODUCTION

With the enormous growth of data, understanding the complex interactions and the underlying structures is becoming more important than ever. Representing signals in graph domain, helps us to understand these sophisticated relations between the signals [1].

The recent developments in graph signal processing enable us to represent signals in more efficient ways. *Dong et al* and *Kalofolias et al* proposed graph learning techniques for signals that have smooth variations [2, 3]. *Egilmez et al* proposed a graph construction method for signals under Laplacian and structural constraints [4]. However, these graph construction techniques have smoothness assumption or structural constraints. Moreover, these works, like many others, target at static graphs and they do not consider the temporal evolution of the graphs.

Biological signals, financial information, environmental data and social networks are some of the fields where graph signal processing methods are utilized [5, 6, 7, 8, 9]. Most real-world signals are time-varying signals by their nature. Modeling these data in time-varying graph domain is advantageous in analyzing and demonstrating the spatiotemporal connections.

There were many attempts in graph construction for time-varying signals [10, 11, 12, 13, 14, 15]. *Kalofolias et al* proposed a time-varying graph learning method, which assumes smooth variation in time [10, 3]. *Liu et al* also proposed a time-varying graph construction framework for smooth signals [13, 16]. *Yamada et al* proposed a time-varying graph learning for signals that show slow variation in time and for the signals that have switching behaviour [14]. *Grassi et al* proposed a joint framework for time-varying graphs [15].

Previously, k Nearest Neighbor (kNN) graphs were used for time-series signals to avoid computation of inverse covariance matrices and Laplacians [17]. KNN graphs are distance based graphs, which are sensitive to redundant features and number of neighbors and K has to be tuned.

Non Negative Kernel regression (NNK) is a new graph construction technique that aims at efficient graph representation with non-negative basis pursuit [18, 19]. In the NNK algorithm, sparsity is a function of the geometry of the data. NNK graph construction is geometrically explicable and intuitive. A direct extension of NNK graphs is time-varying NNK graphs.However, the adaptation of NNK for time-series signals requires finding a suitable similarity metric for the application, determining the time window and taking time delays into an account. In addition, for shorter time windows, the signals become highly correlated, in that case, obtaining a sparse graph can be challenging.

To address the challenges of dealing with highly correlated time-series signals, we used Visibility Graphs (VG) to obtain a meaningful pairwise similarity matrix between the signals in sliding window fashion. In addition, time-shifted signals of the same nodes are included in the NNK dictionary. Incremental NNK is modified to avoid unstable matrices and to enforce consistency between the time-shifted signals.

The rest of the paper is organized as follows. Section 2 explains Visibility Graphs (VG), Section 3 reviews the NNK algorithm, in Section 4 time-varying NNK graphs are introduced and in Section 5 we present and discuss our method applied on temperature dataset.

2. VISIBILITY GRAPHS

The naïve approach for finding the pairwise similarities between the time-series signals in time-domain is to compare direct cosine similarity, or using the statistical features, such as mean, standard deviation, minimum or maximum. However, these approaches are not robust to noise, they are sensitive to scaling and they cannot characterize the nonlinear features in the time-series signals. Visibility Graphs(VG) were proposed by *Lacasa et al* to uncover the scale-free relations in time series signals [20]. Visibility Graphs transform time series signals into a graph, preserving the geometric relations. For chaotic, complex and non-periodic signals like biological signals or financial data, this method has been used as a feature extraction tool to extract nonlinear characteristics of the signals [21, 22, 23, 24, 25] .

Visibility criterion is defined as follows:

$$y(\mathbf{t}_k) < y(\mathbf{t}_j) + (y(\mathbf{t}_i) - y(\mathbf{t}_j)).\frac{\mathbf{t}_j - \mathbf{t}_k}{\mathbf{t}_j - \mathbf{t}_i}, \tag{1}$$

where $y(\mathbf{t}_i)$,$y(\mathbf{t}_j)$ and $y(\mathbf{t}_k)$ are the data points at time $\mathbf{t}_i$,$\mathbf{t}_j$ and $\mathbf{t}_k$.

Data points $y(\mathbf{t}_i)$ are $y(\mathbf{t}_j)$ visible if $\mathbf{t}_k$ is between $\mathbf{t}_i$ and $\mathbf{t}_j$ and if $y(\mathbf{t}_i)$,$y(\mathbf{t}_j)$ and $y(\mathbf{t}_k)$ satisfy this criterion. The idea is depicted in Figure 1.

.

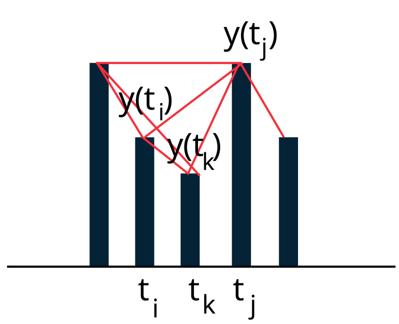


Fig. 1. Visibility Graph criterion is satisfied by data points $y(\mathbf{t}_i)$, $y(\mathbf{t}_j)$ and $y(\mathbf{t}_k)$ in this example.

After VGs are obtained for each time window, degree distribution and clustering coefficient are used as metrics to characterize and summarize the VGs.

degree distribution is given below:

$$DD \tag{2}$$

clustering coefficient is given below:

$$CC \tag{3}$$

3. NON NEGATIVE KERNEL ALGORITHM

Non Negative Kernel (NNK) algorithm aims at constructing a sparse graph by non negative basis pursuit. The key idea behind the NNK algorithm is to approximate each node by a linear combination of atoms from a dictionary formed by its neighboring nodes. By this technique, we can eliminate redundant connections [18, 19] .

NNK algorithm is also useful for understanding the graphs from a geometric perspective. After initialization with Gaussian Kernel, a polytope is formed with Kernel Ratio Interval (KRI) condition, which was described in [18]. Gaussian Kernel is given as

$$G(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\text{dist}(\mathbf{x}_i, \mathbf{x}_j)}{\sigma^2}} \tag{4}$$

where $\text{dist}(\mathbf{x}_i, \mathbf{x}_j)$ is the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. However, different similarity measures can be used, which are beyond the scope of this work. In this approach, $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to a subset of nodes, that is defined by a kNN neighborhood, the size of which is controlled by the parameter K. Here, σ is also chosen as a function of K to define the neighborhood.

In our adaptation, we use the average of pairwise degree similarity and clustering coefficient as $\mathbf{x}_i$ and $\mathbf{x}_j$ in Gaussian Kernel computations, as given in 2 and 2.

It is important to note that, for an accurate graph construction, intuitively it is better to start with the largest neighborhood possible and then to shrink the graph with NNK algorithm. However, under some conditions, this selection can lead to unstable graphs when it is applied to time-varying signals.

4. TIME-VARYING NNK GRAPHS

We define $G_t(V_t, E_t)$ as the graph for the timestamp t, with vertices at V_t and edges E_t, both of them are the functions of the timestamp t. As introduced in [10], we also segment the time-series sequences into smaller non-overlapping windows, with window size w.Each graph $G_t(V_t, E_t)$ is obtained from the segments of the sequences of size w in a sliding window manner.

Time window w is a design parameter and just as it was discussed in [10], there is a trade-off between the temporal resolution of time-varying graphs and the number of samples available for learning the graph, $G_t(V_t, E_t)$. In addition to that, the choice of w also affects the optimal K for NNK algorithm. If we apply NNK algorithm for short window sizes and if we start with a large value for K, this will result in ill-conditioned Gaussian Kernel, as the sequences will be very similar to each other. Condition number measures how poorly the matrix is conditioned for inversion, formulated as in (5). The greater condition number indicates that solutions with this matrix will be more sensitive to the perturbations.

$$\kappa(\mathbf{A}) = ||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}|| \quad (5)$$

To overcome this problem, a modified incremental NNK is used instead. Potential neighbors are sorted using kNN, and they are added one by one until the number of edges stays constant. Before each optimization step, condition number as given in 5 of Gv_i is checked to avoid ill-conditioned matrix. If the condition number exceeds the threshold, search for a new edge is terminated.

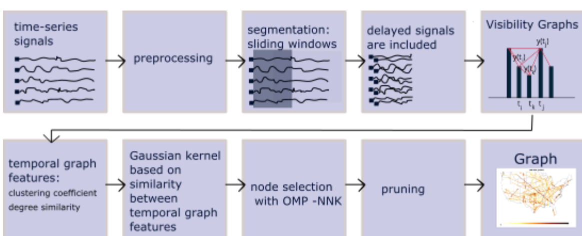
For time-varying graphs, it's also important to analyze how time-series signals propagate in time.

For this purpose, time-shifted signals of the nodes, i.e. signals that are shifted by delays $1, 2, .., d$ are added to the dictionary, as if they were separate nodes.

We treat additional delayed time-series signals equally and to find the basis nodes one-by-one and then do the pruning.

To enforce the consistency between the delayed signals, we used pruning. In pruning, the goal is to find the strongest connections between the time-shifted signals of the same nodes.

In Figure 4, flowchart is given.



The modified OMP-NNK Algorithm is described for the second approach in [**?**].

KRI Condition with Gaussian Kernels:

Lemma 4.1. *Kernel Ratio Interval (KRI) for NNK algorithm to form an edge between node i wtih time series signal $f(t)$ and node j_0 with time series signal $f(t)$ with noise $n(t)$, but not with node j_1, where j_1 represents the same signal as in j_0 but delayed by 1, is given in terms of the angle between the distance vectors $d(i, j_0)$ and $d(i, j_1)$ which is denoted as α, noise and the difference between $f(t+1)$ and $f(t)$ denoted as d below:*

$$\frac{||n||}{||n+d||} < cos\alpha < \frac{||n+d||}{||n||}$$

Proof. Let node i be a function $f(t)$ and let nodes j_0 and j_1 represent $f(t)$ with additive noise $n(t)$ and delayed j_0 by one sample, respectively.

Distances between i and j_0 ,and i and j_1 are denoted as

Algorithm 1 Multinode OMP-NNK algorithm

Require: $Kernel : K$
Ensure: $ConditionNumber for K_{S,S} < threshold$
for $i = 0, 1, ..., N$ **do**
  $j_1(t, d) \leftarrow \arg\max_j K_{i(t,0),j(t,d)}$
  $\theta_1 \leftarrow K_{i(t,0),j_1(t,d)}$
  $S = \{j_1(t), j_1(t+1), ..., j_1(t+d)\}$ {add all delayed signals of the same node}
  for $s = 2, 3, ..., k$ **do**
    $j_s(t) \leftarrow \arg\max_j K_{i(t,0),j(t,d)} - K^T_{S,j(t,d)}\theta_{s-1}$
    if $K_{i(t,0),j(t,d)} - K^T_{S,j(t,d)}\theta_{s-1} < 0$ **then**
      $break$
    end if
    $S \leftarrow S \cup \{j_s(t), j_s(t+1), ..., j_s(t+d)\}$
    $\theta_s = \min \frac{1}{2}\theta^T K_{S,S}\theta - K^T_{S,i(t,0)}\theta$
  end for
  $J_{i(t,0)} = \frac{1}{2}\theta^T K_{S,S}\theta - K^T_{S,i(t,0)}\theta + \frac{1}{2}K_{i(t,0),i(t,0)}$
  $W_{i(t,0),S} = \theta_S, W_{i(t,0),S^c} = 0$
  $E_{i(t,0),s} = J_{i(t,0)}1_k$, $E_{i_{t,0},S^c} = 0$
end for
W ,E =0

$d(i, j_0)$ and $d(i, j_1)$.

$$d(i, j_0) = d(f(t) - (f(t) + n(t))) = ||n|| := a$$
$$d(i, j_1) = d(f(t) - (f(t+1) + n(t+1))) := ||n+d|| = b$$

where d is the difference between functions $f(t)$ and $f(t+1)$

Kernels $K_{i,j0}$,$K_{i,j1}$ and K_{j_0,j_1} are given below:

$$K_{i,j_0} = exp(-\frac{1}{2\sigma^2}(a^2)$$
$$K_{i,j_1} = exp(-\frac{1}{2\sigma^2}(b^2)$$
$$K_{j_0,j_1} = exp(-\frac{1}{2\sigma^2}(bcos\alpha)^2 - (bsin\alpha)^2)$$
$$K_{j_0,j_1} = exp(-\frac{1}{2\sigma^2}(b^2 + a^2 - 2abcos\alpha)$$

KRI condition 1:

$$\frac{K_{i,j_0}}{K_{i,j_1}} > K_{j_0,j_1}$$

$$exp(-\frac{1}{2\sigma^2}(b^2-a^2)) > exp(-\frac{1}{2\sigma^2}(b^2+a^2-2abcos\alpha)$$

$$\frac{-a^2}{2\sigma^2} > \frac{a^2-2abcos\alpha}{2\sigma^2}$$

$$b^2-a^2 > -(b^2+a^2-2abcos\alpha)2b^2 > 2abcos\alpha$$

$$b > acos\alpha$$

$$||n+d|| > ||n||cos\alpha \Leftrightarrow \theta_{ij_0} > 0$$

$$cos\alpha < \frac{||n+d||}{||n||} \Leftrightarrow \theta_{ij_0} > 0$$

KRI condition 2:

$$\frac{K_{i,j_0}}{K_{i,j_1}} < \frac{1}{K_{j_0,j_1}}$$

$$exp(-\frac{1}{2\sigma^2}(b^2-a^2)) < exp(\frac{1}{2\sigma^2}(b^2+a^2-2abcos\alpha)$$

$$\frac{-a^2}{2\sigma^2} < \frac{a^2-2abcos\alpha}{2\sigma^2}$$

$$2abcos\alpha < 2a^2$$

$$bcos\alpha < a$$

$$bcos\alpha > a \Leftrightarrow \theta_{ij_1} = 0$$

$$d(i,j_1)cos\alpha < d(i,j_0)$$

$$||n+d||cos\alpha > ||n|| \Leftrightarrow \theta_{ij_1} = 0$$

$$cos\alpha > \frac{||n||}{||n+d||} \Leftrightarrow \theta_{ij_1} = 0$$

$$\frac{||n||}{||n+d||} < cos\alpha < \frac{||n+d||}{||n||} \Leftrightarrow \theta_{ij_0} > 0, \theta_{ij_1} = 0$$

□

5. RESULTS

The proposed framework is applied on temperature dataset. Temperature dataset consists of daily temperature measurements by 160 US stations over 366 days ($N = 160, T = 366$) [1]. Each time-series signal is detrended by a 4th order polynomial. For these experiments, window size is chosen as 12 days (w=12).

Fixed delays of 1,2 days are considered in both kNN and NNK graphs. To have a fair comparison, all other parameters are constant. The results are visualized in Figure 2.

[1] "National Climactic Data Center," 2020. Available: ftp://ftp.ncdc.noaa.gov/pub/data/gsod

It can be seen in Fig 2 that, pruned NNK graphs, both singlenode and multinode solutions can achieve sparser representation for temperature dataset than a pruned kNN graph. Singlenode and multinode solutions look very similar after pruning, which suggests that multinode OMP-NNK can be a good alternative for OMP-NNK. There are common edges in kNN and NNK graphs, as expected. In addition, although physical distances between the stations were not used, mostly physically neighbouring nodes are connected, and edges between the closely located stations are stronger; which is consistent with our geographical knowledge. Less connections are observed in the Northwest, as also reported in [9], due to the Rocky Mountain. Predominant west-to-est direction of the wind as described in [9] is also visible in all three maps in Figure 2. Unlike the entire time series as in [9], a shorter window is used. Therefore, edges are affected more by the meteorological features than the geographical distances.

Sparsity comparison is given for all time frames, when sliding window size is 12 and sliding step is 12 in Figure 3.

6. CONCLUSION

We presented a new method to represent time-series signals using NNK algorithm. The key novelty of this work is, we have introduced a new framework for constructing time-varying graphs with NNK using VGs and including delayed representations of the signals. Then we demonstrated the proposed graph construction on temperature dataset and we compared it to one of the state-of-the-art methods. We have shown that, for the temperature dataset, where time sequences are highly correlated, NNK can construct sparser graphs that are also consistent with the geographical information.

(a) Pruned kNN (b) Pruned OMP-NNK

Fig. 2. Pruned kNN, pruned OMP-NNK, pruned Multinode OMP-NNK. w=12, $d = 0, 1, 2$. For one time window

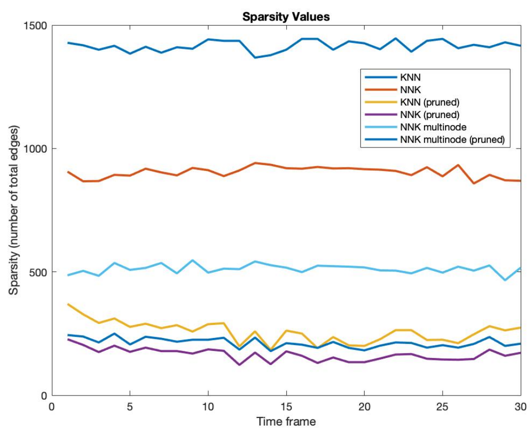


Fig. 3. Sparsity Comparison

7. REFERENCES

[1] Antonio Ortega, Pascal Frossard, Jelena Kovacevic, Jose M.F. Moura, and Pierre Vandergheynst, "Graph Signal Processing: Overview, Challenges, and Applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[2] Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst, "Learning Laplacian Matrix in Smooth Graph Signal Representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[3] Vassilis Kalofolias, "How to learn a graph from smooth signals," vol. 51, pp. 920–929, 09–11 May 2016.

[4] Hilmi E. Egilmez, Eduardo Pavez, and Antonio Ortega, "Graph Learning from Data under Laplacian and Structural Constraints," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.

[5] Weiyu Huang, Thomas A. W. Bolton, John D. Medaglia, Danielle S. Bassett, Alejandro Ribeiro, and Dimitri Van De Ville, "Graph signal processing of human brain imaging data," pp. 980–984, 2018.

[6] José Vinícius de Miranda Cardoso and Daniel P. Palomar, "Learning Undirected Graphs in Financial Markets," *Conference Record - Asilomar Conference on Signals, Systems and Computers*, vol. 2020-Novem, no. 1, pp. 741–745, 2020.

[7] Nicolas Eckert, E. Parent, R. Kies, and H. Baya, "A spatio-temporal modelling framework for assessing the fluctuations of avalanche occurrence resulting from climate change: Application to 60 years of data in the Northern French Alps," *Climatic Change*, vol. 101, no. 3, pp. 515–553, 2010.

[8] Benjamin Renoust, Gene Cheung, and Shin'Ichi Satoh, "Estimating political leanings from mass media via graph-signal restoration with negative edges," pp. 1009–1014, 2017.

[9] Jonathan Mei and José M.F. Moura, "Signal Processing on Graphs: Causal Modeling of Unstructured Data," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2077–2092, 2017.

[10] Vassilis Kalofolias, Andreas Loukas, Dorina Thanou, and Pascal Frossard, "Learning Time Varying Graphs," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2017*, pp. 2826–2830, 2017.

[11] Bo Jiang, Yuming Huang, Ashkan Panahi, Yiyi Yu, Hamid Krim, and Spencer L Smith, "Dynamic Graph Learning : A Structure-Driven Approach," pp. 1–20, 2021.

[12] Kai Qiu, Xianghui Mao, Xinyue Shen, Xiaohan Wang, Tiejian Li, and Yuantao Gu, "Time-Varying Graph Signal Reconstruction," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 870–883, 2017.

[13] Yueliang Liu, Student Member, Lishan Yang, and Student Member, "Graph Learning Based on Spatiotemporal Smoothness for Time-Varying Graph Signal," *IEEE Access*, vol. 7, pp. 62372–62386, 2019.

[14] Koki Yamada, Yuichi Tanaka, and Antonio Ortega, "Time-Varying Graph Learning with Constraints on Graph Temporal Variation," pp. 1–13, 2020.

[15] Francesco Grassi, Andreas Loukas, Nathanaël Perraudin, and Benjamin Ricaud, "A Time-Vertex Signal Processing Framework," pp. 1–16, 2017.

[16] Yueliang Liu, Lishan Yang, Kangyong You, Wenbin Guo, and Wenbo Wang, "Graph Learning Based on Spatiotemporal Smoothness for Time-Varying Graph Signal," *IEEE Access*, vol. 7, pp. 62372–62386, 2019.

[17] Wei Dong, Moses Charikar, and Kai Li, "Efficient K-nearest neighbor graph construction for generic similarity measures," *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*, pp. 577–586, 2011.

[18] Sarath Shekkizhar and Antonio Ortega, "Graph Construction from Data using Non Negative Kernel regression (NNK Graphs)," , no. 1, pp. 1–8, 2019.

[19] Sarath Shekkizhar and Antonio Ortega, "Efficient Graph Construction for Image Representation," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2020-Octob, no. 1, pp. 1956–1960, 2020.

[20] Lucas Lacasa, Bartolo Luque, Fernando Ballesteros, Jordi Luque, and Juan Carlos Nuño, "From time series to complex networks: The visibility graph," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 13, pp. 4972–4975, 2008.

[21] Ritika Jain and Ramakrishnan Angarai Ganesan, "An Efficient Sleep Scoring Method using Visibility Graph and Temporal Features of Single-Channel EEG," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2021, pp. 6306–6309, 2021.

[22] Michail Vamvakaris, Athanasios A. Pantelous, and Konstantin Zuev, "Time Series Analysis of SP 500 Index: A Horizontal Visibility Graph Approach," *SSRN Electronic Journal*, 2017.

[23] Yishu Wang, Ye Yuan, Yuliang Ma, and Guoren Wang, "Time - Dependent Graphs : Definitions , Applications , and Algorithms," *Data Science and Engineering*, vol. 4, no. 4, pp. 352–366, 2019.

[24] Zeynab Mohammadpoory, Mahda Nasrolahzadeh, and Javad Haddadnia, "Epileptic seizure detection in EEGs signals based on the weighted visibility graph entropy," *Seizure*, vol. 50, pp. 202–208, 2017.

[25] Hongduo Cao, Tiantian Lin, Ying Li, and Hanyu Zhang, "Stock Price Pattern Prediction Based on Complex Network and Machine Learning," *Complexity*, vol. 2019, 2019.