Diffusion convolution recurrent neural network – a comprehensive survey

K Tamil Selvi¹*, R Thamilselvan² and S Mohana Saranya¹

¹Assistant Professor, Dept of CSE, Kongu Engineering College, India
²Professor, Dept of MCA, Kongu Engineering College, India
*EMAIL: ktamilselvikec@gmail.com

Abstract. Graphs find its place in many applications like social network analysis, computer vision and bioinformatics. It has the ability to capture the structural relationship among the data, thus provides more insight. Graph Neural Network (GNN) has a deep learning way of analyzing the graph. The target nodes representation is obtained by iterative propagation of neighbour information until the stability is reached. Representational learning is widely used for capturing the insight of graph representation model. The complex structure of graph is hidden by representational learning results in shallow learning mechanism. Convolutional Neural Network (CNN) exploits the stationary properties and hierarchical pattern of the data which are in Euclidean space. Non-Euclidean characteristics of the graph can be captured precisely using graph convolutional neural network. In graph convolution, vertex domain is represented as aggregation of neighbour node’s information. In order to encompass the dynamics of graph, diffusion process is used, in which spatial dependency and temporal dependency are considered simultaneously. In Diffusion Convolution Recurrent Neural Network (DCRNN) uses diffusion convolution to capture spatial dependency and Gated Recurrent Unit (GRU) to capture temporal dependency. DCRNN is capable of handling long-term dependencies. In this survey, we conduct comprehensive survey on diffusion convolutional operations on graph, which is one of the most prominent deep learning models for forecasting in time series domain. First, we categorize the variants of graph convolutional models and its convolution operations on graph. Then based on application, graph convolutional models are categorized with their applications. Finally, open challenges in the area of graph convolutional network and future directions for research are discussed.

Keywords: Diffusion, Recurrent Neural Network, Convolution, Graph Neural Network, Spatio-temporal dependency

1. Introduction

Graph analysis find its place in many real world applications including social network analysis [1], traffic forecasting [2], computer vision [3], life science [4] and many more. The data can be represented as graphs and with structural information, relationship among the entities can be modelled. This model
provides the insight on the given data. The complex structure of graph limits on exploration of insight extraction from data. The non-Euclidean nature of graph data limits the learning of complex pattern in it. In order to learn complex patterns, graphs are transformed into low dimensional Euclidean space using embedding techniques. The most commonly used graph representation learning methods are node embedding and subgraph embedding [5].

Shallow learning limits the learning of complex patterns of the irregular structures like graph. To enhance the representation learning on graph, deep learning models plays an important role. CNN models can capture the high-level features using hierarchical pattern structure. It operates on data which are Euclidean or grid-like structure in nature using convolution and pooling operations. But in graph which is non-Euclidean or irregular structure, convolution and filtering operations are not viable. So many modes of convolutions on graph have been proposed with varying characteristics. Spectral graph convolution and spatial graph convolution are the two predominant ways of employing convolution in the graph network.

In spectral convolutional model, Fourier transform performs convolution operations on the graph. The main limitation of the spectral model is it takes the whole graph for processing simultaneously. And also, spectral method fits well for fixed graph models while it does not support graph dynamic. Graph convolution is modeled using aggregation of neighbour node information in spatial convolutional model. Sampling methods can be accompanied for batch processing of nodes to improve computation efficiency. Since convolution operation is done locally in spatial model, weight sharing can be easily done across locations and structures. Thus, spatial graph convolutional is more efficient than spectral graph convolutional method.

Several surveys exist on graph neural network [6] and graph convolutional network [7]. There is another type of graph convolutional network called Diffusion Convolutional Graph Neural Network (DCGNN). In DCGNN, recurrent neural network-based encoder-decoder architecture captures the temporal dependency. The input to the encoder is historical time series data. The encoder encodes the historical time series data sequence into fixed length vector. The decoder will predict the next timestamp output from the encoded vector. Fully connected network, LSTM and GRU are the available network for the design of encoder-decoder architecture. Recurrent neural network like LSTM and GRU can be used to capture the spatial dependency. In specific, the diffusion operation captures the spatial dependency which is a random walk in the graph. In this survey, existing literature on DCRNN are reviewed and recent progress are covered. The main contribution of this survey are as follows:

1. Generalized taxonomy of diffusion graph convolutional network is introduced. Then its application domains are explored with spectral and spatial convolutional models. The limitation of the convolutional graph models is explored.
2. Diffusion process and its implication in DCRNN are reviewed. Application domains and recurrent models used for diffusion process are explored.
3. And also challenges of the DCRNN models that are to be addressed are summarized and some promising future directions are highlighted.

The rest of the survey article is organized as follows. First, a general introduction on diffusion convolution network is given. Then for temporal dependency modeling, recurrent neural network with diffusion convolution operations are explored. Applications of DCRNN in highway traffic prediction, network traffic prediction and other domains are reviewed. The available metrics for evaluation of performance of the model are summarized. Following is the discussion on some research challenges and scope for future exploration in graph convolution area. Finally, the survey concludes with conclusion part.

2. Diffusion Convolution Network
Convolution Neural Network (CNN) modeled for images cannot be applied directly for irregular structures like graph. The Convolution framework for graph capture the neighbour nodes information. Graph convolution can be achieved using spectral model and spatial model. The spatial dependency
explores the neighborhood information and its influence in the given task. The diffusion process is characterized by random walk using transition matrix which can be induced by spectral convolution ChebNet [8]. Attention mechanism helps to estimate the contribution of neighbour node with respect to the node in consideration. Thus spatial correlation among different nodes in different timestamp can be captured [9].

Graph wavelet method capture the localized features of vertex and applicable for static graph network [10]. The amount of information that can be shared for diffusion of neighborhood can be controlled using gating mechanism [11]. Long-term dependencies and short term dependencies of the sequential data can be captured using LSTM and GRU respectively [12, 13]. Long sequence of data suffers from vanishing or exploding gradient which cannot be handled by recurrent neural network. These problems can be addressed by time gating [14]. Figure 1 showcases the available methods for spatial and dependency in diffusion convolutional network.

Let $G = (V, E, A)$ be a graph with $V$ be set of vertices with N nodes, $E$ be set of edges, $E \in V \times V$. $A$ be the mapping between weights and edges. The connectivity between the nodes and edges represents topology of the graph and is given by matrix. Depends on the application, matrix can be Laplacian matrix, Adjacency matrix, Random walk and Normalized matrix. The degree matrix, $D$ is given by $D(i,i) = \sum_{j=1}^{n} A(i,j)$. The graph diffusion process is given by a matrix, $S$ in Equation (1).

$$S = \sum_{k=0}^{\infty} \theta_k T^k$$  \hspace{1cm} (1)

**Figure 1.** Taxonomy of Diffusion Convolution Network

Graph diffusion is derived by diffusion of node of consideration from the starting node. The process is repeated until the transition matrix, $T$, which defines the continuously weighted graph. Here $\theta_k$ are coefficients can be heat kernel or personalized page rank. Diffusion convolutional operation given by
[15] specifies k-step diffusion process by $k^{th}$ power of transition matrix $T^k$. In general, diffusion convolution operation can be formulated as shown in Equation (2).

$$ Z(u, k, i) = \sigma \left( \Theta(k, i) \sum_{v=1}^{n} T^k(u, v) Y(v, i) \right) \quad (2) $$

Based on $T^k$, aggregation for node $u$ on the $i^{th}$ output feature is represented as $Z(u, k, i)$. $\sigma(.)$ is the non-linear activation function. Graph convolution for node $v$ at the $i^{th}$ layer is given by $Y(v, i)$. MoNet model proposed by [16] uses patch operation for diffusion process. Let $x$ be vertex of graph and $y$ be neighbour of $x$. With each neighbour node $y$, associate a pseudo-coordinate $\delta(x, y)$. The model is associated with kernel function $G_u(u)$, with $\theta$ be the learnable parameters. The generalized patch operation is given by Equation (3).

$$ D_j(x)f = \sum_{y \in N(x)} G_j( \delta(x, y))f(y) \quad j = 1, \ldots, J \quad (3) $$

The dimensionality of extracted patch is given by $J$. The process of diffusion convolution on non-Euclidean structure is formulated as

$$ (f * g)(x) = \sum_{j=1}^{J} g_j D_j(x)f \quad (4) $$

The diffusion process is modeled on propagation of traffic within the given network topology graph. It is characterized by random walk on the given graph $G$, with the restart probability $\in (0, 1)$. The state transition matrix, $D_0^{-1}$, with $D_0$ be the out-degree diagonal matrix of Graph $G$. The core process is to use diffusion convolutional operator with gated recurrent unit to learn parameters $\theta$ using back propagation process. The convolution process is repeated for $k$ steps to converge. Between the graph signal $X$ and the convolution filter $\beta_\theta$, the $k$ steps convolution process is given by [17] in Equation (5).

$$ X \odot \beta_\theta = \sum_{k=0}^{K-1} (\theta_{k,1} (D_0^{-1} A)^k + \theta_{k,2} (D_0^{-1} A^T)^k) X \quad (5) $$

The matrix multiplication operation is replaced by diffusion convolutional operation. In Equation (5), the topology information is given by $A$ and its transpose and $X$ represents information about nodes of graph. The diffusion convolution layer can be trained with a function mapping from feature matrix $F$ and output $H$. Mathematically represented as [17] Equation (6).

$$ H(q) = \sigma \left( \sum_{p=1}^{P} (X_p \odot \beta_{\theta,A}) \right) \quad \forall q \in \{1, \ldots, Q\} \quad (6) $$

3. Diffusion Convolution Recurrent Neural Network

To model temporal dependency, recurrent neural network is used. In Graph Convolution Network (GCN), the convolution layer propagates the attributes of node $h$ using the adjacency matrix $A$ with function $f(A)$. The output is given by

$$ C(A, h) = \sigma(f(A). h. W + b) \quad (7) $$

In Equation (7), $W$ is the weight matrix, $f$ is the propagation rule, $b$ is the bias. GCN fails in learning the graph moments due to permutation invariance constraints [18]. Diffusion convolutional layer can be used to learn the graph representation and trained using stochastic gradient method. In order to
accommodate long-term dependencies, GRU is used with diffusion convolutional layer. GRU recursively performs the following operations given by Equations (8,9,10,11):

\[
r(t) = \sigma(\theta_r \odot [X(t), H(t - 1)] + b_r)
\]

\[
C(t) = \tanh(\theta_c \odot [X(t), (r(t) \odot H(t - 1))] + b_c)
\]

\[
u(t) = \sigma(\theta_u \odot [X(t), H(t - 1)] + b_u)
\]

\[
H(t) = u(t) \odot H(t - 1) + (1 - u(t)) \odot C(t)
\]

Reset, update and cell gates are given by \(r, u, C\) respectively. The kernel parameters are denoted by \(\theta\) and relative bias by \(b\). \(X(t)\) represents input at time \(t\). DCGRU can be trained using backpropagation in time. It processes the input using graph convolutional layer so that GRU simultaneously receives past information from the previous time step and information about neighborhood from graph convolution.

4. Application Domain

Graph convolution network modeled with spatial and temporal dependency find its application in various domains like highway traffic prediction, network traffic prediction, wind speed forecasting, earthquake epicenter prediction and so on. Figure 2 provides the list of applications using DCRNN. In highway traffic prediction, the vehicle diffusions are modeled using convolution layer and sequence to sequence learning framework provides temporal dependency modeling [19]. Table1 provides an overview on the proposed techniques for traffic prediction in highways and roads for urban traffic management.

![Figure 2. Application domains of DCRNN](image)

For optimized resource management and to provide specified Quality-of-Service (QoS) in the network, traffic prediction plays a crucial role. Table 2 lists the review on network traffic prediction. In wireless domain, dynamic nature of the entities of the network is also considered for prediction of network traffic [11].
| Task                      | Technique                  | Spatial Dependency | Temporal Dependency | Additional features                      | Problem Addressed                                      | Dataset used                              | Ref  |
|--------------------------|----------------------------|--------------------|---------------------|------------------------------------------|--------------------------------------------------------|-------------------------------------------|------|
| Road Traffic Forecasting  | DCRNN                      | Random Walk        | GRU                 | Scheduled sampling with probability      | Complex spatial dependency and non-linear temporal     | METR-LA, PEMS-BAY                          | [8]  |
| Large Highway Traffic Forecasting | DCRNN                      | Random Walk        | Encoder – decoder architecture | Graph Partitioning with node overlapping | DCRNN experiences computational and memory bottlenecks | PeMs - California Highway Network         | [20] |
| Road Traffic Forecasting  | DCRNN                      | Diffusion process  | Fully connected network | Rank Influence Factor                    | Complicated dependencies – not captured by GCN         | METR-LA, PEMS-BAY, SZ-taxi               | [21] |
| Highway traffic prediction | Optimized GCRNN             | Graph Convolution network (GCN) | GRU                  | Normalized and parameterized graph matrices | Factors affecting traffic prediction is not clear      | PeMSD4                                    | [22] |
| Short term highway traffic forecasting | DCRNN                      | Diffusion process  | GRU                 | Graph partition-based transfer learning  | Need for large amount of data for training the model   | PeMs - California Highway Network         | [23] |
| Highway traffic prediction | Iterative Spatial-temporal DGCN | Diffusion process  | Diffusion process & state information | Neighbor and state information – diffused by vertices | Spatial and temporal features mutually dependent, separation - inaccurate results. | METR-LA, PEMS-BAY                         | [24] |
| Road Traffic Forecasting  | Spatio-Temporal Graph Convolutional Networks | Graph CNN          | Gated CNN           | Parallelization of Convolution structures | Regular convolutional and recurrent units - complex and more training time and parameters | BJer4, PeMSD7                             | [25] |
| Road Traffic Forecasting  | Dynamic Spatio-temporal Graph based CNN | Graph CNN          | Convolution         | Graph & Flow prediction stream           | Dynamics of sequential data are not considered         | METR-LA, TaxiBJ                           | [26] |
| Road traffic prediction   | Temporal Graph Convolutional Network | Graph CNN          | GRU                 | Short-term and long-term prediction tasks | Complex spatial dependencies are not considered         | SZ-taxi dataset and Los-loop dataset       | [27] |
| Road traffic prediction   | Spatio-Temporal graph attention network | Multiple flow attention heads | Temporal attention  | Spatial and temporal attention, spatial sentinel vectors. | RNNs - limitation in capturing long temporal dependencies | METR-LA, PEMS-BAY                         | [28] |
| Road traffic prediction   | A Spatio-Temporal U-Network | Spatial Pooling    | Temporal Down Sampling | Spatio-temporal pooling and unpooling operators | Unable to extract dynamic complex features from Spatio-temporal structures. | METR-LA and PeMS-M datasets.             | [29] |
Attention based mechanism can be embedded with the graph convolution for enhanced context aware prediction. Table 3 shows the extended list of DCRNN application in various mutually exclusive domain. Rough set theory proved to be the unique technique for extraction of features based on spatial and temporal dependencies [12].

**Table 2. Network Traffic Prediction**

| Task                          | Technique                  | Spatial Dependency | Temporal Dependency | Additional features                          | Problem Addressed                               | Dataset used                     | Ref               |
|-------------------------------|----------------------------|--------------------|---------------------|---------------------------------------------|-----------------------------------------------|----------------------------------|-------------------|
| Network Traffic prediction    | DCRNN                      |                    |                     | Threshold                                   | Topological properties - diffusion of traffic | Abilene Network trace            | [17]              |
| Forecast traffic in research WAN | Dynamic diffusion graph recurrent neural network | Diffusion process | GRU                 | Non-autoregressive graph-based neural network for multistep network traffic forecasting | Weighted adjacency matrix - static in DCRNN | Real time dataset - ESnet traffic traces | [30]              |
| Node Classification           | Sparse Diffusion Convolution Neural Network | GCN                | GCN                 | Pre-thresholding and Post-thresholding       | More computational and memory complexity of DCRNN | Cora Dataset                 | [31]              |
| Network-wide traffic forecasting | Graph Wavelet Gated Recurrent Neural Network | Graph Wavelet      | GRU                 | Graph wavelet weight matrix sparsity analysis and traffic hotspot detection | Absence of localization of feature extraction | Freeway Traffic Dataset, Urban Traffic Dataset | [10]              |
| Traffic flow prediction       | Spatio-Temporal Networks - Multitask Deep Learning | Fully connected network | Temporal correlation | Prediction of node flow and edge flow         | Model multiple correlation and external factors for accurate prediction | TaxiBJ, TaxiNYC          | [32]              |

Highly dynamic traffic demands can be predicted more accurately using graph convolutional network. The varying spatial-temporal pattern of the taxi demands are captured using convolutional recurrent neural network [8]. The ride-hailing demand prediction helps to reduce traffic congestion, improved vehicle utilization, low waiting time and enhanced fleet organization [19]. Graph convolution recurrent neural network is also used for accurate prediction of nodes close to the epicenter of seismic waves [21].

5. **Performance Evaluation Metrics**

The prediction or classification model built using diffusion convolution neural network can be tested and its performance can be measured using the following metrics.

1. **Root Mean Squared Error (RMSE):**

$$RMSE = \sqrt{\frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} (y_i^j - \bar{y_i^j})^2}$$
2. Mean Absolute Error (MAE):

\[ MAE = \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} |y_i^j - \hat{y}_i^j| \]

3. Accuracy:

\[ Accuracy = 1 - \frac{\|Y - Y^-\|F}{\|Y\|F} \]

4. Coefficient of Determination (R²):

\[ R^2 = 1 - \frac{\sum_{j=1}^{M} \sum_{i=1}^{N} (y_i^j - \hat{y}_i^j)^2}{\sum_{j=1}^{M} \sum_{i=1}^{N} (y_i^j - Y^-)^2} \]

5. Explained Variance Score (var):

\[ var = 1 - \frac{Var\{Y - Y^-\}}{Var\{Y\}} \]

| Task                        | Technique                                                   | Spatial Dependency | Temporal Dependency | Additional features                                                                 | Problem Addressed                                      | Dataset used                              | Ref |
|-----------------------------|--------------------------------------------------------------|--------------------|---------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------|-------------------------------------------|-----|
| Identifying epicenter of    | Gated Graph Convolutional Recurrent Neural Network            | GCN                | GCN                 | Linear shift-invariant graph filters, learnable parameters independent of graph size | Long term dependencies result in vanishing gradients    | GeoNet, Synthetic dataset                 | [11]|
| earthquake and prediction   |                                                              |                    |                     |                                                                                     |                                                        |                                           |     |
| of weather                  |                                                              |                    |                     |                                                                                     |                                                        |                                           |     |
| Correlated Time series      | Graph Attention Recurrent Neural Network                      | Multi-head         | GRU                 | p-step ahead forecasting                                                             | Adjacency matrices - static does not capture the spatial-temporal correlation | METR-LA                                   | [9] |
| forecasting                 |                                                              | attention mechanism|                     |                                                                                     |                                                        |                                           |     |
| k-step prediction           | Gated Graph Recurrent Neural Network                         | Node gating        | time gating          | Gated mechanism - Vanishing gradient in space domain                                 | Imbalance between spatial and temporal dependencies     | GeoNet, METR-LA                          | [14]|
| Taxi demand prediction      | Attention based Convolutional recurrent neural network        | Local convolution   | GRU                 | Context aware attention module                                                       | Absence of multi-view feature extraction                | NYC dataset, Chengdu                      | [33]|
| Ride Hailing Demand         | Spatio-temporal Multi-Graph Convolution Network               | Multi-graph         | contextual           | Encode the non-Euclidean correlations among regions into multiple graphs              | Only Euclidean correlations among spatially adjacent regions are considered | Real world datasets: Beijing and Shanghai | [19]|
| Forecasting                  |                                                              | convolution         | gated recurrent     |                                                                                     |                                                        |                                           |     |

Table 3. Application in other domains
In the evaluation metrics, $y_i^t$ represents the true value and $\hat{y}_i^t$ is the predicted value on $t^{th}$ time for $i^{th}$ sample. $M$ is the number of time samples and $N$ is the number of entities. Prediction errors can be estimated using RMSE and MAE. Prediction precision is given by accuracy. $R^2$ and $\text{var}$ provides correlation coefficient for input data on prediction result.

6. Comparative Analysis of different Neural Network Architecture with DCRNN

The various neural network architectures like CNN, RNN, LSTM and GRU provides variant applications in many emerging fields. Mostly for image processing and image-oriented applications, CNN is widely used, whereas for time series forecasting RNN is applied. CNN explores spatial dependencies whereas RNN explores temporal dependencies. Spatio-temporal forecasting is provided by DCRNN. Table 4 discusses the analysis of different neural network architectures and its applicational aspects.

| Neural Network Architecture | Spatial Dependency | Temporal Dependency | Convolution Operation | Domain of Application          |
|-----------------------------|--------------------|---------------------|-----------------------|-------------------------------|
| CNN                         | ✓                  | -                   | Kernel based          | Image processing              |
| RNN                         | -                  | ✓                   | Historic data         | Prediction                    |
| LSTM                        | -                  | ✓                   | Historic data with gates with more memory units | Natural Language processing (NLP) |
| GRU                         | -                  | ✓                   | Historic data with gates with less memory unit | Prediction, NLP               |
| DCRNN                       | ✓                  | ✓                   | Flow of data as diffusion process | Traffic forecasting           |

Compared to RNN, LSTM and GRU incorporates internal gates to control the flow of information. The internal memory cell provides the correlation of the current data with historic data, thus prediction accuracy is improved. All the standard architectures work well on regular or Euclidean pattern of data. However, non-Euclidean structures like graph cannot be handled by CNN, RNN and its associated architectures. DCRNN opens way for exploration of data pattern as graph for capturing the spatial and temporal dependencies.

7. Research challenges and Future directions

The spectral and spatial approaches are used for performing convolutions on graph. However, Spectral model has a hard constraint of samples to be homogeneous structure. But spatial model can accept heterogeneous structures. Before learning starts, all the heterogeneous structures are mapped to fixed size output. To overcome this, graph embed pooling is employed. For the further enhancement of diffusion convolutional network and wide applicability in broad domain, the following aspects can be further explored to the depth:

- A high-level uniform pooling method can be formulated for graph convolution network.
• The computational and memory complexities of DCRNN can be reduced by employing optimized convolution operations

8. Conclusion

Graph convolution network can operate on non-Euclidean space and find its application in various domains. In this survey, two important taxonomical aspects of the graph convolution are reviewed with respect to spatial and temporal dependencies. Diffusion operation for graph convolution in capturing spatial dimensions of the entities are explored exhaustively with its application domain also. We also provided some research challenges and scope for future directions.

References

[1] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang and Dawei Yin 2019 Graph neural networks for social recommendation The World Wide Web Conference 417-426
[2] Yuqing Wang, Dingde Jiang, Liuwei Huo and Yong Zhao 2019 A New Traffic Prediction Algorithm to Software Defined Networking. Mobile Networks and Applications 1-10.
[3] Shi Lei, Zhang Yifan, Cheng Jian and Lu Hanqing 2019 Skeleton-based action recognition with directed graph neural networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 7912-7921
[4] Ohue Masahito, Li Ryota, Yanagisawa Keisuke and Akiyama Yutaka 2019 Molecular activity prediction using graph convolutional deep neural network considering distance on a molecular graph arXiv preprint arXiv:1907.01103
[5] Hamilton William L, Ying Rex and Leskovec Jure 2017 Representation learning on graphs: Methods and applications arXiv preprint arXiv:1709.05584
[6] Wu Zonghan, Pan Shirui, Chen Fengwen, Long Guodong, Zhang Chengqi and Philip S Yu 2020 A comprehensive survey on graph neural networks IEEE Transactions on Neural Networks and Learning Systems 1-21
[7] Zhang S, Tong H, Xu J and Maciejewski R 2019 Graph convolutional networks: a comprehensive review Computational Social Networks 6(1)11.
[8] Li Yaguang, Rose Yu, Cyrus Shahabi and Yan Liu 2017 Diffusion convolutional recurrent neural network: Data-driven traffic forecasting arXiv preprint arXiv:1707.01926
[9] Cirstea, Razvan Gabriel, Yang B and Guo C 2019 Graph attention recurrent neural networks for correlated time series forecasting MileTS19@ KDD
[10] Cui, Zhiyong, Ruimin Ke, Ziyuan Pu, Xiaolei Ma and Yinhai Wang 2020 Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction Transportation Research Part C: Emerging Technologies 115 102620.
[11] Ruiz, Luana., Fernando Gama and Alejandro Ribeiro 2019 Gated graph convolutional recurrent neural networks 27th European Signal Processing Conference (EUSIPCO) IEEE 1-5
[12] Khodayar, Mahdi and Jianhui Wang 2018 Spatio-temporal graph deep neural network for short-term wind speed forecasting IEEE Transactions on Sustainable Energy. 10(2) 670-681
[13] Zhang, Zhengchao, Meng Li, Xi Lin, Yinhai Wang and Fang He 2019 Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. Transportation research part C: emerging technologies. 105 297-322.
[14] Ruiz, Luana, Fernando Gama and Alejandro Ribeiro 2020 Gated graph recurrent neural networks arXiv preprint arXiv:2002.01038
[15] Atwood, James and Don Towsley 2016 Diffusion-convolutional neural networks. in Advances in neural information processing systems 1993-2001
[16] Monti, Federico, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda and Michael Bronstein 2017 Geometric deep learning on graphs and manifolds using mixture model cnns Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 5115-5124
[17] Andreoletti, Davide, Sebastian Troia, Francesco Musumeci, Silvia Giordano, Guido Maier and Massimo Tornatore 2019 Network traffic prediction based on diffusion convolutional recurrent neural networks IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) 246-251

[18] Dehmamy, Nima, Albert Laszlo Barabasi and Rose Yu 2019 Understanding the representation power of graph neural networks in learning graph topology Advances in Neural Information Processing Systems 15413 - 15423

[19] Geng, Xu, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye and Yan Liu 2019 Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting in Proceedings of the AAAI Conference on Artificial Intelligence. 33 3656-3663

[20] Mallick, Tanwi, Prasanna Balaprakash, Eric Rask and Jane Macfarlane 2019 Graph-Partitioning-Based Diffusion Convolution Recurrent Neural Network for Large-Scale Traffic Forecasting. arXiv preprint arXiv:1909.11197

[21] Huang, Yujun, Yunpeng Weng, Shuai Yu and Xu Chen 2019 Diffusion Convolutional Recurrent Neural Network with Rank Influence Learning for Traffic Forecasting 18th IEEE International Conference on Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) 678-685

[22] Guo, Kan, Yongli Hu, Zhen Qian, Hao Liu, Ke Zhang, Yanfeng Sun, Junbin Gao and Baocai Yin 2020 Optimized graph convolution recurrent neural network for traffic prediction IEEE Transactions on Intelligent Transportation Systems

[23] Mallick, Tanwi, Prasanna Balaprakash, Eric Rask and Jane Macfarlane 2020 Transfer Learning with Graph Neural Networks for Short-Term Highway Traffic Forecasting arXiv preprint arXiv:2004.08038

[24] Xie, Yi, Yun Xiong and Yangyong Zhu 2020 ISTD-GCN: Iterative Spatial-Temporal Diffusion Graph Convolutional Network for Traffic Speed Forecasting arXiv preprint arXiv:2008.03970

[25] Yu, Bing, Haoteng Yin and Zhanxing Zhu 2017 Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting arXiv preprint arXiv:1709.04875

[26] Wang, Menglin, Baisheng Lai, Zhongming Jin, Yufeng Lin, Xiaojin Gong, Jianqiang Huang and Xiansheng Hua 2018 Dynamic spatio-temporal graph-based cnns for traffic prediction arXiv preprint arXiv:1812.02019

[27] Zhao, Ling, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng and Haifeng Li 2019 T-gen: A temporal graph convolutional network for traffic forecasting. IEEE Transactions on Intelligent Transportation Systems

[28] Park, Cheonbok, Chunggi Lee, Hyojin Bahng, Kihwan Kim, Seungmin Jin, Sungahn Ko and Jaegul Choo 2019 Sgrat: A spatio-temporal graph attention network for traffic forecasting arXiv preprint arXiv:1911.13181

[29] Yu, Bing, Haoteng Yin and Zhanxing Zhu 2019 ST-UNet: A spatio-temporal U-network for graph-structured time series modeling arXiv preprint arXiv:1903.05631

[30] Mallick, Tanwi, Mariam Kiran, Bashir Mohammed and Prasanna Balaprakash 2020 Dynamic Graph Neural Network for Traffic Forecasting in Wide Area Networks arXiv preprint arXiv:2008.12767

[31] Atwood, James, Siddharth Pal, Don Towsley and Ananthram Swami 2017 Sparse Diffusion-Convolutional Neural Networks. arXiv preprint arXiv:1710.09813

[32] Zhang, Junbo, Yu Zheng, Junkai Sun and Dekang Qi 2019 Flow prediction in spatio-temporal networks based on multitask deep learning IEEE Transactions on Knowledge and Data Engineering. 32(3) 468-478.

[33] Liu, Tong, Wenbin Wu, Yannin Zhu and Weiqin Tong 2020 Predicting taxi demands via an attention-based convolutional recurrent neural network. Knowledge-Based Systems. 206 106294