Network System Forecasting Despite Topology Perturbation

Ramzi Dakhmouche [1] **Ivan Lunati** [2] **Hossein Gorji** [2]

Abstract

Many real-world dynamical systems consist of sparsely interacting components and hence exhibit an underlying graph structure as in logistical, epidemic, and traffic networks. Yet, because of their high dimensionality, their forecasting presents major computational challenges, which are often exacerbated by measurement noise or uncertainty. We propose to partly address this problem, focusing on computationally efficient and robust forecasting under network topology perturbation. The latter represents a singular aspect of network systems entailing specific challenges, such as break in synchronization and cascade failures, hence calling for tailored forecasting algorithms. Specifically, in the limit of large number of nodes, we uncover distinct noise regimes in which the underlying system is either predictable with arbitrary accuracy, predictable only up to a limited accuracy, or entirely unpredictable. Furthermore, we propose a network forecasting approach based on a probabilistic representation of the system under study, that leverages a Bayesian coreset approximation for efficient and robust dimensionality reduction. Numerical experiments demonstrate the competitiveness of the proposed method.

1. Introduction

Network structure is ubiquitous in real-world interconnected systems, making the study of robustness of network forecasting a topic of major importance. It spans applications ranging from epidemic spread prediction to logistics and traffic network forecasting, among many others. Yet, robustness studies have so far been restricted to static network tasks(Ni et al., 2024; Zügner and Günnemann, 2019; Wang et al., 2021; Li et al., 2024; Wang et al., 2024), such as node classification, edge classification or graph regression... As for data-driven modeling of non-linear network dynamical systems, it has been mostly focused on proposing various time-varying graph neural network architectures (Liu and Zhang, 2024; Lan et al., 2022; Shao et al., 2022; Yan et al., 2024b) for settings with given noiseless topologies, besides signal recovery schemes (Sardellitti et al., 2021; Ceci and Barbarossa, 2018). Hence, we consider in this work the problem of forecasting network systems under misspecification of the topology, in a high-dimensional setting i.e. for networks with a large number of nodes. Such types of noise or uncertainty may be encountered due to several reasons, including partial observability such as in social networks, abrupt changes such as road closures due to accidents, or equipment failure in electrical grids. Addressing this problem entails designing a suitable scheme that aims to circumvent the sensitivity to topology perturbation by probabilistic estimation of the main components of the considered state system evolution, i.e. a robust model reduction scheme. Specifically, the key contributions are organized as follows:

- We provide a theoretical analysis of the impact of topology perturbation on the system trajectory, for a class of common network systems
- We propose a simple yet effective graph time series forecasting scheme combining a Bayesian coreset approximation of a Graph Convolution Network (GCN) embedding, for robustness enhancement, with a Recurrent Neural Network (RNN) to model the time evolution
- We conduct numerical experiments on real-world traffic data, demonstrating the competitiveness of the proposed approach.

2. Problem Setting

Consider a network system under random initial condition with node states $(x_1, \dots, x_n)$ such that

$$\frac{dx_i(t)}{dt} = f(x_i(t)) + \sum_{j=1}^{n} a_{ij} g(x_i(t), x_j(t)), \quad i = 1, \dots, n \tag{1}$$

[1]Institute of Mathematics, EPFL, Switzerland [2]Computational Engineering Lab, Empa, Switzerland. Correspondence to: Ramzi Dakhmouche <ramzi.dakhmouche@epfl.ch>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

where f describes the self-dynamics of x_i, g captures the interactions with its neighbors, and $A = (a_{ij})_{1\leq i,j\leq n}$ non-negative weights encoding the network topology. A large number of real-world dynamic networks can be modeled as such, including epidemic (Gao and Yan, 2022), traffic (Ding et al., 2019), and gene regulation(Aubin-Frankowski and Vert, 2020) networks, among many others (Prasse and Van Mieghem, 2022). Assume f and g to be unknown, but instead trajectories of the system nodes $(x_1(t_j), \ldots, x_n(t_j))_{j\leq m}$ are given, for different initial conditions. The task we address in this work is the forecasting of node states for unseen initial conditions as well as for $t > t_m$, as long as $(x_1(t), \ldots, x_n(t))$ is defined, under perturbation or misspecification of the network topology. We consider discrete random perturbations represented as a matrix $(\varepsilon_{i,j})_{1\leq i,j\leq n}$ of i.i.d. Bernoulli random variables with success parameter $p \in (0, 1)$.

3. Forecasting Sensitivity to Noise

In this section, we investigate the predictability of network systems with noisy topologies. Specifically, in the limit of a large number of nodes, we identify distinct noise regimes. Under negligible noise, the system remains arbitrarily predictable. With weak noise, predictability persists but is limited in accuracy. As expected, higher levels of random perturbations render the system effectively unpredictable. more precisely, we present two results analyzing discrete and continuous perturbations respectively.

Proposition 3.1. *(Discrete Noise)*
Consider a binary adjacency matrix $A = (a_{i,j})_{1\leq i,j\leq n}$ and a discrete noise matrix $\varepsilon = (\varepsilon_{i,j})_{1\leq i,j\leq n}$ of independent and identically distributed (i.i.d.) Bernoulli random variables, with success probability $p \in (0, 1)$. Assume f and g are continuously differentiable and the system trajectory to be supported on a space of lower dimension $m \ll n$, where n is the number of nodes. Then, denoting by $y_1, \ldots, y_m$ the spatial modes and $(x_\varepsilon(t))_{t\in T}$ the trajectory over a compact time domain of the perturbed system

$$\frac{dx_i(t)}{dt} = f(x_i(t)) + \sum_{j=1}^{n} (a_{ij} + \varepsilon_{i,j})\ g(x_i(t), x_j(t)) \quad (2)$$

- *If $p \leq \frac{(\max_k \|y_k\|_\infty)^{-1}}{n^{1+\alpha}}$ with $\alpha > 0$, then $\lim_{n\to+\infty} \mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|] = 0$*
- *If $p = \frac{(\max_k \|y_k\|_\infty)^{-1}}{n}$, then there exists $M > 0$, such that for all $n \geq 2$, $\mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|] \leq M$*
- *If $p > \frac{(\max_k \|y_k\|_\infty)^{-1}}{n}$ and $\|g\|_\infty > \delta$, then for all $n > 2$, $\mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|] > \delta$.*

Proof. The proof is postponed to Appendix A.1. □

Proposition 3.2. *(Gaussian Noise)*
Consider a Gaussian noise matrix $\varepsilon = (\varepsilon_{i,j})_{1\leq i,j\leq n}$ of i.i.d. standard Gaussian random variables, with variance $\sigma^2 > 0$. Assume f and g are continuously differentiable and the system trajectory to be supported on a space of lower dimension $m \ll n$, where n is the number of nodes. Then, denoting by denoting by $y_1, \ldots, y_m$ the spatial modes and $(x_\varepsilon(t))_{t\in T}$ the trajectory over a compact time domain of the perturbed system (2), we have

- *If $\sigma \leq \frac{(\max_k \|y_k\|_\infty)^{-1}}{n^{1+\alpha}}$ with $\alpha > 0$, then $\lim_{n\to+\infty} \mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|] = 0$*
- *If $\sigma = \frac{(\max_k \|y_k\|_\infty)^{-1}}{n}$, then there exists $M > 0$, such that for all $n \geq 2$, $\mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|] \leq M$*
- *If $\sigma > \frac{(\max_k \|y_k\|_\infty)^{-1}}{n}$ and $\|g\|_\infty > \delta$, then for all $n > 2$, $\mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|] > \delta$.*

Proof. The proof is postponed to Appendix A.2. □

Remark.
Given the invariance encoded in network systems, the set of their node trajectories typically lives (approximately) on a space of much lower dimensionality than the number of nodes. In particular, several common network models satisfy this property (Prasse and Van Mieghem, 2022).

4. Network Coreset Forecasting

In order to achieve scalability and robustness in network forecasting, we propose to evolve in time only a subset of node embeddings. Hence, the key problem that we aim to solve is how to efficiently down-sample the processed node features for accurate prediction, while keeping sensitivity to network topology perturbation as small as possible. Classical approaches such as max-pooling (Hamilton et al., 2017) or low-rank approximations (Savas and Dhillon, 2011) either suffer from poor performance or high computational cost. Indeed, the nodes with the highest values at different screen-shots, might be very different from the ones that should be tracked to approximate the system trajectory. As for combinations of nodes with high eigenvalues, they can be quite sensitive to random perturbations, in addition to being costly to compute i.e. they require $O(n^3)$ operations. Consequently, we propose to leverage a probabilistic representation of the network time series by selecting the nodes which approximate the best, the distribution of the whole system trajectory. Specifically, we consider a Bayesian setting where the network time series constitute realizations of a given distribution and identify a subset of nodes $\{i_k, k \leq m\}$ such that the posterior given node embeddings

$\{i_k, k \leq m\}$ is the closest to the full posterior. That is, we identify a Bayesian coreset (Campbell and Broderick, 2019; Huggins et al., 2016). In the following, we describe with greater detail the different components of the proposed method, illustrated in figure **??** and report a pseudo-code of the proposed algorithm.

4.1. Bayesian Coreset Approximation

Bayesian coresets have been developed to reduce the cost of Bayesian inference with a large amount of data, without compromising accuracy (Campbell and Broderick, 2019). More precisely, considering a data set $(x_i)_{i\leq n}$ of n observations, a likelihood $p(x_i|\theta)$ for each observation given the parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, and a prior density π_0 on Θ, the Bayesian posterior is given, by $\pi(\theta) := \frac{1}{Z}\exp\left(\mathcal{L}(\theta)\right)\pi_0(\theta)$, where the log-likelihood $\theta \mapsto \mathcal{L}(\theta)$ and the marginal likelihood Z are defined by

$$\mathcal{L}(\theta) := \sum_{i=1}^{n} \mathcal{L}_i(\theta), \text{ s.t. } \mathcal{L}_i(\theta) := \log p(x_i|\theta)$$

and

$$Z := \int \exp(\mathcal{L}(\theta))\pi_0(\theta)\mathrm{d}\theta.$$

The aim of the Bayesian coreset framework is then to find a set of weights $(w_i)_{i\leq n}$ such that

$$\min_{w\in\mathbb{R}^n} \left\| \mathcal{L} - \sum_{i=1}^{n} w_i \mathcal{L}_i \right\|_{\pi,\mathcal{L}} \quad \text{with} \quad w \geq 0,$$

and

$$\sum_{i=1}^{n} \mathbb{I}_{[w_i>0]} \leq m,$$

where $\|\cdot\|_{\pi,\mathcal{L}}$ is a functional norm that involves a scaling by the full likelihood $\mathcal{L}$ and the posterior π. This problem can be solved by a Frank-Wolfe type algorithm (Jaggi, 2013; Campbell and Broderick, 2019). We refer to (Campbell and Broderick, 2019) for more details. In our setting, we consider the (embedded) trajectory $(x_i(t_1), \ldots, x_i(t_\ell))$ of each node $i \leq n$ to be a multi-dimensional realization of a probability distribution, where ℓ is predicted sequence length, and we look for the subset of nodes that summarizes the best the whole network trajectory-segment. To simplify the algorithm and reduce the computational burden, we approximate the data distribution as Gaussian and adopt a Gaussian prior in the implementation.

Remark. (Computational Complexity)
The computational complexity of the Bayesian coreset reduction component may scale like $O(n^2)$ which is already an improvement from classical low-rank approximations which scales like $O(n^3)$. Hence, to improve scalability further, we use *random projections* for the computation of the norms, following (Campbell and Broderick, 2019). This results in a complexity of $O(nq)$, where q is the dimension of the projection space.

4.2. Graph Structure Embedding

Following previous work (Liu and Zhang, 2022; 2024), we extract node embeddings incorporating the graph information with a graph convolution network (Kipf and Welling, 2017). Specifically, we train a GCN encoder-decoder network, where the encoder performs message passing between nodes leveraging the network topology information and the decoder performs a diffusion of the updated states of the coreset selected nodes. For that matter, we first compute the coresets -kept nodes in reduced space, of the training data considering trajectories divided according to forecasting range target -forecasting sequence length, then we train the GCN decoder to estimate the updates of nodes discarded during the space reduction. Once this graph autoencoder is trained, we use its encoder to embed the node features before recomputing a new set of coresets that will be evolved in time by training a RNN, as we describe in the next subsection.

4.3. Latent Representation Temporal Evolution

Once the embedded node subsets are extracted, we simply train a long short-term memory (LSTM) network (Cho et al., 2014) to model the time evolution. As a result of the reduction in dimensionality, the temporal evolution computational cost is considerably reduced allowing for scalability to large networks. Note that such a reduction is necessary since the complexity of effective sequence-to-sequence models (e.g. RNNs, Transformers) is in $O(n^2)$, where n is the dimension of the feature space. In the case of RNNs, that comes from the standard choice of a hidden-state dimension that scales similarly to input space dimension, resulting in an output scaling in $O(n^2)$. Once

Figure 1. Time Evolution in Reduced Space

the predictions in reduced space are obtained, the node updates are diffused using the GNN decoder. Note that, given that probabilistic formulation of the model reduction component, the number of nodes which are kept slightly varies across trajectory segments. Hence, we consider the size of largest coreset as RNN input dimension, and pad the RNN input with the mean encoded value for smaller coresets.

we present numerical results of the network coreset

forecasting approach on a real-world traffic forecasting dataset, made publicly available by (Guo et al., 2021). Traffic forecasting represents a major practical problem that requires robust forecasting schemes, given the different noise sources leading to topology changes that it can be subject to, such as weather incidents, traffic accidents as well as closure due to road maintenance. We note from

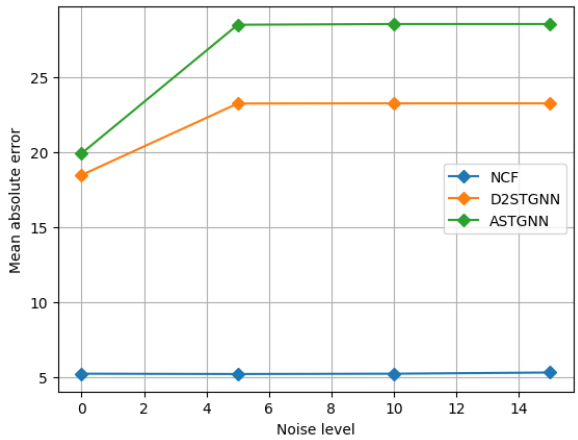

Figure 2. Test error evolution for PEMS04 dataset - 30% corrupted test data with discrete noise

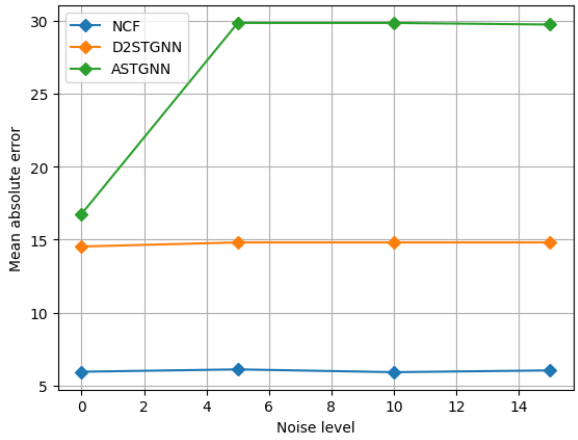

Figure 3. Test error evolution for PEMS08 dataset - 30% corrupted test data with discrete noise

figures 2 and 3 that NCF significantly outperforms the competitor methods. This is partly due to the highly noisy nature of the considered dataset, illustrated in figure 4, as is often the case in real-world settings.

5. Discussion

5.1. Limitations & Outlook

Network Coreset Forecasting performs competitively, but also raises limitations that point to opportunities for future research. First, it relies on a two-stage training approach

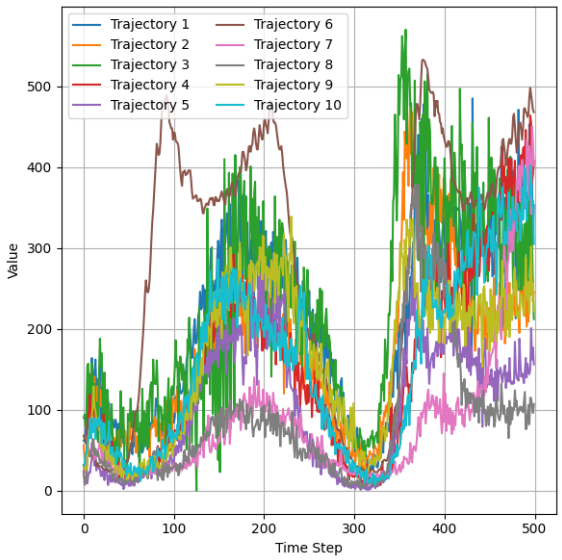

Figure 4. State trajectory evolution from the PEMS04 dataset

as opposed to an end-to-end training formulation. Besides, since it relies on a probabilistic formulation of the model reduction component, it might under-perform under perfect knowledge of the topology, as compared to other approaches. However, that is not a surprising aspect given that all methods are bound by the accuracy-robustness trade-off (Owhadi et al., 2015). More generally, we have mostly focused in this study on *independent identically distributed* (i.i.d.) noise, hence extending the analysis to account for structured or correlated noise represents a promising direction for future work.

5.2. Conclusion

In this work, we analyzed the predictability of a common class of network systems under topology perturbation. While we focused on random perturbations, we also explored structured changes via sparsification of the network topology. Furthermore, we proposed a novel probabilistic network forecasting scheme, employing a Bayesian coreset approach in combination with GNN encoding and RNN temporal modeling, to robustly capture the principal components of the underlying network signal. We demonstrated the competitiveness of the proposed approach with respect to the state-of-the-art, for different types of noise. We believe our results constitute an important initial step towards robust modeling of real-world network systems.

References

Pierre-Cyril Aubin-Frankowski and Jean-Philippe Vert. Gene regulation inference from single-cell rna-seq data with linear differential equations and velocity inference.

Bioinformatics, 36(18):4774–4780, 2020.

Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

Aleksandar Bojchevski and Stephan Günnemann. Certifiable robustness to graph perturbations. *Advances in Neural Information Processing Systems*, 32, 2019.

Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12):1727–1738, 2021.

Trevor Campbell and Tamara Broderick. Automated scalable bayesian inference via hilbert coresets. *Journal of Machine Learning Research*, 20(15):1–38, 2019.

Elena Ceci and Sergio Barbarossa. Small perturbation analysis of network topologies. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4194–4198. IEEE, 2018.

Kyunghyun Cho, Bart Van Merrienboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *EMNLP*, 2014.

Rui Ding, Norsidah Ujang, Hussain Bin Hamid, Mohd Shahrudin Abd Manan, Rong Li, Safwan Subhi Mousa Albadareen, Ashkan Nochian, and Jianjun Wu. Application of complex networks theory in urban traffic network researches. *Networks and Spatial Economics*, 19:1281–1317, 2019.

Negin Entezari, Saba A Al-Sayouri, Amirali Darvishzadeh, and Evangelos E Papalexakis. All you need is low (rank) defending against adversarial attacks on graphs. In *Proceedings of the 13th international conference on web search and data mining*, pages 169–177, 2020.

Ting-Ting Gao and Gang Yan. Autonomous inference of complex network dynamics from incomplete and noisy data. *Nature Computational Science*, 2(3):160–168, 2022.

Simon Geisler, Tobias Schmidt, Hakan Şirin, Daniel Zügner, Aleksandar Bojchevski, and Stephan Günnemann. Robustness of graph neural networks at scale. *Advances in Neural Information Processing Systems*, 34:7637–7649, 2021.

Shengnan Guo, Youfang Lin, Huaiyu Wan, Xiucheng Li, and Gao Cong. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 34 (11):5415–5428, 2021.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Jan S Hesthaven and Tim Warburton. *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*. Springer Science & Business Media, 2007.

Jonathan Huggins, Trevor Campbell, and Tamara Broderick. Coresets for scalable bayesian logistic regression. *Advances in neural information processing systems*, 29, 2016.

Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *International conference on machine learning*, pages 427–435. PMLR, 2013.

Yoshihito Kazashi, Fabio Nobile, and Fabio Zoccolan. Dynamical low-rank approximation for stochastic differential equations. *Mathematics of Computation*, 94(353): 1335–1375, 2025.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Shiyong Lan, Yitong Ma, Weikang Huang, Wenwu Wang, Hongyu Yang, and Pyang Li. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In *International conference on machine learning*, pages 11906–11917. PMLR, 2022.

Kuan Li, YiWen Chen, Yang Liu, Jin Wang, Qing He, Minhao Cheng, and Xiang Ao. Boosting the adversarial robustness of graph neural networks: An ood perspective. In *The Twelfth International Conference on Learning Representations*, 2024.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, Anima Anandkumar, et al. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Jörg Liesen and Zdenek Strakos. *Krylov subspace methods: principles and analysis*. Numerical Mathematics and Scientific Computation, 2013.

Aoyu Liu and Yaying Zhang. Spatial-temporal interactive dynamic graph convolution network for traffic forecasting. *arXiv preprint arXiv:2205.08689*, 2022.

Aoyu Liu and Yaying Zhang. Spatial–temporal dynamic graph convolutional network with interactive learning for traffic forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 2024.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Franco Manessi, Alessandro Rozza, and Mario Manzo. Dynamic graph convolutional networks. *Pattern Recognition*, 97:107000, 2020.

Eleonora Musharbash, Fabio Nobile, and Eva Vidličková. Symplectic dynamical low rank approximation of wave equations with random parameters. *BIT Numerical Mathematics*, 60(4):1153–1201, 2020.

Dandan Ni, Sheng Zhang, Cong Deng, Han Liu, Gang Chen, Minhao Cheng, and Hongyang Chen. Exploring robustness of gnn against universal injection attack from a worst-case perspective. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 1785–1794, 2024.

Houman Owhadi, Clint Scovel, and Tim Sullivan. On the brittleness of bayesian inference. *siam REVIEW*, 57(4): 566–582, 2015.

Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 5363–5370, 2020.

Bastian Prasse and Piet Van Mieghem. Predicting network dynamics without requiring the knowledge of the interaction graph. *Proceedings of the National Academy of Sciences*, 119(44):e2205517119, 2022.

Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo. Online learning of time-varying signals and graphs. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5230–5234. IEEE, 2021.

Berkant Savas and Inderjit S Dhillon. Clustered low rank approximation of graphs in information science applications. In *Proceedings of the 2011 SIAM International Conference on Data Mining*, pages 164–175. SIAM, 2011.

Peter J Schmid. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*, 54(1):225–254, 2022.

Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured sequence modeling with graph convolutional recurrent networks. In *Neural information processing: 25th international conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, proceedings, part I 25*, pages 362–373. Springer, 2018.

Zezhi Shao, Zhao Zhang, Wei Wei, Fei Wang, Yongjun Xu, Xin Cao, and Christian S Jensen. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *Proceedings of the VLDB Endowment*, 15(11): 2733–2746, 2022.

Binghui Wang, Jinyuan Jia, Xiaoyu Cao, and Neil Zhenqiang Gong. Certified robustness of graph neural networks against adversarial structural perturbation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1645–1653, 2021.

Bohao Wang, Jiawei Chen, Changdong Li, Sheng Zhou, Qihao Shi, Yang Gao, Yan Feng, Chun Chen, and Can Wang. Distributionally robust graph-based recommendation system. In *Proceedings of the ACM Web Conference 2024*, pages 3777–3788, 2024.

Huijun Wu, Chen Wang, Yuriy Tyshetskiy, Andrew Docherty, Kai Lu, and Liming Zhu. Adversarial examples for graph data: deep insights into attack and defense. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 4816–4823, 2019.

Yi Yan, Jiacheng Hou, Zhenjie Song, and Ercan Engin Kuruoglu. Signal processing over time-varying graphs: A systematic review. *arXiv preprint arXiv:2412.00462*, 2024a.

Yi Yan, Jiacheng Hou, Zhenjie Song, and Ercan Engin Kuruoglu. Signal processing over time-varying graphs: A systematic review. *arXiv preprint arXiv:2412.00462*, 2024b.

Han Yang, Binghui Wang, Jinyuan Jia, et al. Gnncert: Deterministic certification of graph neural networks against adversarial perturbations. In *The Twelfth International Conference on Learning Representations*, 2024a.

Yuxin Yang, Qiang Li, Jinyuan Jia, Yuan Hong, and Binghui Wang. Distributed backdoor attacks on federated graph learning and certified defenses. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 2829–2843, 2024b.

Xiang Zhang and Marinka Zitnik. Gnnguard: Defending graph neural networks against adversarial attacks. *Advances in neural information processing systems*, 33: 9263–9275, 2020.

Daniel Zügner and Stephan Günnemann. Certifiable robustness and robust training for graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 246–256, 2019.

A. Related Works

Time-Varying Graph Neural Networks.
In order to leverage the expressive power of graph neural networks (Kipf and Welling, 2017) for graph time series forecasting, several architectures have been proposed. First approaches were based on evolving the graph extracted features using recurrent neural networks (Manessi et al., 2020; Seo et al., 2018) or leveraging a RNN to evolve the weights of a graph convolution network (Pareja et al., 2020). More recently, various adaptive attention-based mechanisms for spatial or temporal modeling have been proposed (Guo et al., 2021; Yan et al., 2024a). For instance, (Lan et al., 2022) combine a multi-order Chebyshev polynomial GCN with an adaptive self-attention mechanism to leverage the dynamic spatial correlation within multi-scale neighborhoods, whereas (Shao et al., 2022) combines a self-attention layer with a GRU (Cho et al., 2014) to model the non-diffusive component in traffic forecasting. We refer to (Yan et al., 2024a) for an extensive review. Yet, these approaches assume the graph to be given. In contrast, we propose a forecasting scheme that is designed to be robust to graph topology perturbation or misspecification.

Robustness of Graph Neural Networks. For static tasks such as node classification or edge classification, the robustness of GNNs has been extensively studied (Zügner and Günnemann, 2019; Wang et al., 2021; Bojchevski and Günnemann, 2019; Yang et al., 2024a;b; Geisler et al., 2021), mostly focusing on adversarial defense to node or edge attacks. Specifically, (Entezari et al., 2020; Wu et al., 2019) propose pre-processing techniques to overcome adversarial perturbations via low-rank approximation of the graph adjacency matrix or gradient averaging, while (Zhang and Zitnik, 2020) propose a mechanism to tackle adversarial training by assigning higher weights to edges connecting similar nodes. On the other hand, (Wang et al., 2021; Zügner and Günnemann, 2019) propose certifiable defenses against bounded adversarial attacks, via convex relaxation of the robust optimization target or random smoothing. More recently, (Yang et al., 2024a) proposed the first deterministic certificate defense leveraging a majority vote among sub-graphs defined via an unperturbed hash function. Nonetheless, most of these approaches are too conservative for non-adversarial noise settings and do not scale to time-varying graphs, given the exponential explosion in the number of sub-graphs to consider when a temporal component is introduced.

Reduced Order Modeling.
Model reduction of parametric evolution equations governing physical systems has been extensively studied (Benner et al., 2015), given the high computational cost associated with full-scale resolution. Specifically, two main families of methods have emerged: model-based projections and data-driven surrogates. Model-based approaches leveraging the structure of the equations include Galerkin projections (Hesthaven and Warburton, 2007), Krylov subspace methods (Liesen and Strakos, 2013), and dynamic low-rank approximations (Kazashi et al., 2025; Musharbash et al., 2020). They are based on different identification schemes of the underlying low-dimensional manifold capturing most of the variability of the system. More recently, several deep learning-based approaches have demonstrated competitive performance, among which Dynamic Mode Decomposition (Schmid, 2022), Physics-Informed neural networks (Cai et al., 2021) and Neural Operator Learning (Lu et al., 2021; Li et al., 2021) are most notable. Nonetheless, most work has been restricted to low-dimensional noiseless dynamical systems. Alternatively, we consider noisy high-dimensional network systems.

B. Proofs

B.1. Proof of proposition 1

Recall that we would like to show that a Bernoulli perturbed network system features different behaviors depending the success parameter $p \in (0,1)$ of the Bernoulli noise. Specifically,

$$p \leq \frac{(\max_k \|y_k\|_\infty)^{-1}}{n^{1+\alpha}} \text{ with } \alpha > 0 \quad \Rightarrow \quad \lim_{n\to+\infty} \mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|] = 0$$

Denote by Y the matrix of reduced space basis of the system trajectory, and E the noise matrix i.e.

$$Y = \begin{pmatrix} y_1^\top \\ \vdots \\ y_m^\top \end{pmatrix}, \quad \text{and} \quad E = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \text{ with } \quad \varepsilon_i = (\varepsilon_{i,j})_{j\leq n}$$

Given a perturbed adjacency matrix $\hat{A} = A + E$, note that we have for $k \leq m$,

$$\|\hat{A}y_k - Ay_k\|_1 \leq \|Ay_k - Ay_k\|_1 + \|Ey_k\|_\infty = \|Ey_k\|_1$$

Furthermore, setting without loss of generality $x(1) = 0$ and $T = (1, 2)$, we have

$$\begin{aligned}
\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|_1 &= \sup_{t\in T} \left\| \int_1^t \sum_{i,j} \varepsilon_{i,j} g(x_i(s), x_j(s))\, ds \right\|_1 \\
&\leq \sup_{t\in T} \int_1^t \left\| \sum_{i,j} \varepsilon_{i,j} g(x_i(s), x_j(s)) \right\|_1 ds \\
&\leq \int_1^2 \left\| \sum_{i,j} \varepsilon_{i,j} g(x_i(s), x_j(s)) \right\|_1 ds
\end{aligned}$$

leading to

$$\mathbb{E}\left[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|_1\right] \leq \mathbb{E}\left[\int_1^2 \left\| \sum_{i,j} \varepsilon_{i,j} g(x_i(s), x_j(s)) \right\|_1 ds\right]$$

By regularity of g, x, x_ε and Fubini theorem, we get

$$\mathbb{E}\left[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|_1\right] \leq \int_1^2 \mathbb{E}\left\| \sum_{i,j} \varepsilon_{i,j} g(x_i(s), x_j(s)) \right\|_1 ds$$

Hence, by proposition 2 in section C of (Prasse and Van Mieghem, 2022), it is enough to show for each $k \leq m$ that

$$\mathbb{E}[\|Ey_k\|_1] \underset{n\to+\infty}{\longrightarrow} 0$$

That is equivalent to

$$\forall\, i \geq 1, \quad \mathbb{E}[\|Y\varepsilon_i\|_1] \underset{n\to+\infty}{\longrightarrow} 0$$

However, by Holder inequality

$$\mathbb{E}[\|Y\varepsilon_i\|_1] \leq np(\max_k \|y_k\|_\infty) \leq \frac{1}{n^\alpha} \underset{n\to+\infty}{\longrightarrow} 0$$

And, we can also have the result for $\|\cdot\|_2$ by the fact that $\|\cdot\|_2 \leq \|\cdot\|_1$.

Similar reasoning leads to the result of the case $p = \frac{(\max_k \|y_k\|_\infty)^{-1}}{n}$.

Last, for the case $p > \frac{(\max_k \|y_k\|_\infty)^{-1}}{n}$ and $\|g\|_\infty > \delta$, let

$$\tau = \inf\{s > 1;\ |g(x_1(s), x_2(s))| > \delta\}$$

Then, $1 < \tau < +\infty$, thanks to the regularity of g and the fact that the considered differential system is autonomous. Hence,

$$\begin{aligned}
\mathbb{E}[\sup_{t\in T} \|x(t) - x_\varepsilon(t)\|_1] &\geq \mathbb{E}\left[\int_1^\tau \sum_{i,j} \varepsilon_{i,j} |g(x_i(s), x_j(s))|\, ds\right] \\
&\geq \tau\delta \sum_{i,j} \mathbb{E}[\varepsilon_{i,j}] \\
&\geq \delta(\max_k \|y_k\|_\infty)^{-1} \\
&\geq \delta.
\end{aligned}$$

B.2. Proof of proposition 2

The proof proceeds similarly as for proposition 1, except this time the noise is not non-negative. Hence, using the properties of the absolute value and Gaussian distributions, we get

$$\mathbb{E}[\|Y\varepsilon_i\|_1] \leq \left(\sum_j |\varepsilon_{i,j}|\right)(\max_k \|y_k\|_\infty) = n\sigma(\max_k \|y_k\|_\infty) \leq \frac{1}{n^\alpha} \underset{n\to+\infty}{\longrightarrow} 0$$

since $\mathbb{E}|\varepsilon_{i,j}| = \sigma$. The remaining cases can be shown in an analogus way.