

CONFORMAL INFERENCE FOR TIME SERIES OVER GRAPHS

Sonakshi Dua[‡], Gonzalo Mateos[†], and Sundeep Prabhakar Chepuri[‡]

[‡]Indian Institute of Science, Bangalore, India

[†]University of Rochester, Rochester, NY, USA

ABSTRACT

Trustworthy decision making in networked, dynamic environments calls for innovative uncertainty quantification substrates in predictive models for graph time series. Existing conformal prediction (CP) methods have been applied separately to multivariate time series and static graphs, but they either ignore the underlying graph topology or neglect temporal dynamics. To bridge this gap, here we develop a CP-based sequential prediction region framework tailored for graph time series. A key technical innovation is to leverage the graph structure and thus capture pairwise dependencies across nodes, while providing user-specified coverage guarantees on the predictive outcomes. We formally establish that our scheme yields an exponential shrinkage in the volume of the ellipsoidal prediction set relative to its graph-agnostic counterpart. Using real-world datasets, we demonstrate that the novel uncertainty quantification framework maintains desired empirical coverage, while achieving markedly smaller (up to 80% reduction) prediction regions than existing approaches.

Index Terms— Conformal prediction, Diffusion, Graph time series, Graph filters, Uncertainty quantification

1. INTRODUCTION

Graph time series, a time-evolving multivariate signal defined on the nodes of a graph, arise in diverse application domains such as electricity demand forecasting, network anomaly detection, traffic flow prediction, and epidemic modelling, to name a few. State-of-the-art methods developed for graph time-series forecasting include graph neural networks with recurrent neural networks (GNN-RNN) [1–3] or graph transformers (GTs) [4, 5]. Such neural models generate point predictions, and they typically lack uncertainty quantification (UQ), which is a critical component for trustworthy and robust decision making in complex, dynamic environments.

Conformal prediction (CP) has emerged as a widely applicable, distribution-free UQ paradigm in modern machine learning [6]. Given a pre-trained predictive model, CP constructs a model-agnostic wrapper using a calibration dataset and a suitable non-conformity score, to estimate efficient prediction sets for unseen test data. Interestingly, these prediction sets offer user-prescribed coverage guarantees without imposing any strong distributional assumptions, but require exchangeability between the calibration and test data. Unlike for i.i.d. samples that are customary in statistical learning, the exchangeability assumption is often not tenable for time series and graph data (where sample ordering or node labeling conveys valuable inductive bias).

Related Works. To circumvent said exchangeability requirement, [7] extends the vanilla CP to non-exchangeable data through the notion of weighted exchangeability. Further, sequential predictive conformal inference (SPCI) [8] as well as sequential distribution-free ensemble batch prediction intervals (EnbPI) [9] have been pro-

posed for time series, wherein prediction intervals are constructed sequentially without using any calibration data. For the multivariate setting, [10] developed ellipsoidal sets for prediction regions. CP methods have been developed for *static* graph data as well. Diffused adaptive prediction sets (DAPS) [11] advocates diffusing the node-wise conformity scores over the graph, to incorporate neighborhood information. The conformalized graph neural network (ConfGNN) [12] provides coverage guarantees for transductive node classification and regression in static graphs. Exchangeability holds for base models that are permutation invariant, which means that the model output/non-conformity score is invariant to the ordering of calibration and test nodes on the graph. However, this does not apply to graph time series, where signals evolve dynamically and the train–test split is determined by temporal ordering rather than by node subset selection.

Proposed Approach and Contributions. We study time-evolving graph signals for which neither exchangeability nor permutation invariance hold, thus the need arises for a new framework that integrates CP with graph-structured time series prediction pipelines. This paper addresses the gap of constructing valid and sufficiently tight prediction regions for graph time-series forecasting, where both the graph topology and temporal dependencies play a key role. We propose a novel methodology to compute *graph-aware* nonconforming scores, in which the residuals obtained from a base model are first filtered using a graph convolutional operator. Drawing inspiration from [10], we form ellipsoids that contain a fraction of the graph-filtered residuals, and a quantile regressor is used to predict the quantiles of the unseen data. At the heart of the proposed approaches lies the homophily assumption, i.e., neighboring nodes tend to have similar residuals. Graph filtering fruitfully incorporates local structural information, which provably yields an exponential shrinkage in the volume of the ellipsoidal prediction set relative to its graph-agnostic counterpart, without sacrificing coverage.

In summary, the major contributions of the paper are as follows.

- We develop a novel CP framework for graph time series, which exploits homophily via graph-filtered residual scores.
- Leveraging graph structure leads to efficiency. We establish exponential shrinkage in the volume of the ellipsoidal prediction set, while guaranteeing a user-prescribed coverage.
- Comprehensive tests using real-world data corroborate that our UQ scheme attains the target coverage, with markedly smaller ellipsoids than a graph-agnostic CP baseline.

2. PRELIMINARIES

Notation. Throughout the paper, we use lowercase (uppercase) boldface letters to denote vectors (respectively, matrices) and calligraphic letters to denote sets. We denote an ellipsoid with radius r

and center $\mathbf{c} \in \mathbb{R}^N$ as $\mathcal{B}(r, \mathbf{c}, \mathbf{\Sigma}) := \{\mathbf{x} \in \mathbb{R}^N \mid (\mathbf{x} - \mathbf{c})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{c}) \leq r\}$, where $\mathbf{\Sigma}$ is a positive definite matrix that determines the shape of the ellipsoid. The volume of the ellipsoid follows $\text{Vol}(\mathcal{B}(r, \mathbf{c}, \mathbf{\Sigma})) \propto r^{N/2} \sqrt{\det(\mathbf{\Sigma})}$. Next, we provide a brief background of conformal inference, particularly for regression tasks.

Conformal Inference. CP is a distribution-free method to build a wrapper around a pre-trained model that forms prediction sets to guarantee finite-sample coverage. Given a base model (e.g., a classifier or regressor) and a non-conformity score function, CP constructs a region $\mathcal{C}(1 - \alpha)$ that contains the true prediction with a user-specified probability $1 - \alpha$, i.e., for a feature \mathbf{x} and true label \mathbf{y} , we have the following conditional coverage guarantee

$$\mathbb{P}(\mathbf{y} \in \mathcal{C}(1 - \alpha) \mid \mathbf{x}) \geq 1 - \alpha.$$

The size of the prediction set determines the *inefficiency* of the model, with a smaller prediction set size indicating more confidence. The uncertainty prediction sets, as small as possible, are constructed to include all potential observations whose non-conformity scores are within the empirical quantiles of the calibration data, which is assumed to be exchangeable with the test data. Calibration data can be held out from the training set, or, in the case of sequential data, the training data can be repurposed with temporal windowing.

Graph Filters. Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of N nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. The graph structure is captured by an $N \times N$ symmetric matrix having non-zero (i, j) -th entry whenever nodes i and j are connected. Examples of such matrices are the adjacency matrix $\mathbf{A} \in \mathbb{R}_+^N$ and the normalized adjacency matrix $\mathbf{D}^{-1}\mathbf{A} \in \mathbb{R}_+^N$, to name a few. Here, \mathbf{D} is the diagonal nodal degree matrix. Note that the eigenvalues of $\mathbf{D}^{-1}\mathbf{A}$, denoted by λ_i , are bounded as $\lambda_i \in [0, 1]$. Henceforth, we consider $\mathbf{D}^{-1}\mathbf{A}$ to represent \mathcal{G} .

A *graph convolutional filter* is defined as the matrix polynomial $\mathbf{H} = \sum_{l=0}^{L-1} \tau_l (\mathbf{D}^{-1}\mathbf{A})^l$, where $\{\tau_l\}_{l=0}^{L-1}$ denote the filter coefficients and L is the filter order. In this paper, we adopt a simple first-order graph filter of the form $\mathbf{H} = (1 - \tau)\mathbf{I} + \tau\mathbf{D}^{-1}\mathbf{A}$, with $\tau \in [0, 1]$. Let $\mathbf{e} \in \mathbb{R}^N$ be a *graph signal*, i.e., a signal supported in \mathcal{V} where e_i is the value at node $i \in \mathcal{V}$. Then, we can diffuse \mathbf{e} over the graph by computing $\mathbf{H}\mathbf{e}$ via one-hop local aggregations.

3. PROBLEM STATEMENT

We consider a forecasting problem from graph time-series data. The graph \mathcal{G} is static, i.e., the graph structure (nodes and edges) remains fixed, but the graph signals (or nodal features) evolve over time.

Let us denote the graph time-series data at time t as $\mathcal{G}_t = (\mathbf{x}_t, \mathcal{G})$, which arrives sequentially for $t \geq 1$. Here, $\mathbf{x}_t \in \mathbb{R}^N$ contains the graph signal and $\mathbf{y}_t \in \mathbb{R}^N$ is the prediction target. For instance, for a one-step ahead predictor, we use $\mathbf{y}_t = \mathbf{x}_{t+1}$. We are given a sequential neural model, denoted by \mathcal{M} , whose goal is to predict \mathbf{y}_t from $(\mathbf{x}_t, \mathcal{G})$ at each t , i.e., it provides a *point estimate* $\hat{\mathbf{y}}_t = \mathcal{M}(\mathcal{G}_t)$. For instance, \mathcal{M} could be a GNN-RNN or GTs. Given a user-specified miscoverage level $\alpha \in [0, 1]$, the main goal of the paper is to estimate a prediction set $\mathcal{C}_{t-1}(\alpha)$ that contains the true target \mathbf{y}_t with probability of at least $1 - \alpha$, that is

$$\mathbb{P}(\mathbf{y}_t \in \mathcal{C}_{t-1}(\alpha)) \geq 1 - \alpha, \text{ for each } t.$$

This is referred to as marginal coverage. Specifically, given training data $\{\mathcal{G}_t, \mathbf{y}_t\}_{t=1}^T$, the goal is to construct prediction regions $\mathcal{C}_t(\alpha)$ sequentially for all $t \geq T + 1$. In addition to guaranteeing coverage, it is crucial to construct efficient prediction regions (with smallest possible volume), and to do so, we exploit the graph structure.

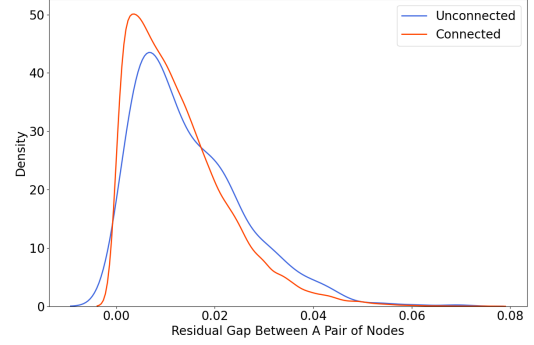


Fig. 1: Residuals are smooth over the graph. Connected nodes have a smaller gap in the residuals than unconnected nodes. The plot is generated using prediction residuals from a model \mathcal{M} on the Wiki Maths dataset (see also Section 6), averaged across all timesteps.

4. GRAPH-AWARE ELLIPSOIDAL PREDICTION SETS

In this section, we describe our construction of prediction sets guaranteeing marginal coverage and provable efficiency. We were inspired by [10], which we extend to the graph setting. In Section 5 we show that, by exploiting the graph structure, the volume of the ellipsoid shrinks relative to a comparable graph-agnostic baseline.

4.1. Graph-aware Nonconformity Score

Consider a pre-trained graph time-series prediction model \mathcal{M} . Suppose we have T training samples $\{\mathcal{G}_t, \mathbf{y}_t\}$ for $t = 1, \dots, T$. We can compute the prediction residuals as

$$\boldsymbol{\varepsilon}_t(\mathbf{y}_t) := \mathbf{y}_t - \mathcal{M}(\mathcal{G}_t), \quad t = 1, \dots, T. \quad (1)$$

Assuming homophily, i.e., prediction targets are smooth over the graph and the model exploits this property, the prediction residuals will also be smooth along the edges \mathcal{E} . See Fig. 1, where we can see that residuals are smaller on connected nodes than on disconnected pairs (a similar observation for static graph data can be found in [13]). To capitalize on this favorable structure, we filter the residuals using a graph filter \mathbf{H} . The filtered residuals are given by

$$\mathbf{e}_t(\mathbf{y}_t) = \mathbf{H}\boldsymbol{\varepsilon}_t = [(1 - \tau)\mathbf{I} + \tau\mathbf{D}^{-1}\mathbf{A}]\boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T. \quad (2)$$

Consider the mean residual vectors $\bar{\boldsymbol{\varepsilon}} = \mathbb{E}[\boldsymbol{\varepsilon}_t]$ and $\bar{\mathbf{e}} = \mathbb{E}[\mathbf{e}_t]$. Accordingly, the covariance matrices are $\mathbf{\Sigma} = \mathbb{E}[(\boldsymbol{\varepsilon}_t - \bar{\boldsymbol{\varepsilon}})(\boldsymbol{\varepsilon}_t - \bar{\boldsymbol{\varepsilon}})^\top]$ and $\mathbf{\Sigma}_{\mathcal{G}} = \mathbb{E}[(\mathbf{e}_t - \bar{\mathbf{e}})(\mathbf{e}_t - \bar{\mathbf{e}})^\top] = \mathbf{H}\mathbf{\Sigma}\mathbf{H}$. We define a *graph-aware* nonconformity score function using the (squared) Mahalanobis distance as [10]

$$s_t(\mathbf{y}_t) := (\mathbf{e}_t - \bar{\mathbf{e}})^\top \mathbf{\Sigma}_{\mathcal{G}}^{-1} (\mathbf{e}_t - \bar{\mathbf{e}}). \quad (3)$$

The scores $\{s_t\}$ can be used to compute an empirical α -quantile radius of the ellipsoid that contains the fraction of residual vectors $\{\mathbf{e}_t\}$. In practice, the ensemble covariance matrices and mean vectors are replaced with their respective sample estimates.

4.2. Ellipsoidal Uncertainty Sets

To predict the quantiles of future unseen nonconformity scores (say, the next score $p_{t'} = s_{t'+1}$), we use quantile regression [14] on a temporal window of non-conformity scores $\mathbf{s}_{t'} = [s_{t'+w-1}, \dots, s_{t'}]^\top$ obtained from training samples $t' = 1, \dots, T - w$, for some window

length w [8]. That is, we fit a quantile regressor $Q_t(1 - \alpha)$ using samples $\{\hat{p}_{t'} = Q_{t'}(1 - \alpha), p_{t'}\}_{t'=1}^{T-w}$. All in all, we have [10]

$$\begin{aligned} C_t(\alpha) &= \{\mathbf{y}_t \mid s_t(\mathbf{y}_t) \leq Q_t(1 - \alpha)\} \\ &= \mathcal{M}(\mathcal{G}_t) + \mathcal{B}(\sqrt{Q_t(1 - \alpha)}, \bar{\mathbf{e}}, \Sigma_{\mathcal{G}}). \end{aligned} \quad (4)$$

To implement the regression step, one can use off-the-shelf quantile regressors that optimize the pinball loss or quantile random forests.

5. THEORETICAL ANALYSIS

Here we offer theoretical results on coverage and volume shrinkage.

5.1. Conditional Coverage

Our construction enjoys the same coverage guarantees as [10], which we reproduce here in the interest of self-containment.

Theorem 1. [10, Corollary 4.14]. *Assume the true covariance matrix $\Sigma_{\mathcal{G}}$ is known and positive definite with minimum eigenvalue at least $\lambda > 0$, the filtered residuals $\{\mathbf{e}_t\}$ are i.i.d. over time, and the CDF of the true nonconformity score is Lipschitz. Then, we have*

$$\begin{aligned} &|\mathbb{P}(\mathbf{y}_{T+1} \in \mathcal{C}_{T+1}(\alpha) \mid \mathcal{G}_{T+1}) - (1 - \alpha)| \\ &\leq 12\sqrt{\frac{\log(16T)}{T}} + L \left(\frac{\delta_T}{\sqrt{\lambda}} + \delta_T \right). \end{aligned} \quad (4)$$

where T is the training data size, δ_T is the bound on the residual error, and L depends on the Lipschitz constant.

Notice that the effect of \mathcal{G} is captured in the constants δ_T and λ .

5.2. Ellipsoid Volume Shrinkage

Let us define $Q'_t(1 - \alpha)$ as the $(1 - \alpha)$ -quantile computed from the graph-agnostic scores $\psi_t(\mathbf{y}_t) := (\boldsymbol{\varepsilon}_t - \bar{\boldsymbol{\varepsilon}})^\top \Sigma^{-1}(\boldsymbol{\varepsilon}_t - \bar{\boldsymbol{\varepsilon}})$ [cf. (3)], using the same windowing procedure described at the end of Section 4.2. For convenience, let $\mathcal{B}_{\mathcal{G}} := \mathcal{B}(\sqrt{Q_t(1 - \alpha)}, \bar{\mathbf{e}}, \Sigma_{\mathcal{G}})$ and $\mathcal{B}' := \mathcal{B}(\sqrt{Q'_t(1 - \alpha)}, \bar{\mathbf{e}}, \Sigma)$ be the resulting graph-aware and graph-agnostic ellipsoids, respectively.

Under mild assumptions on the graph-agnostic and graph-aware quantiles, i.e., we want to approximately include the same fraction of points in the respective ellipsoids, we show that the volume of the prediction sets in (4) shrinks exponentially in the filter coefficient.

Theorem 2. *Suppose $\frac{Q_t(1 - \alpha)}{Q'_t(1 - \alpha)} \approx 1$. Then we have*

$$\text{Vol}(\mathcal{B}_{\mathcal{G}}) \leq e^{-\eta\tau} \text{Vol}(\mathcal{B}'),$$

for some positive number η , where τ is the graph filter coefficient.

Proof. We can write the log ratio of the volumes of two ellipsoids as

$$\log \left[\frac{\text{vol}(\mathcal{B}_{\mathcal{G}})}{\text{vol}(\mathcal{B}')} \right] = \log \det(\mathbf{H}) + \frac{N}{2} \log \left(\frac{Q_t(1 - \alpha)}{Q'_t(1 - \alpha)} \right). \quad (5)$$

When the quantiles before and after graph filtering are approximately the same, i.e., $\frac{Q_t(1 - \alpha)}{Q'_t(1 - \alpha)} \approx 1$, the second term in the right-hand-side of (5) vanishes. Thus, to establish ellipsoid volume

Datasets	nodes	timesteps
Chickenpox Hungary	20	513
MontevideoBus	675	740
Wiki Maths	1068	717

Table 1: Dataset statistics.

shrinkage we must show that $\det(\mathbf{H}) < 1$. To that end, we have

$$\begin{aligned} \log \det(\mathbf{H}) &= \log \det((1 - \tau)\mathbf{I} + \tau\mathbf{D}^{-1}\mathbf{A}) \\ &= \sum_{i=1}^N \log(1 - \tau(1 - \lambda_i)) \\ &\leq -\tau \sum_{i=1}^N (1 - \lambda_i) \end{aligned}$$

as $\log(1 - x) \leq -x$. Since $\lambda_i \in [0, 1]$, then $\log \det(\mathbf{H}) \leq -\tau\eta$ or equivalently $\det(\mathbf{H}) \leq e^{-\eta\tau}$, where $\eta := \sum_{i=1}^N (1 - \lambda_i) > 0$. \square

In a nutshell, Theorem 2 asserts that the volume of the graph-aware ellipsoid $\mathcal{B}_{\mathcal{G}}$ shrinks exponentially with the filter coefficient τ and the spectrum of the underlying graph η , wherein τ can be appropriately designed to get a minimum volume ellipsoid.

6. NUMERICAL EXPERIMENTS

We conduct comprehensive numerical experiments on real-world datasets to demonstrate the advantages of employing the novel CP method over its graph-agnostic counterpart [10], particularly with regard to efficiency improvements. Furthermore, we demonstrate that the proposed graph-aware approach achieves the desired coverage guarantees with a smaller ellipsoid prediction set volume.

Datasets. We evaluate the performance of our method on three real-world graph time-series datasets [15]. These datasets span diverse network types and tasks, including traffic forecasting (MontevideoBus), web traffic prediction (Wiki Maths), and epidemic modeling (Chickenpox Hungary). Additional dataset statistics are summarized in Table 1.

Graph Time Series Prediction Model. For the Wiki Maths and MontevideoBus datasets we employed the diffusion convolutional recurrent neural network (DCRNN) [1], while for the Chickenpox Hungary dataset, a Chebyshev graph convolutional gated recurrent unit cell (GConvGRU) [2] was used to obtain the point predictions. To fully utilize the training data, we trained 15 bootstrap models.

Results and Discussion. To systematically evaluate robustness across different levels of uncertainty and forecasting horizons, we perform the following experiments:

1. *Coverage for different significance level:* We consider $\alpha \in \{0.05, 0.1\}$ to study the trade-off between coverage and ellipsoid volume. The results in Tables 2 and 3 show that the proposed method maintains coverage across different confidence levels. We observe that the prediction set volumes from the proposed method are smaller in both cases, but the ellipsoid prediction regions become larger as the significance level decreases, as expected.

Method	Metric	Wiki Maths	MontevideoBus	Chickenpox Hungary
Graph-agnostic [10]	Coverage	0.903 ± 0.006	0.91 ± 0.0122	0.89 ± 0.0167
	Volume	$5.19 \times 10^3 \pm 2010.298$	$3.09 \times 10^3 \pm 247.325$	$2.74 \times 10^2 \pm 14.842$
Graph-aware (proposed)	Coverage	0.897 ± 0.010	0.912 ± 0.008	0.89 ± 0.018
	Volume	$1.46 \times 10^3 \pm 135.769$	$1.56 \times 10^3 \pm 880.327$	$1.25 \times 10^2 \pm 70.851$

Table 2: Empirical coverage and inefficiency for one-step-ahead prediction, for $\alpha = 0.1$ and window length $w = 10$. Results are averaged over five runs with a train/test split of 0.7.

Method	Metric	Wiki Maths	MontevideoBus	Chickenpox Hungary
Graph-agnostic [10]	Coverage	0.954 ± 0.005	0.948 ± 0.004	0.916 ± 0.011
	Volume	$8.51 \times 10^3 \pm 1458.303$	$1.406 \times 10^4 \pm 205.985$	$1.6 \times 10^2 \pm 14.153$
Graph-aware (proposed)	Coverage	0.952 ± 0.004	0.952 ± 0.008	0.924 ± 0.005
	Volume	$2.04 \times 10^3 \pm 238.579$	$2.7 \times 10^3 \pm 112.472$	$1.29 \times 10^2 \pm 20.144$

Table 3: Empirical coverage and inefficiency for one-step-ahead prediction, for $\alpha = 0.05$, window length $w = 10$. Results are averaged over five runs with a train/test split of 0.7.

Wiki Maths Dataset	$r = 1$		$r = 5$		$r = 10$	
	Coverage	Volume	Coverage	Volume	Coverage	Volume
Graph-agnostic [10]	0.903 ± 0.006	$5.19 \times 10^3 \pm 2010.29$	0.889 ± 0.012	$1.20 \times 10^4 \pm 754.91$	0.875 ± 0.008	$9.40 \times 10^3 \pm 5218.74$
Graph-aware (proposed)	0.897 ± 0.010	$1.46 \times 10^3 \pm 135.77$	0.885 ± 0.0129	$2.40 \times 10^3 \pm 694.09$	0.867 ± 0.002	$3.50 \times 10^3 \pm 595.01$

Table 4: Coverage and volume for the Wiki Maths dataset, for different multi-step prediction settings. Target coverage is $1 - \alpha = 0.9$.

Wiki Maths Dataset	$w = 10$		$w = 50$		$w = 100$	
	Coverage	Volume	Coverage	Volume	Coverage	Volume
Graph-agnostic [10]	0.903 ± .006	$5.19 \times 10^3 \pm 2010.298$	0.892 ± 0.005	$4.80 \times 10^3 \pm 1135.54$	0.885 ± 0.014	$2.85 \times 10^3 \pm 1260.33$
Graph-aware (proposed)	0.897 ± 0.01	$1.46 \times 10^3 \pm 135.769$	0.896 ± 0.009	$1.27 \times 10^3 \pm 182.67$	0.886 ± 0.002	$1.22 \times 10^3 \pm 345.89$

Table 5: Coverage and volume for the Wiki Maths dataset, for three different window lengths w . Target coverage is $1 - \alpha = 0.9$.

2. *Multiple window lengths:* Recall that the window length w is the number of past residuals used to predict the quantile of the future residual. We use $w \in \{10, 50, 100\}$ to analyze how sensitive the method is to temporal context. From Table 5, we observe that coverage stays similar with increasing window lengths, while the volume decreases as w increases. This behavior can be attributed to the quantile predictor having access to more data and the fact that the residuals evolve smoothly over time without abrupt changes.
3. *One-step-ahead vs. multi-step-ahead prediction:* Multi-step prediction refers to predicting multiple future values beyond the immediate next step. Given past observations $\{\mathbf{y}_t\}_{t=1}^T$, an r -step predictor estimates \mathbf{y}_{T+r} for $r > 1$. In practice, the models provide multiple future predictions at once, i.e., for different values of r . Results in Table 4 indicate that coverage diminishes as the number r of future steps to be predicted increases.

In summary, the proposed CP method achieves the desired coverage for one-step-ahead prediction for both $\alpha = 0.05$ and $\alpha = 0.1$, while yielding substantially smaller ellipsoid volumes, up to 80% reduction compared to the graph-agnostic baseline [10]. Our approach consistently outperforms the graph-agnostic counterpart across different window lengths, although the relative gain diminishes as w increases. In the case of multi-step-ahead prediction, we observe a

dip in coverage, indicating the need for further methodological improvements to address longer forecasting horizons.

7. CONCLUSIONS AND FUTURE WORK

In this work, we have developed a conformal inference framework for graph time-series forecasting. Given a base predictive model, our sequential framework produces efficient ellipsoids that can be used for uncertainty quantification with guaranteed user-specified coverage. To account for the graph structure in the nonconformity scores, we exploit the homophilic nature of the residuals via (low-pass) graph convolutional filtering. This mechanism results in provably exponentially smaller (hence, efficient) uncertainty quantifying ellipsoids relative to their graph-agnostic counterparts, while guaranteeing the desired coverage. Specifically, we show that the volume of the ellipsoid shrinks exponentially with the filter coefficient and spectrum of the underlying graph. We conducted experiments on several real-world graph time-series datasets in order to corroborate the benefits of incorporating the graph structure in the proposed non-conformity scores. Since here we compute the prediction regions from scratch at each time instant, it would be an interesting future research direction to adapt the ellipsoids through a recursive update procedure. Extending the algorithm for better coverage in multi-step prediction and heterophilic settings has been left for future work.

8. REFERENCES

- [1] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *International Conference on Learning Representations*, 2018.
- [2] Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson, "Structured sequence modeling with graph convolutional recurrent networks," in *International conference on neural information processing*, 2016.
- [3] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.
- [4] Dexiong Chen, Leslie O'Bray, and Karsten Borgwardt, "Structure-aware transformer for graph representation learning," in *Proceedings of the 39th International Conference on Machine Learning*, 2022.
- [5] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J. Kim, "Graph transformer networks," in *Neural Information Processing Systems (NeurIPS)*, 2019, 2020.
- [6] Denis Volkhonskiy, Evgeny Burnaev, Ilia Nouretdinov, Alexander Gammerman, and Vladimir Vovk, "Inductive conformal martingales for change-point detection," in *Proceedings of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications*, 2017.
- [7] Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel J. Candès, and Aaditya Ramdas, "Conformal prediction under covariate shift," in *arXiv*, 2020.
- [8] Chen Xu and Yao Xie, "Sequential predictive conformal inference for time series," in *International Conference on Machine Learning, 2023 (ICML 2023)*, 2023.
- [9] Chen Xu and Yao Xie, "Conformal prediction for time series," in *Proceedings of the 38th International Conference on Machine Learning*, 2023.
- [10] Chen Xu, Hanyang Jiang, and Yao Xie, "Conformal prediction for multi-dimensional time series by ellipsoidal sets," in *Forty-first International Conference on Machine Learning (ICML 2024)*, 2024.
- [11] Soroush H. Zargarbashi, Simone Antonelli, and Aleksandar Bojchevski, "Conformal prediction sets for graph neural networks," in *Proceedings of the 40th International Conference on Machine Learning*, 2023.
- [12] Kexin Huang, Ying Jin, Emmanuel Candès, and Jure Leskovec, "Uncertainty quantification over graph with conformalized graph neural networks," in *Advances in Neural Information Processing Systems*, 2023.
- [13] Junteng Jia and Austion R. Benson, "Residual correlation in graph neural network regression," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2020.
- [14] Roger Koenker and Kevin F. Hallock, "Quantile regression," *Journal of Economic Perspectives*, 2001.
- [15] "Pytorch temporal datasets," <https://pytorch-geometric-temporal.readthedocs.io/en/latest/modules/dataset.html>.