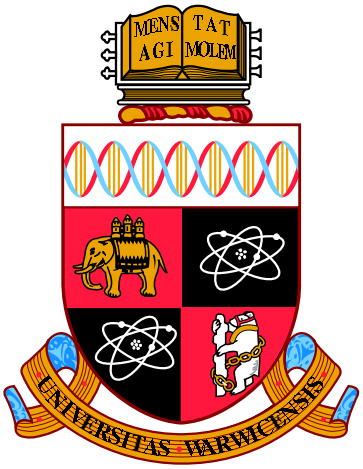

Sampling of Time-Varying Network Signals from Equation-Driven to Data-Driven Techniques

by

Zhuangkun Wei

Thesis

submitted to the University of Warwick

for the degree of

Doctor of Philosophy

School of Engineering

December 2020

THE UNIVERSITY OF WARWICK

Contents

List of Figures

Acknowledgments

At first, I would like to appreciate my two supervisors Prof. Weisi Guo and Prof. Yunfei Chen at School of Engineering, University of Warwick. Weisi was my first PhD supervisor, who helped me apply for the PhD and the scholarships to cover my tuition fees. During my first year at Warwick, Weisi guided me to identify my research of interest, helped me deduce equations, and taught me how to write academic papers; his professional and forward-looking eyesight on signal processing and complex networks, and excellent writing skills helped me a great deal and are very worth learning. After Weisi left Warwick, Prof. Yunfei Chen became my second PhD supervisor, who offered me a great amount of helps for my second and third years' study. Chen arranged meeting with me once/twice a month, discussing the research progress, and tutoring me to improve my scientific writing, especially how to reply to reviewers that had negative comments. Also, Chen's specialty on numerical optimisation and digital signal processing provided me a great help for the algorithm and scheme designs in my researches.

Besides, I appreciate Dr. Alessio Pagani for his expertise on water-distribution networks, Prof. Wei Chen for his help on compressed sensing theory, and Prof. Bin Li and Prof. Chenglin Zhao for their professional guidance on Koopman linearization theory and molecular communications. Also, I appreciate Prof. Jérôme Charmet, Prof. Tianhua Xu, Ms. Wenxiu Hu, and Mr. Schyler Chengyao Sun for helps.

Finally, I owe a great deal to the School of Engineering, University of Warwick who offer me the chance to pursue my PhD study here, and who cover my tuition fees, and I sincerely appreciate China Scholarship Council to cover my living expenses.

Declarations

I, Zhuangkun Wei confirm that, this thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in Engineering. It has been composed by myself and has not been submitted for any degree at other universities.

List of Publications

Inclusion of Published Works During PhD

1. Z. Wei, B. Li, C. Sun, and W. Guo, "Sampling and Inference of time-varying network signals using Log-Koopman Nonlinear Graph Fourier Transform", IEEE Transactions on Signal Processing, vol. 68, pp. 6187–6197, Oct. 2020.

2. Z. Wei, W. Guo, B. Li, J. Charmet and C. Zhao, "High-Dimensional Metric Combining for Non-Coherent Molecular Signal Detection," IEEE Transactions on Communications, vol. 68, no. 3, pp. 1479-1493, Mar. 2020.

3. Z. Wei, A. Pagani, G. Fu, I. Guymer, W. Chen, J. McCann, W. Guo, "Optimal Sampling of Water Distribution Network Dynamics Using Graph Fourier Transform," IEEE Transactions on Network Science and Engineering, vol. 7, no. 3, pp. 1570-1582, Sep. 2020.

4. Z. Wei, A. Pagani, B. Li and W. Guo, "Monitoring Embedded Flow Networks Using Graph Fourier Transform Enabled Sparse Molecular Relays," IEEE Communications Letters, vol. 24, no. 5, pp. 986-990, May 2020.

5. Z. Wei, B. Li, W. Hu, W. Guo, C. Zhao, "Hamming–Luby Rateless Codes for Molecular Erasure Channels", Nano Communication Networks, vol. 23, p. 100280, 2020.

6. Z. Wei, B. Li and W. Guo, "Optimal Sampling for Dynamic Complex Networks With Graph-Bandlimited Initialization," IEEE Access, vol. 7, pp. 150294-150305, 2019.

7. Z. Wei, A. Pagani and W. Guo, "Monitoring Networked Infrastructure with Minimum Data via Sequential Graph Fourier Transforms," 2019 IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, 2019, pp. 703-708.

Other Publications During PhD

8. Z. Wei, B. Li, W. Guo, W. Hu and C. Zhao, "On the Accuracy and Efficiency of Sensing and Localization for Robotic," IEEE Transactions on Mobile Computing, doi: 10.1109/TMC.2020.3038146.

9. Z. Wei, B. Li, W. Guo, W. Hu and C. Zhao, "Sequential Bayesian Detection of Spike Activities From Fluorescence Observations," IEEE Transactions on Molecular, Biological and Multi-Scale Communications, vol. 5, no. 1, pp. 3-18, Oct. 2019.

10. W. Hu, Z. Wei, S. Popov, M. Leeson, M. Zhang and T. Xu, "Non-Coherent Detection for Ultraviolet Communications With Inter-Symbol Interference," Journal of Lightwave Technology, vol. 38, no. 17, pp. 4699-4707, Sep. 2020.

11. S. Liu, Z. Wei, B. Li and C. Zhao, "Unsupervised Clustering-Based Non-Coherent Detection for Molecular Communications," IEEE Communications Letters, vol. 24, no. 8, pp. 1687-1690, Aug. 2020.

12. W. Guo, Z. Wei and B. Li, "Secure Internet-of-Nano Things for Targeted Drug Delivery: Distance-based Molecular Cipher Keys," 2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME), Amman, Jordan, 2020, pp. 1-6.

Abstract

Sampling and recovering the time-varying network signals via the subset of network vertices is essential for a wide range of scientific and engineering purposes. Current studies on sampling a single (continuous) time-series or a static network data, are not suitable for time-varying network signals. This will be even more challenging when there is a lack of explicit dynamic models and signal-space that indicate the time-evolution and vertex dependency.

The work begins by bridging the time-domain sampling frequency and the network-domain sampling vertices, via the eigenvalues of the graph Fourier transform (GFT) operator composed by the combined dynamic equations and network topology. Then, for signals with hidden governing mechanisms, we propose a data-driven GFT sampling method using a prior signal-space. We characterize the signal dependency (among vertices) into the graph bandlimited frequency domain, and map such bandlimitedness into optimal sampling vertices.

Furthermore, to achieve dynamic model and signal-space independent sensor placement, a Koopman based nonlinear GFT sampling is proposed. A novel data-driven Log-Koopman operator is designed to extract a linearized evolution model using small ($M = O(N)$) and decoupled observables defined on N original vertices. Then, nonlinear GFT is proposed to derive sampling vertices, by exploiting the inherent nonlinear dependence between M observables (defined on $N < M$ vertices), and the time-evolved information presented by Log-Koopman evolution model.

The work also informs the planned future work to formulate an easy-to-use and explainable neural network (NN) based sampling framework, for real-world industrial engineering and applications.

Abbreviations

APM	Arrival Probability Matrix
CS	Compressed Sensing
EDMD	Extended Dynamic Mode Decomposition
DCT	Discrete Cosine Transform
Deep-DMD	Deep Dynamic Mode Decomposition
DMD	Dynamic Mode Decomposition
FT	Fourier Transform
GFT	Graph Fourier Transform
JFT	Joint Fourier Transform
LT	Luby-Transform
MAK	Mass-Action Kinetics
MIP	Mixed-Integer Program
MRDD	Molecular Communication Relay Data-Driven
MSE	Mean Squared Error
NN	Neural Network
NRMSE	Normalized Root Mean Squared Error
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
PD	Population Density
PGA	Progressive Generic Algorithm
RMSE	Root Mean Squared Error
WDN	Water Distribution Network
RIP	Restricted Isometric Property

SVD	Singular Value Decomposition
X-AI	Explainable Artificial Intelligence

Chapter 1

Introduction

Dynamical element (e.g., the continuous signals, or time-series) underpins a variety of complex systems, ranging from the single and individual systems, to the complex and high-dimensional networks (e.g., epidemic spreading [1], the social network [2], the urban structure [3], and the engineering infrastructure [4]). Such networked elements, coupled with each other and exhibiting complex behaviours, are required to be monitored and controlled for a wide range of academic and industrial purposes, including but not limited to the system modelling [5], contaminant alarming [6], smart-grid controlling [7], and the Digital Twin informed maintenance [8,9]. Among these, one important application is to monitor the contaminant spread in water-distribution network (WDN), which serves as the fundamentals for the welfare of the society (e.g., more than $350,000$ kilometers of water pipes in the UK).

Ideally, an installation of sensors on all network vertices would be a straightforward option to monitor the whole time-varying network signals (states). However, when the network scale (the number of vertices) is large, this is often not possible due to the expenses of both sensors and their high-dense deployments (e.g., £160 for pH sensor each [10], and £270 for dissolved oxygen sensor each [10]), or even the operation difficulty with networks that are hard to be accessed (e.g., the buried underground water-distribution network). This therefore raises the studies on the effective sensor deployment through a small subset of network vertices.

1.1 Motivation

Optimal sensor placement for sampling (compressing) and recovering the time-varying network signals is challenging. Here, the time-varying network signals are the stacked time-series on all network vertices. Current compression schemes either

rely on the signal dependency among different network vertices (e.g., the compressed sensing CS, and the graph spectral analysis), or exploit the dynamic time-evolution information (e.g., the graph observability analysis). In the context of network sampling for signal recovery, the challenges lie in two aspects.

- First, signal dependency analysis requires an operator that is able to uncover the dependencies of the time-varying network signals among all vertices (e.g., the sparsity for CS and the bandlimitedness for graph spectral analysis). This is of difficulty for current graph Fourier transform (GFT) operators, which involve only the network topological information, thereby rendering its inability to characterize and discover the signal dependency governed by the underlying dynamical mechanism.

- For the graph observability analysis, an explicit linear/linearized dynamic model serves as a prerequisite to understand and exploit the time-evolved information for sensor placement and signal recovery. This thereby blocks its usages for the monitoring of the nonlinear time-varying network signals, not to mention if such dynamic governing models are unknown.

These two challenges constitute the motivations of this thesis, in which four network sampling schemes aiming at signal recovery are proposed, for different scenarios. We will first construct the GFT operators via the combinations of the dynamic mechanism and the network topology, with and without the explicit dynamic governing equations (models). Then, the sampling vertices can be selected corresponding to the signal dependency among different vertices discovered by our proposed dynamic-topology combined GFT operator. Second, if the signal dependency property (e.g., the sparsity and graph bandlimitedness to a designed operator) is hard to find, we will develop a Log-Koopman nonlinear GFT sampling framework, where the Log-Koopman operator is designed and extracted from the experimental data, in order to derive a linearized dynamic time-evolution model, and the nonlinear GFT concept and sampling theory will determine the time-invariant sampling vertices for complete signal recovery. The detailed introduction of the existing works and their advantages/drawbacks are provided in the following approach outlines in Section 1.2, followed by the approaches and contributions of this thesis in Section 1.3.

1.2 Outline of Existing Approaches

From academic perspectives, existing network sampling and compression approaches can be categorized by their different objectives. These include but are not limited to the resilience analysis [11–13], the minimization of event detection-time [14–16], the minimization of event affected area [17, 18], and the complete recovery of time-varying network signals [19]. Each objective requires a subset of dynamic information, which should be mapped from the sampled and compressed results for understanding. With the desired information truncated, less vertices and their signals should be monitored, and are unable to posses compatibility with higher required objectives. This therefore blocks the rich literature of the lower objective-oriented optimization approaches (e.g., the mixed-integer program, the genetic algorithms [18], and the randomized contamination matrix [14]) being used for the purpose of the time-varying network signal recovery, which requires the highest amount of dynamic information without truncated.

On the purpose of network sampling and complete signal recovery, the solutions can be categorized into two groups: (i) vertex dependency approaches, and (ii) time-evolution analysis.

Vertex dependency approaches characterize the whole dynamic information (of a network with N vertices) by an operator and the signal transformation to such operator. The compression then can be pursued on the latter if the operator is able to uncover the dependency of signals on different vertices. Typical vertex dependency approaches include the compressed sensing (CS) [20–26], and the graph spectral analysis [27–38]. CS relies on a complete dictionary (a transforming matrix) to sparsely represent a network signal (vector), so that any vector whose transformation is r-sparse (i.e., r nonzero elements in transformed vector) can be recovered by $> 2r$ samples [20, 23, 24]. The graph spectral analysis resorts to a network topology based graph Fourier transform (GFT) operator (typically the eigenvector space of graph Laplacian operator, or of graph adjacent matrix [27–29, 39]). Then, any network signal belonging to a vector subspace spanning by $r < N$ eigenvectors (called graph bandlimited to such GFT operator) can be sampled and recovered by r independent vertices [29–38].

These two schemes perform well in the static network data (vector). However, when it comes to address the sampling and recovery of the time-varying network signals, they will result in either a time-varying sensor placement strategy, or a large amount of time-invariant sampling points. This is due to the difficulty to find operators that are able to characterize the vertex dependency of network signals

for all continuous or discrete time, and different vertex dependency property on different time will lead to the changing of sampling vertices. For example, typical CS operators using the discrete cosine transform (DCT) [24], and the data-driven principal component analysis (PCA) [24] are hard to transform every network signal at different time into r-sparse representation. This is further proved by the study in [40], which provides the average samples needed for each discrete time as $(N+K-r) \times r/K$ if a network signal matrix has N vertices and K discrete-times with rank as r. This is not to mention for the case of a full row-rank signal matrix, in which all vertices have to be monitored at each discrete-time for CS framework. Similarly, the well-studied GFT sampling is also unable to characterize time-evolved network signals as graph bandlimited, as no underlying dynamic mechanism is involved in the topology-based GFT operators. Further researches on joint Fourier transform (JFT) in [41], and graph smoothness batch in [19], analyze the vertex dependency of both graph and time-domain bandlimited signals, however, they become less attractive for the most of the real-world complex systems where the signals are not bandlimitd, let alone the extreme cases where the vertex dependency is not existed.

To address the vertex in-dependency cases, the alternative group is to use the time-evolution information. The typical framework is referred to as the graph observability analysis [42–45], leveraging either the observability gramian or the linear time-evolution analysis. Here, observability gramian is a model-relevant operator mapping from the selected samples to a scalar-valued energy, which if been maximized, can lead to an optimal sampling subset. Linear time-evolved analysis is to find the sampling points by checking the rank conditions of the linear time-evolved model. Leveraging this idea, the work in [46] further combines the time-evolution information with the vertex dependency for sampling some of the specific time-varying network signals (e.g., the auto-regressive moving average graph process, the wave propagation, and the signal diffusion). These graph observability analysis methods all provide promising performances for sampling and recovering the time-varying network signals, under the important prerequisite of an exactly known and linear/linearized dynamic time-evolution mechanism. However, when such dynamic equations are nonlinear or even non-existed, the schemes will lose the compass and become malfunctioned.

Recent studies focus on a combination of Koopman linearization schemes with the graph observability analysis [47], whereby a Koopman operator has been designed to extract and linearize an unknown and nonlinear time-varying network signals, which thereby paves the way for graph observability analysis using the rich standard linear algebra theory. Whilst many of classical Koopman observables

(e.g., the Polynomial based observables [47], and the deep learning based observables [48]) have been designed for analyzing the invariant dynamic modes, or stability of a nonlinear systems, they are not suitable for network sampling tasks. For one thing, the polynomial based Koopman observables designed by [47, 49], although capable of characterizing a linear time-evolution model for small-scale network ($N < 30$), suffer from size explosion, due to the multi-elemental multiplicative observable elements of Talyor series, rendering inapplicably for large-scale network sampling. For another, even if such size explosion can be partially alleviated by the deep learning framework in [48], the scheme results in a group of coupled observables, each of which involves multiple signals on different vertices. This, when combined with the graph observability analysis, will lead to redundant sampling vertex selection, as selecting one leading observable may require a placement of sensors on multiple vertices. Moreover, direct usages of graph observability methods on the Koopman linearized time-evolution model overlook the inherent nonlinear dependency among different observable elements, as all observable elements are defined on the lower-sized signals on original network vertices. This will also cause redundant sampling vertex selection by mapping the samples to the super-set of observables.

1.3 Contributions of this Thesis

In this thesis, we study how to approach a time-invariant sensor placement of network vertices (called sampling vertex subset) for the recovery of the time-varying network signals. The contributions are categorized into network sampling using the vertex dependency and the time-evolution information.

1.3.1 Sampling using Vertex Dependency

We extend the GFT sampling framework into the analysis of time-varying network signals, and derive two GFT sampling schemes for two cases (i.e., explicit model-driven, and data-driven when model is unavailable).

(1) In the face of an explicit dynamic governing equation (model), we develop an equation-driven GFT sampling scheme, which is able to understand (i) how to discretize the continuous network signals, and (ii) where to place sensors for recovering the time-varying network signals. A combining dynamic equation and network topology based GFT operator is proposed, which is able to characterize all the transient behaviours of the continuous network signals as graph $\mathcal{B}$-bandlimited. Then, leveraging such an equation-topology based GFT operator, a joint time-domain and network-domain sampling is proposed, which maps the sampling vertex sub-

set for sensor placement and the discretizing rate (i.e., the time-domain sampling frequency) to the graph frequencies determined by the GFT operator. More importantly, we deduce an explicit relation between the optimal sampling locations and the time-domain sampling frequency, leveraging the graph bandlimited property and the governing dynamic equations. As such, this novel equation-driven GFT sampling framework offers the dynamical system insight unavailable from the CS and the previous GFT researches (that resort to the topology-based GFT operator), and provides a time-invariant optimal sampling vertex subset for immediate recovery of time-varying network signals.

(2) In the absence of an explicit dynamic governing model, we develop a data-driven GFT sampling scheme, which is able to learn the hidden dynamic mechanism from the experimental data. By exploiting the dynamic data matrix, we identify the graph Fourier basis (i.e., the GFT operator), which enables us to determine the optimal network vertex subset for the full recovery of the time-varying network signals. Compared with the topology-based GFT operators (e.g., the Laplacian operator and adjacent matrix [27–29, 39]), the proposed data-driven GFT operator is able to characterize the network signals from all discrete-times into the graph bandlimited region, thereby making it possible to achieve a time-invariant sampling vertex subset for signal recovery. Compared with the CS approaches, we are able to achieve a smaller number of sampling vertices, as the data-driven GFT operator is able to uncover a more compact subspace (i.e., the graph bandlimited space), which is a subset of that composed of all the r-sparse signals. As such, the data-driven GFT sampling framework can provide a more compact and time-invariant sampling vertex subset, used for immediate and real-time signal recovery.

1.3.2 Sampling using Time-Evolution Information

In the absence of signal dependency among network vertices, a linear/linearized dynamic time-evolution model can provide more information for network sampling and signal recovery. One challenge, as aforementioned, lies in the non-linearity or even the unavailability of the explicit dynamic evolution model. To address this, we propose two sampling schemes that can adopt or extract the hidden dynamic time-evolution mechanism. We briefly summarize the contributions in the following.

(3) We propose a sequential data-driven GFT sampling scheme in the absence of signal dependency among vertices and the explicit dynamic time-evolution model. At each discrete-time, we compute a GFT operator and a time-varying graph bandwidth set via the principal component analysis (PCA) of the previous recovered signals. Then, a sampling vertex subset is achieved using the GFT sampling theory

and a proposed signal prediction method. When compared with the CS approaches, our scheme provides a lower-sized sampling vertex subset, as the sequential data-driven GFT operator is able to characterize the current dynamic into a predictable graph bandwidth set, rather than the loose set composed of all the r-sparse signals.

(4) To further exploit the hidden time-evolution information for time-invariant network sensor placement, we propose a novel logarithm-based Koopman operator and non-linear GFT scheme (abbreviated as Log-Koopman nonlinear GFT), whereby the Koopman operator is to derive a linearized dynamic evolution model of observable defined on original time-varying network signals, and the nonlinear GFT is to exploit the nonlinear dependency of observable. In order to address the observable size-explosion and dynamic coupling mentioned before, we design a logarithm based vector-valued observable to approximate the multi-elemental multiplicative terms of Taylor series in the manner of logarithm summation. In this view, the size of vector-valued observable can be reduced to $M = O(N)$, as a smaller number of logarithm terms can be used and linearly combined for a large number of polynomial-based observables in [47] (which requires a size of $M = O(N^2)$). Then, leveraged on the Log-Koopman operator, we propose the concept and theory of nonlinear GFT to exploit the nonlinear dependence between the M elements of the logarithm vector-valued observable defined on the lower-sized N original signals. Compared to the graph observability analysis [47, 50] that regard the M observable elements independent, our proposed Log-Koopman nonlinear GFT sampling scheme is able to derive a more compact sampling vertex subset by mapping it to a lower-sized observable signal-space. Besides, when compared with the sampling schemes relying on the vertex dependency, our scheme can achieve a time-invariant sampling strategy when such vertex dependency does not exist, but at the expense of a latency signal recovery.

1.4 Organization of the Rest of this Thesis

The rest of this thesis is organized as follows. Chapter 2 is to review and analyze the published literature in theoretical manner. Chapters 3-4 correspond to the equation-driven and the data-driven sampling method using the vertex dependency. In Chapter 3, the explicit dynamic equation driven GFT sampling framework is proposed, whereby the sampling vertex subset and the discretizing rate for continuous network signals are provided. In Chapter 4, we elaborate the proposed data-driven GFT sampling scheme, in the absence of the explicit dynamic model.

The sampling schemes relying on the time-evolution information are elabo-

rated in Chapters 5-6. In Chapter 5, in the absence of both the dynamic governing equations, and the signal dependency among vertices, we propose the sequential data-driven GFT sampling method, which can achieve a time-varying sensor activation strategy for network signal recovery. In Chapter 6, to further exploit the time-evolution mechanism and to achieve a time-invariant sampling vertex subset, we propose the Log-Koopman nonlinear GFT scheme, including the Log-Koopman observable and operator designs, and the nonlinear GFT concept and theories for the selection of the fixed sampling vertices.

In Chapter 7, we conclude this thesis, and describe the potential future works on an explainable and trustworthy machine learning framework.

Chapter 2

Review of Published Literature

2.1 Objective Oriented Optimization Approaches

Objective oriented optimization methods consider a number of single or hybrid objectives for sensor placement and network monitoring. Such objectives range from the sensing or tracking of the dynamic events, the network accessibility, to the resource and complexity aspects of the cyber-physical interface [51]. We list some of the typical aims and explain them in a real WDN contaminant monitoring and alarming systems.

2.1.1 Minimizing Detection Time

Detection time is referred to as the time elapsed from a start of a dynamic event (an injection of a contaminant on WDN) to its first detection by any of the deployed sensors [14]. To reduce the detection time, the work in [14] identifies the best sensor placement locations in a greedy manner by constructing and using travel time matrix. Then, various heuristic frameworks in [15, 16] have been designed to search the best monitoring locations by maximizing the sensing coverage under the condition of time to detect. However, these sensor placement approaches, only interested in the detection of a dynamic event (contaminant) over the network, are unable to track or recover the full propagation of the signals from the samples.

2.1.2 Minimizing Affections

Another aspect focuses on how to place sensors for an event alarm in order to minimize the affected populations. For example, leveraging the awareness of the population distribution around a WDN system, the work in [17] formulates a sensor placement strategy, with aim to minimize the expected proportion of population

affected by a potential contaminant spread. Similar approaches include but are not limited to the mixed-integer program (MIP), the genetic algorithms [18] formulation, and the randomized contamination matrix [14]. However, apart from the inability for their sensors to recover the complete time-varying network signals, these approaches render less feasible for large-scale network (e.g., a network with a total of $N > 50$ vertices), especially for multiple or diverse contaminant signals in WDN (e.g., each would require a different sensor placement solution).

2.1.3 Computational Efficiency

Computational efficiency has been regarded as another important factor when monitoring a large scale network. For instance, a progressive generic algorithm (PGA) has been designed in [52, 53], to solve models for large-scale WDNs. Another group of common schemes aiming to optimally identify sampling locations is via the formulations of an optimization task by multiple objectives. This indeed provides the guideline to compress the dimensionality of a complex network through sensitivity-awareness analysis [54], which is also able to incorporate the uncertainties of the network's demands and services for Early Winning System operation [55]. These numerical approaches cannot construct a mapping to the optimal sampling locations, given their lack of an explicit bridge between the network topology (e.g., the adjacent matrix) and the underlying dynamical mechanisms, let alone for the purpose of time-varying network signal recovery that requires in depth the dependencies from both the network vertex domain and dynamic time-evolution.

2.2 Nyquist Sampling Rate from Time-Domain

When dealing with the sampling/discretizing of an individual dynamic element on time-domain, Nyquist sampling and interpolation theory serves as the fundamentals. By defining the cut-off angular frequency, denoted as $\omega_{\text{cue-off}}$, as the highest frequency of the individual signal, Nyquist sampling can avoid the aliasing if assigning the sampling frequency as $\omega_{\text{s}} \geq 2\omega_{\text{cut-off}}$. As such, the time-domain single and continuous signal $x(t)$ (t denotes the continuous time) can be sampled as $x(2\pi k/\omega_{\text{s}})$, and then recovered as:

$$\hat{x}(t) = \sum_{k\in\mathbb{Z}} x\left(\frac{2\pi}{\omega_{\text{s}}}\cdot k\right)\cdot sinc\left(\frac{\omega_{\text{s}}}{2\pi}\cdot t - k\right), \tag{2.1}$$

where $sinc(t) = \sin \pi t/(\pi t)$ is the interpolation function.

It is noteworthy that when compared with some more sophisticated sampling

studies, e.g., [56,57], Nyquist sampling theory may not be the optimal one, and may result in redundant selections of discrete-time sampling. But, what we are interested in is to compute and measure the rationality of the time-frequency domain cut-off frequency, and how this can be extended for continuous network signals. Without this, one cannot derive the discretized network signals (matrix) for further digital domain signal processing. Other sophisticated theories and interpolation methods can be adopted based on this baseline.

2.3 Sampling for Network Signal Recovery

On the purpose of understanding the sensor placement for the recovery of the time-varying network signals, we categorize current researches into two groups. The first one considers the sampling (compression) via the usages of signal dependencies among vertices (e.g., the compressed sensing CS [20–26] and the graph spectral analysis [27–38]), whereby only vertices with independent signals will be selected for the recovery of other dependent ones. The second leverages the dynamic time-evolution models, using the time-evolved samples to recover the time-varying network signals (e.g., the graph observability analysis [42–46, 58, 59]). We will introduce these in details in the following.

2.3.1 Compressed Sensing Approaches: using Vertex Dependency

One sampling approach that relies on the signal dependencies among different vertices is the compressed sensing, which is a framework to compress the (transformed) sparse signals by the measurements (or samples) that are linearly independent [20–26]. Given a network signal (vector) $\mathbf{x} \in \mathbb{R}^N$ defined on the network with N vertices, CS uses a designed $N \times N$ transforming operator (a reversible matrix), denoted as $\mathbf{D}$, in order to achieve a sparse representation of $\mathbf{x}$, i.e., [20, 23, 24]

$$\mathbf{x} = \mathbf{D} \cdot \mathbf{s}, \tag{2.2}$$

where $\mathbf{s}$ of size $N \times 1$ is r-sparse representation of $\mathbf{x}$ with $r = \|\mathbf{s}\|_{l_0}$, the number of non-zero elements in $\mathbf{s}$. In Eq. (2.2), typical designs of the transforming operator $\mathbf{D}$ are the discrete cosine transform (DCT) [24], and the data-driven principal component analysis (PCA) based operator [24].

Then, a sampling vertex subset $\mathcal{C} \subset \mathcal{N} = \{1, 2, \cdots, N\}$ is selected by satisfying the restricted isometric property (RIP) criteria [20, 23, 24]. In the context of network sampling, we provide in the following the sampling theory deduced from

the works in [25, 26].

Theorem 1 *Denote the network signal* $\mathbf{x} \in \mathbb{R}^N$*, and the transforming matrix* $\mathbf{D}$ *(of size* $N \times N$*) that* $\mathbf{x} = \mathbf{D} \cdot \mathbf{s}$ *with* $r = \|\mathbf{s}\|_{l_0} < N$*. Then the sampling vertex subset* $\mathcal{C} \subset \mathcal{N}$ *to ensure the recovery of any* $\mathbf{x}$ *that has* r*-sparse representation from* $\mathbf{x}_{\mathcal{C}}$ *is*

$$rank\,(\mathbf{D}_{\mathcal{CN}}) > 2 \cdot r, \tag{2.3}$$

where $\mathbf{D}_{\mathcal{CN}}$ *is the sub-matrix of* $\mathbf{D}$ *with rows' indices selected from set* $\mathcal{C}$ *and columns' indices selected from set* $\mathcal{N} = \{1, 2, \cdots, N\}$*.*

Proof 1 *Otherwise, if* $rank\,(\mathbf{D}_{\mathcal{CN}}) = 2r$*, there exists a* $2r$*-sparse vector* $\mathbf{s}^{(0)}$ *of size* $N \times 1$ *satisfying* $\mathbf{D}_{\mathcal{CN}} \cdot \mathbf{s}^{(0)} = \mathbf{0}$*. This suggests two different* r*-sparse vectors* $\mathbf{s}^{(1)}$ *and* $\mathbf{s}^{(2)}$ *of size* $N \times 1$ *exist but* $\mathbf{D}_{\mathcal{CN}} \cdot (\mathbf{s}^{(1)} - \mathbf{s}^{(2)}) = \mathbf{0}$*. Therefore,* $\mathbf{s}^{(1)}$ *and* $\mathbf{s}^{(2)}$ *cannot be recovered using the samples from sampling vertex subset* $\mathcal{C}$*, as* $\mathbf{D}_{\mathcal{CN}} \cdot \mathbf{s}^{(1)} = \mathbf{D}_{\mathcal{CN}} \cdot \mathbf{s}^{(2)}$ *results in the same samples. This contradicts the statement in the Theorem that any* $\mathbf{x} \in \mathbb{R}^N$ *that can be transformed by* $\mathbf{D}$ *into* r*-sparse representation, can be recovered by samples from* $\mathcal{C}$*.*

After the determination of the sampling vertex subset $\mathcal{C}$ from Theorem 1, the recovery of $\mathbf{x}$, denoted as $\hat{\mathbf{x}}$, is pursued by computing the r-sparse representation $\hat{\mathbf{s}}$ via convex optimisation or orthogonal matching pursuit (OMP), i.e., [20, 23, 24]

$$\begin{aligned} \hat{\mathbf{s}} &= \underset{\mathbf{s} \in \mathbb{R}^N}{argmin} \|\mathbf{s}\|_{l_1}, \text{ such that } \mathbf{x}_{\mathcal{C}} = \mathbf{D}_{\mathcal{CN}} \cdot \mathbf{s}, \\ \hat{\mathbf{x}} &= \mathbf{D} \cdot \hat{\mathbf{s}}. \end{aligned} \tag{2.4}$$

When adopting CS for sampling the time-varying network signals, one needs to consider the compression task on a data matrix, denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K]$ of size $N \times K$ where N is the number of total network vertices, and K is the total discrete-time of interest. The work in [40] proved that, given $rank(\mathbf{X}) = r$, CS framework requires at least $(N + K - r) \times r$ samples for the recovery of time-varying network signals. This indicates for each discrete-time, the average $(N + K - r) \times r/K$ sensors are needed.

The advantages/disadvantages of using CS for network sampling and signal recovery are provided as follows.

- One advantage lies in that the exploitation of signal dependency guarantees an immediate recovery of the time-varying network signals, as no further time-evolved sample is needed. This is important for those requiring real-time and

low-latency network monitoring (e.g., the real-time and low detection time monitoring of contaminant spread in WDN [14–16]).

- For one drawback, the CS framework in [40] may approach time-varying sampling strategy for network signals at different discrete times, which as mentioned before is less attractive for some of the sensor placement and network monitoring applications, e.g., the WDN.

- For another, even if other CS approaches in [20, 23] can offer time-invariant sensor placement for all discrete-time, the homogeneous $(N + K - r) \times r/K$ sampling vertices are still too large.

- Moreover, either the sparse property of time-varying network signals $\mathbf{X}$ and/or the desired operator $\mathbf{D}$ for sparse representation may not exist; if the data matrix is full row-rank (i.e., $rank(\mathbf{X}) = N$), according to [40], entire N vertices are required to place sensors for signal recovery.

2.3.2 Graph Spectral Analysis: using Vertex Dependency

Another framework analyzing the signal dependency among different vertices is graph spectral analysis, which has been proposed in [27], and paves the way for optimally sampling on a combinatorial graph or network. In [27–29], several network topology based operators are adopted to analyze the independent graph frequency components. Here, we list two popular topology based operators, denoted as $\boldsymbol{\mathcal{L}}$ in the following [27–29]:

$$\boldsymbol{\mathcal{L}} = \begin{cases} \mathbf{A} & \text{graph weight operator,} \\ \mathbf{Deg}^{-\frac{1}{2}} \cdot (\mathbf{Deg} - \mathbf{A}) \cdot \mathbf{Deg}^{-\frac{1}{2}} & \text{graph Laplacian operator,} \end{cases} \tag{2.5}$$

where $\mathbf{A}$ of size $N \times N$ is the adjacent matrix of a network with N vertices, of which the element $a_{n,m}$ represents the weight of the direct link from vertex m to vertex n, and $\mathbf{Deg} = diag([deg_1, deg_2, \cdots, deg_N])$ is the diagonal vertex degree matrix with $deg_n = \sum_{m=1}^{N} a_{n,m}$. As such, a transforming operator, denoted as $\mathbf{P}^{-1}$, has been derived from the eigenvectors of the topology-based operator $\boldsymbol{\mathcal{L}}$, i.e., [27–29]

$$\boldsymbol{\mathcal{L}} = \mathbf{P} \cdot diag([\lambda_1, \lambda_2, \cdots, \lambda_N]) \cdot \mathbf{P}^{-1}, \tag{2.6}$$

where $\lambda_1, \lambda_2 \cdots, \lambda_N$ are the topology-based graph frequencies. Such an operator in Eq. (2.6) is referred to as the graph Fourier transform (GFT) operator. Then,

given a graph signal defined on N vertices as $\mathbf{x} \in \mathbb{R}^N$, the GFT process from $\mathbf{x}$ to its graph frequency response, and the inverse GFT process are given as [27–29]:

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{P}^{-1} \cdot \mathbf{x}, \qquad \text{GFT process}, \\ \mathbf{x} &= \mathbf{P} \cdot \tilde{\mathbf{x}}, \qquad \text{inverse GFT process}, \end{aligned} \tag{2.7}$$

where $\tilde{\mathbf{x}}$ is defined as the graph frequency response to the operator $\mathbf{P}^{-1}$, and the nth element of $\tilde{\mathbf{x}}$, i.e., $\tilde{x}_n$ is the magnitude for nth graph frequency.

Leveraging these foundations, the concept and theories of graph signal processing has been proposed and studied in [28–30,32,33,35–38,60] to understand how to optimally select samples from a graph signal $\mathbf{x}$ that maps to the total graph frequency response $\tilde{\mathbf{x}}$, which subsequently ensures the complete recovery of the original graph signal $\mathbf{x}$ by inverse GFT process.

Graph Bandlimited Signal

The concept of graph bandlimited signal is firstly proposed in [28], which introduces a Paley-Wiener subspace of $\mathbb{R}^N$, and analyzes the graph frequencies of signals that belong to such space. Then, the work in [30] generalizes the graph bandlimited concept to any subspace of $\mathbb{R}^N$ determined by a GFT operator (a reversible matrix of size $N \times N$) $\mathbf{P}^{-1}$ of the network Laplacian operator, and following definitions are given:

Definition 1 *[30] A network signal $\mathbf{x}$ of size $N \times 1$ is called graph bandlimited to a GFT operator (a reversible matrix of size $N \times N$) $\mathbf{P}^{-1}$, if its graph frequency response $\tilde{\mathbf{x}} = \mathbf{P}^{-1}\mathbf{x}$ has less than N nonzero elements. Denote the graph bandwidth set $\mathcal{B}$ composed by all the indices of non-zero elements of the graph frequency response $\tilde{\mathbf{x}}$. Then, $\mathbf{x}$ is called graph $\mathcal{B}$-bandlimited to the GFT operator $\mathbf{P}^{-1}$.*

Definition 2 *[30] Denote a subspace as $BS(\mathcal{B}, \mathbf{P}^{-1}) \subset \mathbb{R}^N$ composed of all graph $\mathcal{B}$-bandlimited signals to the GFT operator $\mathbf{P}^{-1}$.*

Optimal Sensor Placement for Graph Bandlimited Signal

Based on the graph bandlimitedness definitions, the works in [30,32,35–37,60] further develop theories and approaches to find the optimal vertex subset for sensor placement on N network vertices (i.e., sampling vertex subset $\mathcal{C} \subset \mathcal{N} = \{1, 2, \cdots, N\}$). They do so, by maintaining an one-to-one mapping from the samples to the graph bandlimited signals. Their graph sampling theory is provided in the following:

Theorem 2 *[29–38] For any* $\mathbf{x} \in BS(\mathcal{B}, \mathbf{P}^{-1})$*, if a subset* $\mathcal{C} \subset \mathcal{N}$ *(*$\mathcal{N} = \{1, 2, \cdots, N\}$ *is the set of* N *vertex indices) satisfies*

$$rank\,(\mathbf{P}_{\mathcal{CB}}) = |\mathcal{B}|, \tag{2.8}$$

then $\mathbf{x}$ *can be recovered as* $\hat{\mathbf{x}}$ *with free of error by the samples* $\mathbf{x}_{\mathcal{C}}$ *as*

$$\hat{\mathbf{x}} = \mathbf{P}_{\mathcal{NB}} \cdot (\mathbf{P}_{\mathcal{CB}}^T \cdot \mathbf{P}_{\mathcal{CB}})^{-1} \cdot \mathbf{P}_{\mathcal{CB}}^T \cdot \mathbf{x}_{\mathcal{C}}, \tag{2.9}$$

where $|\mathcal{B}|$ *represents the number of elements of the set* $\mathcal{B}$*.* $\mathbf{P}_{\mathcal{CB}}$ *(*$\mathbf{P}_{\mathcal{NB}}$*) denotes the sub-matrix of* $\mathbf{P}$ *whose rows are selected by their indices/subscripts in* $\mathcal{C}$ *(*$\mathcal{N}$*), and whose columns are selected by their indices/subscripts in* $\mathcal{B}$*.* $\mathbf{x}_{\mathcal{C}}$ *is the samples of* $\mathbf{x}$ *by selecting indices/subscripts in* $\mathcal{C}$*.*

As such, sensor placement for a graph bandlimited signal is suggested in Theorem 2, which is equivalent to identify the sampling vertex subset $\mathcal{C}$ to ensure the full-column rank of the sub-matrix $\mathbf{P}_{\mathcal{CB}}$.

The advantages/disadvantages of using GFT for network sampling and signal recovery are provided as follows.

- Similar to CS, an advantage is the immediate signal recovery, as only independent signals on vertices are sampled and no time-evolved sample is needed. This is important for those requiring real-time and low-latency network monitoring (e.g., the real-time and low detection time monitoring of contaminant spread in WDN [14–16]

- Compared with CS framework introduced in Section 2.3.1, the GFT framework can achieve a more compact sampling vertex subset for sensor placement. We illustrate this conceptually in Fig. 2.1, and explain this in the following. At first, it is noticed that both the GFT and the CS exploit the non-zeros of a transformed version of networked signal $\mathbf{x}$. As such, by denoting the number of non-zeros as r (i.e., $|\mathcal{B}| = r$ for graph $\mathcal{B}$-bandlimited signals), it is noteworthy that the set composed of all the graph $\mathcal{B}$-bandlimited signals, i.e., $BS(\mathcal{B}, \mathbf{P}^{-1})$ from Definition 2, is actually a real subset of that composed of all r-sparse signals (as GFT knows the indices of the non-zeros, i.e., $\mathcal{B}$). See Fig. 2.1 for illustration. This indicates (also provided by Theorems 1-2) that, GFT requires a smaller $|\mathcal{C}|$ than that of CS (i.e., $r < 2r$), as the samples from the GFT framework is to recover the signal in $BS(\mathcal{B}, \mathbf{P}^{-1})$, the subset of the one the CS requires to recover. This also has another equivalent explanation.

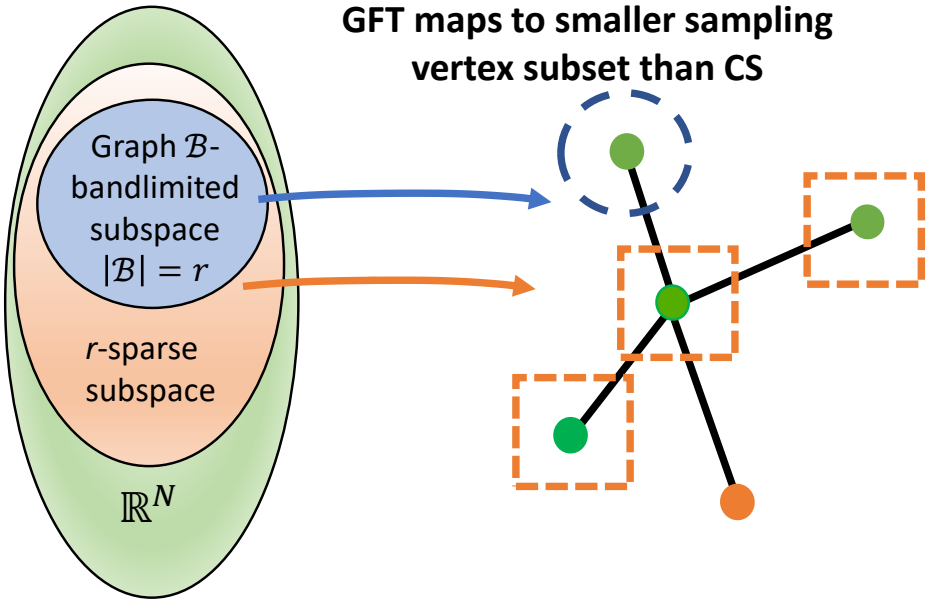

Figure 2.1: Comparison between GFT and CS from conceptual perspective.

Compared to the CS only knowing the sparsity of the transformation, the GFT framework knows the positions/indices/subscripts (that constitute set $\mathcal{B}$) of the non-zero elements in the transformation of the network signal $\mathbf{x}$, and therefore the samples of its sampling vertex subset only need to map to the whole graph $\mathcal{B}$-bandlimited frequency response.

- Whilst these studies contribute a lot to the advancement on how to select sampling vertex subset for a graph bandlimited signal, they are not suitable for sampling the time-varying network signals over the network. This is because the topology-based operator does not involve any information from the dynamic governing time-evolution model, thereby rendering its inability to maintain the graph bandlimited property for all time-evolved network signals. Recent work in [41] studies the JFT for time-varying network signals, whereby a Fourier transform (FT) operator combining with the GFT operator are used to characterize the graph bandlimited properties from the joint time and network domains. However, the sampling vertex subset derived by [41] is not time-invariant, rendering its less useful for sensor deployment in some real-world network monitoring applications (e.g., the WDN buried underground is hard to be penetrated for changing the locations of sensors with time).

2.3.3 Graph Observability Analysis: using Time-Evolved Information

In order to exploit in-depth the dynamic evolution model, network sampling driven by explicit dynamic governing equations are studied (which is also called the graph observability, i.e., to recovery every transient behaviours/states/signals from the samples). Popular approaches range from the convex optimisation [58], the causal modeling [59], and the observability analysis using linear evolution models [42–46]. Given the time-varying network signals $\mathbf{x}_k \in \mathbb{R}^N$ at $k \in \mathbb{Z}^+$ discrete-time, and its time-evolution model as $\mathbf{x}_{k+1} = \mathbf{L} \cdot \mathbf{x}_k$ ($\mathbf{L}$ of size $N \times N$ is the time-evolution matrix), the graph observability analysis is equivalent to recover the initial network signal $\mathbf{x}_1$ from the samples, so that $\mathbf{x}_k$ for any $k \in \mathbb{Z}^+$ can be recovered, i.e.,

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} = \begin{bmatrix} \mathbf{L}^0 \\ \mathbf{L}^1 \\ \vdots \\ \mathbf{L}^{K-1} \end{bmatrix} \cdot \mathbf{x}_1, \tag{2.10}$$

where K is the total discrete times that represents how long the time-evolution information is required for sampling and signal recovery.

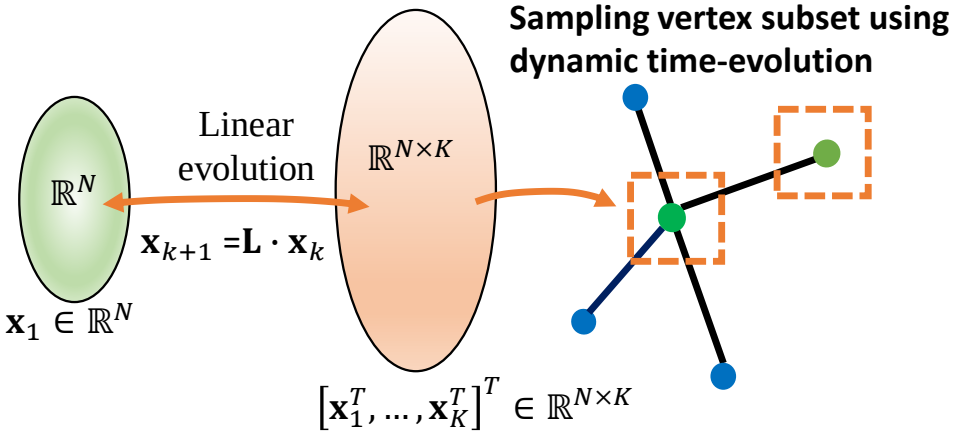

Figure 2.2: Comparison between GFT and CS from conceptual perspective.

According to Eq. (2.10), we show the selection of sampling vertex subset illustration in Fig. 2.2, where the selection of sampling vertex subset is to construct a reversible map from the time-evolution space to the sampling space. To achieve this, the selection of the sampling vertex subset $\mathcal{C} \subset \mathcal{N} = \{1, 2, \cdots, N\}$ for sensor

placement is equivalent to maintain the full-column rank, i.e., [42–46]

$$rank\left(\begin{bmatrix}\mathbf{L}_{\mathcal{CN}}^{0}\\ \mathbf{L}_{\mathcal{CN}}^{1}\\ \vdots\\ \mathbf{L}_{\mathcal{CN}}^{K-1}\end{bmatrix}\right)=N, \tag{2.11}$$

where $\mathbf{L}_{\mathcal{CN}}$ is the sub-matrix of $\mathbf{L}$ with rows' indices in set $\mathcal{C}$ and columns' indices in set $\mathcal{N}=\{1,2,\cdots,N\}$. By checking the full-column rank condition of the linear evolution model, the initial network signal $\mathbf{x}_1$ can be recovered by the samples, i.e., [42–46]

$$\hat{\mathbf{x}}_1=pinv\left(\begin{bmatrix}\mathbf{L}_{\mathcal{CN}}^{0}\\ \mathbf{L}_{\mathcal{CN}}^{1}\\ \vdots\\ \mathbf{L}_{\mathcal{CN}}^{K-1}\end{bmatrix}\right)\cdot\begin{bmatrix}\mathbf{x}_{\mathcal{C}\{1\}}\\ \mathbf{x}_{\mathcal{C}\{2\}}\\ \vdots\\ \mathbf{x}_{\mathcal{C}\{K\}}\end{bmatrix}, \tag{2.12}$$

where $pinv(\cdot)$ is the generalized inverse operator, and $\mathbf{x}_{\mathcal{C}\{k\}}$ denotes the samples from sampling vertex subset $\mathcal{C}$ at kth discrete-time.

Further studies in [46] assume a graph $\mathcal{B}$-bandlimited $\mathbf{x}_1$ to a GFT operator, and further compress the sampling vertex subset by making the rank condition in Eq. (2.11) equals $|\mathcal{B}|$, i.e., the number of elements in $|\mathcal{B}|$. We will go through this in detail in Section 6.4.2.

The advantages/disadvantages of the graph observability analysis are listed as follows.

- Given the use of time-evolved information, the graph observability analysis can be used for the cases, where signal dependency among vertices is hard to be uncovered or even does not exist (where the current vertex dependency approaches are useless). This is attributed to the usages of the time-evolved samples, which can provide more information for the recovery of time-varying network signals.

- This also constitutes the drawback, as a high recovery latency is inevitable when forward samples are collected and used. Such a drawback blocks its usages for some of the network monitoring applications, where the real-time and low-latency requirements are vital (e.g., the real-time and low detection time monitoring of contaminant spread in WDN [14–16]).

- Another obvious disadvantage is their inability to address the nonlinear dynamic governing equations, as the standard linear algebra theory cannot be

adopted to analyze the characteristics of a nonlinear evolution model, e.g., $\mathbf{x}_{k+1} = \mathbf{L}(\mathbf{x}_k)$ with nonlinear $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^N$. This is not to mention if the nonlinear equations are unavailable due to the complex underlying dynamic mechanisms (e.g., the high-dimensional and highly nonlinear Navier-Stokes dynamics with unpredictable Reynolds numbers in WDN [61]).

2.3.4 Sampling and Modeling Time-Evolution from Data

In the absence of exact dynamic models, an alternative idea is to formulate the dynamic evolution model from experimental data, so that the time-evolution information can be used for network sampling. Popular approaches include but are not limited to the machine learning, the sparsity-promoting optimization [5], and the Koopman operator [47, 50, 62–64]. Compared with the machine learning and sparsity-promoting optimization, Koopman linearization theory can derive a linearized dynamic evolution model, which is more preferable, as the time-evolved dependency can be easily analyzed using standard linear algebra theory.

Koopman operator is a linear but infinite dimensional operator that governs the time-evolution of scalar-value observables (functions) defined on the original signal space of a nonlinear dynamical system. To adopt the Koopman operator in real engineering applications, one need to design appropriate Koopman observables that can maintain the linear (quasi) time-evolution between two successive observables. This is still an open challenge, but has attracted a wide-range of researches for different nonlinear dynamical systems. The methods can be categorized as the dynamic mode decomposition (DMD), the extended DMD (EDMD) [47, 49], and the deep DMD [48]. For example, the work in [47] designs the Koopman observables using $M = O(N^2)$ key polynomial terms of Taylor expansion (e.g., the multi-elemental multiplicative terms of vertex 1 and vertex 2, $x_1 \cdot x_2$), based on which a promising Koopman operator is derived. Similarly, using the multiplications of the Logistic functions defined on each vertex, the work in [49] develops a state-inclusive vector-valued observable with proved error-bound. To further reduce the size of observables, deep-DMD is recently proposed by Yeung, Hodas, and Kundu, relying on deep Neural networks (NN). Leveraging this idea, the work in [48] further develops an auto-encoder and an auto-decoder for the reversible original signal-observable and observable-signal transformations. They do so by minimizing the mean squared errors (MSEs) of the reversible mapping between observables and original states, and of observable and original signal predictions.

Whilst many remarkable Koopman observable designs have been proposed, they are rarely adopted for network sampling applications. One work that combines

the Koopman linearization and network sampling has been performed in [50], which relies on their produced polynomial-based Koopman operator in [47]. They select a minimum number of samples using the graph observability analysis in Eqs. (2.10)-(2.11), by treating the $M = O(N^2)$ elements of the vector-valued observable as independent as in $\mathbb{R}^M$.

However, the sampling scheme in [50] and the current Koopman observable designs in [47–49] have three drawbacks, which block their usages in network sampling applications.

- First, to guarantee the linearization accuracy, the polynomial-based and logistic-based Koopman operator lead to a size explosion (i.e., $M = O(N^2)$) when addressing large-scale networks, due to their multi-elemental multiplicative term based observable design (we further analyze this in Section 6.2).

- Second, even if the deep-DMD designs in [48] can reduce the size of vector-valued observable, the learned observable may involve coupling signals on different vertices. For example, one learned observable element in [48] is $x_2 - bx_1^2$, which containing the signals on both vertex 1 and vertex 2. This is however not applicable for network sensor deployment, as a selection of leading observable elements may require sensors located on every vertices.

- Third, directly utilizing graph observability analysis (e.g., rank analysis in Eq. (2.11) on vector-valued observable (of size $M \times 1$ defined on the original signal of size $N \times 1$) neglects the intrinsic nonlinear dependency between the elements of the vector-value observable, which are all determined by the originally lower-sized signal-space. For example, x_1, x_2, x_1x_2, $x_1^2x_2^2$ are all elements of the poly-based vector-valued observable in [47], which cannot be treated as 4 independent elements, as they are determined by original signals x_1 and x_2. Therefore, treating them as $\mathbb{R}^4$ does not make sense, and will lead to extra redundant sampling vertices for signal recovery (we will explain this in greater detail in Sections 6.2 and 6.5).

2.4 Conclusions & Discussions

This chapter reviews the current sampling and compression approaches, which were categorized by relying on the signal dependency among network vertices, and the time-evolution information. The former requires an operator to uncover and characterize the dependencies of the signals on different vertices, rendering a huge difficulty for current topology-only GFT framework. The latter regards the linear/linearized

time-evolution dynamic model as the prerequisite for further exploration of sequential information and sampling selection, thereby making it less practical for most of the real-world applications without explicit dynamic models. These two challenges then motivated us to develop (i) dynamic mechanism-topology combined GFT sampling, and (ii) linearized dynamic time-evolution modelling and sampling, which will be elaborated in the following chapters.

Chapter 3

Equation Driven Sampling

In this chapter, we elaborate the proposed dynamic equation-driven network sampling that relies on the signal dependency among vertices. For this chapter, we assume a Lyapunov stability that is maintained by the time-varying network signals whose systems work on stable area. Leveraging this, we study how to place sensor and how to discretize time for sampling and recovering the continuous network signals.

The structure of the rest of this chapter is as follows. We first introduce the network dynamic mechanisms governed by the explicit dynamic equations in Section 3.1. Then, in Section 3.2, we provide the equation-driven sampling method on both time-domain and network-domain, and give a relationship between such two domains. Sections 3.3-3.4 provide the simulation and experimental results. We finally conclude this chapter in Section 3.5.

3.1 System Model & Problem Formulation

Signal processing on time-varying network signals is concerned with the analysis and processing on a dynamic signal-space, where individual signals on vertices are interacted with each other with respect to both the network topology and the governing dynamic mechanisms (shown in Fig. 3.1). Here, we denote the network topology as $G(\mathcal{N}, \mathbf{A})$. $\mathcal{N} = \{1, 2, \cdots, N\}$ represents the index set of total N vertices in network. $\mathbf{A}$ gives the adjacent matrix, in which the (n, m)th element $a_{n,m} \in \{0, 1\}$ represents the existence of a direct link from vertex m to vertex n (i.e., $a_{n,m} = 1$ means the link from m to n exists, otherwise $a_{n,m} = 0$). In this view, the signal $x_n(t)$ on vertex n with continuous time t can be expressed as a differential-type evolution with respect to the self-dynamic function $f(\cdot) : \mathbb{R} \to \mathbb{R}$, the mutualistic coupling function with

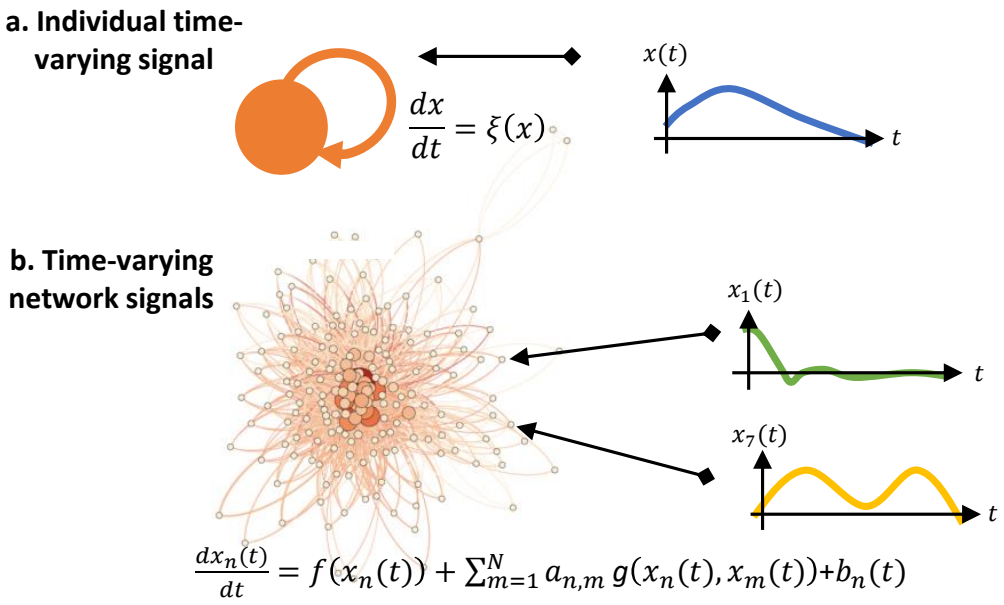

Figure 3.1: Illustration of signals governed by differential equations: (a) individual signal, and (b) time-varying network signals.

link (n, m), i.e., $g(x_n(t), x_m(t)) : \mathbb{R}^2 \to \mathbb{R}$, and an unknown input $b_n(t)$. We express such differential equations in the following:

$$\frac{dx_n(t)}{dt} = f(x_n(t)) + \sum_{m=1}^{N} a_{n,m} \cdot g(x_n(t), x_m(t)) + b_n(t). \tag{3.1}$$

Here, the input is assumed to be a group of Dirac delta functions. Such an input can be interpreted as a control signal for automation, or an information signal transmitted via the dynamic channel for communications. We express the aforementioned input as:

$$b_n(t) = \sum_{i=1}^{+\infty} b_{n,i} \cdot \delta(t - t_i), \tag{3.2}$$

where $\delta(\cdot)$ represents the Dirac delta function, and $t_i, i \in \mathbb{N}^+$ is the specific time for an input $b_{n,i}$ with random amplitude. We stack $\mathbf{b}_i = [b_{1,i}, b_{2,i}, \cdots, b_{N,i}]^T$.

Given that the dynamic mechanism in Eq. (3.1) defined on both time-domain

and network domain, we hereby denote the continuous network signals at time t as:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix}, \tag{3.3}$$

and the simplified dynamic governing equations of Eq. (3.1) as:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{\Xi}(\mathbf{x}(t)) + \mathbf{b}(t), \tag{3.4}$$

where $\mathbf{\Xi} : \mathbb{R}^N \to \mathbb{R}^N$ is the vector-valued time-evolution function, composed by $\mathbf{\Xi}(\cdot) = [\xi_1(\cdot), \xi_2(\cdot), \cdots, \xi_N(\cdot)]^T, \xi_n(\cdot) : \mathbb{R}^N \to \mathbb{R}$, and $\mathbf{b}(t) = [b_1(t), b_2(t), \cdots, b_N(t)]^T$ is the vector-valued inputs. Here, different from the traditional graph studies that consider static network signal (i.e., a fixed data on each vertex) [28–30,32–38], we in this chapter consider the continuous network signals on both the time and network -domains, given the knowledge of the explicit dynamic equations in Eq. (3.1).

The purpose of this chapter is to analyze how to determine the sampling vertex subset for sensor placement from network domain and the discretizing rate (i.e., the sampling frequency) from time-domain, so that the discretized and sampled signals can ensure the recovery of the original time-varying network signals. In mathematical manner, we define the sampling (angular) frequency as ω_{s} from the time-domain, and the time-invariant sampling vertex subset $\mathcal{C} \subset \mathcal{N}$ from the network domain. The aim of this chapter is then converted to compute the suitable ω_{s} and $\mathcal{C}$. Here, it is noteworthy that the time-invariant property of sampling vertex subset $\mathcal{C}$ is important; otherwise, one should change the sensor deployment with time, which is challenging and even impractical in some real network sensing applications (e.g., the pollutant surveillance in WDN). As such, the traditional graph sampling Theorem 2 cannot be directly used, since the topology based GFT operator is unable to characterize the dependency of the continuous signals on different vertices, and thereby will lead to time-varying sampling strategy. Thus, in this chapter, we will construct a combined dynamic-topology GFT operator, and extend the traditional static graph signals to time-varying network signals.

3.2 Sampling for Dynamic Network Signal

In this section, we elaborate our joint time and network domains sampling methods for the sampling and recovery of the time-varying network signals. The purpose here

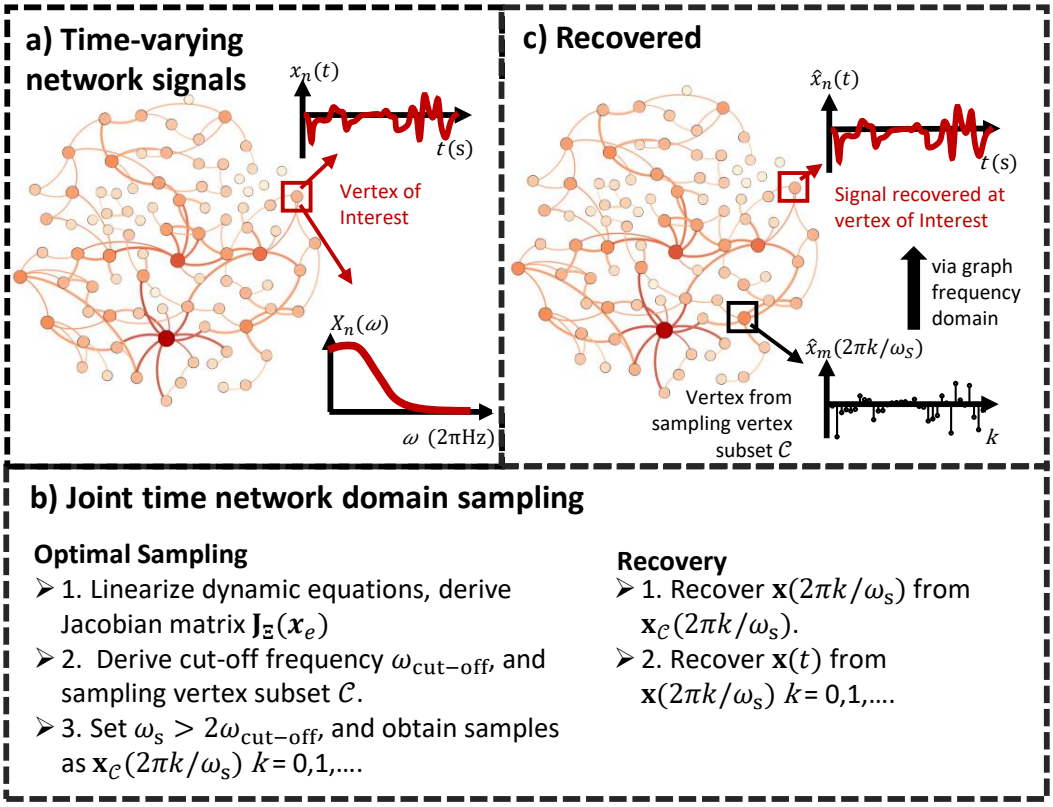

Figure 3.2: Illustration of (a) time-varying network signals; (b) the joint time and network domain sampling and recovery; and (c) the recovered time-varying network signals.

is to (i) identify the time-invariant sampling vertex subset $\mathcal{C} \subset \mathcal{N}$ from the network domain, and (ii) compute the cut-off frequency $\omega_{\text{cut-off}}$ from the time-frequency domain. The schematic flow of the designs are illustrated in Fig. 3.2

By doing so, one can recover the time-varying network signals as:

$$\hat{\mathbf{x}}(t) = \boldsymbol{\Theta} \cdot \mathbf{X}_{\mathcal{C}} \cdot \mathbf{R}. \tag{3.5}$$

In Eq. (3.5), $\mathbf{R} = [sinc(\omega_s t/(2\pi)), \cdots, sinc(\omega_s t/(2\pi) - K)]^T$ is the Nyquist interpolation matrix. $\boldsymbol{\Theta}$ denotes the recovery matrix from the network domain, which will be specified later in this chapter. $\mathbf{X}_{\mathcal{C}}$ represents the the discrete samples on vertices

in sampling vertex subset $\mathcal{C}$, read:

$$\mathbf{X}_{\mathcal{C}} = \begin{bmatrix} x_{n_1}\left(\frac{2\pi}{\omega_{\mathrm{s}}} \cdot 0\right) & \cdots & x_{n_1}\left(\frac{2\pi}{\omega_{\mathrm{s}}} \cdot K\right) \\ \vdots & \ddots & \vdots \\ x_{n_{|\mathcal{C}|}}\left(\frac{2\pi}{\omega_{\mathrm{s}}} \cdot 0\right) & \cdots & x_{n_{|\mathcal{C}|}}\left(\frac{2\pi}{\omega_{\mathrm{s}}} \cdot K\right), \end{bmatrix} \tag{3.6}$$

with $\omega_{\mathrm{s}} \geq 2\omega_{\text{cut-off}}$ the sampling angular frequency, and $2\pi K/\omega_{\mathrm{s}}$ the total discrete-times of interest.

3.2.1 Assumptions of Network Stability

In this work, we consider the dynamic systems that work on the stable area. To characterize this, we assume that the system in Eq. (3.1) works on the Lyapunov stability area of an equilibrium point $\mathbf{x}_e$.

Assumption 1 *[65] Continuous network signals $\mathbf{x}(t)$ work on the Lyapunov stability area of point $\mathbf{x}_e$, if and only if the following condition is satisfied. For any $\epsilon > 0$, if $||\mathbf{x}(0) - \mathbf{x}_e|| < \delta$ for some positive δ, then*

$$||\mathbf{x}(t) - \mathbf{x}_e|| < \epsilon.$$

By assuming the Lyapunov stability of a dynamic system, the corresponding non-linear time-varying network signals can be approximated via the linearized parts. We next discuss the linearizing approximation process and measure the caused error.

Linearize Dynamics

Given the Lyapunov stability in Assumption 1, $\mathbf{x}(t)$ is converging to $\mathbf{x}_e$ with time $t \to +\infty$, i.e., $\mathbf{x}(\infty) = \mathbf{x}_{\mathrm{e}}$. For convenience, we write:

$$\mathbf{z}(t) = \mathbf{x}(t) - \mathbf{x}_{\mathrm{e}}, \tag{3.7}$$

and thus $\lim_{t\to+\infty} \mathbf{z}(t) = \mathbf{z}(+\infty) = \mathbf{0}$. As such, the further analysis of the time-varying network signals governed by the nonlinear differential equations in Eq. (3.1) can be converted to the study of its linear approximations, i.e.,

$$\begin{aligned} \frac{d\mathbf{z}(t)}{dt} &= \mathbf{J}_{\Xi}(\mathbf{x}_e) \cdot \mathbf{z}(t) + o\left(\|\mathbf{z}(t)\|\right) + \mathbf{b}(t) \\ &\approx \mathbf{J}_{\Xi}(\mathbf{x}_e) \cdot \mathbf{z}(t) + \mathbf{b}(t), \end{aligned} \tag{3.8}$$

in which $o(\|\mathbf{z}(t)\|)$ represents the high-order terms of $\mathbf{z}(t)$ that converge to zero faster than the first-order term when $t \to +\infty$. $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ represents the Jacobian first-order derivative matrix valued on $\mathbf{x}_e$, i.e.,

$$\mathbf{J}_{\Xi}(\mathbf{x}_e) \triangleq \left[\begin{array}{ccc} \frac{\partial \xi_1(t)}{\partial x_1(t)} & \cdots & \frac{\partial \xi_1(t)}{\partial x_N(t)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \xi_N(t)}{\partial x_1(t)} & \cdots & \frac{\partial \xi_N(t)}{\partial x_N(t)} \end{array}\right] \Bigg|_{\mathbf{x}(t)=\mathbf{x}_e}, \tag{3.9}$$

where functions $\boldsymbol{\Xi}(\cdot)$ and $\xi_n(\cdot)$ are defined in Eq. (3.4).

For the convenience to observe and analyze the structure of $\mathbf{J}_{\Xi}(\mathbf{x}_e)$, we re-write Eq, (3.9) in an intuitive form, by taking Eq. (3.1) into Eq. (3.9), i.e.,

$$\mathbf{J}_{\Xi}(\mathbf{x}_e) = \mathbf{A} \circ \mathbf{J}_g(\mathbf{x}_e) + diag\left\{\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_N}\right\}\Bigg|_{\mathbf{x}_e}, \tag{3.10}$$

where $\circ$ denotes the *Hadamard product*, and $\mathbf{J}_g(\mathbf{x}_e)$ is the Jacobian matrix of function $g(\cdot,\cdot)$ valued on $\mathbf{x}_e$. From Eq. (3.10), it is observed that $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ is composed of both the topology of the network (as is represented by the adjacency matrix $\mathbf{A}$), and the governing dynamic equations, i.e., the coupling functions $g(\cdot,\cdot)$ and the self-dynamics $f(\cdot)$ in Eq. (3.1).

Linearization Error

After the derivation of the linearized model, we measure the accuracy of such linearized signals. We do so by the use of the effectiveness equation proposed in [11,12]. In essence, effectiveness equation of a dynamical system is to compress the N-dimension signals (indexed on N vertices) into a single dimension. For example, given the time-varying network signals as $\mathbf{z}(t)$, the effectiveness is computed as [11]:

$$\frac{dz_{\text{eff}}(t)}{dt} = \beta_{\text{eff}} \cdot z_{\text{eff}}(t), \tag{3.11}$$

where $z_{\text{eff}}(t) = \mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{z}(t)/(\mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{1})$, and $\beta_{\text{eff}} = \mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{d}^{(\text{in})}$, with the nth entry of $\mathbf{d}^{(\text{in})}$ as $d_n = \sum_{m=1}^{N} a_{n,m}$, and $\mathbf{1} \triangleq [1, \cdots, 1]^T$ of size $N \times 1$.

As such, using the single dimensional effectiveness $z_{\text{eff}}(t)$ as the compression of the N-dimensional dynamic $\mathbf{z}(t)$, we are able to quantify the linearization error. In mathematical manner, we denote $d\mathbf{z}_{\text{l}}(t)/dt = \mathbf{J}_{\Xi}(\mathbf{x}_e)\mathbf{z}_{\text{l}}(t)$. Then, the l_1-norm linearization error and its measurement via effectiveness can be expressed as follows:

$$err = \int_0^{+\infty} \|\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t)\|_{l_1} dt, \tag{3.12}$$

$$c_1 \cdot err \leq \int_0^{+\infty} |z_{\text{l,eff}}(t) - z_{\text{eff}}(t)| dt \leq c_2 \cdot err, \tag{3.13}$$

where c_1 and c_2 are some positive constants, and $z_{\text{l,eff}}(t) = \mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{z}_{\text{l}}(t)/(\mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{1})$ can be computed via Eq. (3.11).

The proof of Eq. (3.13) is given in the following, separated by its left-hand side and right-hand side. We first prove the right-hand side by:

$$\begin{aligned} |z_{\text{l,eff}}(t) - z_{\text{eff}}(t)| &= \frac{|\mathbf{1}^T \cdot \mathbf{A} \cdot (\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t))|}{|\mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{1}|} \\ &\leq \frac{\|\mathbf{A} \cdot (\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t))\|_{l_1}}{|\mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{1}|} \\ &\leq \frac{\|\mathbf{A}\|_{l_1}}{|\mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{1}|} \cdot \|\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t)\|_{l_1}. \end{aligned} \tag{3.14}$$

Then, the left-hand side of Eq. (3.13) can be proved by evaluating the minimum of

$$\begin{aligned} c(t) &= \frac{|z_{\text{l,eff}}(t) - z_{\text{eff}}(t)|}{\|\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t)\|_{l_1}} \\ &= \frac{1}{|\mathbf{1}^T \cdot \mathbf{A} \cdot \mathbf{1}|} \cdot \frac{|\mathbf{1}^T \cdot \mathbf{A} \cdot (\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t))|}{\|\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t)\|_{l_1}}. \end{aligned} \tag{3.15}$$

From Eq. (3.15), we notice that the minimal value of $c(t)$ is zero if $\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t)$ takes vectors from the null-space of $\mathbf{A}$, denoted as $null(\mathbf{A})$. However, it is noteworthy that $\mathbf{z}_{\text{l}}(t) - \mathbf{z}(t)$ cannot belong to $null(\mathbf{A})$ for all $t \in (0, +\infty)$. Therefore, $c(t) > 0$ holds for some $t > 0$, and subsequently a positive c_1 exists to maintain the left-hand side of Eq. (3.13).

Given the liearization of the differential model for the time-varying network signals, we will then analyze the Jacobian matrix $\mathbf{J}_{\Xi}(\mathbf{x}_e)$, and derive the optimal sampling vertex subset $\mathcal{C}$ and the cut-off frequency $\omega_{\text{cut-off}}$ for sampling and discretizing the time-varying network signals.

3.2.2 Sampling from Network Domain

In this part, we study how to determine the sampling vertex subset $\mathcal{C}$ from the network domain, so that the time-varying network signals can be recovered from the samples. We first analyze the time-varying network signals whose initialization and inputs are graph bandlimited to a GFT operator. Then, the general cases with arbitrary initialization and inputs are studied.

Before we start, we construct the equation-driven GFT operator, denoted as

$\mathbf{Q}^{-1}$. To do so, we decompose $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ by eigen-decomposition:

$$\mathbf{J}_{\Xi}(\mathbf{x}_e) = \mathbf{Q} \cdot diag\{\mu_1, \cdots, \mu_N\} \cdot \mathbf{Q}^{-1}, \tag{3.16}$$

where $\mu_i, i \in \{1, \cdots, N\}$ is the ith eigenvalue, and $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_N]$ is the reversible matrix, whose columns are the corresponding eigenvectors. As such, we assign $\mathbf{Q}^{-1}$ as the GFT operator. This is reasonable, as $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ characterizes the first-order evolution of the time-varying network signals, and embraces the information from both the governing dynamic equations and the network topological structure (see Eq. (3.10)).

Graph Bandlimited Signals

We study the case where the initial network signal $\mathbf{z}(0)$ and the unknown inputs $\mathbf{s}_i$ are $\mathcal{B} \subsetneq \{1, 2, \cdots, N\}$-bandlimited with respect to the GFT operator $\mathbf{Q}^{-1}$. In fact, this case holds for a wide range of the time-varying network signals [46], among which we can list the auto-regressive moving average graph process, the wave propagation, and the signal diffusion.

Recalling to the Definitions 1-2, we can construct a subspace, i.e., $BS(\mathcal{B}, \mathbf{Q}^{-1})$, and have $\mathbf{z}(0), \mathbf{b}_i = [b_{1,i}, b_{2,i}, \cdots, b_{N,i}]^T \in BS(\mathcal{B}, \mathbf{Q}^{-1})$ given the $\mathcal{B}$-bandlimitedness property of the initialization $\mathbf{z}(0)$ and the inputs $\mathbf{b}_i$. Then, we provide in the following Lemma 1 and Theorem 3 to show that the whole time-varying network signals belong to the subspace, i.e., $\mathbf{z}(t) \in BS(\mathcal{B}, \mathbf{Q}^{-1})$.

Lemma 1 *Denote* $d(\mathbf{z}(t))/dt = \mathbf{J}_{\Xi}(\mathbf{x}_e) \cdot \mathbf{z}(t)$. *If* $\mathbf{z}(0) \in BS(\mathcal{B}, \mathbf{Q}^{-1})$, *then* $\mathbf{z}(t) \in BS(\mathcal{B}, \mathbf{Q}^{-1})$.

Proof 2 *At first,* $\mathbf{z}(t)$ *has a closed-form expression as:*

$$\mathbf{z}(t) = e^{t \cdot \mathbf{J}_{\Xi}(\mathbf{x}_e)} \cdot \mathbf{z}(0). \tag{3.17}$$

Then, the graph Fourier transformation of $\mathbf{z}(t)$ *with respect to the GFT operator*

$\mathbf{Q}^{-1}$ *is:*

$$\begin{aligned}\tilde{\mathbf{z}}(t) =&\mathbf{Q}^{-1}\cdot e^{t\cdot\mathbf{J}_{\Xi}(\mathbf{x}_e)}\cdot\mathbf{z}(0)\\ =&\mathbf{Q}^{-1}\cdot\sum_{k=0}^{\infty}\frac{t^k}{k!}\cdot\mathbf{J}_{\Xi}(\mathbf{x}_e)^k\cdot\mathbf{z}(0)\\ =&\sum_{k=0}^{+\infty}\frac{t^k}{k!}\cdot diag\{\mu_1,\cdots,\mu_N\}^k\cdot\mathbf{Q}^{-1}\cdot\mathbf{z}(0)\\ =&\sum_{k=0}^{+\infty}\frac{t^k}{k!}\cdot diag\{\mu_1^k,\cdots,\mu_N^k\}\cdot\tilde{\mathbf{z}}(0),\end{aligned}\tag{3.18}$$

where $\tilde{\mathbf{z}}(0)=\mathbf{Q}^{-1}\cdot\mathbf{z}(0)$ *and* $\mathbf{J}_{\Xi}(\mathbf{x}_e)^0=\mathbf{I}_{N\times N}$ *is the identity matrix of size* $N\times N$*. From Eq. (3.18), we can observe that the positions of the non-zero elements in* $\tilde{\mathbf{z}}(t)$ *is consistent with those in* $\tilde{\mathbf{z}}(0)$*, suggesting that the non-zero subscripts all belong to* $\mathcal{B}$*. Hence, according to Definition 2, we prove that* $\mathbf{z}(t)\in BS(\mathcal{B},\mathbf{Q}^{-1})$*, suggesting that the whole time-varying network signals* $\mathbf{z}(t)$ *are* $\mathcal{B}$*-bandlimited with respect to the GFT operator* $\mathbf{Q}^{-1}$*.*

Theorem 3 *Denote* $d(\mathbf{z}(t))/dt=\mathbf{J}_{\Xi}(\mathbf{x}_e)\mathbf{z}(t)+\mathbf{b}(t)$*. If* $\mathbf{z}(0),\mathbf{b}_i=[b_{1,i},\cdots,b_{N,i}]^T\in BS(\mathcal{B},\mathbf{Q}^{-1})$*, then* $\mathbf{z}(t)\in BS(\mathcal{B},\mathbf{Q}^{-1})$*.*

Proof 3 *Recall from Eq. (3.2) that* $\mathbf{b}(t)=\sum_{i=1}^{+\infty}\mathbf{b}_i\circ\boldsymbol{\delta}(t-t_i)$ *with* $\boldsymbol{\delta}(t-t_i)=[\delta(t-t_i),\cdots,\delta(t-t_i))]^T$ *of size* $N\times 1$*. The proof is equivalent to prove* $\mathbf{z}(t)\in BS(\mathcal{B},\mathbf{Q}^{-1})$ *for any* $t\in[0,t_1)\cup\cdots\cup[t_{+\infty},+\infty)$*. According to Lemma 1, it is proved the* $\mathbf{z}(t)\in BS(\mathcal{B},\mathbf{Q}^{-1})$ *for any* $t\in[0,t_1)$*. Then, we notice that* $\mathbf{z}(t_1)=e^{t_1\cdot\mathbf{J}_{\Xi}(\mathbf{x}_e)}\cdot\mathbf{z}(0)+\mathbf{b}_1$*. Thus, if* $\mathbf{b}_1\in BS(\mathcal{B},\mathbf{Q}^{-1})$*, then* $\mathbf{z}(t_1)\in BS(\mathcal{B},\mathbf{Q}^{-1})$*. Taking* $\mathbf{z}(t_1)$ *as the graph bandlimited initialization, the proof of* $\mathbf{z}(t)\in BS(\mathcal{B},\mathbf{Q}^{-1})$ *for interval* $t\in[t_1,t_2)$ *is straightforward according to Lemma 1. Similarly, we can extend this fact for all the intervals, and therefore prove* $\mathbf{z}(t)\in BS(\mathcal{B},\mathbf{Q}^{-1})$*.*

As we complete the proof of the $\mathcal{B}$-bandlimited property of the whole time-varying network signals $\mathbf{z}(t)$, we can determine the sampling vertex subset $\mathcal{C}$ from the network domain via the existing Theorem 2, i.e., $rank(\mathbf{Q}_{\mathcal{CB}})=|\mathcal{B}|$. To implement this, we compute $\mathcal{C}$ by finding the minimum condition number of the matrix $\mathbf{Q}_{\mathcal{CB}}$, i.e.,

$$\mathcal{C}=\underset{\mathcal{C}\subset\mathcal{N}}{\operatorname{argmin}}\,cond\left(\mathbf{Q}_{\mathcal{CB}}\right),\tag{3.19}$$

where $cond(\cdot)$ denotes the condition number of a matrix. Leveraging Eq. (3.19), a greedy algorithm can be adopted by finding and adding the row, i.e., $\mathcal{C}\leftarrow\mathcal{C}\cup\{n\}$, such that $n=\operatorname{argmin}_{i\in\mathcal{N}\setminus\mathcal{C}}cond(\mathbf{Q}_{(\mathcal{C}\cup\{i\})\mathcal{B}})$. Then, the recovery matrix $\boldsymbol{\Theta}$ in Eq. (3.5) can be derived as $\boldsymbol{\Theta}=\mathbf{Q}_{\mathcal{NB}}\cdot(\mathbf{Q}_{\mathcal{CB}}^T\cdot\mathbf{Q}_{\mathcal{CB}})^{-1}\cdot\mathbf{Q}_{\mathcal{CB}}^T$, and the recovery process can

be pursued by $\hat{\mathbf{z}}(t) = \mathbf{\Theta} \cdot \mathbf{z}_{\mathcal{C}}(t)$, with $\mathbf{z}_{\mathcal{C}}(t)$ the sampled continuous signals selected from the vertices in $\mathcal{C}$.

General Case with Arbitrary Initialization

It is noteworthy that an existence of $\mathcal{B} \subsetneq \{1, 2, \cdots, N\}$ to make initialization and inputs $\mathcal{B}$-bandlimited may not be easily satisfied for some network monitoring scenarios. To address this, we assign a bandwidth set $\mathcal{B}$ and approximately regard the initialization and inputs as $\mathcal{B}$-bandlimited. Here, $\mathcal{B}$ is selected by the indices of the $|\mathcal{C}|$ smallest magnitudes of the real parts of eigenvalues, i.e.,

$$\mathcal{B} = \left\{ n_i | i \in \{1, 2, \cdots, |\mathcal{C}|\}, |Re[\mu_{n_1}]| \leq |Re[\mu_{n_2}]| \leq \cdots \leq |Re[\mu_{n_{|\mathcal{C}|}}]| \right\}, \tag{3.20}$$

where $|\mathcal{C}|$ is pre-defined according to the limit number of sensors or accuracy indicators.

We explain the reason in the following. According to Eq. (3.17), the graph Fourier transform of $\mathbf{z}(t)$ for jth graph-frequency component is:

$$\tilde{z}_j(t) = e^{\mu_j \cdot t} \cdot \tilde{z}_j(0) + \sum_{t_i \leq t} e^{\mu_j \cdot (t - t_i)} \cdot \tilde{b}_{j,i}, \tag{3.21}$$

where $\tilde{z}_j(t)$ is the jth element in $\tilde{\mathbf{z}}(t) = \mathbf{Q}^{-1} \cdot \mathbf{z}(t)$, and $\tilde{b}_{j,i}$ is the jth element in $\tilde{\mathbf{b}}_i = \mathbf{Q}^{-1} \cdot \mathbf{b}_i$. Then, the energy measured by the integral on t is:

$$E_j = \int_0^{+\infty} |\tilde{z}_j(t)| dt = \frac{|\tilde{z}_j(0) + \sum_i \tilde{b}_{j,i}|}{|\mathrm{Re}[\mu_j]|}, \tag{3.22}$$

where we have $\mathrm{Re}[\mu_j] \leq 0$ given the Lypunov stability assumption [65]. As such, given the non-bandlimited property of both $\mathbf{z}(0)$ and $\mathbf{b}_i$, one option for the selection of the graph sampling bandwidth set $\mathcal{B}$ is to minimize the energies of the un-selected graph-frequency components. This thereby make us omit the $N - |S|$ largest $|\mathrm{Re}[\mu_j]|$, since $|\mathrm{Re}[\mu_j]|$ constitutes the denominator of the energy according to Eq. (3.22). Also, an upper-bound recovery error of such approximation can be deduced as follows:

$$\int_0^{+\infty} \|\hat{\mathbf{z}}(t) - \mathbf{z}(t)\|_2 dt \leq \frac{\|\mathbf{z}(0)\|_2 + \sum_i \|\mathbf{b}_i\|_2}{|\mathrm{Re}[\mu_{n_{|\mathcal{C}|+1}}]|} \cdot \frac{N - |\mathcal{C}|}{N}, \tag{3.23}$$

where $|\mathrm{Re}[\mu_{n_{|\mathcal{C}|+1}}]|$ represents the $(|\mathcal{C}|+1)$th smallest $|\mathrm{Re}(\mu_n)|$ for $n \in \{1, 2, \cdots, N\}$.

After the derivation of an approximated $\mathcal{B}$, we regard the initialization and inputs as $\mathbf{z}(0), \mathbf{b}_i \in BS(\mathcal{B}, \mathbf{Q}^{-1})$, and further from Lemma 1 and Theorem 3, we

have $\mathbf{z}(t) \in BS(\mathcal{B}, \mathbf{Q}^{-1})$. As such, we can select the sampling vertex subset $\mathcal{C}$, and the recovery matrix $\mathbf{\Theta}$, as was stated by Eq. (3.19). Such a derivation is for the arbitrary initialization and inputs for sampling and recovering the time-varying network signals.

Here, it is highlighted that such an equation-driven sampling method does not rely on the time-evolved information for signal recovery. It is true that as one derives a closed-form formula of $d\mathbf{z}(t)/dt = \mathbf{J}_{\Xi}(\mathbf{x}_e)\mathbf{z}(t)$ as $\mathbf{z}(t) = e^{t\cdot\mathbf{J}_{\Xi}(\mathbf{x}_e)} \cdot \mathbf{z}(0)$ in Eq. (3.17), one can determine the sampling vertex subset such that the time-evolved samples (e.g., $\mathbf{z}_{\mathcal{C}}(t), t > 0$) can be used for the recovery of $\mathbf{z}(0)$ and then take $\mathbf{z}(0)$ into Eq. (3.17) to derive the whole time-varying network signals. However, such a case is an ideal one without the unknown input $\mathbf{b}(t)$, which if added, makes the dynamic governing equation as $d\mathbf{z}(t)/dt = \mathbf{J}_{\Xi}(\mathbf{x}_e)\mathbf{z}(t)+\mathbf{b}(t)$. When we try to address the unknown spatial-temporal patterns of the input (i.e., the exact time of the Dirac delta functions and the amplitudes), we have to select the sampling vertex subset $\mathcal{C}$ which requires only the current-time samples for the recovery of the time-varying network signals. The time-evolved information is no longer reliable, as we do not know whether there is an input at next monitoring time.

In the following, we will analyze when to monitor the network from the time-domain, i.e., the discrete time (or sampling frequency $\omega_{\rm s}$) that can ensure the discrete sampled data to recover the continuous network signals.

3.2.3 Sampling from Time Domain

After the derivation of the sampling vertex subset $\mathcal{C}$, and the recovery matrix $\mathbf{\Theta}$, we will elaborate how to determine the time-domain cut-off frequency $\omega_{\text{cut-off}}$.

Lemma 2 *Denote $d\mathbf{z}(t)/dt = \mathbf{J}_{\Xi}(\mathbf{x}_e) \cdot \mathbf{z}(t)$ satisfying Lyapunov stability, and a threshold ε to truncate the time-domain frequency components that are less than ε. Then, the time-domain cut-off frequency $\omega_{cut\text{-}off}$ is:*

$$\omega_{cut\text{-}off} = \max_{j\in\mathcal{B}} |Im[\mu_j]| + \sqrt{\frac{\|\mathbf{z}(0)\|_2^2}{\varepsilon^2} - \min_{j\in\mathcal{B}} Re^2[\mu_j]}. \tag{3.24}$$

Proof 4 *With the help of Eq. (3.17), we write the signal on nth vertex, i.e., $z_n(t)$ of $\mathbf{z}(t)$, as:*

$$z_n(t) = \sum_{j\in\mathcal{B}} q_{n,j} \cdot \tilde{z}_j(0) \cdot e^{\mu_j t}, \tag{3.25}$$

where $q_{n,j}$ is the (n,j)th entry of the matrix $\mathbf{Q}$. Given that $\mathbf{z}(t)$ is assumed to have Lyapunov stability, the eigenvalues of $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ have non-positive real values [65], i.e., $Re[\mu_j] \leq 0$. As such, we deduce the time-frequency Fourier transform of $z_n(t)$ as follows:

$$\begin{aligned} Z_n(\omega) =& \int_0^{+\infty} \sum_{j\in\mathcal{B}} q_{n,j} \cdot \tilde{z}_j(0) \cdot e^{\mu_j t} \cdot e^{-i\omega t} dt \\ =& \sum_{j\in\mathcal{B}} q_{n,j} \cdot \tilde{z}_j(0) \int_0^{+\infty} e^{Re[\mu_j]t - i(\omega - Im[\mu_j])t} dt \\ =& \sum_{j\in\mathcal{B}} \frac{q_{n,j} \cdot \tilde{z}_j(0)}{-Re[\mu_j] + i\,(\omega - Im[\mu_j])}. \end{aligned} \tag{3.26}$$

We then extract the magnitude of $Z_n(\omega)$ from Eq. (3.26) as:

$$\begin{aligned} |Z_n(\omega)| =& \left| \sum_{j\in\mathcal{B}} \frac{q_{n,j} \cdot \tilde{z}_j(0)}{-Re[\mu_j] + i\,(\omega - Im[\mu_j])} \right| \\ \leq& \sum_{j\in\mathcal{B}} \left| \frac{q_{n,j} \cdot \tilde{z}_j(0)}{-Re[\mu_j] + i\,(\omega - Im[\mu_j])} \right| \\ \leq& \sum_{j\in\mathcal{B}} \frac{|q_{n,j} \cdot \tilde{z}_j(0)|}{\sqrt{Re^2[\mu_j] + (\omega - Im(\mu_j))^2}}. \end{aligned} \tag{3.27}$$

It is observed from Eq. (3.27) that the imaginary parts of the eigenvalues contribute to the left/right shift of ω. Therefore, we can deduce that, for any $\omega > \max_{j\in\mathcal{B}} |Im[\mu_j]|$,

an upper-bound of the the magnitude $|Z_n(\omega)|$ *as:*

$$\begin{aligned}
|Z_n(\omega)| &< \frac{\sum_{j=1}^{N} |q_{n,j} \cdot \tilde{z}_j(0)|}{\sqrt{\min_{j\in\mathcal{B}} Re^2[\mu_j] + \left(\omega - \max_{j\in\mathcal{B}} |Im[\mu_j]|\right)^2}} \\
&< \frac{\sqrt{\sum_{j=1}^{N} q_{n,j}^2 \sum_{j=1}^{N} \tilde{z}_j(0)^2}}{\sqrt{\min_{j\in\mathcal{B}} Re^2[\mu_j] + \left(\omega - \max_{j\in\mathcal{B}} |Im[\mu_j]|\right)^2}} \\
&= \frac{\|\tilde{\mathbf{z}}(0)\|_2}{\sqrt{\min_{j\in\mathcal{B}} Re^2[\mu_j] + \left(\omega - \max_{j\in\mathcal{B}} |Im[\mu_j]|\right)^2}} \\
&= \frac{\|\mathbf{Q}\cdot\tilde{\mathbf{z}}(0)\|_2}{\sqrt{\min_{j\in\mathcal{B}} Re^2[\mu_j] + \left(\omega - \max_{j\in\mathcal{B}} |Im[\mu_j]|\right)^2}} \\
&= \frac{\|\mathbf{z}(0)\|_2}{\sqrt{\min_{j\in\mathcal{B}} Re^2[\mu_j] + \left(\omega - \max_{j\in\mathcal{B}} |Im[\mu_j]|\right)^2}}.
\end{aligned} \tag{3.28}$$

As such, we make such upper-bound of $|Z_n(\omega)|$*, i.e., the right-hand side of Eq. (3.28), smaller than the given threshold* ε*, and therefore derive the cut-off frequency* $\omega_{cut\text{-}off}$ *given by Lemma 2.*

Theorem 4 *Denote* $d\mathbf{z}(t)/dt = \mathbf{J}_\Xi(\mathbf{x}_e)\mathbf{z}(t) + \mathbf{b}(t)$ *satisfying the Lyapunov stability, and a threshold* ε *to truncate the time-domain frequency components that are less than* ε*. Then, the time-domain cut-off frequency* $\omega_{cut\text{-}off}$ *is:*

$$\omega_{cut\text{-}off} = \max_{j\in\mathcal{B}} |Im[\mu_j]| + \sqrt{\frac{\|\mathbf{z}(0)\|_2^2 + \sum_i \|\mathbf{b}_i\|_2^2}{\varepsilon^2} - \min_{j\in\mathcal{B}} Re^2[\mu_j]}. \tag{3.29}$$

Proof 5 *With the help of Lemma 2, we compute the time-domain Fourier transform of* $\mathbf{z}(t)$ *via the summation of Eq. (3.27) corresponding to different inputs* $\mathbf{b}_i$*. As such, the upper-bound of the magnitude* $|Z_n(\omega)|$ *is*

$$|Z_n(\omega)| < \frac{\|\mathbf{z}(0)\|_2 + \sum_i \|\mathbf{b}_i\|_2}{\sqrt{\min_{j\in\mathcal{B}} Re^2[\mu_j] + \left(\omega - \max_{j\in\mathcal{B}} |Im[\mu_j]|\right)^2}}. \tag{3.30}$$

Then, let such upper-bound in Eq. (3.30) be less than the provided threshold ε, we therefore prove Theorem 4.

Given the deduction of Lemma 2 and Theorem 4, the cut-off frequency $\omega_{\text{cut-off}}$ is determined to discretize the continuous network signals. Then, we can adopt the Shannon sampling interpolation in Eq. (3.5) to recover the continuous signals from the discretized ones under the cut-off frequency $\omega_{\text{cut-off}}$.

3.2.4 Explicit Relationship between Optimal Sampling and Graph Dynamics

It is highlighted that a key contribution of our equation-driven GFT sampling framework is the creation of an **explicit relationship** between the time- and network-domain cut-off frequencies, the networked topological properties, and the governing nonlinear modes. We discuss this from the following three perspectives.

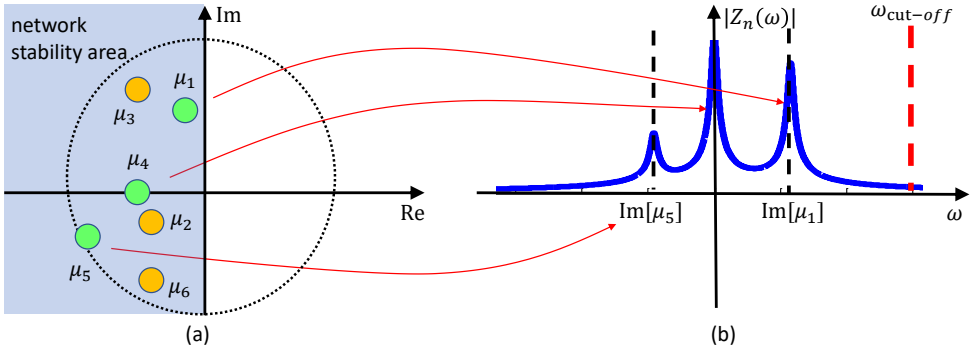

Figure 3.3: Illustration of the relations between graph bandwidth set $\mathcal{B} = \{1, 4, 5\}$, and the time-domain cut-off frequency $\omega_{\text{cut-off}}$.

- First, we build a bridge between the dynamical governing equations (i.e., the self-dynamic function and the mutualistic coupling equations), and the linearized matrix $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ from Eq. (3.8). The latter that combines the topological structure and the governing equations, is able to (i) govern the first-order (main part) time and network domain evolution of the time-varying network signals, and (ii) give an interpretation of system stability via the real parts of its eigenvalues (seen from Fig. 3.3(a)).
- Second, we analyze the initialization and inputs that are graph $\mathcal{B}$-bandlimited (quasi) to the GFT operator determined by the linearized matrix $\mathbf{J}_{\Xi}(\mathbf{x}_e)$, and prove that the whole time-varying network signals are $\mathcal{B}$-bandlimited (quasi).

Such a set $\mathcal{B}$ further maps to the eigenvalues (graph frequencies) illustrated in Fig. 3.3(a), and only the corresponding eigenvectors whose indices belong to $\mathcal{B}$ have non-zero contributes. Also, the graph bandwidth set $\mathcal{B}$ maps to an optimal sampling vertex subset $\mathcal{C}$ that ensures complete signal recovery.

- Third, this graph bandwidth set $\mathcal{B}$ further leads to the computation of the time-domain cut-off frequency. It is noteworthy that, only the indices of eigenvalues of $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ that belong to $\mathcal{B}$ affect the shape of the time-frequency Fourier transform (seen from Fig. 3.3(b)). To be specific, a direct relation between $\omega_{\text{cut-off}}$ and $\mathcal{B}$ is shown in Theorem 4. In summary, as is illustrated in Fig. 3.3, the time-domain cut-off frequency $\omega_{\text{cut-off}}$ is related to the eigenvalues whose indices belong to $\mathcal{B}$, which in turn is related to the optimally sampled graph structure and the underlying dynamics.

3.3 Compared with Two State-of-the-Arts

It is noteworthy that for this chapter, it is hard to pursue any performance comparison with the existing sampling methods. This is because most of the referenced works (e.g., [41,46]) concentrated on discrete-time network signals, and are different from our scope (i.e., the continuous network signals). In this view, our work in this chapter provides them a discretization method, after which they can perform their analysis on discretized data matrix (we will compare with these methods in following chapters that consider the network sampling on discrete signal matrix). In addition, most of the existing works offered the time-varying sampling vertex subsets, which therefore are not suitable for sensor placement applications requiring fixed sampling vertex subset for sensor deployment.

As such, we here just render the comparisons with their works by describing the major differences in conceptual manner. For the work in [41], the author designed a joint time and network domains GFT sampling method (called JFT). The key of their JFT method is to rely on the topological structure based Laplacian operator, and the Discrete FT matrix. The difference therefore lies in that, they did not consider the dynamic mechanism (i.e., the governing dynamic equations) in their JFT framework. This leads to their inability to analyze the signal property, e.g., they cannot keep an $\mathcal{B}$-bandlimited property for the whole time-varying network signals. As such, it is hard for them to derive a fixed sampling vertex subset for sensor placement. For the work in [46], the authors proposed a sequential Kalman filter to track the discrete network signals, under the assumptions of the known inputs and a $\mathcal{B}$-bandlimited initialization to a GFT operator (i.e., composed of eigenvectors of

the linear evolution matrix). However, such scheme becomes less attractive when addressing the unknown input challenges. This is because such unknown inputs lead to an unavailability of their transition probability density mapping, which if lost, will block the predict-stage of the current network signal from the previously recovered ones. More importantly, they did not analyze the case when the initialization and the inputs are not $\mathcal{B}$-bandlimited (as what we do in Section 3.2.2).

3.4 Simulations & Experimental Results

In the section, we evaluate the sampling and recovery performances of our proposed equation-driven GFT sampling method. First, we analyze the case when the unknown initialization and inputs, i.e., $\mathbf{z}(0)$ and $\mathbf{b}_i$ are graph $\mathcal{B}$-bandlimited to the GFT operator $\mathbf{Q}^{-1}$, with a known $\mathcal{B}$. Then, the general cases with arbitrary initialization and inputs, i.e., $\mathbf{z}(0), \mathbf{b}_i \notin BS(\mathcal{B}, \mathbf{Q}^{-1})$ are tested.

In the following simulations, the impacts of both the network-domain sampling vertex subset, and the time-domain sampling frequency are considered. For the former, we use the size of the sampling vertex subset $|\mathcal{C}|$ to measure how dense the sensors are deployed. For different sizes $|\mathcal{C}|$, the selection of the sampling vertex subset $\mathcal{C}$ follows the method in Eq. (3.19). For the time-domain sampling frequency, we use a very small sampling interval Δ_t to simulate the continuous network signals, i.e., $2\pi/\Delta_t = 8\omega_{\text{cut-off}}$. Then, we provide the root mean square error (RMSE) of the recovered time-varying network signals $\hat{\mathbf{x}}(t), t \in [0, K\Delta_t)$ to measure the recovery accuracy, i.e.,

$$\text{RMSE} = \sqrt{\frac{1}{N \cdot K} \sum_{k=0}^{K-1} \|\hat{\mathbf{x}}(k \cdot \Delta_t) - \mathbf{x}(k \cdot \Delta_t)\|_{l_2}^2}, \tag{3.31}$$

where $\|\cdot\|_{l_2}$ denotes the l_2-norm.

For the simulation configurations, we use two typical governing dynamic equations [66]:

$$\frac{dx_n(t)}{dt} = -B \cdot x_n(t) + R \cdot \sum_{m=1}^{N} a_{n,m} \cdot x_m(t) + \sum_{i=1}^{+\infty} b_{n,i} \cdot \delta(t - t_i), \tag{3.32}$$

$$\frac{dx_n(t)}{dt} = F - B \cdot x_n(t) + R \cdot \sum_{m=1}^{N} a_{n,m} \cdot x_n(t) \cdot x_m(t) + \sum_{i=1}^{+\infty} b_{n,i} \cdot \delta(t - t_i). \tag{3.33}$$

Eq. (3.32) is referred to as the linear networked population density (PD) dynamics

representing the evolution of population density, where population density on each vertex has a self growth rate $-B$ and is also affected by the migrated strength R from neighbouring connected vertices. Eq. (3.33) is referred to as the non-linear biochemical protein-protein interaction dynamics governed by the mass-action kinetics (MAK). The detailed explanations for the governing dynamic equations are in [66]. In Eqs. (3.32)-(3.33), we assign the number of vertices $N = 500$, and assign other parameters to satisfy the Lypunov stability assumed in Assumption 1.

3.4.1 Performance with Graph Bandlimited Initialization and Inputs

We first evaluate the performance of the proposed equation-drive GFT sampling method, when the graph $\mathcal{B}$-bandlimited property (with the known $\mathcal{B}$) of the initialization and inputs are assumed.

Linear Governing Dynamic Equation

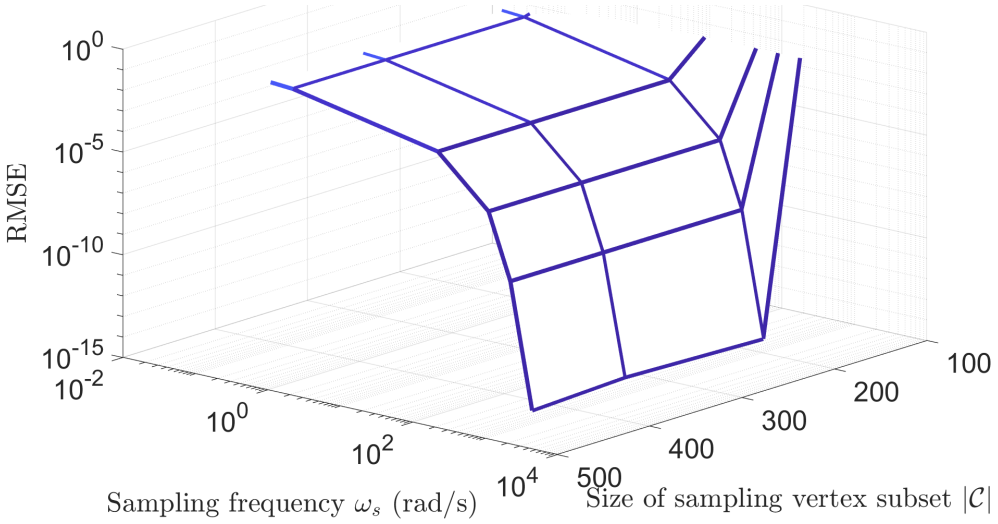

Figure 3.4: PD linear dynamic model with graph bandlimited initialization and inputs: recovery RMSE versus both the sampling frequency $\omega_{\rm s}$, and the size of sampling vertex subset $|\mathcal{C}|$.

In Figs. 3.4-3.6, we provide the signal recovery performance of the proposed equation-driven GFT sampling method for the linear dynamic model in Eq. (3.32). Fig. 3.4 illustrates the recovery RMSE versus the joint time-domain sampling frequency $\omega_{\rm s}$, and the network-domain size of sampling vertex subset $|\mathcal{C}|$. It is observed that as both $\omega_{\rm s}$ and $|\mathcal{C}|$ become larger, the recovery RMSE decreases. This indicates

that the performance of the signal recovery gets better, when more samples from both the time- and the network- domains are used. We also demonstrate this in Fig. 3.5 and Fig. 3.6, whereby the two tangent planes of Fig. 3.4, i.e., RMSE versus time-domain sampling frequency $\omega_{\rm s}$, and RMSE versus network-domain size of sampling vertex subset $|\mathcal{C}|$, are given.

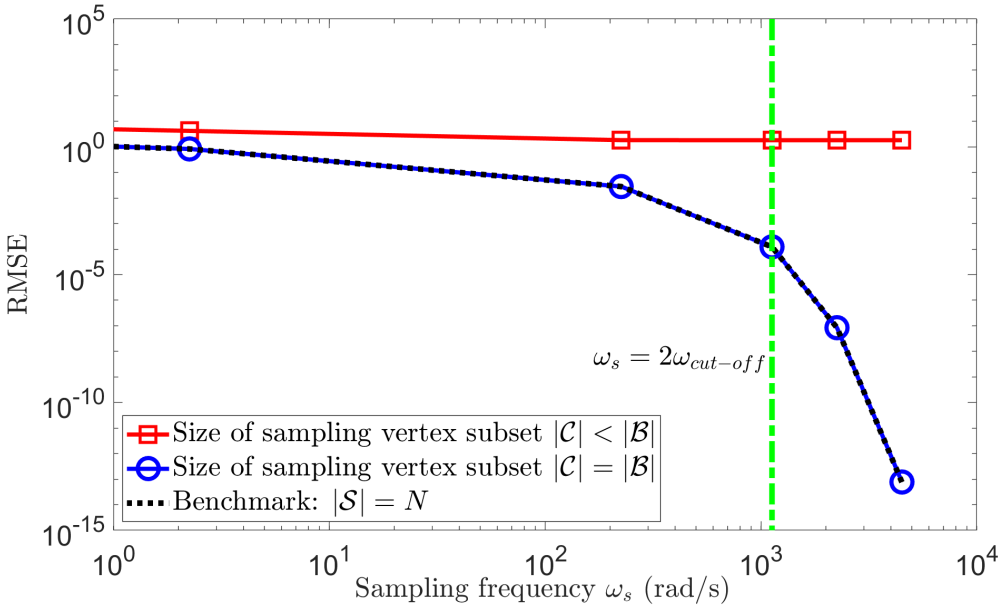

Figure 3.5: PD linear dynamic model with graph bandlimited initialization and inputs: recovery RMSE versus sampling frequency $\omega_{\rm s}$.

Fig. 3.5 illustrates the recovery RMSEs versus the time-domain sampling frequency $\omega_{\rm s}$, given different and fixed sizes of sampling vertex subsets $|\mathcal{C}|$. It is seen from Fig. 3.5 that, as $\omega_{\rm s}$ increases, the recovery RMSEs for all fixed $|\mathcal{C}|$ decrease. This is because that, the larger the time-domain sampling frequency $\omega_{\rm s}$, the more number of samples from the time-domain, suggesting an improvement of the recovery accuracy according to the Nyquist sampling theory. Also, we compare in Fig. 3.5 the cases whether the fixed size of sampling vertex subset $|\mathcal{C}|$ is larger than size of graph bandwidth set, i.e., $|\mathcal{C}| \gtrless |\mathcal{B}|$. For the case $|\mathcal{C}| > |\mathcal{B}|$, the recovery RMSE can converge to a very low order (i.e., from 10^{-1} to 10^{-15} as $\omega_{\rm s}$ increases). Moreover, in such a case, the trends of the RMSE match the benchmark whereby all the vertices are selected and sampled (i.e., $|\mathcal{C}| = N$). This is because that in such a linear dynamic model, e.g., the PD model, the Jacobian linearized matrix $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ is actually the linear evolution model, and there is no linearized error described by Eqs. (3.12)-(3.13). In this view, the perfect signal recovery from the network domain can be achieved only if $rank\,(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{B}|$ is approached. Then, we consider

the case with $|\mathcal{C}| < |\mathcal{B}|$. We can see that the recovery RMSE decreases slightly with the increases of the time-domain sampling frequency ω_{s}, since the perfect signal recovery cannot be achieved if the size of sampling vertex subset is lower than that of the graph bandwidth set.

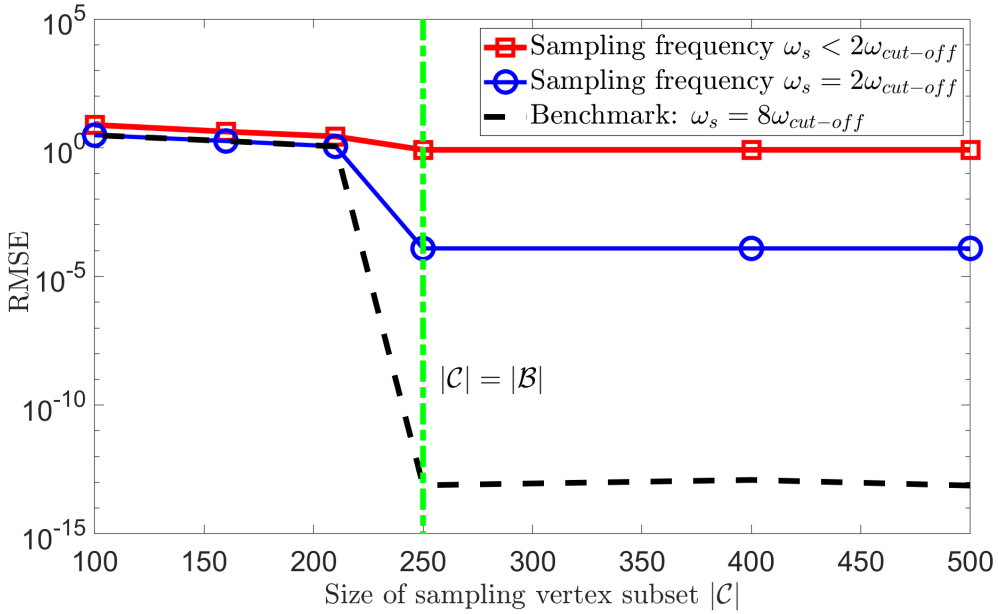

Figure 3.6: PD linear dynamic model with graph bandlimited initialization and inputs: recovery RMSE versus the size of sampling vertex subset $|\mathcal{C}|$.

Fig. 3.6 illustrates the recovery RMSEs versus the varied sizes of sampling vertex subsets $|\mathcal{C}|$, given different and fixed time-domain sampling frequencies ω_{s}. Likewise, it is observed from Fig. 3.6 that the recovery RMSEs get lower as the size of sampling vertex subset $|\mathcal{C}|$ increases, due to the fact that the more samples from the network domain are involved. Then, it is seen from Fig. 3.6 that, the recovery RMSE for the case $\omega_{\text{s}} = 2\omega_{\text{cut-off}}$ is lower than that with $\omega_{\text{s}} < 2\omega_{\text{cut-off}}$. This is because the former represents the case when time-domain sampling frequency is larger than the cut-off frequency, and the latter accounts for the under-sampling from the time-domain. As such, the recovery performance of the under-sampling are deteriorated by the missing samples from the time-domain. Also, it is noteworthy that the threshold ε for the magnitudes of the transformed frequency components in Theorem 4 matters, since the computation of the time-domain cut-off frequency $\omega_{\text{cut-off}}$ omits the frequency components with magnitudes smaller than ε. This therefore results in the gap between the benchmark with larger $\omega_{\text{s}} = 8\omega_{\text{cut-off}}$ and the recovery RMSE with $\omega_{\text{s}} = 2\omega_{\text{cut-off}}$. Furthermore, it is noticed that, after the size of the sampling

vertex subset $|\mathcal{C}|$ reaches the size of graph bandwidth set $\mathcal{B}$ i.e., $|\mathcal{C}| = |\mathcal{B}|$, the recovery RMSEs converges to a very low bound. We explain this in the following. In the case of the linear dynamic model, Theorem 3 holds perfectly and suggests a sampling vertex subset $\mathcal{C}$ for completely signal recovery if $rank\,(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{B}|$.

Nonlinear Governing Dynamic Equations

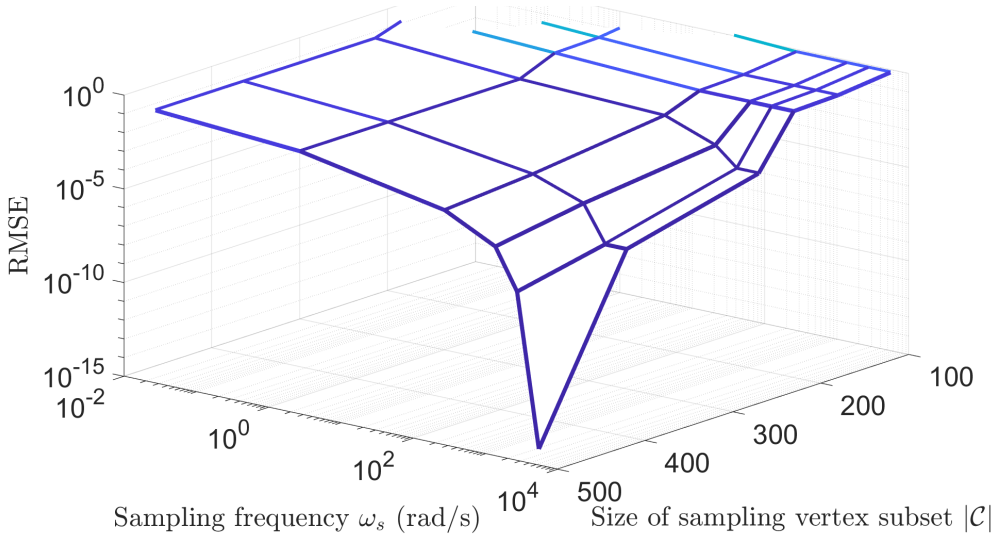

Figure 3.7: MAK nonlinear time-varying network signals with graph bandlimited initialization and inputs: recovery RMSE versus both the sampling frequency ω_{s} and the size of sampling vertex subset $|\mathcal{C}|$.

We next evaluate the proposed equation-driven GFT sampling in the context of the nonlinear dynamic model, i.e., Eq (3.33). Fig. 3.7 provides the recovery RMSE versus both the time-domain sampling frequency ω_{s} and the network-domain size of sampling vertex subset $|\mathcal{C}|$. Likewise, it is straightforward that with both the increases of ω_{s} and $|\mathcal{C}|$, the recovery RMSE becomes lower, due to the fact that more samples from the both domains can help improve the accuracy of the signal recovery.

Fig. 3.8 provides one tangent plane of Fig. 3.7, whereby the recovery RMSEs versus the time-domain sampling frequency ω_{s} are tested. Here, we fix the network domain size of sampling vertex subset $|\mathcal{C}|$ into 3 different values. In Fig. 3.8, it is observed that as ω_{s} gets larger, the recovery RMSEs of all fixed $|\mathcal{C}|$ decrease, as the larger the sampling frequency ω_{s} from time-domain, the more samples can be used for a better recovery accuracy. Moreover, we compare the cases whether the fixed size of the sampling vertex subset $|\mathcal{C}|$ is larger than that of the graph bandwidth

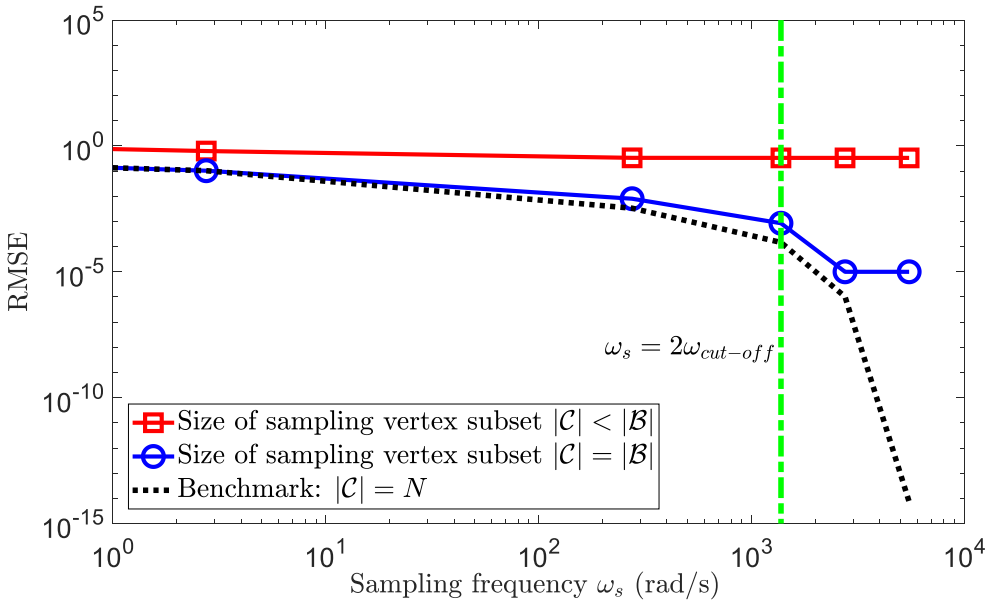

Figure 3.8: MAK nonlinear time-varying network signals with graph bandlimited initialization and inputs: recovery RMSE versus the sampling frequency $\omega_{\rm s}$.

set, i.e., $|\mathcal{C}| \gtrless |\mathcal{B}|$. When $|\mathcal{C}| > |\mathcal{B}|$, the recovery RMSE decreases first (from 10^{-1} to 10^{-5}), and then converges to a very low limitation (i.e., 10^{-5}), as exceeding the 2 times of the Nyquist sampling frequency (i.e., $\omega_{\rm s} > 2\omega_{\text{cut-off}}$). This is different from the linear dynamic scenarios in Fig. 3.5. We explain this in the following. Given the linearized matrix $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ that tries to approximate the nonlinear dynamics, an error (gap) from Eqs. (3.12)-(3.13) exists and cannot be alleviated just by increasing the time-domain sampling frequency $\omega_{\rm s}$. Such an error caused by linearization, i.e., $10^{-5} - 10^{-15}$ in Fig. 3.8, is illustrated by the benchmark whereby all network-domain vertices are selected for samples i.e., $|\mathcal{C}| = N$. It is also noteworthy that after the time-domain sampling frequency $\omega_{\rm s}$ exceeds 2 times of the cut-off frequency $\omega_{\text{cut-off}}$, the recovery RMSE are still decreasing. This is because the computation of $\omega_{\text{cut-off}}$ in Theorem 4 omits the frequency components with magnitudes lower than the given threshold ε. Then, we consider the case with $|\mathcal{C}| < |\mathcal{B}|$. We can see that the recovery RMSE decreases slightly with the increases of the time-domain sampling frequency $\omega_{\rm s}$, since the perfect signal recovery cannot be achieved if the size of sampling vertex subset is lower than that of the graph bandwidth set.

Fig. 3.9 illustrates the recovery RMSEs versus the network-domain size of sampling vertex subset $|\mathcal{C}|$, given the different and fixed time-domain sampling frequencies $\omega_{\rm s}$. As aforementioned, the recovery RMSEs of all time-domain sampling frequencies decrease when $|\mathcal{C}|$ increases, due to the fact that more samples from the

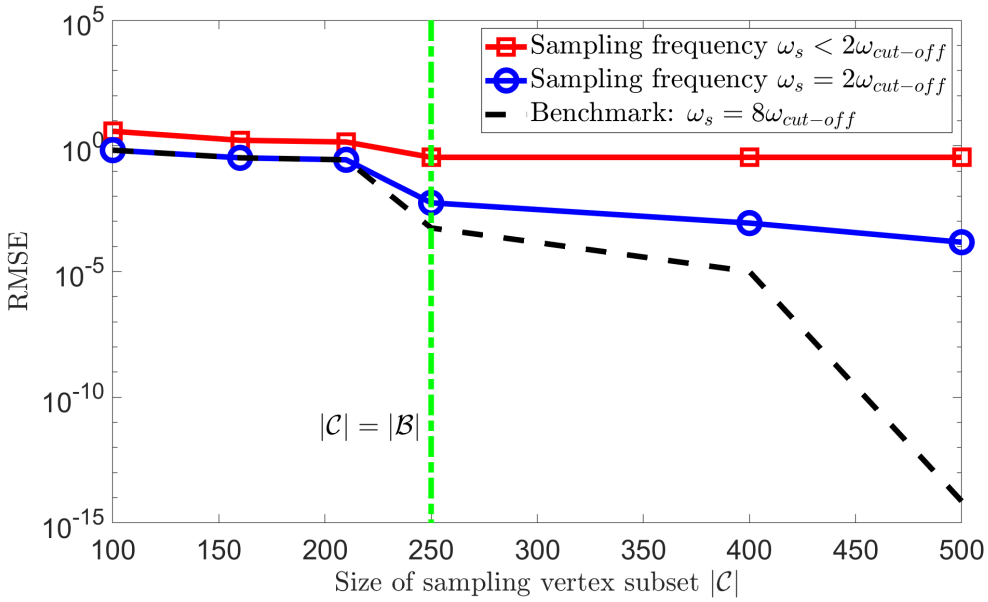

Figure 3.9: MAK nonlinear time-varying network signals with graph bandlimited initialization and inputs: recovery RMSE versus the size of sampling vertex subset $|\mathcal{C}|$.

network domain gives rise to a more accurate signal recovery. Then, it is observed that the recovery RMSE from the case where Nyquist sampling theory is approached i.e., $\omega_{\text{s}} = 2\omega_{\text{cut-off}}$ outperforms the one of $\omega_{\text{s}} < 2\omega_{\text{cut-off}}$ accounting for the under-sampling from time-domain, since the latter has less samples from the time-domain for signal recovery. Also, similar to the linear dynamic scenarios, the threshold for transformed time-domain frequency components in Theorem 4 matters. This is because the computed time-domain cut-off frequency $\omega_{\text{cut-off}}$ in Theorem 4 omits those frequency components with lower magnitudes than the provided threshold. This therefore leads to the gap between the benchmark with a larger $\omega_{\text{s}} = 8\omega_{\text{cut-off}}$ and the recovery RMSE with $\omega_{\text{s}} = 2\omega_{\text{cut-off}}$. In addition, we can notice that unlike the linear dynamic shown in Fig. 3.6 that has a recovery convergence after $|\mathcal{C}| \geq |\mathcal{B}|$, the recovery RMSEs for nonlinear time-varying network signals are still decreasing. This is because equation-driven sampling method we deduced in Theorem 3 leverages the linear system. As such, for nonlinear scenarios, the linearized error measured by Eqs. (3.12)-(3.13) cannot be eliminated until all the vertices from the network domain are selected for samples, i.e., $|\mathcal{C}| = N$. We demonstrate this gap via the Benchmark whose size of the sampling vertex subset is $|\mathcal{C}| = N$.

3.4.2 Performance with Arbitrary Initialization and Inputs

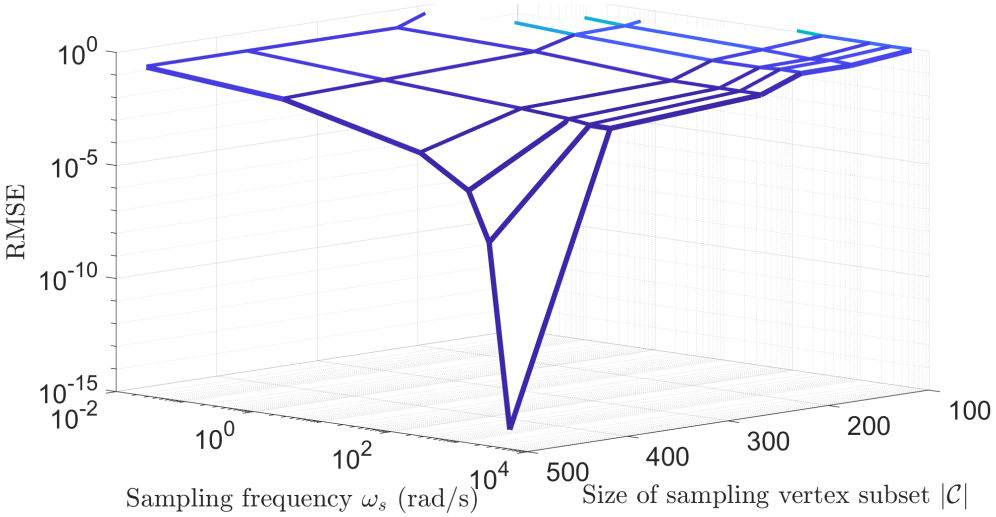

Figure 3.10: MAK nonlinear time-varying network signals with arbitrary initialization and inputs: recovery RMSE versus both the sampling frequency $\omega_{\text{cut-off}}$ and the size of sampling vertex subset $|\mathcal{C}|$.

We finally evaluate our proposed equation-driven GFT sampling method for arbitrary initialization and inputs in the context of the nonlinear time-varying network signal model, i.e., the MAK dynamic in Eq. (3.33). The performance of signal recovery is illustrated in Fig. 3.10, in which the recovery RMSE versus the joint time-domain sampling frequency and network-domain size of sampling vertex subset $|\mathcal{C}|$ is provided. Similar to the graph bandlimited scenarios, the recovery RMSE decreases when ω_{s} and $|\mathcal{C}|$ grow, which indicates a higher recovery accuracy of the time-varying network signals as more samples are involved and utilised.

Fig. 3.11 illustrates the recovery RMSEs versus the time-domain sampling frequency ω_{s}, as different and fixed sizes of the sampling vertex subset $|\mathcal{C}|$ are studied. We can see from Fig. 3.11 that, when the time-domain sampling frequency ω_{s} increases, the recovery RMSEs of all fixed $|\mathcal{C}|$ get smaller, since the larger the time-domain sampling frequency ω_{s}, more samples from time-domain can be used for a more accurate signal recovery. Then, we compare between different sizes of sampling vertex subsets, i.e., $|\mathcal{C}| = 4N/5$ and $|\mathcal{C}| = 2N/5$. It is straightforward that a larger $|\mathcal{C}| = 4N/5$ provides a greater recovery performance as opposed to a lower $|\mathcal{C}| = 2N/5$, since a larger $|\mathcal{C}|$ can embrace more information from the graph frequency domain. Furthermore, we observe the case with larger size of sampling vertex subset, i.e., $|\mathcal{C}| = 4N/5$. We can see that the recovery RMSE decreases at

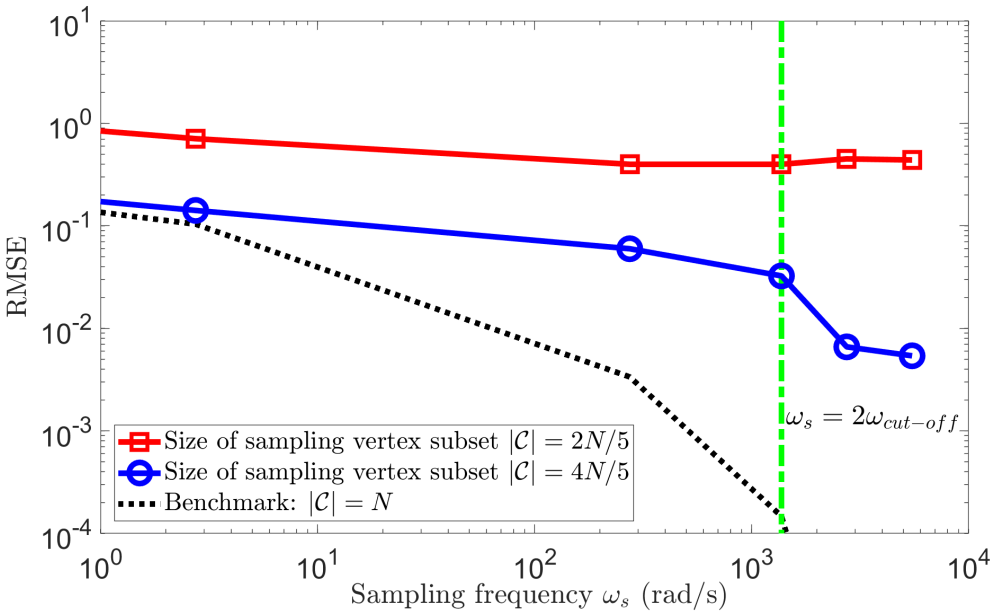

Figure 3.11: MAK nonlinear time-varying network signals with arbitrary initialization and inputs: recovery RMSE versus the sampling frequency $\omega_{\rm s}$.

first (from 10^{-1} to 10^{-2}), and then converges to a limit (as 10^{-2}), which is higher than that of the graph-bandlimited scenarios from Fig. 3.8. We categorize the reason into two aspects. First, just like the graph bandlimited cases, the recovery accuracy improves when a larger time-domain sampling frequency $\omega_{\rm s}$ is used, but the aforementioned linearized error by $\mathbf{J}_{\Xi}(\mathbf{x}_e)$ determines its converged limitation. Second, different from the graph bandlimited scenarios, a further error occurs as we approximate the signals as $\mathcal{B}$-bandlimited and omit parts of the components from the graph frequency domain. This limitation (computed as $10^{-2} - 10^{-15}$) is also shown by the benchmark in Fig. 3.11 whereby all vertices are used for samples i.e.,$|\mathcal{C}| = N$. We should also note that after $\omega_{\rm s}$ exceeds 2 times of the time-domain cut-off frequency $\omega_{\text{cut-off}}$, the recovery RMSE is still lowering, due to the omission of the smaller time-domain frequency components that are lower than the provided threshold ε in Theorem 4.

Fig. 3.12 illustrates the recovery RMSEs versus the size of sampling vertex subset $|\mathcal{C}|$, as we examine two fixed time-domain sampling frequencies $\omega_{\rm s}$. As aforementioned, the recovery RMSEs of all fixed $\omega_{\rm s}$ become smaller as the size of sampling vertex subset $|\mathcal{C}|$ increases, due to the fact that an increasing $|\mathcal{C}|$ can embrace more samples from the network domain for better recovery performance. It is noteworthy here the factors that affect the recovery RMSE are two, i.e., i) the linear approxi-

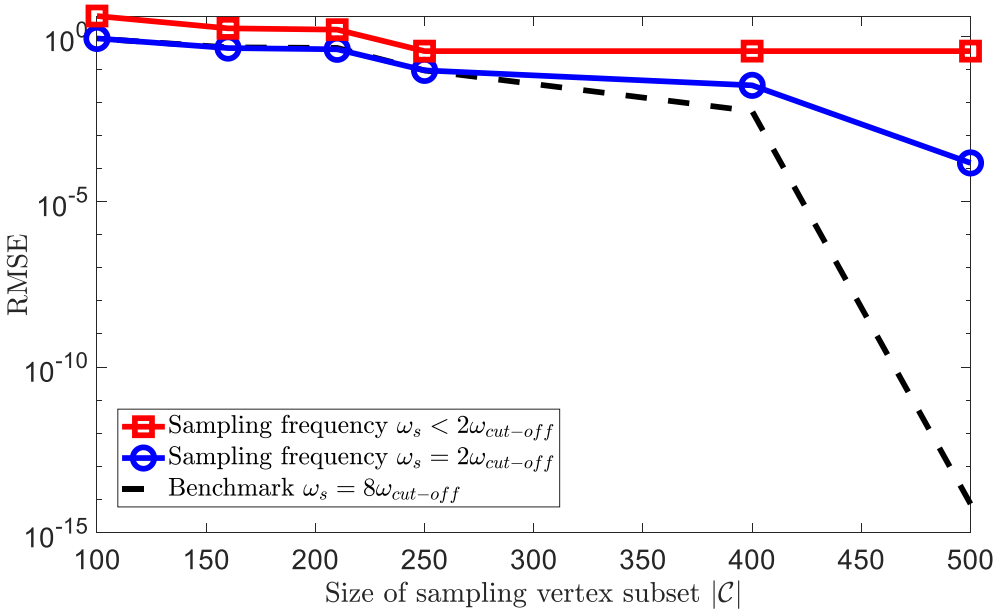

Figure 3.12: MAK nonlinear time-varying network signals with arbitrary initialization and inputs: recovery RMSE versus the size of sampling vertex subset $|\mathcal{C}|$.

mation error caused by $\mathbf{J}_{\Xi}(\mathbf{x}_e)$, and ii) the error caused by the approximated graph bandlimited property using Eq. (3.20). Then, it is seen that the recovery RMSE from the case $\omega_{\text{s}} = 2\omega_{\text{cut-off}}$ outperforms that of the time-domain under-sampling case i.e., $\omega_{\text{s}} < 2\omega_{\text{cut-off}}$. This is because the time-domain under-sampling leads to a lack of samples for signal recovery. Also, the gap between the benchmark with a larger $\omega_{\text{s}} = 8\omega_{\text{cut-off}}$ and the recovery RMSE with $\omega_{\text{s}} = 2\omega_{\text{cut-off}}$ is shown. This is mainly because the computation of the cut-off frequency $\omega_{\text{cut-off}}$ that omits the time-domain frequency components with smaller magnitudes than the given threshold ε in Theorem 4.

3.5 Conclusions & Discussions

For this chapter, the equation-driven GFT sampling framework relying on vertex signal dependency has been proposed, capable of sampling the time-varying network signals from the joint time (discretizing rate) and network (sampling vertices) domains. We first characterized the nonlinear time-varying network signals using the first-order Jacobian matrix of the dynamic differential equations. Leveraging the eigenvectors of the linearzation matrix, we have constructed the dynamic-topology combined GFT operator, which is able to uncover the dependency of the continuous network signals on different vertices (i.e., the graph $\mathcal{B}$-bandlimitedness). Then, we

have proved the existence of the time-invariant sampling vertex subset for recovering the continuous network signals under a graph bandlimited initialization and inputs.

Unlike the traditional compression frameworks (e.g., CS and topological based GFT), we combine the network topology with nonlinear dynamical mechanism (explicit dynamic governing equations) that has explicit causal relations among vertices. Therefore, our sampling framework is able to indicate a direct mapping from the network sampling vertices and time-domain discretizing rate to the network structure and governing nonlinear dynamics, and reversibly the changing in the underlying dynamics or the network topology will be able to inform the network sampling processes (see Fig. 3.3). In this view, this chapter provides a straightforward understanding on how the network topology and underlying dynamic mechanism affect the information sampling from both the time and the network domains.

The limitation of the dynamic equation-driven GFT sampling scheme, as its name suggests, lies in its heavily relying on the dynamic governing equations. This therefore renders its unsuitability for many complex systems, whose explicit governing mechanisms are unavailable, e.g., a multiplex of various network and dynamics [13], or with dynamics in higher dimensions and higher order differential equations [4, 6], or having non-Markovian extended-memory dynamics [1]. These further motivate us to design new sampling framework that can learn the hidden dynamical mechanism from the experimental data, and we will introduce our data-driven GFT sampling in the next chapter.

Chapter 4

Signal-Space Dependent Data-Driven Sampling

In this chapter, we elaborate the signal-space dependent vertex selections for sensor placement, and the signal recovery scheme. The motivation is from the real-world water-distribution network surveillance application, where the governing dynamic equations are not available. In such a case, the equation-driven sampling methods will lose their compasses, rendering the difficulty of constructing the topology-equation based compress operator for sampling vertices selection and dynamic signal recovery. To overcome this, an alternative way is to use the prior property of the signal-space (e.g., the sparsity or bandlimitedness to a given operator). As such, we provide in this chapter the data-driven Graph Fourier Transform operator, and how it can be used for network sampling tasks.

The structure of this chapter is given as follows. We first formulate a data-driven modelling of time-varying network signals and analyze the problem in Section 4.1. Then, in Section 4.2, we elaborate the detailed sampling and recovery processes. In Section 4.3, we consider the pathway constraints for sampling vertex subset selection. In Section 4.4, we provide the theoretical error bounds of the proposed GFT sampling scheme. In Section 4.5, we compare our scheme with other two state-of-the-art data-driven methods. The simulation and experimental results are provided in Section 4.6. We finally conclude this chapter in Section 4.7.

4.1 Water Distribution Network Formulation and Problem Analysis

In this section, we describe the water-distribution network (WDN), and formulate a data-driven modelling for an injected contaminant propagation over the WDN.

Like Chapter 3, we configure the WDN by a static graph denoted as $G(\mathcal{N}, \mathbf{A})$. $\mathcal{N} = \{1, 2, \cdots, N\}$, $N \in \mathbb{N}^+$ is a set of indices of the total WDN vertices. $\mathbf{A}$ represents the WDN adjacent matrix, where the element $a_{n,m} = 1/0$ represents an existence/nonexistence of a directed edge from vertex m to vertex n. In WDN, the vertices can be represented as the junctions, the reservoirs, or the tanks, while links can be the pump, the valve, and the pipe [67]. For each a WDN vertex $n \in \mathcal{N}$, signals such as the water demands, the head-loss, and the water-quality are time-varying and interacted with its neighbouring vertices.

In the rest of this chapter, we are only interested in the water-quality of the WDN, measured by the amount of contaminant propagated over the network. The contaminant if deliberately/accidentally injected in some vertices will be spread over the WDN and finally expelled by the water-demands (e.g., the usages and drinking) of other vertices, which is needed for monitoring and surveillance in order to protect the majority. As such, sensors are required to deploy in some of the vital vertices for collecting the contaminant signals and for the recovery of those on unsampled vertices. The aforementioned WDN topology and contaminant propagation are illustrated via Fig. 4.1.

For the contaminant propagation over WDN, we characterize the time-varying network signals via a discrete-time matrix of size $N \times K$, i.e.,

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K], \tag{4.1}$$

where $\mathbf{x}_k$ represents the network signal at discrete time k. $N = |\mathcal{N}|$ is the number of vertices in WDN, and $\mathcal{K} = \{1, \cdots, K\}$ is the set of total discrete times for monitoring. We also denote $\mathcal{C} \subset \mathcal{N}$ as the sampling vertex subset. The objective of this chapter is to identify the appropriate sampling vertex subset $\mathcal{C} \subset \mathcal{N}$, such that the samples derived from $\mathcal{C}$ can recover the whole time-varying network signals:

$$\mathbf{X} = \mathbf{\Theta} \cdot \mathbf{X}_{\mathcal{C}\mathcal{K}}, \tag{4.2}$$

where $\mathbf{\Theta}$ denotes the recovery matrix. $\mathbf{X}_{\mathcal{C}\mathcal{K}}$ represents the samples of $\mathbf{X}$, whose rows are selected with indices/subscripts in set $\mathcal{C}$, and columns are selected with indices/subscripts in set $\mathcal{K}$. Here, we consider an immediate signal recovery, as any

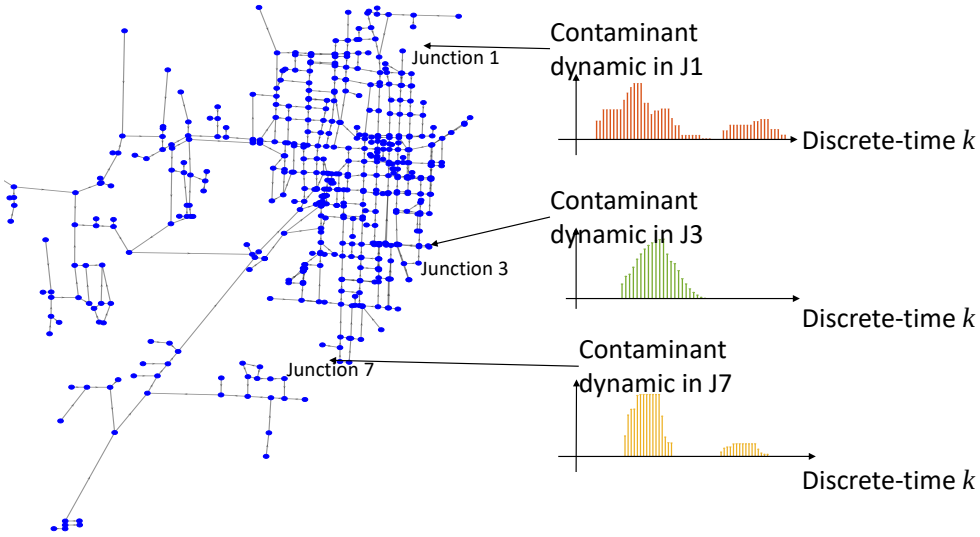

Figure 4.1: Illustration of the WDN topology and the contaminant time-varying network signals. Contaminant signals in three example junctions (vertices) are provided.

latency on contaminant monitoring for WDN may cause catastrophe. As such, only the signal dependency among different WDN vertices are used for sampling vertex subset selection and signal recovery; we do not rely on any time-evolved information as this will cause latency.

Different from the previous Chapter 3, the governing dynamic equations are not available due to the complex partial differential manner for hydraulic signal evolution. Instead, to discover the signal dependency among vertices, we assume a prior knowledge of the contaminant signal-space, denoted as $\mathbb{S}$, spanned by $r < N$ independent supports:

$$\mathbb{S} = span\{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r\} \subset \mathbb{R}^N, \tag{4.3}$$

and we have for any discrete time k, the time-varying network signals belongs to the signal-space, i.e., $\mathbf{x}_k \in \mathbb{S}$. We explain the derivation of the signal-space in the following. Given a specific region, the potential sources (vertices) of the contamination are known (e.g., one can know exactly the locations of main factories and their potential pollutant injections). As such, we can simulate the WDN, and the contaminant propagation flows with fixed injection vertices via EPANET engine [67], and derive the contaminant signal-space.

4.2 Data-Driven Sampling and Recovery Process

In this section, we elaborate our proposed data-driven network sampling and signal recovery methods leveraging the signal dependency among the network vertices. The idea is borrowed from traditional graph sampling theory 2 in Chapter 2. To be specific, we first derive a data-driven GFT operator using the prior knowledge of signal-space, to keep the whole network signals in $\mathbf{X}$ graph bandlimited. Then, we select the optimal sampling vertex subset $\mathcal{C}$ based on the graph bandlimited property. We finally recover the whole time-varying network signals via the samples derived from the sampling vertices in $\mathcal{C}$.

Before we start, we extend from the graph bandlimitedness concept of the static network signal ($N \times 1$ vector) in Definition 1, and provide the graph bandlimitedness definition to a joint time and network -domains matrix (of size $N \times K$).

Definition 3 *Consider a set $\mathcal{B} \subset \mathcal{N}$, a GFT operator $\mathbf{Q}^{-1}$, a joint time and network domain data matrix $\mathbf{X}$ of size $N \times K$, and its graph-frequency response $\tilde{\mathbf{X}} = \mathbf{Q}^{-1} \cdot \mathbf{X}$. We call $\mathbf{X}$ graph $\mathcal{B}$-bandlimited with respect to $\mathbf{Q}^{-1}$, if in $\tilde{\mathbf{X}}$, only the rows with indices in $\mathcal{B}$ are nonzero vectors. We call $\mathcal{B}_{cut\text{-}off} = \mathcal{B}$ as the cut-off graph bandwidth set of $\mathbf{X}$.*

Actually, the graph $\mathcal{B}$-bandlimited of the time-varying network signals $\mathbf{X}$ characterizes the signal dependency among vertices in $\mathbf{X}$ (i.e., the linear row dependency). Given $r = rank(\mathbf{X})$, the GFT operator $\mathbf{Q}^{-1}$ proceeds elementary row transformation till r independent and linearly combined rows left.

4.2.1 Data-Driven GFT Operator

Given the joint time and network domain contaminant data matrix $\mathbf{X}$ of size $N \times K$, we aim to generate a GFT operator (a reversible matrix) $\mathbf{Q}^{-1}$ to make $\mathbf{X}$ graph $\mathcal{B}$-bandlimited with some $\mathcal{B}$ whose cardinality $|\mathcal{B}| < N$. By doing so, the graph sampling theory can be adopted to identify the optimal sampling vertex subset $\mathcal{C}$ satisfying $|\mathcal{C}| = |\mathcal{B}|$.

Given that the columns of $\mathbf{X}$ belong to the signal-space, i.e., $\mathbf{x}_k \in \mathbb{S}$, we compute the GFT operator $\mathbf{Q}^{-1}$ using the independent supports $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r$. To be specific, the singular value decomposition (SVD) is adopted on $[\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r]$, and we have:

$$[\mathbf{\Gamma}, \mathbf{\Sigma}, \mathbf{V}] = svd\left([\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r]\right). \tag{4.4}$$

As such, the data-driven GFT operator $\mathbf{Q}^{-1}$ is derived by assigning

$$\mathbf{Q}^{-1} = \mathbf{\Gamma}^{-1}. \tag{4.5}$$

After the derivation of the GFT operator $\mathbf{Q}^{-1}$, we analyze the graph bandlimtiedness of $\mathbf{X}$ with respect to $\mathbf{Q}^{-1}$. At first, we show that $[\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r]$ is graph $\mathcal{B}$-bandlimited to $\mathbf{Q}^{-1}$, given $\mathcal{B} = \{1, 2, \cdots, r\}$. This is straightforward given the SVD and the in-dependency of the r supports $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r$, i.e.,

$$\begin{aligned} &\mathbf{Q}^{-1} \cdot [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r] = \mathbf{\Sigma} \cdot \mathbf{V} \\ &= \begin{bmatrix} diag\left([\sigma_1, \cdots, \sigma_r]^T\right) \cdot \mathbf{V} \\ \mathbf{0}_{(N-r)\times r} \cdot \mathbf{V} \end{bmatrix}, \end{aligned} \tag{4.6}$$

where $\sigma_1, \sigma_2, \cdots, \sigma_r$ are the r non-zero singular values of $[\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r]$. Then, given the fact that each column of data matrix $\mathbf{X}$ belongs to the signal-space $\mathbb{S} = span(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r)$, it is easy to show that $\mathbf{X}$ is also graph $\mathcal{B}$-bandlimited with respect to $\mathbf{Q}^{-1}$. We compute the graph frequency response matrix of $\mathbf{X}$ as follows:

$$\begin{aligned} \tilde{\mathbf{X}} =& \mathbf{Q}^{-1} \cdot \mathbf{X}, \\ \overset{(a)}{=}& \mathbf{Q}^{-1} \cdot \Big[[\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r] \cdot \mathbf{\Pi}\Big], \\ \overset{(b)}{=}& \begin{bmatrix} diag\left([\sigma_1, \cdots, \sigma_r]^T\right) \cdot \mathbf{V} \cdot \mathbf{\Pi} \\ \mathbf{0}_{(N-r)\times r} \cdot \mathbf{V} \cdot \mathbf{\Pi} \end{bmatrix}. \end{aligned} \tag{4.7}$$

In Eq. (4.7), (a) holds for the fact that each column of $\mathbf{X}$ can be expressed by the independent supports spanning for the signal-space, and therefore there exists a matrix $\mathbf{\Pi}$ of size $r \times K$, such that $\mathbf{X} = [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r] \cdot \mathbf{\Pi}$. (b) shows that only the first r rows of $\tilde{\mathbf{X}}$ are non-zero vectors, suggesting the graph $\mathcal{B} = \{1, 2, \cdots, r\}$-bandlimited property of $\mathbf{X}$ with respect to the GFT operator $\mathbf{Q}^{-1}$.

4.2.2 Sampling and Recovery Designs

After the derivation of the data-driven GFT operator from Eqs. (4.4)-(4.7), we here elaborate how to determine an appropriate sampling vertex subset $\mathcal{C}$. In essence, the idea is to identify $\mathcal{C}$, such that there exists an one-to-one transformation between the whole time-varying network signals $\mathbf{X}$ and the sampled matrix $\mathbf{X}_{\mathcal{C}\mathcal{K}}$. To achieve this, we bridge the two using the graph-frequency response $\tilde{\mathbf{X}}$, and try to find the reversible mapping between the whole time-varying network signals $\mathbf{X}$ and graph-frequency response $\tilde{\mathbf{X}}$, and the sampled data $\mathbf{X}_{\mathcal{C}\mathcal{K}}$.

Given the graph $\mathcal{B} = \{1, 2, \cdots, r\}$-bandlimited property of $\mathbf{X}$, we use $\mathcal{B}$ as the graph sampling bandwidth set, selecting the graph frequency indices (rows) of the graph frequency response matrix. As such, we have $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$ as the sub-matrix of the frequency response $\tilde{\mathbf{X}}$ with rows (corresponding graph-frequency indices) selected from $\mathcal{B}$, and columns (discrete time) selected from $\mathcal{K} = \{1, 2, \cdots, K\}$. Then, the sampling and recovery can be therefore converted to search two reversible transformations, i.e., between $\mathbf{X}$ and $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$, and between $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$ and $\mathbf{X}_{\mathcal{C}\mathcal{K}}$.

For $\mathbf{X}$ and $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$, given the orthogonal property of the GFT operator (i.e., $\mathbf{Q}^{-1} = \mathbf{Q}^T$), we can extract all non-zero graph-frequency responses by selecting the rows in $\mathcal{B}$ of $\mathbf{Q}^T$, i.e.,

$$\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}} = \mathbf{Q}^T_{\mathcal{N}\mathcal{B}} \cdot \mathbf{X}, \tag{4.8}$$

and reversibly derive the time-varying network signals $\mathbf{X}$ from $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$, i.e.,

$$\mathbf{X} = \mathbf{Q}_{\mathcal{N}\mathcal{B}} \cdot \tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}. \tag{4.9}$$

As such, the one-to-one mapping between the whole time-varying network signals $\mathbf{X}$ and the $\mathcal{B}$-bandlimited graph-frequency response is established.

Then, for $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$ and $\mathbf{X}_{\mathcal{C}\mathcal{K}}$, we establish the following equation. For any selection $\mathcal{C} \subset \mathcal{N}$, $\mathbf{X}_{\mathcal{C}\mathcal{K}}$ can be derived by selecting the rows in $\mathcal{C}$ of the left and right hand sides of Eq. (4.9), i.e.,

$$\mathbf{X}_{\mathcal{C}\mathcal{K}} = \mathbf{Q}_{\mathcal{C}\mathcal{B}} \cdot \tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}. \tag{4.10}$$

From Eq. (4.10), we notice that $rank(\mathbf{X}_{\mathcal{C}\mathcal{K}}) \leq \min\{rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}}), rank(\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}})\}$, which indicates that an existence of the reversible computation between $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$, if $\mathbf{Q}_{\mathcal{C}\mathcal{B}}$ has full column rank, i.e.,

$$rank\left(\mathbf{Q}_{\mathcal{C}\mathcal{B}}\right) = |\mathcal{B}| = r. \tag{4.11}$$

As such, the computation from $\mathbf{X}_{\mathcal{C}\mathcal{K}}$ to $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$ can be pursued using the pseudo inverse of $\mathbf{Q}_{\mathcal{C}\mathcal{B}}$, i.e.,

$$\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}} = (\mathbf{Q}^T_{\mathcal{C}\mathcal{B}} \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}})^{-1} \cdot \mathbf{Q}^T_{\mathcal{C}\mathcal{B}} \cdot \mathbf{X}_{\mathcal{C}\mathcal{K}}. \tag{4.12}$$

Finally, by combining Eqs. (4.8)-(4.12), we derive the one-to-one mapping between the whole time-varying network signals $\mathbf{X}$ and its sample $\mathbf{X}_{\mathcal{C}\mathcal{K}}$, and the recovery of $\mathbf{X}$, denoted as $\hat{\mathbf{X}}$ can be computed as:

$$\hat{\mathbf{X}} = \mathbf{Q}_{\mathcal{N}\mathcal{B}} \cdot (\mathbf{Q}^T_{\mathcal{C}\mathcal{B}} \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}})^{-1} \cdot \mathbf{Q}^T_{\mathcal{C}\mathcal{B}} \cdot \mathbf{X}_{\mathcal{C}\mathcal{K}}. \tag{4.13}$$

After the derivation of the sampling and signal recovery process, we provide

in the following an intuitive explanation for understanding. At first, for any graph $\mathcal{B}$-bandlimited time-varying network signals $\mathbf{X}$ with respect to the GFT operator $\mathbf{Q}^{-1}$, the graph sampling bandwidth set from the graph-frequency domain should contain all the non-zero graph-frequency responses, so that the information from the graph frequency domain will not be lost. In other words, the reversible transformation between the whole time-varying network signals $\mathbf{X}$ and its truncated graph-frequency response $\tilde{\mathbf{X}}_{\mathcal{BK}}$ exists. Then, Eq. (4.11) provides the criterion for sampling vertex subset selection, which ensures the one-to-one transformation between the graph-frequency response $\tilde{\mathbf{X}}_{\mathcal{BK}}$ and the sampled data $\mathbf{X}_{\mathcal{CK}}$. In this view, we can select the sampling vertices satisfying Eq. (4.11), and the recovery process can be pursued by the combination of the above two mappings. For clarity, we provide the sampling vertex subset selection algorithm and the signal recovery algorithm in the following.

Sampling Algorithm Flow

To implement Eq. (4.11), we provide the sampling process in Algorithm 1. The input is the prior knowledge of the signal-space $\mathbb{S}$, and the time-varying network signals $\mathbf{X}$ that is waiting to be sampled.

Algorithm 1 Sampling Vertex Subset Selection Algorithm

Input: Prior signal-space $\mathbb{S} = span\{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r\}$.

1: Derive the GFT operator $\mathbf{Q}^{-1}$ using Eqs. (4.4)-(4.5).
2: Initialize $\mathcal{C} = \emptyset$.
3: **while** $|\mathcal{C}| < r$ **do**
4: Select vertex $n = \text{argmax}_{m \in \mathcal{N} \setminus \mathcal{C}} \, \sigma_{\min} \left(\mathbf{Q}_{(\mathcal{C}+\{m\})\mathcal{B}} \right)$
5: $\mathcal{C} = \mathcal{C} \cup \{n\}$
6: **end while**

Output: GFT operator $\mathbf{Q}^{-1}$, and the sampling vertex subset $\mathcal{C}$.

Step 1 is to compute the data-driven GFT operator using the prior knowledge of the signal-space. Step 2 is to initialize the sampling vertex subset. Steps 3-6 are to select the sampling vertex subset $\mathcal{C}$ under the condition provided by Eq. (4.11). We can notice that various selections of $\mathcal{C}$ can hold the condition in Eq. (4.11). In order to achieve the robust sampling results for signal recovery, we select $\mathcal{C}$ by maximizing the minimum singular of $\mathbf{Q}_{\mathcal{CB}}$. Denoting $\sigma_{\min}(\cdot)$ as the minimum singular of a matrix, we express the the sampling vertex subset selection process as:

$$\mathcal{C}_{\text{opt}} = \underset{\mathcal{C} \subset \mathcal{N}}{\text{argmax}} \, \sigma_{\min} \left(\mathbf{Q}_{\mathcal{CB}} \right). \tag{4.14}$$

As such, a greedy algorithm is adopted to realize Eq. (4.14), and is provided by

Steps 3-6. Finally, the output is the data-driven GFT operator $\mathbf{Q}^{-1}$, and the selected sampling vertex subset $\mathcal{C}$.

Signal Recovery using Data-Driven GFT Operator

After the determination of the sampling vertex subset $\mathcal{C}$, samples can be collected as $\mathbf{X}_{\mathcal{CK}}$. We provide the recovery algorithm by Algorithm 2.

Algorithm 2 Data-Driven Signal Recovery Algorithm.

Input: Sampled signals $\mathbf{X}_{\mathcal{CK}}$, the GFT operator $\mathbf{Q}^{-1}$, and sampling vertex subset $\mathcal{C}$.

1: Compute $\mathbf{Q}_{\mathcal{CB}}$ by selecting the rows of $\mathbf{Q}_{\mathcal{NB}}$ with indices/subscripts in $\mathcal{C}$.

2: Derive the recovered time-varying network signals $\tilde{\mathbf{X}}$ via Eq. (4.13).

Output: The recovered signals $\tilde{\mathbf{X}}$.

In Algorithm 2, the input is the sampled signals $\mathbf{X}_{\mathcal{CK}}$, the data-driven GFT operator $\mathbf{Q}^{-1}$, and the sampling vertex subset $\mathcal{C}$. Step 1 is to construct $\mathbf{Q}_{\mathcal{CB}}$. Step 2 is to recover the whole time-varying network signals $\hat{\mathbf{X}}$ using Eq. (4.13).

4.3 Molecular Relay Data-Driven GFT

In WDN monitoring applications, the sensors on selected vertices will transmit their sampled data (signals) to a hub for the recovery of the whole time-varying network signals, which has been overlooked by our previous studies. To address this, existing works rely on either the ground penetrating technique based underground wireless communication systems (e.g., the ultra-low frequency, underground wave-guide [68], and the magnetic induction [69]), or the fixed line access. However, the former tends to use a vulnerable and bulky system, due to the potential antenna damage and its inability to embed enough to enable large-scale and continuous data gathering. The latter may encounter damages to cables due to high underground pressure. This therefore motivates us to design new techniques for data transmission.

Recent developments on molecular communications [70, 71] have opened up a possibility of using messenger molecules (e.g., the encoded DNA molecule [72]) for data transmission in WDN. In terms of the molecular communication propagation channel, most of the current studies focus on the nano-scale diffusion channels, where mass diffusion, other than the flow or advection propagation, dominates the propagation spread and the communications [70, 73, 74]. These, unfortunately, are not very consistent with the WDN water propagation where water-flow advection dominates the process. To study the macro-scale molecular communications relying

on advection propagation, the work in [75] develops a point-to-point erasure channel, characterizing the arrival probability of the messenger molecules that may be lost due to the absorption or entrapment in porous media bio-membrane examples. Similar to this bio-membrane cases, water-flow in the WDN provides an advection propagation and water-demand in each junction causes the information loss due to the industrial or domestic water usages. This therefore inspires us to design molecular communication based data transmission systems using the pipes of WDN itself. In the following, we will introduce our proposed novel molecular communication relay data-driven (MRDD) GFT sampling system, which is able to deploy sensors and transmit the sampled signals to the hub using the water-flow of the WDN itself, therefore avoiding both the complex ground penetrating methods and the extra fixed lines for communications.

4.3.1 Molecular Relay

The sketch of the MRDD GFT sampling system is provided as follows. We select one WDN vertex as the hub for gathering the transmissions of the sampled data, and for recovering the whole time-varying network signals. Each deployed sensor is equipped with one specific type of (harmless) DNA molecules [72], which aims to encode the sampled data embedded into the molecular structure at each discrete time k. Here, the synchronization of sensors is realized by the low-rate blind synchronization technique in [76]. Then, the encoded DNA molecules will be transmitted to the hub vertex via the propagation of the water-flow in WDN itself. From the above process, one prerequisite for sampling vertex subset selection and the hub vertex selection is to ensure the existence of paths from each selected vertex to the hub, leveraging which the hub can receive the molecular reports from each sensor for signal recovery.

For the modelling of the molecular communication channel, we construct an erasure channel for the following two reasons. First, given the dynamical water flows and different water-demands in junctions (vertices) and pipes, the transmitted molecules may be lost, which will subsequently lead to a low signal recovery accuracy at the hub. Second, we measure and notice that the molecular diffusion rate (an order of $10^{-9}\mathrm{m}^2/\mathrm{s}$ [73, 74]) is trivial when compared to the velocity of the water flow (typically an order of $1\mathrm{m}^2/\mathrm{s}$ [67]). In this view, we remain the leading fluid parameters for arrival probability computation, and omit the weak diffusion mechanism.

As such, the erasure channel modelling with an arrival probability matrix can be specified as

$$\mathbf{APM} = [P_{i,j}], \quad i, j \in \mathcal{N}, \tag{4.15}$$

where $P_{i,j}$ denotes the arrival probability for one molecule transmitted from vertex j and received by vertex i. To compute $P_{i,j}$, we need to traverse all the paths from vertex j to vertex i. We first define the arrival probability of one molecule transmitted from j to i using link $a_{i,j}$ (if existed), i.e.,

$$p_{i,j} = \frac{E_{i,j}}{\sum_{n\in\mathcal{N}} E_{n,j} + R_j}, \tag{4.16}$$

where $E_{i,j}$ denotes the average of water-flow in pipe j to i ($E_{i,j} = 0$ of $a_{i,j} = 0$), and R_j is the average of water demands of vertex j. Then, $P_{i,j}$ can be computed by traversing all the paths from vertex j to vertex i, i.e.,

$$P_{i,j} = 1 - \prod_{\forall l} \left(1 - \prod_{l} p_{l_1,j} \cdot p_{l_2,l_1} \cdot \cdots \cdot p_{i,l_m}\right), \tag{4.17}$$

where $l = (j, l_1, \cdots, l_m, i)$ is one path from vertex j to vertex i.

It is noteworthy that the arrival probability matrix **APM** characterizes the relationship between the network topological structure (i.e., the paths) and the arrival probability of each vertex-pair. For further theoretical sampling vertex subset analysis, we assume an extremely large number of molecules for report transmission at each sensor, and the **APM** converges to a simplified binary matrix reflecting the path existence, where $P_{i,j} = 1$ or 0 represents the existence/nonexistence of the path from vertex j to vertex i. Nevertheless, for simulation part, we measure the **APM** in Fig. 4.12, and further demonstrate its influence on the signal recovery performance of the proposed scheme in Fig. 4.13.

4.3.2 Sensor Vertex Selection under Connectivity Constraint

Given the designed molecular relay mechanism, the selection scheme of the sampling vertex subset $\mathcal{C}$ is required to ensure the existence of paths from each sensor vertex in $\mathcal{C}$ to the hub vertex; otherwise, a non-existed path from sensor to hub will definitely cause report loss, and subsequently deteriorate the signal recovery performance.

Based on this, a novel sampling vertex subset selection method is designed, by adding the connectivity constraint, as follows:

$$(\mathcal{C}, n_{\text{hub}}) \;\; s.t.\; rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}}) = |\mathcal{B}|, \;\; \left\| \left(\mathbf{A}_0^L\right)_{\{n_{\text{hub}}\}\mathcal{C}} \right\|_{l_0} = |\mathcal{C}|. \tag{4.18}$$

In Eq. (4.18), we still use the proposed data-driven GFT operator $\mathbf{Q}^{-1}$ in Eqs (4.4)-(4.5), where both $\mathbf{Q}^{-1}$ and $\mathcal{B}$ are derived from the geaph bandwidth set of the prior knowledge of the signal-space. n_{hub} accounts for the vertex of hub selected

from the total WDN vertices. $\mathbf{A}_0 = \mathbf{A} + \mathbf{I}_N$ is the adjacent matrix $\mathbf{A}$ involving the self-loop of each vertex, where $\mathbf{I}_N$ represents the identity matrix of size $N \times N$. Using the L-exponent of $\mathbf{A}_0$ (L is large), we construct the matrix $\mathbf{A}_0^L$ to characterize the existence of any $\leq L$ -length paths between each vertex pair, whereby a nonzero (i, j) element of $\mathbf{A}_0^L$ represents an existed path from vertex j to vertex i. $(\mathbf{A}_0^L)_{\{n_{\text{hub}}\}\mathcal{C}}$ is a vector of $(\mathbf{A}_0)^L$ whose row is with the index n_{hub} and whose columns are selected by their indices in $\mathcal{C}$, where each element indicates the existence/non-existence path status from a vertex in $\mathcal{C}$ to the hub vertex n_{hub}. This combines the use of l_0-norm is to count the number of the existed paths from vertices in $\mathcal{C}$ to the hub vertex n_{hub}. $\|(\mathbf{A}_0^L)_{\{n_{\text{hub}}\}\mathcal{C}}\|_{l_0} = |\mathcal{C}|$ therefore ensures the existence of paths from all sensor vertices in $\mathcal{C}$ to the hub n_{hub}. As such, Eq. (4.18) is to search the pair of sampling vertex subset $\mathcal{C} \subset \mathcal{N}$, to the hub vertex $n_{\text{hub}} \in \mathcal{N}$, i.e., $(\mathcal{C}, n_{\text{hub}})$, such that, the successful sampling and molecular reporting to the hub can be pursued for recovering the whole time-varying network signals.

To solve Eq. (4.18), we provide the algorithm flow in the following.

Step 1: Find all potential $n_{\text{hub}} \in \mathcal{N}$ whose l_0-norm of the n_{hub}th row of $\mathbf{A}_0^L$ is greater than $|\mathcal{B}|$, i.e., $\|(\mathbf{A}_0^L)_{\{n_{\text{hub}}\}\mathcal{N}}\|_{l_0} \geq |\mathcal{B}|$. Otherwise, if one selects the hub vertex n_{hub} with $\|(\mathbf{A}_0^L)_{\{n_{\text{hub}}\}\mathcal{N}}\|_{l_0} < |\mathcal{B}|$, then it violates Eq. (4.18) by providing $\|(\mathbf{A}_0^L)_{\{n_{\text{hub}}\}\mathcal{C}}\|_{l_0} \leq \|(\mathbf{A}_0^L)_{\{n_{\text{hub}}\}\mathcal{N}}\|_{l_0} < |\mathcal{B}|$ even if $|\mathcal{B}| = |\mathcal{C}|$ can be approached deduced by $rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}}) = |\mathcal{B}|$.

Step 2: For each potential hub vertex n_{hub}, a revised topological and data-driven combined GFT operator is designed, via the combination of the network topology structure and the data-driven GFT operator $\mathbf{Q}^{-1}$. Such a revised GFT operator reads:

$$\mathbf{Q}^{(n_{\text{hub}})} = diag\left(\mathbf{1}\left((\mathbf{A}_0^L)_{\{n_{\text{hub}}\}\mathcal{N}}\right)\right) \cdot \mathbf{Q}, \tag{4.19}$$

where $diag(\cdot)$ is to diagonalize a vector into diagonal matrix. $\mathbf{1}(\cdot)$ is a function to assign the non-zeros of a vector as 1, and zeros as 0. By doing so, the rows in such topology-data combined GFT operator $\mathbf{Q}^{(n_{\text{hub}})}$ are zeros if their corresponding vertices do not have paths to the hub vertex n_{hub}. Therefore, using the topological and data-driven GFT operator $\mathbf{Q}^{(n_{\text{hub}})}$, instead of $\mathbf{Q}^{-1}$, is a prerequisite to satisfy the conditions in Eq. (4.18), for the further sampling vertex subset selection and signal recovery.

Step 3: Select the vertex as hub whose topology-data GFT operator $\mathbf{Q}_{\mathcal{N}\mathcal{B}}^{(n_{\text{hub}})}$ has the minimum condition number, i.e.,

$$n_{\text{hub}} = \underset{n \in \mathcal{N}_{\text{potential}}}{\text{argmin}} \; cond\left(\mathbf{Q}_{\mathcal{N}\mathcal{B}}^{(n)}\right), \tag{4.20}$$

where $\mathcal{N}_{\text{potential}} \subset \mathcal{N}$ is the set composed of all potential hub vertices from Step 2, and $cond(\cdot)$ represents the condition number of a matrix. This step is to ensure the signal recovery, as the lower the condition number of the operator, the more accurate and robust the recovery result.

Step 4: For the selected hub vertex n_{hub}, identify $\mathcal{C} \subset \mathcal{N}$ to maintain $rank(\mathbf{Q}_{\mathcal{CB}}^{(n_{\text{hub}})}) = |\mathcal{B}|$, so that a robust inverse matrix of $\mathbf{Q}_{\mathcal{CB}}^{(n_{\text{hub}})}$ exists. We implement this by minimizing the condition number of $\mathbf{Q}_{\mathcal{CB}}^{(n_{\text{hub}})}$, i.e.,

$$\mathcal{C} = \underset{\mathcal{C} \subset \mathcal{N}}{\operatorname{argmin}}\, cond\left(\mathbf{Q}_{\mathcal{CB}}^{(n_{\text{hub}})}\right). \tag{4.21}$$

Given that Eq. (4.21) is a NP-hard problem, we hereby use a *greedy* algorithm, in which we find and add the ith vertex, i.e., $\mathcal{C} \leftarrow \mathcal{C} \cup \{i\}$, where such i satisfies $i = \text{argmax}_{j \in \mathcal{N} \setminus \mathcal{C}}\, cond(\mathbf{Q}_{(\mathcal{C}+\{j\})\mathcal{B}}^{(n_{\text{hub}})})$.

4.3.3 Signal Recovery with Potential Report Loss

The recovery process of the time-varying network signals can be simply pursued by taking $\mathbf{Q}_{\mathcal{CB}}^{(n_{\text{hub}})}$ and $\mathcal{C}$ into Eq. (4.13).

It is noteworthy that in the MRDD GFT sampling system, even if we ensure the connectivity from all sensor vertices in $\mathcal{C}$ to the hub vetex n_{hub}, a report may still be lost given the molecule erasure channel and the **APM** in Eq. (4.15). To cope with this, we use the last arrival report from the corresponding sensor vertex for current signal recovery. To be specific, supposing the report from vertex $v \in \mathcal{C}$ is lost at k discrete-time, we use its last arrival report as a replacement for current lost one, i.e., $\mathbf{X}_{v,k} = \mathbf{X}_{v,k'}$, $k' < k$ the last discrete-time with molecular report arrived from vertex v.

Another idea to prevent the information loss relies on the rateless channel coding. This includes the widely-adopted Luby-Transform (LT) code and cascaded Hamming-LT (Raptor) code for macro-scale molecular erasure channels [75]. To be specific, in a given time period, the LT code can be used for each sensor vertex to encode its time-series samples and transmit via the molecules. As such, at the hub vertex, the entire samples can be received by decoding the corresponding LT codes. One limit is the potential decoding latency at the hub, as obtaining all samples from one sensor vertex requires to wait completely receiving the redundant LT codes (e.g., with code rate lesser than 1).

4.4 Recovery Error Analysis

In this part, we analyze the recovery error of the time-varying network signals in mathematical manners. To be specific, we firstly give a closed-form expression of the signal recovery error. Then, given the graph cut-off bandwidth set as $\mathcal{B}_{\text{cut-off}} = \{1, 2, \cdots, r\}$ from the prior knowledge of the signal-space, i.e., $\mathbb{S} = span\{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r\}$, we consider two types of under-sampling cases, with the graph sampling bandwidth set $\mathcal{B}$ as $\mathcal{B} \subsetneqq \mathcal{B}_{\text{cut-off}}$, and $\mathcal{B} \supset \mathcal{B}_{\text{cut-off}}$ but $|\mathcal{C}| < |\mathcal{B}|$, respectively, where the former accounts for the graph-frequency domain under-sampling, and the latter accounts for the network-domain under-sampling.

Theoretical Recovery Error

The recovery error of the time-varying network signals is measured by the squared difference of two Frobenius norms, denoted as, $\|\cdot\|_{fro}$. Given a sampling bandwidth set $\mathcal{B} \subset \{1, \cdots, N\}$ and sampling vertex subset $\mathcal{C}$, the recovery error (root squared error, RSE), denoted as $rse(\mathcal{B}, \mathcal{C})$ is computed as:

$$
\begin{aligned}
rse(\mathcal{B}, \mathcal{C}) &= \sqrt{\|\mathbf{X}\|_{fro}^2 - \|\hat{\mathbf{X}}\|_{fro}^2} \overset{\text{(c)}}{=} \sqrt{\|\mathbf{X}\|_{fro}^2 - \|\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2} \\
&\overset{\text{(d)}}{=} \begin{cases} \sqrt{\|\mathbf{X}\|_{fro}^2 - tr\left(\mathbf{X}_{\mathcal{C}\mathcal{K}} \cdot \mathbf{X}_{\mathcal{C}\mathcal{K}}^T \cdot \left(\mathbf{Q}_{\mathcal{C}\mathcal{B}} \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}}^T\right)^{-1}\right)} & |\mathcal{C}| < |\mathcal{B}| \\ \sqrt{\|\mathbf{X}\|_{fro}^2 - tr\left(\mathbf{X}_{\mathcal{C}\mathcal{K}} \cdot \mathbf{X}_{\mathcal{C}\mathcal{K}}^T \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}} \cdot \left(\mathbf{Q}_{\mathcal{C}\mathcal{B}} \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}}^T\right)^{-1} \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}}^T\right)} & |\mathcal{C}| \geq |\mathcal{B}| \end{cases},
\end{aligned}
\tag{4.22}
$$

where $tr(\cdot)$ is to compute the trace of a matrix, and $\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}$ is the recovered graph-frequency response from the sampled data $\mathbf{X}_{\mathcal{C}\mathcal{K}}$. In Eq. (4.22), (c) is because $\|\hat{\mathbf{X}}\|_{fro}^2 = \|\mathbf{Q}_{\mathcal{N}\mathcal{B}}\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2 = tr(\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}^T \mathbf{Q}_{\mathcal{N}\mathcal{B}}^T \mathbf{Q}_{\mathcal{N}\mathcal{B}} \hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}) = \|\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2$. (d) is to divide the equation into two cases. For $|\mathcal{C}| \geq |\mathcal{B}|$, we have $\mathcal{C}$ satisfying $rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}}) = |\mathcal{B}|$ according to Algorithm 1, and $\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}$ can be easily derived using Eq. (4.12). For $|\mathcal{C}| < |\mathcal{B}|$, we have $rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}}) < |\mathcal{B}|$, and the recovered graph-frequency response is $\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}} = \mathbf{Q}_{\mathcal{C}\mathcal{B}}^T \cdot (\mathbf{Q}_{\mathcal{C}\mathcal{B}} \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}}^T)^{-1} \cdot \mathbf{X}_{\mathcal{C}\mathcal{K}}$, based on which $\|\hat{\hat{\mathbf{X}}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2$ can be computed as provided in Eq. (4.22).

Two Under-Sampling Cases

After providing the mathematical recovery error in Eq. (4.22), we study specifically two types of under-sampling cases, i.e., Case I: under graph-frequency domain sampling $\mathcal{B} \subsetneqq \mathcal{B}_{\text{cut-off}}$, and Case II: under network domain sampling $\mathcal{B} \supset \mathcal{B}_{\text{cut-off}}$ but $|\mathcal{C}| < |\mathcal{B}|$.

Case I accounts for under graph-frequency domain sampling, characterizing the information loss from the graph-frequency domain. In such a case, the sampling process will definitely omit the information of some important graph-frequency components whose indices belong to $\mathcal{B}_{\text{cut-off}} \setminus \mathcal{B}$. It is noteworthy that, when using the GFT sampling techniques, such an omission from the graph-frequency domain cannot be compensated by any selection of $\mathcal{C}$ (even if $\mathcal{C} = \mathcal{N}$), and therefore a lower-bound of recovery RSE can be specified as:

$$\begin{aligned} rse(\mathcal{B} \subsetneqq \mathcal{B}_{\text{cut-off}}, \mathcal{C}) &\geq \|\tilde{\mathbf{X}}_{\mathcal{B}_{\text{cut-off}} \setminus \mathcal{B}\ \mathcal{K}}\|_{fro}^2 \\ &= \|\tilde{\mathbf{X}}_{\mathcal{B}_{\text{cut-off}}\mathcal{K}}\|_{fro}^2 - \|\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2. \\ &= \|\mathbf{Q}_{\mathcal{N}\mathcal{B}_{\text{cut-off}}}^T \cdot \mathbf{X}\|_{fro}^2 - \|\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2 \\ &= tr\left(\mathbf{X}^T \cdot \mathbf{Q}_{\mathcal{N}\mathcal{B}_{\text{cut-off}}} \cdot \mathbf{Q}_{\mathcal{N}\mathcal{B}_{\text{cut-off}}}^T \cdot \mathbf{X}\right) - \|\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2 \\ &= \|\mathbf{X}\|_{fro}^2 - \|\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2. \end{aligned} \tag{4.23}$$

Eq. (4.23) provides a lower-bound of the recovery RSE caused by graph frequency domain under-sampling. This subsequently suggests two facts. First, the graph-frequency domain under-sampling will cause at least an error as $\|\mathbf{X}\|_{fro}^2 - \|\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}\|_{fro}^2$, which equals the Frobenius-norm (energy) difference between the original matrix $\mathbf{X}$, and its sampled graph-frequency response matrix $\tilde{\mathbf{X}}_{\mathcal{B}\mathcal{K}}$. Second, in GFT based sampling method, the graph sampling bandwidth set $\mathcal{B}$ underpins the recovery performance, and therefore the selection of $\mathcal{B} = \mathcal{B}_{\text{cut-off}}$ serves as the prerequisite to avoid the information loss in graph-frequency domain.

Case II leads to the under-sampling RSE from the network domain, as the graph sampling bandwidth set $\mathcal{B}$ embraces the graph cut-off bandwidth $\mathcal{B}_{\text{cut-off}}$, i.e., $\mathcal{B} = \mathcal{B}_{\text{cut-off}}$. In such a case, the recovery RSE is totally affected by the selection of the sampling vertex subset $\mathcal{C}$. Here, we consider the under-sampling with $rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}}) = |\mathcal{C}| < |\mathcal{B}|$, since Algorithm 1 is still able to identify $\mathcal{C}$ such that the full row-rank property of $\mathbf{Q}_{\mathcal{C}\mathcal{B}}$ is achieved. Therefore, such network domain under-

sampling RSE can be computed by taking $\mathcal{B} = \mathcal{B}_{\text{cut-off}}$ into Eq. (4.22), i.e.,

$$\begin{aligned} &rse(\mathcal{B}_{\text{cut-off}}, |\mathcal{C}| < |\mathcal{B}_{\text{cut-off}}|) \\ =&\sqrt{\|\mathbf{X}\|_{fro}^2 - tr\left(\mathbf{X}_{\mathcal{C}\mathcal{K}} \cdot \mathbf{X}_{\mathcal{C}\mathcal{K}}^T \cdot \left(\mathbf{Q}_{\mathcal{C}\mathcal{B}_{\text{cut-off}}} \cdot \mathbf{Q}_{\mathcal{C}\mathcal{B}_{\text{cut-off}}}^T\right)^{-1}\right)}. \end{aligned} \tag{4.24}$$

Eq. (4.24) suggests two facts. First, an inappropriate selection of $\mathcal{C}$ can lead to the under-sampling recovery error from the network domain, and maintaining $\mathcal{C}$ such that $rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}}) = |\mathcal{B}|$ can avoid such under-sampling. Second, the recovery error caused by the network domain under-sampling is smaller than that of the recovered signals composed directly of the samples, i.e.,

$$rse(\mathcal{B}_{\text{cut-off}}, |\mathcal{C}| < |\mathcal{B}_{\text{cut-off}}|) < \sqrt{\|\mathbf{X}\|_{fro}^2 - \|\mathbf{X}_{\mathcal{C}\mathcal{K}}\|_{fro}^2}. \tag{4.25}$$

This indicates that even if network domain under-sampling may happen, the GFT sampling method is still workable, especially better than the direct collection of samples as the recovered signals.

4.5 Distinguish with Two State-of-the-Arts

In this section, we compare our proposed data-driven GFT sampling method with other two state-of-the-art compression schemes (from the conceptual manner). The first peer method is the topology-based GFT sampling methods in Section 2.3.2 [34, 77]. The second is the data-driven compressed sensing methods revised from the classical CS in Section 2.3.1 [24].

4.5.1 Topological based GFT Sampling Methods

Existing graphs Fourier transform sampling methods rely mostly on the topology based GFT operators, e.g., using the graph Laplacian or adjacent matrix. For example, one popular topological based GFT operator, denoted as $\mathbf{P}^{-1}$, is constructed via the eigenvectors of the graph Laplacian matrix $\boldsymbol{\mathcal{L}}$ (i.e., [34, 77], or in Section 2.3.2). The network signals they are interested in are graph bandlimited to such topological based GFT operator $\mathbf{P}^{-1}$. Given a graph cut-off bandwidth set $\mathcal{B} \subset \{1, 2, \cdots, N\}$, a graph $\mathcal{B}$-bandlimited signal (vector) $\mathbf{x} = [x_1, x_2, \cdots, x_N]^T$ with respect to the GFT operator $\mathbf{P}^{-1}$ is defined to have non-zero coefficients in $\tilde{\mathbf{x}} = \mathbf{P}^{-1}\mathbf{x}$ with indices belong to the set $\mathcal{B}$. The GFT sampling theory in Theorem 2 states that any graph $\mathcal{B}$-bandlimited signal $\mathbf{x}$ can be sampled using the sampling vertex subset $\mathcal{C} \subset \mathcal{N}$, and a complete recover of $\mathbf{x}$ from $\mathbf{x}_{\mathcal{C}}$ can be achieved if $rank\left(\mathbf{P}_{\mathcal{C}\mathcal{B}}\right) = |\mathcal{B}|$.

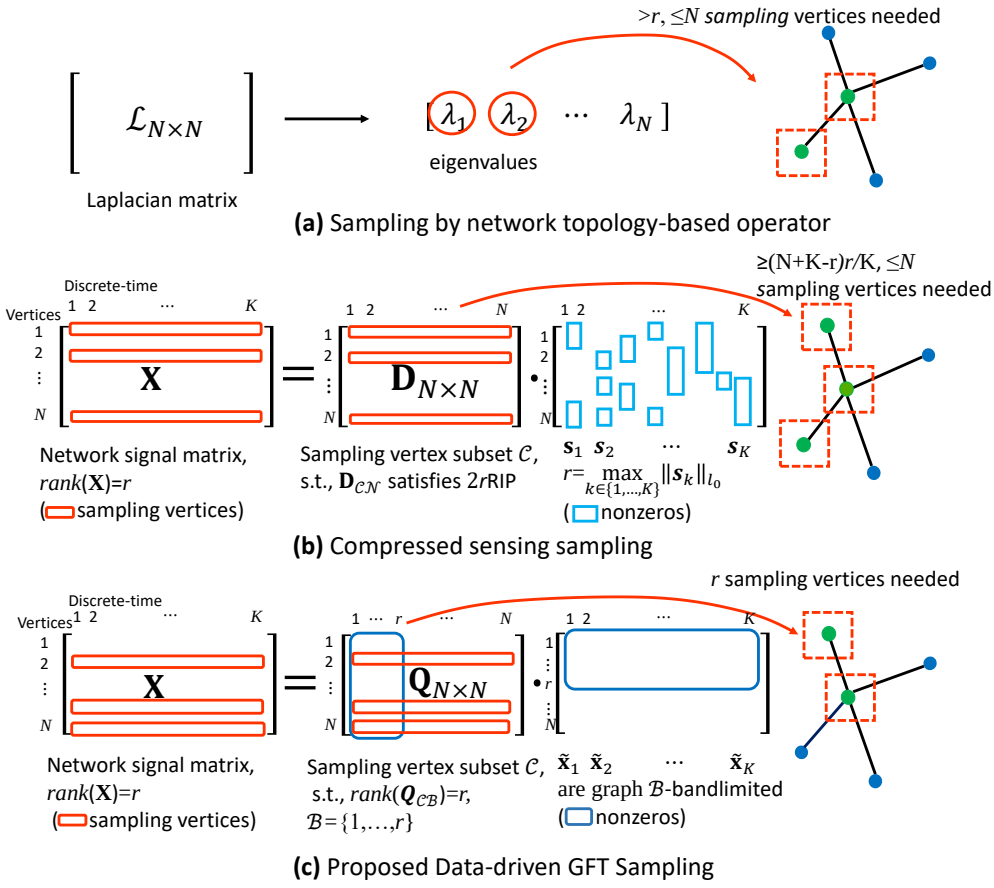

Figure 4.2: Comparison between the proposed data-driven GFT sampling method, the topology based GFT sampling, and the CS.

The main differences between the topological based GFT sampling and the proposed data-driven GFT sampling are listed as follows:

- First, in the topological based GFT sampling, the selection of sampling vertex subset $\mathcal{C}$ is determined solely by the topological structure of the network, since the GFT operator $\mathbf{P}^{-1}$ contains no underlying dynamic mechanism (e.g., dynamic governing equations or hidden dynamic mechanism from data), but is totally constructed by the network topological matrix (e.g., the graph Laplacian matrix or the graph adjacent matrix). This can be observed via Fig. 4.2(a), in which the sampling vertex subset is selected by mapping to the leading eigenvalues of the topology based matrix $\boldsymbol{\mathcal{L}}$.

- Second and more importantly, when dealing with the time-varying network signals (e.g., the WDN contaminant propagation), the direct use of the topological based GFT operator for network sampling and signal recovery is chal-

lenging. The topological GFT operator $\mathbf{P}^{-1}$, derived from either the Laplacian matrix or the adjacent matrix, cannot ensure that the whole time-varying network signals at different discrete time (i.e., $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K$) are graph $\mathcal{B}$-bandlimited within a same $\mathcal{B} \subsetneq \{1, 2, \cdots, N\}$. In other words, the signal dependency among vertices cannot be uncovered by the topology based operator $\mathbf{P}^{-1}$. As such, $\mathcal{B} = \{1, \cdots, N\}$ is inevitable, which will cause $\mathcal{C} \equiv \mathcal{N}$ (we show this in Figs. 4.7-4.10). In contrast, our proposed data-driven GFT sampling method constructs a GFT operator using the signal-space $\mathbb{S} = \{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r\}$. This enables to uncover the vertex signal dependency by characterizing the time-varying network signals at all discrete time as graph $\mathcal{B} = \{1, 2, \cdots, r\}$-bandlimited to the data-driven GFT operator $\mathbf{Q}^{-1}$. Therefore, the time-invariant sampling vertex subset $\mathcal{C}$ can be derived with $|\mathcal{C}| = r < N$ for sensor deployment.

4.5.2 Sampling using Compressed Sensing

CS is a framework to compress the (transformed) sparse signals by a few measurements (or samples). The classical CS theorem has been provided in Section 2.3.1, which should be revised and adjusted for the network sampling scenarios. In the context of the WDN monitoring applications, the idea is to sparsely represent $\mathbf{X}$ by an one-to-one transformation using a designed operator (matrix) $\mathbf{D}$ of size $N \times N$. Then, samples can be selected for the complete recovery of the sparse representation of $\mathbf{X}$, which subsequently can be used to reconstruct $\mathbf{X}$ via the inverse of the transforming operator $\mathbf{D}^{-1}$ [20, 23].

We provide the sampling process via Fig 4.2(b). For each discrete time $k \in \mathcal{K}$, we denote $\mathbf{s}_k$ as the sparse representation of $\mathbf{x}_k$ transformed by the operator $\mathbf{D}$, i.e., [20, 23, 24]

$$[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K] = \mathbf{D} \cdot [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K]. \tag{4.26}$$

Then, to achieve the time-invariant sampling vertex subset $\mathcal{C} \subset \mathcal{N}$, we revise the restricted isometric property (RIP) criteria in [20, 23, 24], and make it hold for the least sparse vector in $[\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K]$, i.e.,

$$1 - \delta \leq \frac{\|\mathbf{D}_{\mathcal{CN}} \cdot \mathbf{s}\|_{l_2}^2}{\|\mathbf{s}\|_{l_2}^2} \leq 1 + \delta, \ r = \max_{k \in \mathcal{K}} \|\mathbf{s}_k\|_{l_0}, \tag{4.27}$$

holds for any sparse $\mathbf{s}$ with $2 \cdot r$ nonzero elements, and some $\delta \in [0, 1]$. The setting of $r = \max_{k \in \mathcal{K}} \|\mathbf{s}_k\|_{l_0}$ is reasonable, since otherwise if $r < \max_{k \in \mathcal{K}} \|\mathbf{s}_k\|_{l_0}$, then $\mathbf{s}_k$ with $k = \text{argmax}_{k \in \mathcal{K}} \|\mathbf{s}_k\|_{l_0}$ cannot be recovered in the CS framework. Eq. (4.27) is

equivalent to select $\mathcal{C}$ to ensure any $2 \cdot r$ columns of $\mathbf{D}_{\mathcal{CN}}$ are independent, according to Theorem 1. After the selection of the time-invariant sampling vertex subset $\mathcal{C}$, the samples $\mathbf{X}_{\mathcal{C}\{k\}}$ is derived, and the sparse representation $\mathbf{s}_k$ can be recovered via the convex optimization: [20, 23, 24]

$$\hat{\mathbf{s}}_k = \underset{\mathbf{s}_k \in \mathbb{R}^N}{argmin} \|\mathbf{s}_k\|_{l_1}, \text{ such that } \mathbf{X}_{\mathcal{C}\{k\}} = \mathbf{D}_{\mathcal{CN}} \cdot \mathbf{s}_k. \tag{4.28}$$

Finally, we compute via the transforming matrix $\mathbf{D}$, and have $\hat{\mathbf{x}}_k = \mathbf{D} \cdot \hat{\mathbf{s}}_k$ and $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \cdots, \hat{\mathbf{x}}_K]$.

We list the similarity and difference between the proposed data-driven GFT sampling method and CS method in the following.

- Regarding the similarity, we observe from Fig. 4.2(b)-(c) that both the schemes determine the sampling vertex subset $\mathcal{C}$ by analyzing the non-zeros of their transformed signals. For the CS in Fig. 4.2(b), such non-zeros are called the sparsity, and they transform the original signal-space into an operator $\mathbf{D}$ determined domain that can sparsely represent the time-varying network signals $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K]$ by $[\mathbf{s}_1, \cdots, \mathbf{s}_K]$. As such, the sampling vertex subset $\mathcal{C}$ can be selected using the revised RIP, by mapping $\mathcal{C}$ to the least sparse vector in $[\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K]$. For the proposed data-driven GFT sampling provided in Fig. 4.2(c), the non-zeros are called the graph bandlimitedness. The scheme transforms the signals into the GFT operator determined graph-frequency bandlimited domain, and maps its graph cut-off bandwidth set $\mathcal{B}_{\text{cut-off}}$ to the sampling vertex subset $\mathcal{C}$ by $rank(\mathbf{Q}_{\mathcal{C}\mathcal{B}_{\text{cut-off}}}) = |\mathcal{B}_{\text{cut-off}}|$.

- The major difference is that whether the positions of such transformed non-zeros (sparsity and graph bandlimitedness) are known. In the proposed data-driven GFT sampling scheme (as is illustrated in Fig. 4.2(c)), we know the positions of the non-zeros, i.e., the graph cut-off bandwidth set $\mathcal{B}_{\text{cut-off}} = \{1, \cdots, r\}$. In other words, the time-varying network signals at each discrete-time compose a signal-space $BS(\mathcal{B}, \mathbf{Q}^{-1})$, which is the subset of that composed by all $r = |\mathcal{B}_{\text{cut-off}}|$ sparse signals. So, the data-driven GFT method can shrink the size of sampling vertex subset $|\mathcal{C}|$ to $|\mathcal{B}_{\text{cut-off}}| = r$, as the samples from $\mathcal{C}$ only needs to map to $BS(\mathcal{B}, \mathbf{Q}^{-1})$, rather than the one composed of all $r = |\mathcal{B}_{\text{cut-off}}|$ sparse signals. In contrast, the CS does not know the exact positions of the non-zeros, but only know the number of non-zeros, i.e., r. In this view, the study in [40] provides a theoretical proof that the CS requires at least $(N + K - r)r$ measurements (samples) in total K discrete times for

complete recovery of the time-varying network signals $\mathbf{X}$ of size $N \times K$ with $r = rank(\mathbf{X})$. This can be used to deduce an approximation of the size of the time-invariant sampling vertex subset as $|\mathcal{C}| = (N+K-r)r/K$, by computing $|\mathcal{C}| \cdot K \geq (N+K-r)r$. As such, it is straightforward that such number is larger than that of the proposed data-driven GFT sampling method, i.e., $|\mathcal{C}| = (N+K-r)r/K > r$. The comparison between CS and the proposed data-driven GFT are provided in Figs. 4.7-4.10.

4.6 Simulations & Experimental Results

In this section, we evaluate the proposed data-driven signal-space dependent GFT sampling method. At first, the performance of signal recovery is tested and analyzed with respect to the changes of the joint graph sampling bandwidth $\mathcal{B}$ and the sampling vertex subset $\mathcal{C}$. Then, recovery performance comparisons are provided, where the compared schemes are the topological Laplacian operator based GFT sampling, and the compressed sensing using PCA [20, 23, 24]. Third, we analyze the proposed molecular relay data-driven GFT sampling scheme, under the molecular connectivity constraint in WDN. The recovery performance is measured by the root mean square error (RMSE) between the recovered data matrix $\hat{\mathbf{X}}$ and the original data matrix $\mathbf{X}$, i.e.,

$$\text{RMSE} = \sqrt{\frac{1}{NK} \cdot \|\mathbf{X} - \hat{\mathbf{X}}\|_{fro}^2}. \tag{4.29}$$

We provide the background setting of WDN and its contaminant propagation as follows. An extended-period hydraulic and water quality simulation is performed using the stat-of-the-art EPANET engine [67], which aims to provide the propagation of the contaminant components through the WDN over time. Here, the WDN works within the pressurized mode and generates pressure-dependent water flows. The topological elements of the WDN contain the pipes, junctions, reservoirs and other hydraulic-related components. For each junction, the water-demand is predefined and varies with the time to simulate different user modes and behaviours in residential and industrial manners.

For our case-study, such a water simulation platform is used to track the spread of a contaminant in each vertex of the WDN. We configure the WDN with $N = 102$ vertices (see Fig. 4.3(a)). We simulate 100 different time-varying contaminant signal matrices. For each matrix, the contaminant is injected in a vertex for a predefined amount of time, spreads over the WDN, and is finally expelled by the water-demands on junctions. Each matrix $\mathbf{X}$ with an injection is simulated for 3

hours in $K = 168$ discrete times.

4.6.1 Influences on Recovery Performance

We first evaluate the recovery accuracy of our proposed data-driven GFT sampling method with respect to the changes of the graph sampling bandwidth set $\mathcal{B}$, and the sampling vertex subset $\mathcal{C}$. For this experiment, we wish to show whether the recovery RMSE approaches to zero under following two conditions: i) if the graph sampling bandwidth set contains the graph cut-off bandwidth set, i.e., $\mathcal{B} \supset \mathcal{B}_{\text{cut-off}}$, and ii) if the sampling vertex subset $\mathcal{C}$ maintains $rank(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{B}|$. We do so by changing the selections of $\mathcal{B}$ and $\mathcal{C}$ using Algorithm 1, and representing these changes by their sizes, i.e., $|\mathcal{B}|$ and $|\mathcal{C}|$. To be specific, for the size of graph sampling bandwidth set $|\mathcal{B}|$, we denote $|\mathcal{B}| \geq |\mathcal{B}_{\text{cut-off}}| = rank(\mathbf{X}) = r$ to represent $\mathcal{B} \supset \mathcal{B}_{\text{cut-off}}$, and denote $|\mathcal{B}| < |\mathcal{B}_{\text{cut-off}}|$ to represent $\mathcal{B} \subsetneqq \mathcal{B}_{\text{cut-off}}$. Likewise, for the size of sampling vertex subset $|\mathcal{C}|$, we use $|\mathcal{C}| \geq |\mathcal{B}|$ to represent $rank(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{B}|$, and use $|\mathcal{C}| < |\mathcal{B}|$ to indicate $rank(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{C}| < |\mathcal{B}|$.

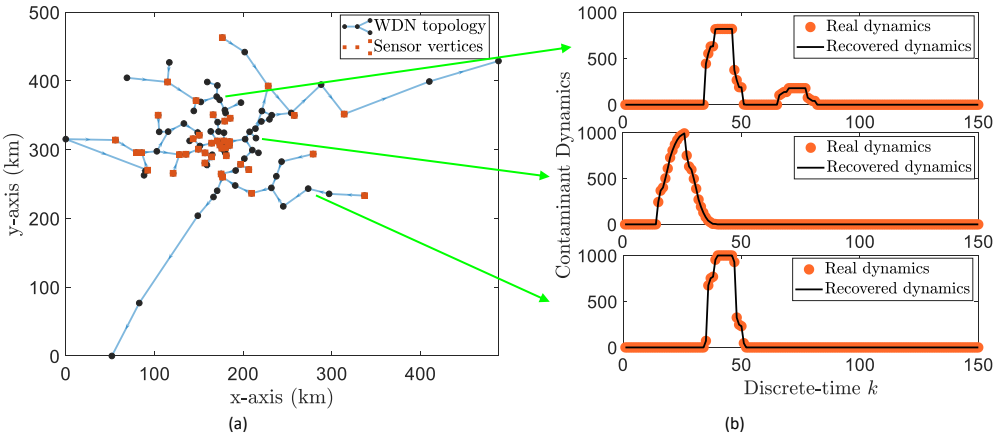

Figure 4.3: Illustration of sensor selections on WDN (a), and the recovery of contaminant signals on 3 un-sampled vertices (b).

One illustration of the proposed data-driven GFT sampling method and its recovery performance is shown in Fig. 4.3, where (a) is the topology of the WDN and the selected sensor vertices, and (b) shows the recovered signals on 3 un-sampled vertices. In this experiment, we assign i) the graph sampling bandwidth set equals the graph cut-off bandwidth set, i.e., $|\mathcal{B}| = |\mathcal{B}_{\text{cut-off}}| = r$, and 2) the sampling vertex subset satisfy $rank(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{B}|$ in Eq. (4.11), i.e., $|\mathcal{C}| = |\mathcal{B}| = |\mathcal{B}_{\text{cut-off}}|$. We figure out that the perfect recovery of the time-varying network signals is achieved.

Then, we evaluate the recovery RMSE by varying the sizes of both the graph

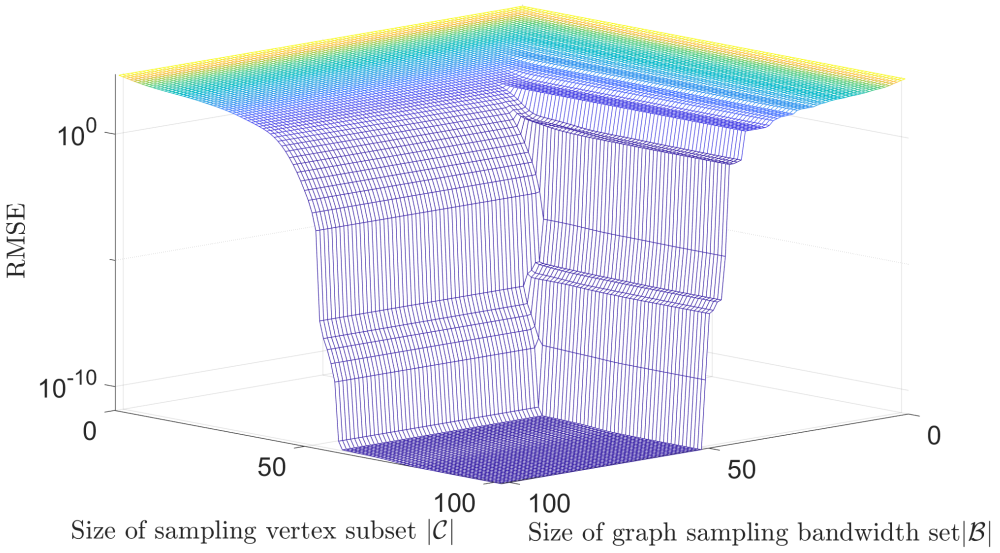

Figure 4.4: Recovery RMSE of the proposed data-driven GFT, versus both the size of sampling vertex subset $|\mathcal{C}|$ and the size of graph sampling bandwidth set $|\mathcal{B}|$.

sampling bandwidth set $|\mathcal{B}|$ and the sampling vertex subset $|\mathcal{C}|$. We can see from Fig. 4.4 that, the recovery RMSE decreases when $|\mathcal{B}|$ and $|\mathcal{C}|$ grow, and then converges to a very low value (an order of 10^{-10}) after $|\mathcal{B}|$ and $|\mathcal{C}|$ meet the conditions, i.e., $|\mathcal{C}| = |\mathcal{B}| = |\mathcal{B}_{\text{cut-off}}|$. We explain this with the analysis of Figs. 4.5-4.6.

RMSE versus Graph Sampling Bandwidth Set

Fig. 4.5 illustrates the recovery RMSE versus the size of the graph sampling bandwidth set $|\mathcal{B}|$. We here consider 3 fixed sizes of sampling vertex subset (i.e., $|\mathcal{C}| = 30, 40, 54$). We can firstly observe that the recovery RMSE stays lower as a larger $|\mathcal{C}|$ is used. This is because a larger size of sampling vertex subset can bring more samples for a more accurate signal recovery.

Secondly, it is seen that for all fixed $|\mathcal{C}|$, the recovery RMSEs get smaller when the sizes of graph sampling bandwidth set $|\mathcal{B}|$ increases from 0 to that of the cut-off set (i.e., $|\mathcal{B}_{\text{cut-off}}| = 54$). Then, the RMSEs keep unchanged after $|\mathcal{B}| > |\mathcal{B}_{\text{cut-off}}| = 54$. We explain this in the following. When $|\mathcal{B}| < |\mathcal{B}_{\text{cut-off}}|$, the time-varying network signals are under-sampled from the graph frequency domain, and therefore cannot be completely recovered given the information loss of some non-trivial graph frequency components. This is equivalent to the non-existence of the reversible mapping between the original data matrix $\mathbf{X}$ and the graph frequency response selected by the graph sampling bandwidth set $\mathcal{B}$, i.e., $\tilde{\mathbf{X}}_{\mathcal{BK}}$, since Eq. (4.9) holds no more if $|\mathcal{B}| < |\mathcal{B}_{\text{cut-off}}|$. In such a case, even if we select more vertices for

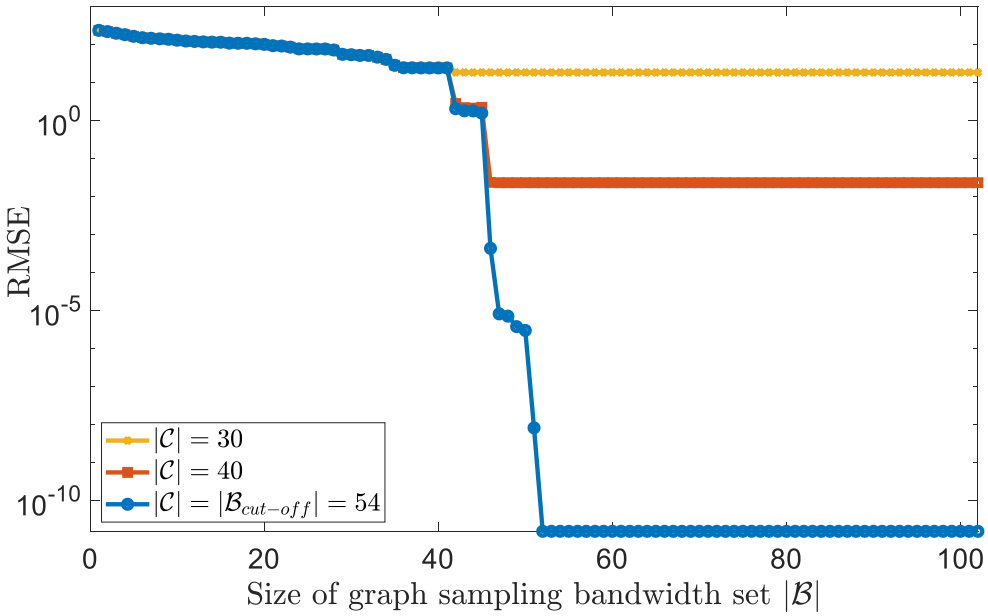

Figure 4.5: Recovery RMSE of the proposed data-driven GFT, versus the size of graph sampling bandwidth set $|\mathcal{B}|$.

$\mathcal{C}$ to ensure the sampling full-column rank condition, i.e., $rank(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{B}| < r$, it is still impossible to use the GFT sampling framework to recover $\mathbf{X}$ from the samples $\mathbf{X}_{\mathcal{CK}}$, given the under-sampling from the graph frequency domain. When $|\mathcal{B}| \geq |\mathcal{B}_{\text{cut-off}}| = 54$, the total information from the graph frequency domain is included, and the reversible mapping between the original data $\mathbf{X}$ and the graph frequency response $\tilde{\mathbf{X}}_{\mathcal{BK}}$ can be ensured. In such a case, the recovery accuracy is affected only by the sampling vertex subset.

RMSE versus Sampling Vertex Subset

Fig. 4.6 illustrates the recovery RMSE versus the size of the sampling vertex subset $|\mathcal{C}|$, given 3 fixed graph sampling bandwidth sets (e.g., $|\mathcal{B}| = 30, 40, 54$). It is observed that the recovery RMSE with a larger fixed $|\mathcal{B}|$ stays lower as opposed to those with smaller $|\mathcal{B}|$. For example, the recovery RMSE of $|\mathcal{B}| = 54$ is lower than that of $|\mathcal{B}| = 40$. We explain this as mentioned above that, the larger is the size of the graph sampling bandwidth set $|\mathcal{B}|$, the more graph frequency domain information can be taken, which leads to a better signal recovery accuracy.

Moreover, the recovery RMSEs of all fixed $|\mathcal{B}|$ become smaller at first when $|\mathcal{C}|$ grows from 0 to $|\mathcal{B}_{\text{cut-off}}| = 54$, and then keep unchanged after $|\mathcal{C}| > |\mathcal{B}_{\text{cut-off}}| = 54$. This is straightforward that, more selections of sensor vertices will enhance the

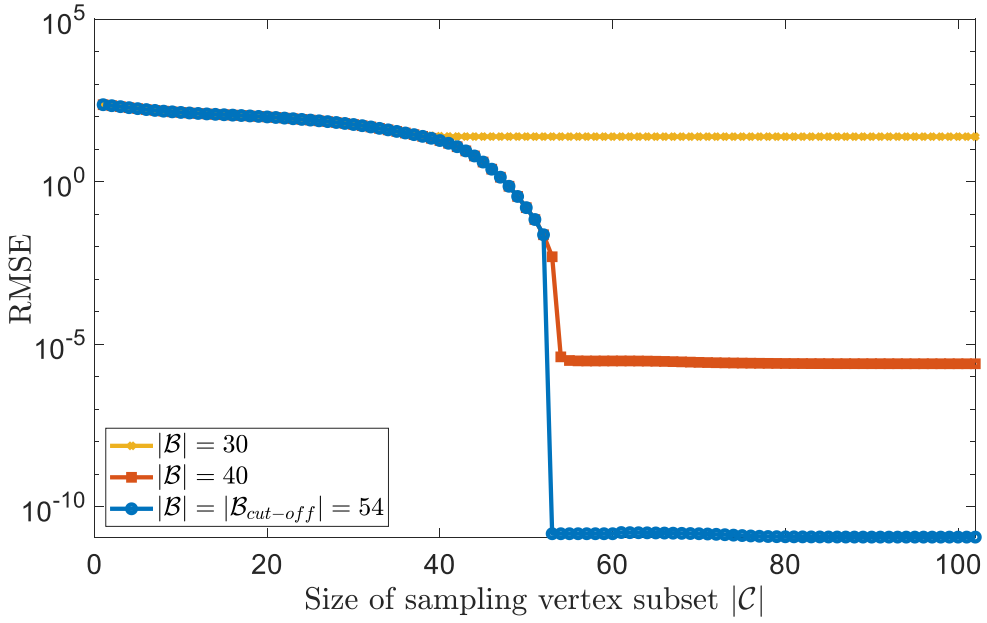

Figure 4.6: Recovery RMSE of the proposed data-driven GFT, versus the size of sampling vertex subset $|\mathcal{C}|$.

recovery accuracy, and will further leads to the perfect recovery when and after $|\mathcal{C}| \geq |\mathcal{B}_{\text{cut-off}}|$.

It is also noteworthy that, in the cases of graph frequency domain undersampling (i.e., $|\mathcal{B}| < |\mathcal{B}_{\text{cut-off}}|$), even if the size of the sampling vertex subset $|\mathcal{C}|$ exceeds the size of the graph cut-off bandwidth set, i.e., $|\mathcal{C}| > |\mathcal{B}_{\text{cut-off}}|$, the recovery accuracy cannot converge to its perfectness. This is due to the fact that the GFT sampling framework relies on the information from the graph frequency domain, which if missed, will definitely deteriorate the recovery accuracy.

4.6.2 Performance Comparisons

In this part, we compare our proposed data-driven GFT sampling method with other two state-of-the-art methods, i.e., the topological Laplacian operator based GFT sampling, and the CS scheme. We illustrate the comparison in Fig. 4.7-4.10. For the experiment illustrated in Figs. 4.7-4.8, the WDN structure is provided in Fig. 4.3(a), and the prior signal-space is derived by simulating the contaminant signals with one specific-vertex injection. In Figs. 4.7-4.8, one data matrix $\mathbf{X}$ is used, whose graph cut-off bandwidth set is measured as $|\mathcal{B}_{\text{cut-off}}| = 41$, i.e., $\mathcal{B}_{\text{cut-off}} = \{1, \cdots, r\}$ with $r = 41$. Different data matrices are used in the experiments illustrated in Figs. 4.9-4.10.

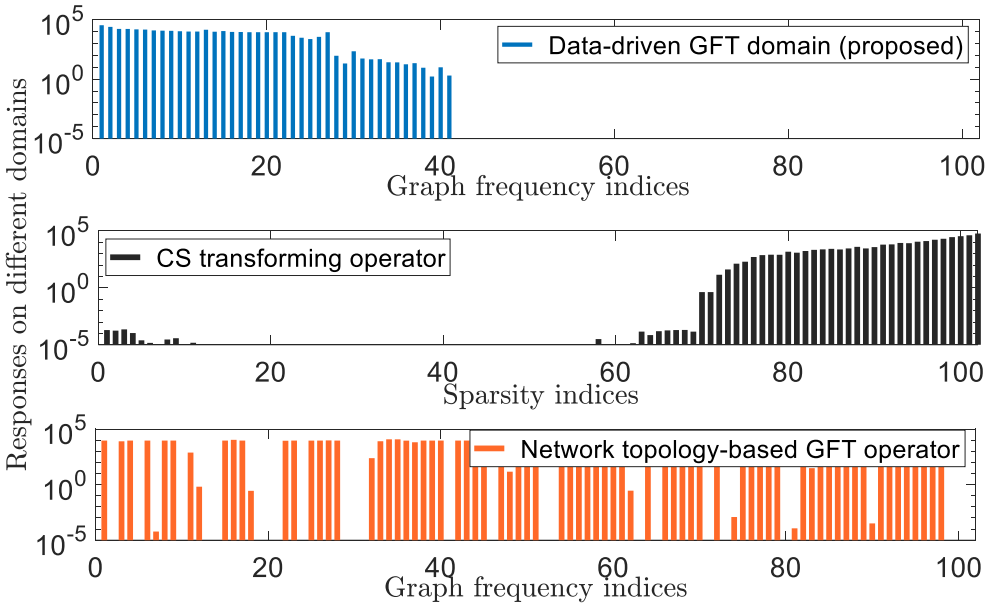

Figure 4.7: Comparison of responses to different operators, i.e., the data-driven GFT operator (top-plot), the CS operator (middle-plot), and the topological based GFT operator (bottom-plot).

In Fig. 4.7, we represent the frequency indices from different domains as x-axis. Here, the frequency domains are from the linear transformations of our proposed data-driven GFT operator, of topological Laplacian operator, and of the PCA-based CS operator. The y-axis accounts for the summations of the graph frequency responses over discrete time, i.e., $\sum_{k=1}^{K} |\tilde{\mathbf{x}}_k|$. It is seen that our proposed data-driven GFT sampling method can concentrate the graph frequency response only within the low-frequency area (i.e., $\mathcal{B} = \{1, \cdots, r\}$ with $r = 41$), as opposed to those using PCA based and topological Laplacian based operators. We explain this in the following. First, the topological Laplacian operator cannot characterize the signal dependency (e.g., the sparsity or the graph bandlimitedness) governed by the hidden dynamic mechanism, as it has only the network topological information. Second, PCA-based operator is limited due to its overlook of the network topology information. In contrast, our proposed data-driven GFT operator is derived from the signal-space, which involves both the hidden dynamic mechanism and the network topology, thereby capable of concentrating the network signals within the very low graph-frequency area $\mathcal{B} = \{1, \cdots, r\}$. This graph bandlimited property with respect to the proposed GFT operator then enables the selection of sampling vertex subset whose size equals that of the graph cut-off bandwidth set. We will show this via Fig. 4.8.

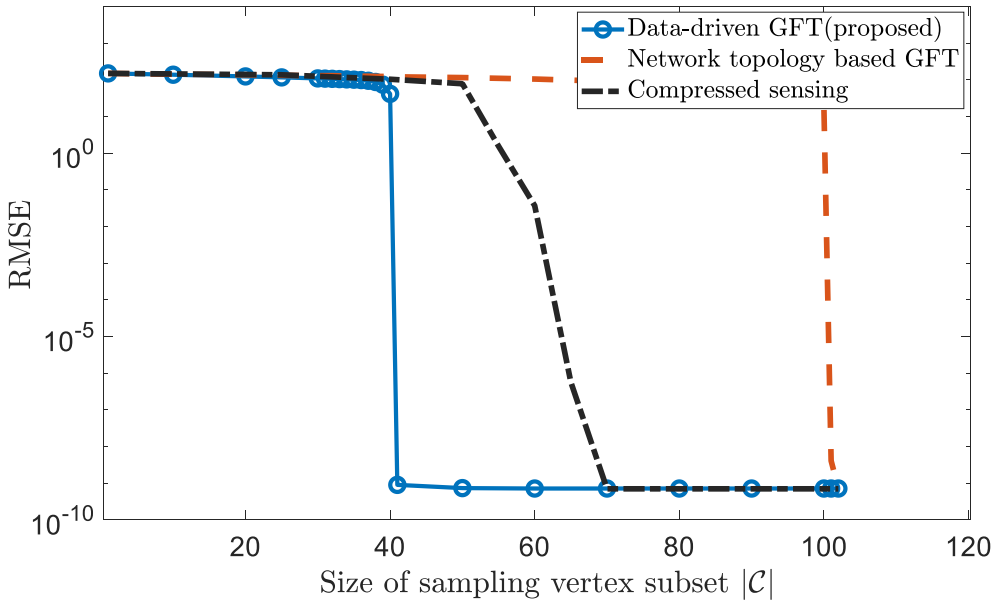

Figure 4.8: Comparison of recovery RMSE among different schemes.

Fig. 4.8 illustrates recovery RMSE comparisons between the proposed data-driven GFT sampling scheme, the topological Laplacian operator based GFT sampling method, and the PCA based CS scheme. Here, the x-axis accounts for the size of the sampling vertex subset $|\mathcal{C}|$. Firstly, it is straightforward that with the growth of $|\mathcal{C}|$, the recovery RMSEs from all schemes decrease, as a larger size of sampling vertex subset $|\mathcal{C}|$ can embrace more samples for more accurate signal recovery. Secondly, it is seen that, the recovery RMSE of the proposed data-driven GFT method becomes smaller till $|\mathcal{C}|$ reaches the size of the graph cut-off bandwidth set, i.e., $|\mathcal{C}| = |\mathcal{B}_{\text{cut-off}}| = 41$, and then goes to a convergence nearly 10^{-8} after $|\mathcal{C}| > |\mathcal{B}_{\text{cut-off}}| = 41$. This outperforms the other two schemes that require more sensor vertices to reach the perfect signal recovery (e.g., $|\mathcal{C}| = 100$ for topological Laplacian based GFT, and $|\mathcal{C}| = 70$ for PCA based CS). The reason lies in two aspects. First, the proposed data-driven GFT operator is capable of transforming the data $\mathbf{X}$ into a more compact graph bandlimited set (see Fig. 4.7), whose size is smaller than those from PCA CS and Laplacian GFT. Then, different from the CS that only knows the number of non-zero elements transformed by the operator (i.e., the sparsity), the GFT sampling framework knows the exact elements of the graph cut-off bandwidth set (i.e., the indices of the non-zero graph frequency response), and thereby can map it to the sampling vertex subset $\mathcal{C}$ with size equaling that of the graph cut-off bandwidth set i.e., $|\mathcal{C}| = |\mathcal{B}_{\text{cut-off}}|$.

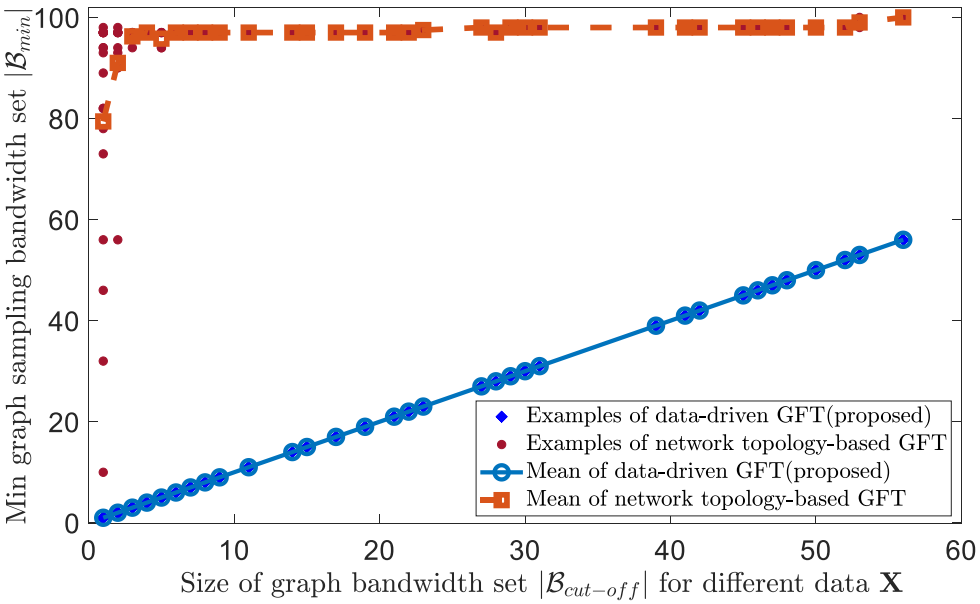

Figure 4.9: Comparison of the minimum graph sampling bandwidth set for the proposed data-driven GFT and the network topology based GFT schemes.

Then, we demonstrate the robustness of our proposed data-driven GFT sampling method. To do so, we measure the minimum graph sampling bandwidth set, denoted as $|\mathcal{B}_{\text{min}}|$, and the minimum sampling vertex subset, denoted as $|\mathcal{C}_{\text{min}}|$ to ensure a low recovery error as RMSE$< 10^{-8}$. Here, different WDN topology and time-varying network signals are considered, and their related signal spaces and graph cut-off bandwidth sets are also different. In Figs. 4.9-4.10, the x-axis denotes the size of the graph cut-off bandwidth set $|\mathcal{B}_{\text{cut-off}}|$ for different data matrices, and the y-axis represents the size of minimum graph sampling bandwidth set $|\mathcal{B}_{\text{min}}|$ and the size of minimum sampling vertex subset $|\mathcal{C}_{\text{min}}|$ respectively. It is firstly seen that, with the increase of $|\mathcal{B}_{\text{cut-off}}|$ for different data matrices, $|\mathcal{B}_{\text{min}}|$ and $|\mathcal{C}_{\text{min}}|$ of all schemes increase, as more information from the graph frequency domain and the vertices are required for a more complicated dynamic scenario with larger graph cut-off bandwidth set. Then, we notice that $|\mathcal{B}_{\text{min}}|$ and $|\mathcal{C}_{\text{min}}|$ from the proposed data-driven GFT sampling method stays at the minimum value (i.e., $|\mathcal{B}_{\text{min}}| = |\mathcal{B}_{\text{cut-off}}|$, and $|\mathcal{C}_{\text{min}}| = |\mathcal{B}_{\text{cut-off}}|$), as opposed to those of the topological Laplacian based GFT and the PCA based CS. This suggests the robustness of our proposed data-driven GFT method when addressing different time-varying network signals in WDN. The advantage of our proposed method is also attributed to the ability of the data-driven GFT operator to transform the signals into a more compact graph bandlimited area, when compared with the other two state-of-the-art methods. As such, the GFT sam-

pling framework that knows exactly the graph cut-off bandwidth set $\mathcal{B}_{\text{cut-off}}$, is able to derive the sampling vertex subset whose size equals that of $\mathcal{B}_{\text{cut-off}}$.

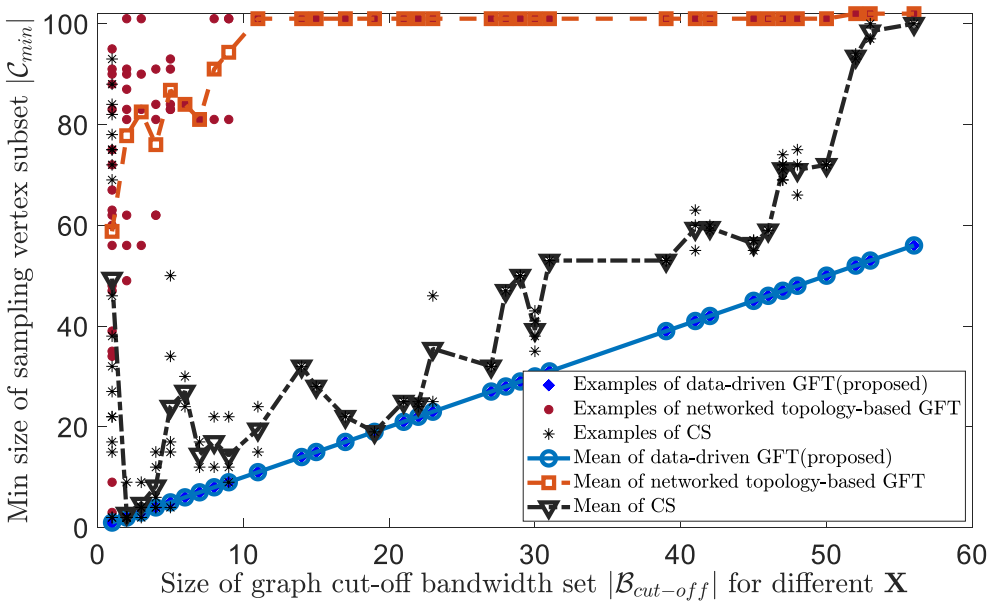

Figure 4.10: Comparison of the minimum size of sampling vertex subset among different schemes.

Finally, we count the average number of the sensor vertices from Fig. 4.10. Given that the minimum size of sampling vertex subset equals the size of graph cut-off bandwidth set, i.e., $|\mathcal{C}_{\text{min}}| = |\mathcal{B}_{\text{cut-off}}|$, the average number of sensor vertices is the average of $|\mathcal{B}_{\text{cut-off}}|$. When the total number of vertices in WDN is 102, this average $|\mathcal{B}_{\text{cut-off}}|$ is nearly $\overline{|\mathcal{B}_{\text{cut-off}}|} = 30$. As such, an average of 30% network vertices are required for the complete recovery of the network contaminant signals in WDN.

Although one may question that sensor deployment on 30% of the WDN might still be expensive (as pipes and junctions are usually buried underground and therefore expensive to be penetrated), we resort to our proposed data-driven GFT sampling method as the baseline for the future sampling optimization techniques (e.g., neural network methods to reduce the sensors by finding and exploiting the nonlinear vertex dependency).

4.6.3 Performance of Molecular Relay Data-Driven GFT Sampling Method

In this part, we evaluate the performance of our proposed MRDD GFT sampling scheme. The key performance indicators contain (i) the connectivity between sensor

vertices to hub vertex for molecular communication, and (ii) the recovery performance measured by RMSE between the recovered signals $\hat{\mathbf{X}}$ and the original signals $\mathbf{X}$. For the simulation of molecular transmission, we simulate the erasure channel over the WDN by computing the arrival probability matrix **APM** from Eqs. (4.15)-(4.17). Then, for each sensor vertex, a total of 10^4 molecules is transmitted for every report to the hub.

Molecular Communication Connectivity

We first examine whether the selection of sampling vertex subset $\mathcal{C}$ can link to a hub vertex, since such a connectivity from all sensor vertices to the hub vertex underpins the successful transmission of the samples for signal recovery. We measure such connectivity by the ratio of sensor vertices that can link to a hub among all selected sensor vertices, denoted as $|\mathcal{C}_{\text{hub}}|/|\mathcal{C}|$. For the experiment in Fig. 4.11, we use the contaminant signal matrix with a graph cut-off bandwidth set as $|\mathcal{B}_{\text{cut-off}}| = 41$.

It is seen from Fig. 4.11 that the ratio $|\mathcal{C}_{\text{hub}}|/|\mathcal{C}|$ of our proposed MRDD GFT sampling method keeps at 1 when the size of sampling vertex subset $|\mathcal{C}|$ increases from 0 to 50. This indicates the existed connectivity from the selected sensor vertices to a hub vertex, for selected sampling vertices satisfying $rank(\mathbf{Q}_{\mathcal{CB}}) = |\mathcal{B}_{\text{cut-off}}|$ that guarantees the complete signal recovery. Compared to the previously proposed data-driven GFT scheme in Fig. 4.11 (proposed in Section 4.2), the difference is that the newly proposed MRDD GFT is able to ensure the connectivity for samples transmitted to the hub vertex by the molecules via the pipes. This is attributed to the deeper combination of the topological information (i.e., the path existence matrix $\mathbf{A}_0^L$ in Eq. (4.18)) with the data-driven GFT operator in Eq. (4.19), and the connectivity constraint used for selecting the sensor vertices in Eq. (4.21). Such molecular connectivity thus underlies the signal recovery at the hub vertex from the transmitted samples, which will be evaluated in the following part.

Recovery Performance of Molecular Relay Data-Driven GFT Sampling Method

The recovery performance is measured by RMSE of the recovered data. We here examine the cases with and without the erasure molecular channel. The results are illustrated via Figs. 4.12-4.13, where Fig. 4.12 gives the arrival probability matrix **APM** in Eq. (4.15), and Fig. 4.13 illustrates the recovery RMSE.

In Fig. 4.13, we can observe that, in the absence of the erasure model (only pure connectivity is considered), the RMSE of the proposed MRDD GFT method

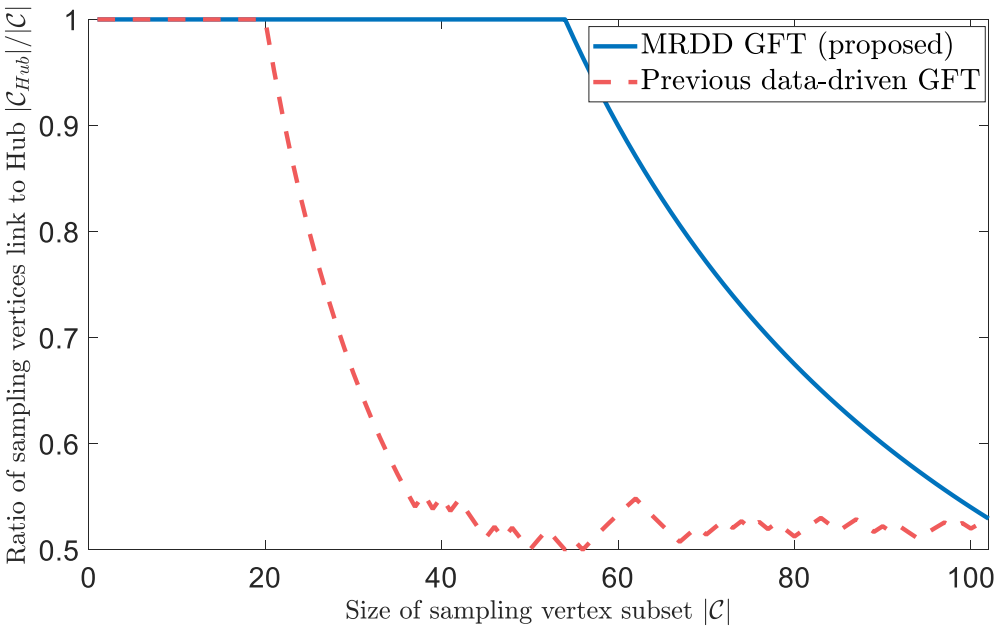

Figure 4.11: Comparison of whether vertices in sampling vertex subset can have a hub between the proposed MRDD GFT and previously proposed data-driven GFT.

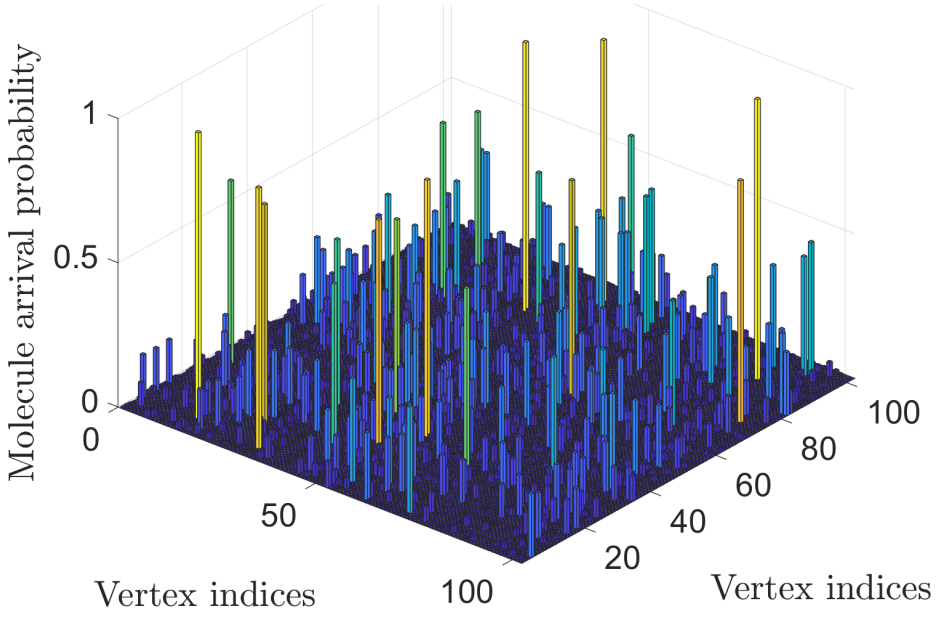

Figure 4.12: Arrival probability matrix **APM** for any vertex pair

converges faster than that of the previously proposed data-driven GFT sampling method in Section 4.2. The MRDD GFT requires only $|\mathcal{C}| = |\mathcal{B}_{\text{cut-off}}| = 41$ sensor

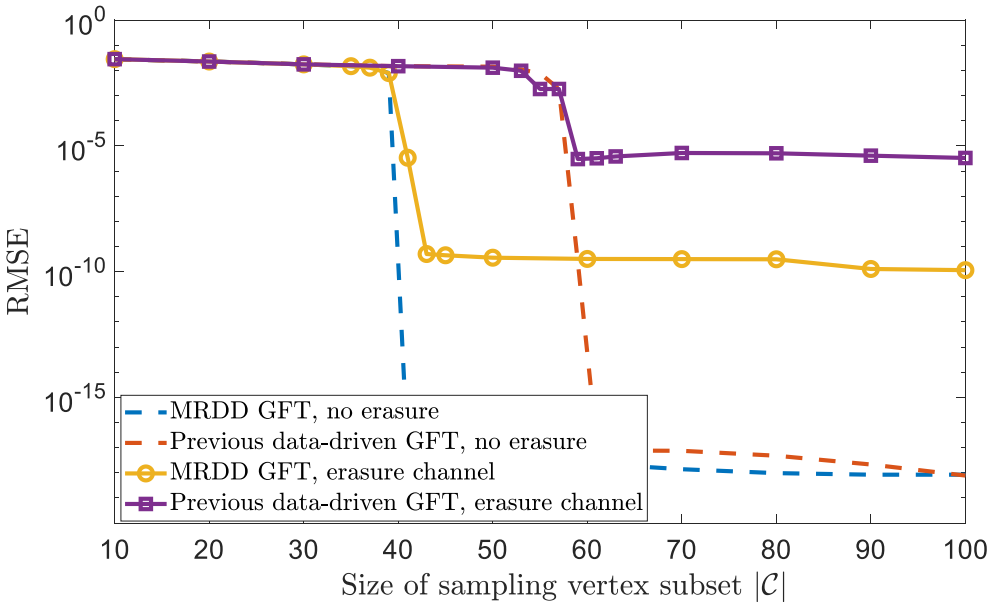

Figure 4.13: Comparison of recovery RMSE versus the size of sampling vertex subset among different schemes in WDN molecular erasure channels.

vertices to achieve perfect signal recovery, much smaller than that of the previously proposed data-driven GFT method. This is because the MRDD GFT is able to guarantee connectivity from all the sensor vertices in $|\mathcal{C}|$ to the hub vertex, which serves as the prerequisites for receiving all the samples at the hub vertex for complete signal recovery.

Then, when the molecular erasure is considered, it is seen that the MRDD GFT sampling scheme can still use a smaller sized sampling vertex subset $|\mathcal{C}|$ for complete signal recovery, as opposed to the previously proposed data-driven GFT scheme. This is attributed to (i) the molecular connectivity guaranteed by our proposed MRDD GFT, and (ii) the utilization of the last arrival reports from the sensor vertices whose current reports are missing.

4.7 Conclusions & Discussions

In this chapter, the data-driven GFT sampling framework has been proposed for the scenarios where the explicit dynamic governing equations are unknown. Here, we choose the WDN and its contaminant propagation as the time-varying network signals for monitoring. Using the construction of the contaminant signal-space as a prior knowledge from the experimental data, the data-driven GFT operator is derived, and is thereby able to characterize the dependency (i.e., the graph ban-

dlimitedness) of the time-varying network signals on different vertices. With the help of the data-driven GFT operator, we then identify the time-invariant sampling vertex subset by maintaining the one-to-one mapping between the samples and the whole graph bandlimited frequency response.

Furthermore, we address the practical challenge on how the sensors transmit their samples to the data centre (hub vertex), for WDNs that are buried in depth the underground and are hard to be penetrated. We have designed a molecular relay data-driven GFT sampling scheme whereby the reports are encoded by the biological structure of the DNA molecules, and transmitted via the pipes of the WDNs, avoiding the penetrating techniques and extra communication networks. For the sampling selection, we have re-determined the optimal sensor locations to ensure the connectivity between all the selected vertices and the central hub vertex.

Compared with the state-of-the-art CS approaches, our data-driven GFT methods show the more compact sampling vertex subsets, i.e., an average 30-40% of the sensor vertices can ensure the complete recovery of the time-varying network signals. The framework is useful and beyond the application of WDNs and can be applied to a wide-range of infrastructure sensing (e.g. railways [78]) where the dynamic governing equations are unavailable.

One limitation lies in the signal-dependency assumption of the time-varying network signal-space. Actually, such a vertex dependency serves as the prerequisite, if one wish to compress (sample) and recover the network signals in an immediate manner, i.e., without using the further time-evolved information (e.g., using $\mathbf{x}_2$ to infer the initial $\mathbf{x}_1$). Then, in the following two chapters, we will consider the sampling using time-evolution information when the signal-dependency among vertices does not exist, but one may note that such an immediate signal recovery cannot be achieved as further time-evolved samples are required to recover the initial network signal.

Chapter 5

Sequential Data-Driven GFT for Network Sensor Activation

In previous chapters, either the exact governing dynamic equations or the prior knowledge of the signal-space is used to analyze and exploit the signal dependency among vertices, for network sampling and signal recovery. Such information, if unavailable, will block their usages, thereby making them less attractive for some of complicated real-world monitoring scenarios. In this chapter, we will introduce our proposed principal component analysis based sequential graph Fourier transform sampling (PCA GFT) method, which does not rely on the governing dynamic equations and the prior signal-space.

Note that, this work is only a preliminary research to achieve the concept of an equation and signal-space independent network sampling. One drawback is its derivation of the time-varying sampling vertices selection strategy. In this view, we categorize this work as sensor activation, and will compare the sampling and recovery performance with other state-of-the-art time-varying sampling schemes. A mature work of equation and signal-space independent scheme is provided in Chapter 6.

The structure of this chapter is shown as follows. In Section 5.1, we introduce the model and the purpose formulations. In Section 5.2, we elaborate our proposed PCA GFT sampling method. In Section 5.3, we compare the proposed PCA GFT sampling method with other two state-of-the-art methods from the conceptual prospective. The simulations and experimental results are provided in Section 5.4. We finally conclude this chapter in Section 5.5.

5.1 Model and Problem Formulation

Similar with Chapter 4, we consider the contaminant monitoring application in water-distribution network. The topology of the WDN is configured by the static directed graph $\mathcal{G}(\mathcal{N}, \mathbf{A})$, with total N vertices indexed by $\mathcal{N} = \{1, 2, \cdots, N\}$, and the binary adjacent matrix $\mathbf{A}$. The time-varying network signals of the interest is the contaminant propagation over the WDN. We characterize it by a discrete-time signal matrix of size $N \times K$, i.e., $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K]$, where the $n \in \mathcal{N}$th row represents the discrete-time signal on vertex n, and $\mathbf{x}_k, k \in \mathcal{K} = \{1, 2, \cdots, K\}$ represents the network signal at discrete-time k. We plot an illustration of the WDN and its time-varying network signals via Fig. 5.1.

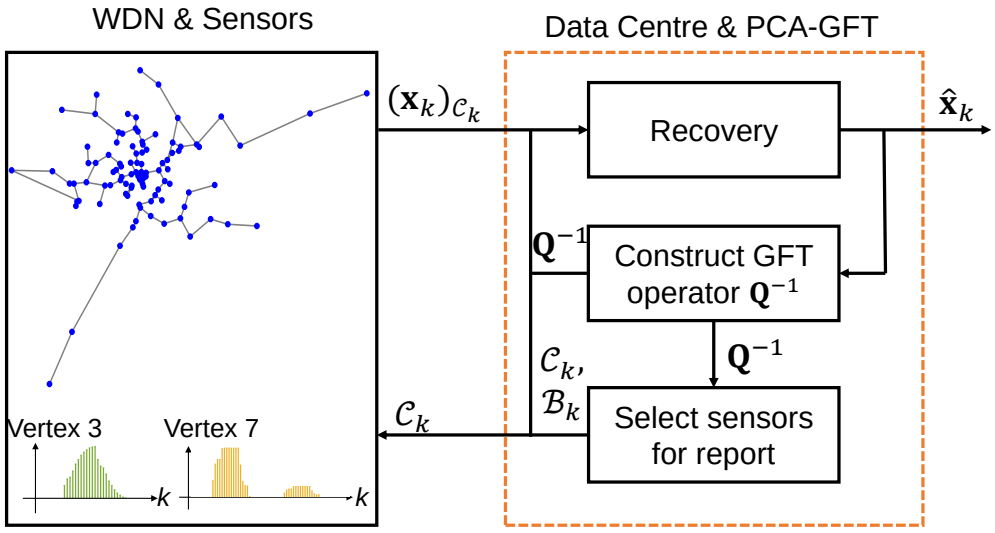

Figure 5.1: Illustration of the WDN, and the schematic flow of the proposed PCA GFT sampling method.

Different from Chapter 4, we here consider the sensor activation strategy for the recovery of the time-varying network signals. To be specific, for each vertex $n \in \mathcal{N}$, we deploy a sensor to (i) sample the signal if such sensor is activated, and (ii) exchange the information with the data centre. The aim of the data centre is to (i) recover the time-varying network signals from the reported samples, and (ii) inform the sensors whether they should be activated for sampling and reporting. The illustration of the WDN sensor activation system is provided in Fig. 5.1.

As such, we denote $\mathcal{C}_k \subset \mathcal{N}$ as the sensor activation set for discrete-time k. The aim of this chapter is to determine $\mathcal{C}_k$, so that the samples from $\mathcal{C}_k$ reported to data centre can ensure the complete recovery of the current network signals $\mathbf{x}_k$. Here, we emphasize in the first place that, different from Chapter 4, we do not

assume and rely on a prior knowledge of the signal-space.

5.2 Principal Component Analysis based GFT Sampling

In this section, we elaborate our proposed PCA GFT sampling scheme, which is able to sample via activated sensors and recover the time-varying network signals $\mathbf{x}_k$ of WDN, without the prior knowledge of the signal-space. This contains two essential steps (as is illustrated in Fig. 5.1). First, we construct a sequential data-driven GFT operator for every discrete-time k, which is able to characterize the current network signal $\mathbf{x}_k$ as graph bandlimitedness. Second, we adopt the static graph sampling theory in Definitions 1-2 and Theorem 2 in Chapter 2, for activation sensor selection and signal recovery.

5.2.1 Sequential PCA GFT Operator

Given the concept of static graph sampling theory in Theorem 2, a sequential GFT operator $\mathbf{Q}^{-1}$ should ensure two aspects. First, $\mathbf{x}_k$ should be graph $\mathcal{B}_k$-bandlimited with respect to the GFT operator $\mathbf{Q}^{-1}$ for some compact $\mathcal{B}_k \subsetneq \{1, 2, \cdots, N\}$. Second, such a $\mathcal{B}_k$ (whose elements are the subscripts of the nonzero elements in the graph frequency response $\tilde{\mathbf{x}}_k = \mathbf{Q}^{-1}\mathbf{x}_k$) should be predictable using the previous information and recovered results. To do so, we resort to the PCA techniques, given their ability to transform the time-varying network signals $\mathbf{x}_k$ into sparse representations [24].

Construction of PCA GFT Operator

For each discrete-time k, we construct the PCA GFT operator $\mathbf{Q}^{-1}$ using the previously recovered signals. Denote the recovered signal at discrete-time k as $\hat{\mathbf{x}}_k$. The mean and co-variance of the previous recoveries from time $k-\tau$ to $k-1$ are computed as:

$$\bar{\mathbf{x}} = \frac{1}{\tau}\sum_{l=1}^{\tau}\hat{\mathbf{x}}_{k-l}, \tag{5.1}$$

$$\mathbf{Cov} = \frac{1}{\tau}\sum_{l=1}^{\tau}(\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}}) \cdot (\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}})^T, \tag{5.2}$$

where τ is the lag accounting for the correlations, i.e., $\mathbf{x}_k = \xi(\mathbf{x}_{k-1}, \cdots, \mathbf{x}_{k-\tau})$. Such τ can be selected based on the data structure and the dynamic mechanism of different network systems, and we provide the effect of τ for the WDN systems in Section 5.4. With the help of Eqs. (5.1)-(5.2), we use SVD to derive the PCA GFT

operator for discrete-time k, i.e.,

$$[\mathbf{\Gamma}, \mathbf{\Sigma}, \mathbf{V}] = svd(\mathbf{Cov}), \tag{5.3}$$

$$\mathbf{Q}^{-1} = \mathbf{\Gamma}^{-1}. \tag{5.4}$$

Graph Bandlimited Analysis

After the design of the sequential PCA GFT operator $\mathbf{Q}^{-1}$ in Eqs. (5.1)-(5.4), we here demonstrate that $\mathbf{Q}^{-1}$ satisfies the aforementioned two properties (i.e., the graph $\mathcal{B}$-bandlimited property of $\mathbf{x}_k$ for some $\mathcal{B} \subsetneq \{1, 2, \cdots, N\}$, and such a $\mathcal{B}$ can be predicted).

Given that the time-varying network signals are spatially and temporally correlated, we express the current k discrete-time signal $\mathbf{x}_k$ using its previous values, i.e.,

$$\mathbf{x}_k = \sum_{l=1}^{\tau} \alpha_l \cdot \hat{\mathbf{x}}_{k-l} + \boldsymbol{\varpi}_k, \tag{5.5}$$

where α_l represents the corresponding coefficients, and $\boldsymbol{\varpi}_k$ accounts for the residual and high-order components. By subtracting the mean $\bar{\mathbf{x}}$ on both sides, we re-write Eq. (5.5) as follows:

$$\mathbf{x}_k - \bar{\mathbf{x}} = \sum_{l=1}^{\tau} \alpha_l \cdot (\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}}) + \left(\sum_{l=1}^{\tau} \alpha_l - 1 \right) \bar{\mathbf{x}} + \boldsymbol{\varpi}_k. \tag{5.6}$$

Note from Eq. (5.2) that $\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}}$ can be described by linearly combining columns in matrix $\mathbf{Cov}$, we hereby characterize $\mathbf{x}_k - \bar{\mathbf{x}}$ in the following:

$$\mathbf{x}_k - \bar{\mathbf{x}} = \mathbf{Cov} \cdot \boldsymbol{\beta} + \rho \cdot \bar{\mathbf{x}} + \boldsymbol{\varpi}_k, \tag{5.7}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_N]^T$ is composed by the coefficients, and $\rho = \sum_{l=1}^{\tau} \alpha_l - 1$. Then, the graph frequency response of $\mathbf{x}_k - \bar{\mathbf{x}}$ with respect to the GFT operator $\mathbf{Q}^{-1}$ is computed as:

$$\begin{aligned} \tilde{\mathbf{x}}_k - \tilde{\bar{\mathbf{x}}} &= \mathbf{Q}^{-1} \cdot (\mathbf{x}_k - \bar{\mathbf{x}}) \\ &= \mathbf{\Gamma}^{-1} \cdot \mathbf{Cov} \cdot \boldsymbol{\beta} + \rho \cdot \mathbf{\Gamma}^{-1} \cdot \bar{\mathbf{x}} + \mathbf{\Gamma}^{-1} \cdot \boldsymbol{\varpi}_k \\ &\overset{(a)}{=} \underbrace{\begin{bmatrix} diag\left([\sigma_1, \cdots, \sigma_r]^T \right) \cdot \mathbf{V} \cdot \boldsymbol{\beta} \\ \mathbf{0}_{(N-r)\times r} \cdot \mathbf{V} \cdot \boldsymbol{\beta} \end{bmatrix}}_{\text{predictable indices of nonzeros}} + \rho \cdot \mathbf{\Gamma}^{-1} \cdot \bar{\mathbf{x}} + \mathbf{\Gamma}^{-1} \cdot \boldsymbol{\varpi}_k. \end{aligned} \tag{5.8}$$

In Eq. (5.8), we can see that the subscripts of the non-zeros in $\rho \cdot \mathbf{\Gamma}^{-1} \cdot \bar{\mathbf{x}}$ can be determined directly via the computation. Then, for the vector $\mathbf{\Gamma}^{-1} \cdot \mathbf{Cov} \cdot \boldsymbol{\beta}$, only the first ($r \leq \tau$) rows are non-zeros. Here, we have $r \leq \tau$, i.e., (a), in Eq. (5.8) since:

$$\begin{aligned} r = rank(\mathbf{Cov}) &= rank\left(\sum_{l=1}^{\tau}(\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}}) \cdot (\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}})^T\right) \\ &\leq \sum_{l=1}^{\tau} rank\left((\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}}) \cdot (\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}})^T\right) \\ &= \tau. \end{aligned} \tag{5.9}$$

These two indicate that, for discrete-time k, there exists a graph bandwidth set $\mathcal{B}_k \subsetneq \{1, 2, \cdots, N\}$ such that $\mathbf{x}_k$ is $\mathcal{B}_k$-bandlimited with respect to the GFT operator $\mathbf{Q}^{-1}$. Further, we denote $\mathcal{B}_{k|k-1}$ as the predicted version of $\mathcal{B}_k$ from previously recovered signals. As according to the "predictable indices of nonzeros" component in Eq. (5.8), $\mathcal{B}_{k|k-1}$ can be derived via $\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}}$, and the indices of nonzero elements in vector $\mathbf{Q}^{-1} \cdot \bar{\mathbf{x}}$, i.e.,

$$\mathcal{B}_{k|k-1} = \left\{ n | \bigcup_{l=1}^{\tau} \left(\mathbf{Q}^{-1} \cdot (\hat{\mathbf{x}}_{k-l} - \bar{\mathbf{x}})\right)_n \neq 0 \bigcup \left(\mathbf{Q}^{-1} \cdot \bar{\mathbf{x}}\right)_n \neq 0 \right\}, \tag{5.10}$$

where $(\cdot)_n$ represents the nth element of vector. According to Theorem 2, we can select the sensor for activation, denoted as $\mathcal{C}_{k|k-1}$ that maps to the predicted graph bandwidth set $\mathcal{B}_{k|k-1}$, and the samples from $\mathcal{C}_{k|k-1}$ can be used to recover the component of $\mathbf{x}_k - \bar{\mathbf{x}}$, i.e., $\mathbf{Cov}\boldsymbol{\beta} + \rho\bar{\mathbf{x}}$ in Eq. (5.7). For the residual and high-order parts, i.e., $\boldsymbol{\varpi}_k$ in Eq. (5.7), an extra set of sensors, denoted as $\mathcal{C}^{\dagger}$ will be activated for signal sampling. In the following, we will elaborate the processes of identifying $\mathcal{C}_{k|k-1}$ and $\mathcal{C}^{\dagger}$.

5.2.2 Selection of Sensor Activation Set

As from Section 5.2.1, we derive the sequential PCA GFT operator $\mathbf{Q}^{-1}$, and the predicted graph bandwidth set $\mathcal{B}_{k|k-1}$. Accordingly, we will design the sensor activation set $\mathcal{C}$ from which the samples will be used for the recovery of the time-varying network signals $\mathbf{x}_k$. Given Eq. (5.8), the sensor activation set $\mathcal{C}_k$ should embrace two subsets, i.e.,

$$\mathcal{C}_k = \mathcal{C}_{k|k-1} \bigcup \mathcal{C}^{\dagger}, \tag{5.11}$$

where $\mathcal{C}_{k|k-1}$ is the predicted sensor activation set that is mapped from the predicted graph bandwidth set $\mathcal{B}_{k|k-1}$, and $\mathcal{C}^{\dagger}$ represents the extra sensor activation set that

aims to collect samples for the residual high-order components $\boldsymbol{\varpi}_k$ in Eq. (5.5).

The derivation of $\mathcal{C}_{k|k-1}$ is detailed in Theorem 2, by satisfying the following equation:

$$rank(\mathbf{Q}_{\mathcal{C}_{k|k-1}\mathcal{B}_{k|k-1}}) = |\mathcal{B}_{k|k-1}|, \tag{5.12}$$

which is conducted by finding the rows corresponding to the first $|\mathcal{B}_{k|k-1}|$ smallest singulars of $\mathbf{Q}_{\mathcal{N}\mathcal{B}_{k|k-1}}$, i.e.,

$$\mathcal{C}_{k|k-1} = \underset{\mathcal{C}_{k|k-1}\subset\mathcal{N}}{\operatorname{argmax}}\, \sigma_{\min}\left(\mathbf{Q}_{\mathcal{C}_{k|k-1}\mathcal{B}_{k|k-1}}\right), \tag{5.13}$$

with $\sigma_{\min}(\cdot)$ the minimum singular of the matrix. To implement Eq. (5.13), we adopt a greedy algorithm by finding and adding the row with minimum singular as follows: $\mathcal{C}_{k|k-1} \leftarrow \mathcal{C}_{k|k-1} \cup \{n\}$, such that $n = \operatorname{argmax}_{i\in\mathcal{N}\setminus\mathcal{C}_{k|k-1}} \sigma_{\min}(\mathbf{Q}_{(\mathcal{C}_{k|k-1}\cup\{i\})\mathcal{B}_{k|k-1}})$.

Then, for the extra sensor activation set $\mathcal{C}^{\dagger}$, the aim is to collect samples for residual $\boldsymbol{\varpi}_k$, whose non-zero elements are equivalent to the burst of contaminant on corresponding vertices (i.e., the vertex m with $(\mathbf{x}_{k-1})_m = 0$, but $(\mathbf{x}_k)_m \neq 0$). To predict such burst vertices, we rely on the topological structure of the WDN, i.e., $G(\mathcal{N}, \mathbf{A})$. We estimate a rough outcome by multiplying the adjacent matrix $\mathbf{A}$ with the previously recovered signals, i.e.,

$$\mathcal{C}^{\dagger} = \left\{ m \middle| \ (\mathbf{A}\cdot\hat{\mathbf{x}}_{k-1})_m \neq 0 \bigcap (\hat{\mathbf{x}}_{k-1})_m = 0, m \in \mathcal{N} \right\}. \tag{5.14}$$

5.2.3 Signal Recovery using PCA GFT

After the derivation of the sequential PCA GFT operator $\mathbf{Q}^{-1}$ and the senor activation set $\mathcal{C}_k$ in Eqs. (5.11)-(5.14), the sensors with indices belonging to $\mathcal{C}_k$ can report their samples to the data centre for the recovery of the network signal $\mathbf{x}_k$. We here denote the samples (vector) from $\mathcal{C}_k$ as $(\mathbf{x}_k)_{\mathcal{C}_k}$. The recovery process of the data centre is described in the following. We divide the process into two parts. The first part is to recover the signal using graph sampling Theorem 2, i.e.,

$$\begin{aligned}\hat{\mathbf{x}}_k =& \mathbf{Q}_{\mathcal{N}\mathcal{B}_{k|k-1}} \cdot \left(\mathbf{Q}^T_{\mathcal{C}_{k|k-1}\mathcal{B}_{k|k-1}} \cdot \mathbf{Q}_{\mathcal{C}_{k|k-1}\mathcal{B}_{k|k-1}}\right)^{-1} \\ & \cdot \mathbf{Q}^T_{\mathcal{C}_{k|k-1}\mathcal{B}_{k|k-1}} \cdot \left((\mathbf{x}_k)_{\mathcal{C}_{k|k-1}} - \bar{\mathbf{x}}_{\mathcal{C}_{k|k-1}}\right) + \bar{\mathbf{x}}.\end{aligned} \tag{5.15}$$

The second part is to replace the corresponding recovered signals with vertices in $\mathcal{C}^{\dagger}$, i.e.,

$$(\hat{\mathbf{x}}_k)_{\mathcal{C}^{\dagger}} = (\mathbf{x}_k)_{\mathcal{C}^{\dagger}}. \tag{5.16}$$

5.3 Distinguish with Two Data-Driven Schemes

In this section, we compare our proposed PCA GFT sensor activation method with other two state-of-the-art data-driven sampling schemes, in conceptual manner. The first one is provided in Chapter 4, i.e., the signal-space dependent data-driven GFT sampling scheme. The second one is the widely adopted PCA based compressed sensing scheme in [24].

5.3.1 Data-Driven Static Graph Sampling

In Chapter 4, a data-driven GFT sampling algorithm has been proposed, leveraging the prior knowledge of the dynamic signal-space, denoted as $\mathbb{S} = span\{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \cdots, \bar{\mathbf{x}}_r\}$. We assumed (i) the real time-varying network signals $\mathbf{x}_k$ (i.e., the columns of data matrix $\mathbf{X}$) belongs to the signal-space, i.e., $\mathbf{x}_k \in \mathbb{S}$, and (ii) such signal-space $\mathbb{S}$ is graph $\mathcal{B} = \{1, 2, \cdots, r\}, r < N$-bandlimited. In such a manner, the data-driven GFT operator has been derived from the SVD of the r-leading vectors of $\mathbb{S}$, and therefore is able to maintain the graph $\mathcal{B}$-bandlimited property of the data matrix $\mathbf{X}$. As such, the time-invariant sampling vertex subset $\mathcal{C}$ is derived by mapping the samples from $\mathcal{C}$ to the complete graph $\mathcal{B}$-bandlimited frequency response.

Compared with our previously proposed signal-space dependent data-driven GFT sampling method, our newly proposed PCA GFT sampling does not rely on the prior knowledge of the signal-space, i.e., $\mathbb{S}$, which if not reliable, may lead to difficulties for the construction of the GFT operator in Chapter 4, and subsequently result in poor monitoring performance. Then, it is highlighted that the proposed PCA GFT sampling method is unable to determine a time-invariant sampling vertex subset for sensor placement. For computational complexity aspect, the proposed PCA GFT sampling method has to compute the sequential GFT operator and sensor activation set at every discrete-time, which thereby requires more energy expenditure for computations and data storage.

5.3.2 PCA CS Sampling Method

PCA CS sampling method is proposed by the work in [24], and is to sample and recover the time-varying network signals via smaller number of samples (measurements). Here, similar to our proposed PCA GFT sampling method, PCA CS is to derive time-varying sampling vertices (i.e., sensor activation) for network signal recovery at different discrete-time. We briefly introduce the steps of the PCA CS as follows. First, they construct the sparse transformation matrix $\mathbf{D}$ by the

eigenvectors of the co-variance matrix **Cov**, i.e., [24]

$$\mathbf{x}_k - \bar{\mathbf{x}} = \mathbf{D} \cdot \mathbf{s}_k, \text{ with } \mathbf{Cov} = \mathbf{D} \cdot \mathbf{\Lambda} \cdot \mathbf{D}^{-1}, \tag{5.17}$$

where $\mathbf{s}_k$ denotes the sparse representation vector. The mean $\bar{\mathbf{x}}$ and the co-variance matrix **Cov** are provided by Eqs. (5.1)-(5.2). Second, they select the sensor activation set $\mathcal{C}_k \subset \mathcal{N}$ to satisfy the RIP, i.e., [24]

$$1 - \delta \leq \frac{\|\mathbf{D}_{\mathcal{C}_k \mathcal{N}} \cdot \mathbf{s}\|_{l_2}^2}{\|\mathbf{s}\|_{l_2}^2} \leq 1 + \delta, \ r = \|\mathbf{s}_k\|_{l_0}, \tag{5.18}$$

for any vector $\mathbf{s}$ with only $2r$ nonzero elements, and some $\delta \in [0, 1]$, where $\|\cdot\|_{l_2}$ represents the l_2-norm, and $\|\cdot\|_{l_0}$ denotes the l_0-norm. Third, given the samples at discrete-time k, they recover $\mathbf{s}_k$ via the standard convex optimization (introduced in Chapter 2.3.1), or the orthogonal matching pursuit. Finally, we have $\hat{\mathbf{x}}_k = \mathbf{D} \cdot \hat{\mathbf{s}}_k + \bar{\mathbf{x}}$.

The main difference from the PCA CS sampling scheme in [24] is that we are able to exploit the positions of the non-zero elements in transformations for the selection of sensor activation set. In PCA CS, the positions of non-zeros are the indices of the sparse elements of $\mathbf{s}_k$, which are unknown. In our proposed PCA GFT, such positions of non-zeros are the indices of non-zero elements in the graph-frequency response $\tilde{\mathbf{x}}_k = \mathbf{Q}^{-1} \cdot \mathbf{x}_k$, and we proved in Eq. (5.8) that such indices are predictable from the previously recovered signals. With this prediction, our PCA GFT sampling method is able to reduce the number of selected sensors to nearly r (validated by the graph sampling Theorem 2). This is greatly smaller than that of the PCA CS scheme in [24], which requires an order of $O(r \log(N/r))$ activated sensors. We further demonstrate this by the simulations in Section 5.4.

5.4 Simulations & Experimental Results

In this section, we evaluate the performance of our proposed PCA GFT sampling method. The key performance indicators include the RMSE of the recovered time-varying network signals, and the average size of the selected sensor activation set, denoted as $|\mathcal{C}|_{\text{avg}}$. We specify the two indicators in the following, i.e.,

$$\text{RMSE} = \frac{\sum_{k=1}^{K} \|\hat{\mathbf{x}}_k - \mathbf{x}_k\|_{l_2}}{\sqrt{N \cdot K}}, \tag{5.19}$$

$$|\mathcal{C}|_{\text{avg}} = \frac{\sum_{k=1}^{K} |\mathcal{C}_k|}{K}, \tag{5.20}$$

where $\|\cdot\|_{l_2}$ denotes the l_2-norm, and $|\mathcal{C}_k|$ denotes the number of elements in set $\mathcal{C}_k$.

For this experiment, the simulation platform is the Microsoft Azure, and the state-of-the-art EPANET2 [67] is adopted to simulate the WDN and the contaminant spreads. For the topology of the WDN network, $N = 102$ vertices are assigned including 100 junctions and 2 reservoirs (see Fig. 5.1). Each junction has a random and unknown water-demand. The directed edges between two vertices are pipes with unknown pressures. We simulate 100 different contaminant data matrices over the WDN. For each matrix $\mathbf{X}$, total $K = 3240$ discrete-times are simulated within 3 hours.

5.4.1 One Illustration of Signal Recovery

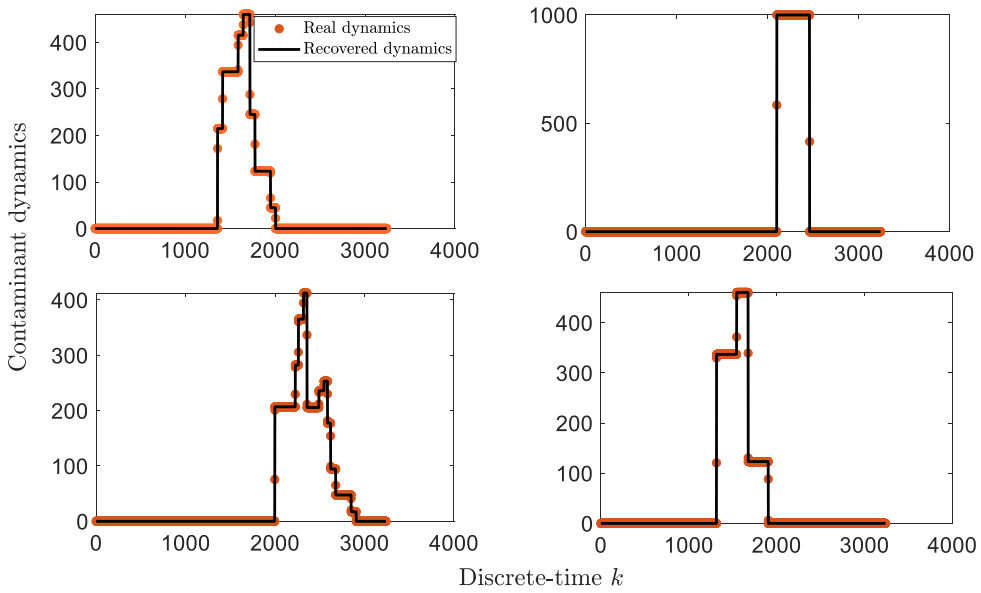

Figure 5.2: Illustration of 4 examples of real and recovered signals of the proposed PCA GFT method.

An illustration of recovered signals from our proposed PCA GFT sampling method is provided in Fig. 5.2, where the comparisons between real and recovered signals on 4 un-sampled vertices are shown. In Fig. 5.2, a perfect signal recovery is realized by an average size of sensor activation set as $|\mathcal{C}|_{\text{avg}} = 46 < N = 102$. Note that such size is bigger than the size of sampling vertex subset provided by Chapter 4 (nearly 30 from $N = 102$), however, we do not require a prior knowledge of the signal-space for our PCA GFT sensor activation scheme.

5.4.2 Performance Comparisons

We next compare the performances between our proposed PCA GFT sampling method with the PCA CS scheme in [24]. Fig. 5.3 illustrates the recovery RMSE versus the average size of sensor activation set $|\mathcal{C}|_{\text{avg}}$ defined in Eq. (5.20). From Fig. 5.3, we can observe that, as the average size of sensor activation set $|\mathcal{C}|_{\text{avg}}$ increases, the recovery RMSEs from all schemes become smaller. For example, the RMSEs of the PCA CS and proposed PCA GFT approaches decrease from 1 to an order of 10^{-14}, when the average size of sensor activation set increases from 30 to 80. This is because that an increasing $|\mathcal{C}|_{\text{avg}}$ leads to more samples and therefore can ensure a more accurate signal recovery.

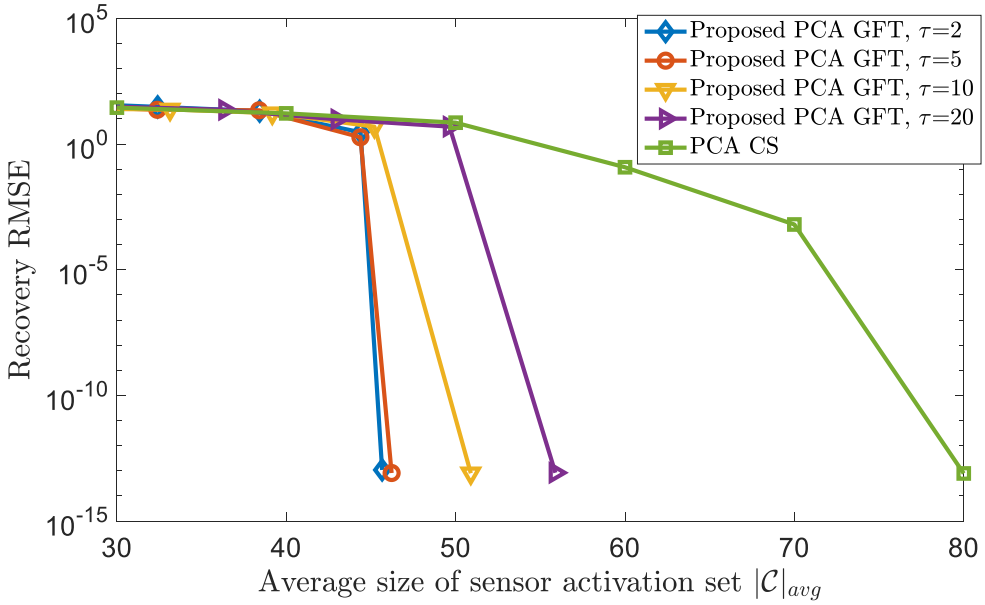

Figure 5.3: Comparison of recovery accuracy between proposed PCA GFT and PCA CS.

Then, we observe from Fig. 5.3 that the lag τ in Eq. (5.5) affects the recovery accuracy of our proposed PCA GFT sampling scheme, and a smaller τ leads to a more accurate recovery performance. For instance, when $\tau = 2$, the proposed PCA GFT approach can reach a RMSE as 10^{-13} faster at $|\mathcal{C}|_{\text{avg}} = 46$, which is smaller that that of $\tau = 10$ (requiring $|\mathcal{C}|_{\text{avg}}$ as 56). We explain this in the following. In Eq. (5.5), τ accounts for how the current k discrete-time dynamic $\mathbf{x}_k$ can be characterized by its previous signals, i.e., $\mathbf{x}_k = \xi(\mathbf{x}_{k-1}, \cdots, \mathbf{x}_{k-\tau})$. As such, due to the fact that current hydraulic contaminant signal is directly evolved from its last signal, i.e., $\mathbf{x}_{k-1} \to \mathbf{x}_k$, we should adopt a smaller τ as $\tau = 2$ in order to hold such

time-evolution. Besides, such a lag τ determines the size of the predicted sensor activation set $\mathcal{C}_{k|k-1}$, i.e., $|\mathcal{C}_{k|k-1}| = \mathcal{B}_{k|k-1} = \tau$ provided by Eq. (5.12). In such a view, $\tau = 2$ gives a minimum $|\mathcal{C}_{k|k-1}|$ and subsequently a minimum $|\mathcal{C}|_{\text{avg}}$ as shown in Fig. 5.3.

Finally, it is seen from Fig. 5.3 that, the proposed PCA GFT sampling method requires less samples for the perfect recovery of the time-varying network signals, as opposed to PCA CS scheme in [24]. The former requires approximately the average size of the sensor activation set as $|\mathcal{C}|_{\text{avg}} = 46$ for the perfect signal recovery, which is much smaller than that from the PCA CS scheme (an overall $|\mathcal{C}|_{\text{avg}} = 80$ needed).

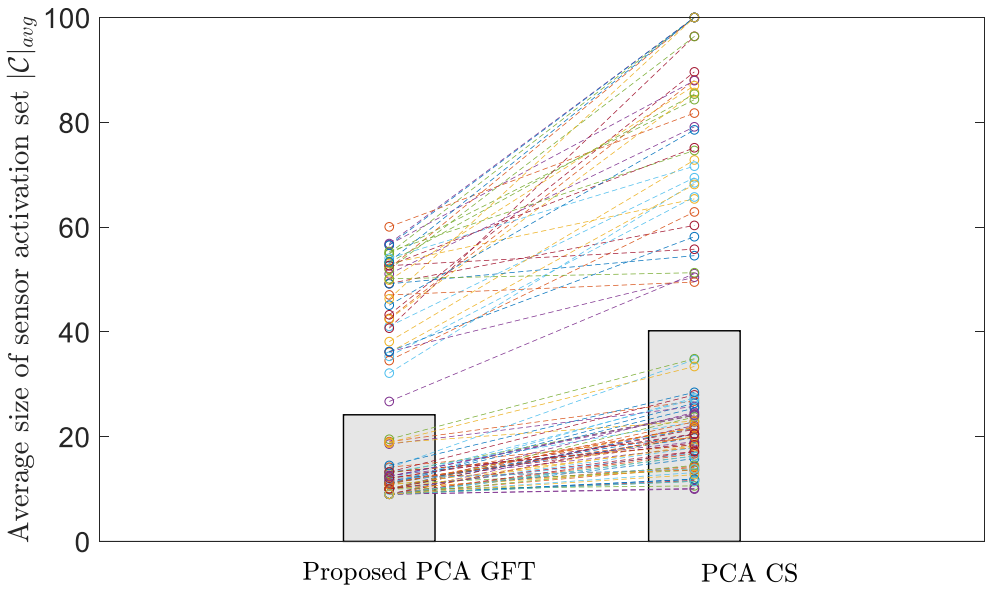

Figure 5.4: Comparison of average size of sensor activation set $|\mathcal{C}|_{\text{avg}}$.

We further demonstrate the sampling reduction advantage of our proposed scheme via Fig. 5.4, where for each of the 100 different data matrices, we record the corresponding minimum average sizes of the sensor activation sets $|\mathcal{C}|_{\text{avg}}$ for perfect recovery of the time-varying network signals. In Fig. 5.4, each pair of dashed line connected points indicates the results of the two competitive schemes processing the same data matrix. It is seen that for each data matrix, the proposed PCA GFT is able to use a smaller size of sensor activation set as opposed to that of the PCA CS method, which demonstrates the robustness of our proposed method. This further indicates a smaller average $|\mathcal{C}|_{\text{avg}}$ of the proposed PCA GFT scheme, which stays lower than that of the PCA CS. The advantage of the sensor activation reduc-

tion derived from our proposed PCA GFT method is attributed to the awareness (predictability) of the indices of the r nonzero elements (that constitute the graph bandwidth set) of the transformation $\tilde{\mathbf{x}}_k = \mathbf{Q}^{-1} \cdot \mathbf{x}_k$, ($r = \|\tilde{\mathbf{x}}_k\|_{l_0}$). As such, the GFT sampling framework is able to construct the sensor activation set $\mathcal{C}_k$ by mapping the samples from it directly to the graph bandlimited frequency response, and therefore derives a smaller $\mathcal{C}_k$ with $|\mathcal{C}_k| = r$, as opposed to the PCA CS that requires an order of $O(r \log(N/r))$ samples to satisfy the RIP criterion.

5.5 Conclusions & Discussion

To address the network sampling when signal dependency among vertices is unavailable (non-existed or unable to be characterized by an unchanged operator), we have proposed a sequential data-driven GFT sampling approach, which is able to achieve a time-varying sampling vertex subset to recover the whole time-varying network signals (for sensor activation applications where sensors are deployed on all network vertices). By analysing the principal components of the previously recovered signals, a sequential data-driven PCA GFT operator has been derived, being able to identify the vertices with independent signals at each discrete time. The experimental simulations demonstrate an average of 40% sensors are required to be activated to ensure the complete recovery of the time-varying network signals. Also, the performance guarantee in this chapter enables us to further reduce the number of activated sensors for a desirable but loosing recovery accuracy.

As a preliminary exploration for sampling without the vertex dependency, the drawbacks of this chapter are obvious. These include the computational complexity for updating the PCA GFT operator at each discrete time, and moreover, the proposed scheme cannot achieve the time-invariant sampling locations for sensor placement.

Chapter 6

Model and Signal-Space Independent Sampling

In this chapter, we will elaborate the proposed model and signal-space independent sampling scheme for network sampling and signal recovery task. The motivation lies in the scenarios where the explicit dynamic model and the prior information of the signal-space (e.g., the sparsity and bandlimitedness to a given operator) is unavailable, which dis-enables both the equation-driven and the signal-space driven methods provided in the previous chapters.

The structure of the rest of this chapter is as follows. We first use a general time-evolution model to characterize the task in Section 6.1. Then, the designed logarithm observable and Koopman operator for the derivation of linear time evolution model is provided in Section 6.2. In Section 6.3, we further elaborate the nonlinear Graph Fourier Transform (nonlinear GFT) concept and sampling theory, and how it can be implemented for network sampling and signal recovery. Section 6.4 is for the distinguishment and explanation of the proposed scheme with other state-of-the-arts. The simulation and experimental results are illustrated in Section 6.5. We finally conclude the whole chapter in Section 6.6.

6.1 Model and Problem Formulation

6.1.1 General Dynamic Model

In the context of dynamic model and signal-space independent sampling and recovery scenarios, both the explicit dynamic governing equations and the prior knowledge of signal dependencies among vertices (i.e., sparsity or graph bandlimitedness of the signal space) are not required. Instead, we employ a general form to describe the

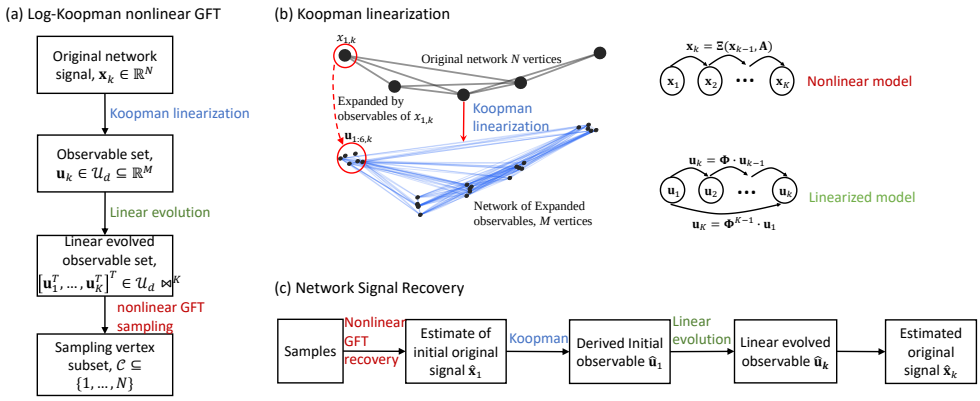

Figure 6.1: Schematic flow of the proposed Log-Koopman nonlinear GFT sampling and recovery method: (a) sampling process; (b) Koopman linearization; (c) signal recovery.

time-varying network signals, as follows:

$$\mathbf{x}_{k+1} = \boldsymbol{\Xi}(\mathbf{x}_k, \mathbf{A}). \tag{6.1}$$

In Eq. (6.1), $\mathbf{x}_k$ is the network signal at discrete time k. The dynamical signal flow is characterized by the underlying graph adjacent matrix $\mathbf{A}$, and a general time evolution function $\boldsymbol{\Xi} : \mathbb{R}^N \to \mathbb{R}^N$. Here, $\mathbf{A}$ of size $N \times N$ is defined as a binary or weighted adjacent matrix on a network (static graph) with N vertices whose indices constitute the index set $\mathcal{N} = \{1, 2, \cdots, N\}$. The (m, n)th element $a_{m,n} \in \{0, 1\}$ of $\mathbf{A}$ represents whether a link from vertex n to vertex m exists.

The functionality of $\boldsymbol{\Xi}(\cdot)$ is to generate the signals $\mathbf{x}_{k+1} = [x_{1,k+1}, \cdots, x_{N,k+1}]^T$ at discrete time $k+1$ ($k \in \{1, 2, \cdots, K\}, K \in \mathbb{N}^+$) from $\mathbf{x}_k$, in accordance with the self-dynamic and the coupling vertex interactions from the adjacent vertices given $\mathbf{A}$. At discrete time $k = 1$, $\mathbf{x}_1 \in \mathbb{R}^N$ is regarded as the initial network signal. It is noteworthy that for the further analysis, $\boldsymbol{\Xi}(\cdot)$ and $\mathbf{x}_1$ are both unknown, falling into the category of model and signal-space independence.

6.1.2 Challenge Formulation

Given the general dynamic time-evolution model in Eq. (6.1), the scope of this chapter is to determine an optimal time-invariant sampling vertex subset for sensor placement, denoted as $\mathcal{C} \subset \mathcal{N}$, such that the whole network signals $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K$ can be reconstructed. Here, two challenges are inevitable for the sampling vertex subset selection and the signal recovery design. First, the explicit dynamic time-evolution model is unavailable, rendering the existing works that require linear/linearized evo-

lution model (e.g., graph observability analysis) [42–46] less implementable. Second, the dependency of signals among different vertices is unknown. That is, we no longer know the specific graph bandlimitedness or the sparsity of the network signals to a given operator, which thereby blocks the compression approaches using vertex dependency [30, 34, 35, 38, 41, 46, 79–83]. To overcome these, this chapter aims to 1) approximate a linear dynamic time-evolution model, and 2) identify time-invariant vertices for sampling and recovering the time-varying network signals.

6.1.3 Sketch of Design

The schematic flow and illustration of the proposed sampling and recovery scheme is shown in Fig. 6.1. First, the Koopman linearization theory is adopted to derive an approximately linearized time-evolution model for time-varying network signals. Then, a novel concept and sampling theory of nonlinear GFT is provided as a guideline, for the algorithm designs of the sampling vertex subset selection and the signal recovery. This is also how we organize the rest of this chapter.

6.2 Koopman Operator and Linearization

Koopman linearization theory is used to generate a linearized time-evolution model to approximate a nonlinear dynamic system, so that the rich linear algebra theories can be adopted for further signal analysis (e.g., the stability, and the leading and un-scaling systematic factors). A Koopman operator specific to a dynamical system is a linear matrix that characterizes the time evolution of the observables (functions) defined on the original signal space. To be specific, for one system, we define the space of all suitable observables as $\mathcal{H} = span(h_1(\cdot), h_2(\cdot), \cdots, h_M(\cdot))$, spanned by the $M \in \mathbb{N}^+$ leading observables $h_m(\cdot) : \mathbb{R}^N \to \mathbb{R}$. As such, a stacked vector-valued observable selected from $\mathcal{H}$ is $\mathbf{h}(\cdot) = [h_1(\cdot), h_2(\cdot), \cdots, h_M(\cdot)]^T$, and its corresponding Koopman operator $\mathbf{\Psi}$ can be expressed as [47, 50, 62]:

$$\mathbf{\Psi} \cdot \mathbf{h}(\mathbf{x}_k) = \mathbf{h}\left(\mathbf{\Xi}(\mathbf{x}_k)\right) = \mathbf{h}(\mathbf{x}_{k+1}). \tag{6.2}$$

As is shown in Eq. (6.2), we linearize the nonlinear time-evolution $\mathbf{x}_{k+1} = \mathbf{\Xi}(\mathbf{x}_k)$ by making the vector-valued observable defined on $\mathbf{x}_k$ a linear evolution, i.e., $\mathbf{h}(\mathbf{x}_{k+1}) = \mathbf{\Psi} \cdot \mathbf{h}(\mathbf{x}_k)$. In this view, one thing required is to design appropriate observable elements to make an existence of the linear evolution matrix $\mathbf{\Psi}$. This is difficult and is still remaining as an open challenge, given the infinite dimension of $\mathcal{H}$, i.e., $M \to +\infty$, rendering an infinite size of the Koopman operator $\mathbf{\Psi}$ and therefore

impeding its practical use in real-world applications. To cope with this, loads of existing works [47, 50, 62] tried to approximate the infinite observable space, via the usages of definite and independent (orthogonal) observables. In the rest of this section, we will first introduce existing observable designs for Koopman operator, with the explanations of why those are limited in the context of network sampling scenarios. Then, the proposed logarithm observable is provided with the analysis of its merits and drawbacks.

6.2.1 Existing Observable Designs for Koopman Operator

Existing observable designs are classified into 3 groups: the dynamic mode decomposition (DMD), the extended DMD (EDMD), and the deep DMD. In the context of network sampling and signal recovery, two challenges for Koopman linearization are faced.

- First, the size of one vector-valued observable, i.e., the number of elements M in $\mathbf{h}(\cdot) = [h_1(\cdot), h_2(\cdot), \cdots, h_M(\cdot)]^T$ cannot be very large; otherwise, a size explosion will occur and result in further computational burden for a $M \times M$ Koopman operator $\mathbf{\Psi}$.
- Second, the observable element $h_m(\mathbf{x}_k)$ should be decoupled with the signals on different vertices; otherwise, selecting some of the coupled observable elements for signal recovery may result in redundant selections of sampling vertices.

Dynamic Mode Decomposition

DMD directly uses the original N time-varying network signals as the $M = N$ observable elements, i.e., $\mathbf{h}(\mathbf{x}_k) = \mathbf{x}_k$ [62]. This is suitable for a linear and quasi-linear dynamical system, however will result in intolerant linearization error for nonlinear dynamic systems.

Extended Dynamic Mode Decomposition

Extended DMD uses the observable elements defined on the original N time-varying network signals [62]. Typical designs of EDMD observable elements include the Fourier extensions, the radial basis functions, the polynomials (e.g., the Legendre polynomials, the Hermite polynomials, and the Taylor polynomials). For linearization accuracy, by selecting appropriate observable elements, EDMD performs better as opposed to the DMD only. However, for the network sampling applications, the

drawbacks lie in two aspects. First, the size explosion of the vector-valued observable $\mathbf{h}(\mathbf{x}_k)$ will cause heavy computational burden (e.g., the inverse computation of the Koopman operator requires an order of M^3 multiplications, which will require substantial resources when using current EDMD methods as $M > O(N^2)$ is needed. Second, the coupling effect between original N signals on different vertices will result in redundant selections of sampling points. We will go through these in the following.

One state-of-the-art EDMD observable design is the Taylor expansion based polynomial observables. The work in [47] has proved a completeness of an observable space $\mathcal{H}$ leveraging the polynomial terms of Taylor expansion. Then, a truncated subspace $\mathcal{H}_{\text{approx}} \subset \mathcal{H}$ can be constructed by spanning from the leading low-ordered and independent Taylor series. Here, we provide the completeness observable space as follows [47]:

$$\mathcal{H} = \left\{ \prod_{n=1}^{N} x_{n,k}^{p_n},\ \forall p_n \in \mathbb{N} \right\}. \tag{6.3}$$

According to Eq. (6.3), the work in [47] used a selection of $\mathbf{h}(\mathbf{x}_k) = [x_{m,k}^{p_m} \cdot x_{n,k}^{p_n}]^T$ with $\forall m, n \in \mathcal{N},\ p_m, p_n \in \{0, 1, 2\}$ to provide the truncated observable space $\mathcal{H}_{\text{approx}}$, based on which an approximated Koopman operator has been proved to have good lineaization performance for small-scale (i.e., number of vertices $N < 10$) time-evolved network signals.

However, when it comes to address the large-scale networks ($N > 50$), the scheme either leads to a size explosion of the vector-valued observable by selecting the multi-elemental multiplicative polynomial Taylor terms (e.g., $x_{m,k}^{p_m} \cdot x_{n,k}^{p_n} \cdot x_{i,k}^{p_i} \cdot x_{j,k}^{p_j}$), or results in low linearization accuracy given the incompleteness of $\mathcal{H}_{\text{approx}}$ caused by their selections of only two-elemental multiplicative terms. We explain this by the following equation, showing why the multi-elemental multiplicative polynomial Taylor terms at discrete time k, are important for the constitution of the existing observable elements at the next $k+1$ discrete time, i.e.,

$$\begin{aligned} x_{m,k+1} \cdot x_{n,k+1} =& \xi_m(\mathbf{x}_k) \cdot \xi_n(\mathbf{x}_k) \\ =& \sum_{\substack{i,j \in \mathcal{N} \\ p_m, p_n, p_i, p_j \in \{0,1,2\}}} \alpha_{i,j,m,n} \cdot x_{m,k}^{p_m} \cdot x_{n,k}^{p_n} \cdot x_{i,k}^{p_i} \cdot x_{j,k}^{p_j}, \end{aligned} \tag{6.4}$$

where $\xi_m(\cdot) : \mathbb{R}^N \to \mathbb{R}$ is the mth time-evolution function in $\boldsymbol{\Xi}$, and $\alpha_{i,j,m,n}$ represents the coefficient. As provided in Eq. (6.4), the impact of 4-element multiplicative terms becomes non-trivial, as their numbers increase with the larger network scale N. In this view, in order to maintain the linearization accuracy of the Koopman

operator, the selection of observables in $\mathbf{h}(\mathbf{x}_k)$ should be expanded accordingly to cover those high-impact multi-elemental multiplicative terms. This will inevitably result in a size explosion of the vector-valued observable, and subsequently lead to a size of $> N^2$ increase for the approximated Koopman operator $\mathbf{\Psi}$. Such a large size $\mathbf{\Psi}$, if implemented in large-scale network (e.g., $N > 50$) sampling applications, will lead to a heavy computational burden. Apart from that, multi-elemental multiplicative terms also cause the coupling effect in observable elements, and subsequently lead to the redundant selections of sampling vertex subset. For instance, a selection of observable $x_{m,k} \cdot x_{n,k}$ for signal recovery will result in a selection on both vertex m and vertex n.

Deep Dynamic Mode Decomposition

With the aims to find lower-sized but more accurate Koopman linearization operator, deep DMD was proposed using the deep learning neural network. Leveraging this concept, one extraordinary work in [48] developed an auto-encoder and an auto-decoder, of which the former is to generate observable and the latter is to inversely map the observable to the original time-varying network signal. Their training process is via the minimization of the errors of forward and backward observable -original signal mapping, and of observable computed by original signal and Koopman time-evolved observable.

The scheme performs very well on operator size-reduction and linearization accuracy. However, one miserable drawback is the existence of coupled signals on observable elements. For example, in their work [48], one learned observable for discrete spectrum dynamic is $x_{2,k} - b \cdot x_{1,k}^2$, which involves the original signals on both vertex 1 and vertex 2. This therefore becomes less attractive for sensor placement applications, since selecting the leading observable elements may require to place sensors on every vertices.

6.2.2 Proposed Logarithm based Observable Design

Given that most of the current observable designs suffer from either size explosion or coupling signal effects that are not suitable for network sampling, we hereby propose a novel observable design, which is able to transform the multi-elemental multiplicative terms (e.g., $x_{m,k}^{p_m} \cdot x_{n,k}^{p_n}$) into decoupled and lower sized summation terms. The essence is to use the logarithm summations to approximate the polynomial multiplicative Taylor series. For example, the multiplicative term of $x \cdot y$ can

be approximated via the following equation:

$$\begin{aligned}\log(1+x)+\log(1+y) &= \log\left((1+x)(1+y)\right)\\ &\approx x+y+xy,\end{aligned} \tag{6.5}$$

which holds, when $x, y \in (0-\delta, 0+\delta)$ given $\delta \to 0$. In this view, if assigned an appropriate parameter η to maintain $sup\{x_{m,k}/\eta, m \in \mathcal{N}, k \in \mathbb{N}^+\} < \delta$, the vector-valued observable based on the logarithm-form expressions can be designed as follows:

$$\mathbf{h}\left(\mathbf{x}_k\right) = \left[1, \frac{x_{m,k}}{\eta}, \log\left(1+\left(\frac{x_{m,k}}{\eta}\right)^{p_m}\right)\right]^T,\ \forall m \in \mathcal{N}, \tag{6.6}$$

where $p_m \subset \mathbb{N}^+$ can be selected based on specific dynamic systems. For the convenience of further elaborations, we use $\mathbf{u}_k$ to describe the designed vector-valued observable of size $M \times 1$ in Eq. (6.6) as:

$$\mathbf{u}_k = \mathbf{h}(\mathbf{x}_k),\ \mathbf{u}_k \in \mathcal{U}_d \subsetneq \mathbb{R}^M. \tag{6.7}$$

It is noteworthy that the range set of $\mathbf{u}_k$, i.e., $\mathcal{U}_d$ is a subset of $\mathbb{R}^M$, which if neglected, will result in redundant selections of sampling vertices. This is because we need more samples to build an one-to-one mapping between the sampling set to $\mathbb{R}^M$, other than to its subset $\mathcal{U}_d$.

After the designs of the logarithm based observable in Eq. (6.6), we demonstrate by the following two equations that the vector-valued observable at discrete time $k+1$, i.e., $\mathbf{u}_{k+1}$ maintains the linear time-evolution of the previous observable.

First, the observable element $x_{m,k+1}/\eta$ can be approximated in the following:

$$\begin{aligned}\frac{x_{m,k+1}}{\eta} &= \frac{\xi_m(\mathbf{x}_k)}{\eta}\\ &= \frac{1}{\eta}\left(\xi_m(\mathbf{0}) + \mathbf{x}_k^T \cdot \bigtriangledown_{\xi_m}(\mathbf{0}) + \frac{1}{2}\mathbf{x}_k^T \cdot \mathbf{H}_{\xi_m}(\mathbf{0}) \cdot \mathbf{x}_k + o^3\right)\\ &= \frac{\xi_m(\mathbf{0})}{\eta} + \sum_{\substack{i,j\in\mathcal{N}\\ p,q\in\mathbb{N}^+}} \alpha_{i,j,p,q} \cdot \left(\frac{x_{i,k}}{\eta}\right)^p \cdot \left(\frac{x_{j,k}}{\eta}\right)^q\\ &\approx \frac{\xi_m(\mathbf{0})}{\eta} + \sum_{i\in\mathcal{N}} \alpha_i \cdot \frac{x_{i,k}}{\eta} + \sum_{\substack{i\in\mathcal{N}\\ p\in\mathbb{N}^+}} \log\left(1+\left(\frac{x_{i,k}}{\eta}\right)^p\right),\end{aligned} \tag{6.8}$$

where $\xi_m(\cdot)$ is the mth element function of the vector-valued function $\mathbf{\Xi}(\cdot)$, $\bigtriangledown_{\xi_m}(\cdot)$ represents the gradient function of $\xi_m(\cdot)$, and $\mathbf{H}_{\xi_m}(\cdot)$ denotes the Hessian matrix.

$\alpha_{i,j,p,q}$ and α_i are coefficients that maintain constant with varied time.

Second, the observable element $\log(1+(x_{m,k+1}/\eta)^{p_m})$ at discrete time $k+1$ can be approximated as:

$$\begin{aligned}
&\log\left(1+\left(\frac{x_{m,k+1}}{\eta}\right)^{p_m}\right) \approx \left(\frac{x_{m,k+1}}{\eta}\right)^{p_m} = \left(\frac{\xi_m(\mathbf{x}_k)}{\eta}\right)^{p_m} \\
=&\frac{\xi_m(\mathbf{0})^{p_m}}{\eta^{p_m}} + \sum_{p_1,\cdots,p_N\in\mathbb{N}^+} \alpha_{p_1,\cdots,p_N}\cdot\prod_{m\in\mathcal{N}}\left(\frac{x_{m,k}}{\eta}\right)^{p_m} \\
\approx&\frac{\xi_i(\mathbf{0})^{p_m}}{\eta^{p_m}} + \sum_{i\in\mathcal{N},p\in\mathbb{N}^+} \alpha_{i,p}\cdot\log\left(1+\left(\frac{x_{i,k}}{\eta}\right)^{p}\right),
\end{aligned} \tag{6.9}$$

where $\alpha_{p_1,\cdots,p_N}$ and $\alpha_{i,p}$ are time-invariant coefficients.

Eqs. (6.8)-(6.9) indicate that all observable elements in Eq. (6.6) can have a linear time-evolution from other previous observables. As such, by using a linear operator, i.e., the approximated Koopman operator $\mathbf{\Psi}$ of size $M\times M$, we can re-write the linear time-evolution of the designed vector-valued observable as:

$$\mathbf{h}\left(\mathbf{x}_{k+1}\right) = \mathbf{\Psi}\cdot\mathbf{h}\left(\mathbf{x}_k\right). \tag{6.10}$$

In Eq. (6.10), the Koopman operator $\mathbf{\Psi}$ is trained by the simulating data of $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$. To be specific, we generate D (e.g., $D=10^4$) groups of simulating data, denoted as $\mathbf{x}_{1:K}^{(d)}$ with $d=1,\cdots,D$, and separate them as two sub-groups, i.e.,

$$\mathbf{Y} = \left[\mathbf{h}(\mathbf{x}_{2:K}^{(1)}),\cdots,\mathbf{h}(\mathbf{x}_{2:K}^{(D)})\right] \tag{6.11}$$

$$\mathbf{X} = \left[\mathbf{h}(\mathbf{x}_{1:K-1}^{(1)}),\cdots,\mathbf{h}(\mathbf{x}_{1:K-1}^{(D)})\right], \tag{6.12}$$

based on which the Koopman operator can be trained by following equation:

$$\mathbf{\Psi} = \text{argmin}\,\|\mathbf{Y}-\mathbf{\Psi}\cdot\mathbf{X}\|_{fro}^2, \quad \|\cdot\|_{fro}:\ \text{Frobenius norm}. \tag{6.13}$$

After the design of the vector-valued observable and the derivation of corresponding Koopman operator, we analyze the advantages and the drawbacks of the proposed logarithm-based Koopman operator.

- First, one important merit when compared with the polynomial based Koopman operator [47] is its ability to reduce the observable size and subsequently the size of Koopman operator, which is attributed to the logarithm summation for the approximation of the multi-elemental multiplicative polynomial terms in observable designs. By using the logarithm based observable design, a size

of $M = O(N)$ observable $\mathbf{h}(\cdot)$ can be derived, as opposed to the polynomial observable that requires at least $M > O(N^2)$ polynomial terms. This is of great significance to reduce the further computational burden when using the corresponding Koopman operator of size $M \times M$ for network sampling and signal recovery.

- Second, the designed logarithm based observable is decoupled, whereby each observable element is determined by only one vertex's signal. This is different from the polynomial design in [47] and the deep DMD design in [48]. Such a property is important when dealing with sampling vertex subset selection, given that selecting one leading observable for signal recovery requires one sensor placing on only one vertex.

- When compared with the polynomial design [47] and the deep DMD designs, the drawback of the proposed logarithm based observable design lies in its less accurate linearization performance, due to the approximation of the multiplicative terms. However, in the context of network sampling scenarios, such linearization accuracy can be further enhanced by the use of samples. For example, we can generate a sample-awareness training data: the initialization ($k = 1$) values corresponding to sensor vertices equal their initial samples, and others remain random.

From this section, we are able to use the designed logarithm observable and the corresponding Koopman operator to linearize an unknown and nonlinear dynamic system. This therefore enables the further sampling and recovery analysis using rich linear algebra theory.

6.3 Sampling with Nonlinear Graph Fourier Transform

For this section, the aim is to introduce the proposed nonlinear Graph Fourier Transform concept and theory, and how those can be implemented for the network sampling and signal recovery. The purpose of this nonlinear GFT sampling is to determine the time-invariant selection of sampling vertex subset, from which the sampled time-evolved signals can help recover the whole time-varying network signals. Recalling that with the Koopman linearization operator for unknown and nonlinear dynamic modelling, we can convert the recovery task into the recovery of the initial observable $\hat{\mathbf{u}}_1$, from which the network signals at any discrete time k can be recovered by $\hat{\mathbf{x}}_k = \mathbf{h}^{-1}(\mathbf{\Psi}^{k-1} \cdot \hat{\mathbf{u}}_1)$. Such a recovery process is illustrated in Fig. 6.1(c).

Here, we wish to emphasize why the nonlinear GFT is required in the first place. As we mentioned in Eq. (6.7), the range set of $\mathbf{u}_1$ is a subset of $\mathbb{R}^M$, i.e., $\mathcal{U}_d \subsetneq \mathbb{R}^M$. This makes the existing graph observability analysis [46,50] that treat the range set as $\mathbb{R}^M$ less attractive. In their views, the M elements in $\mathbf{u}_1$ are independent with each other from $\mathbb{R}$, which contradict the fact that the M observable elements in $\mathbf{u}_1 = \mathbf{h}(\mathbf{x}_1)$ are determined by $N < M$ original elements in $\mathbf{x}_1$. Ignoring this will inevitably result in redundant sampling vertices selection, and this composes the motivation of the proposed nonlinear GFT concept and theory. Compared with those existing methods, the nonlinear GFT method takes into account the structure of the designed logarithm observable which characterizes the M observable elements in $\mathbf{h}(\mathbf{x}_1)$ by its N independent signals in $\mathbf{x}_1$, and aim to determine the sampling vertex subset whose time-evolved samples can map to the N independent signals $\mathbf{x}_1$.

6.3.1 Nonlinear GFT Concept

We extend the linear GFT concept to define what a generalized (nonlinear) GFT operator is and what the graph bandlimited property to the nonlinear GFT operator accounts for.

Definition 4 *The generalized (nonlinear) Graph Fourier Transform operator is a reversible (one-to-one) vector-valued function that transforms between sets $\mathcal{U}_r$ and $\mathcal{U}_d$. Here, we call $\mathcal{U}_r$ the set of graph-frequency response of $\mathcal{U}_d$, and even a bandlimited graph-frequency response if the element number of $\mathbf{x} \in \mathcal{U}_r$ is smaller than that of $\mathbf{u} \in \mathcal{U}_d$.*

Seen from Definition 4, a generalized GFT operator is provided, extended from the linear GFT operator concept [30, 34, 35, 38, 41, 46, 79–83], by replacing the linear matrix with the reversible vector-valued function. In this view, if the reversible vector-valued function is nonlinear, we call such operator the nonlinear GFT operator.

Recalling from the above derivation of Koopman operator $\mathbf{\Psi}$, we enlarged the original network $\mathcal{G}(\mathcal{N}, \mathbf{A})$ by the defined observable in Eq. (6.6), and therefore derived a new network with $M > N$ vertices connected via the Koopman operator $\mathbf{\Psi}$. We illustrate the enlarged process in Fig. 6.1, and denote the newly derived network topological structure as $\mathcal{G}(\mathcal{M}, \mathbf{\Psi})$ with $\mathcal{M} = \{1, 2, \cdots, M\}$. Also, the original time-varying network signals $\mathbf{x}_k$ over $\mathcal{G}(\mathcal{N}, \mathbf{A})$ are linearized by the Koopman process as $\mathbf{u}_k = \mathbf{h}(\mathbf{x}_k)$ over $\mathcal{G}(\mathcal{M}, \mathbf{\Psi})$. As such, we have the new network structure and the dynamical signals over it. Then, according to the Definition 4, a nonlinear

GFT operator can be assigned as the inverse of the vector-valued observable, i.e., $\mathbf{h}^{-1}(\cdot) : \mathcal{U}_d \to \mathcal{U}_r = \mathbb{R}^N$, which is able to transform the time-varying network signals $\mathbf{u}_k \in \mathcal{U}_d \subsetneq \mathbb{R}^M$ into a lower-sized (bandlimited) graph-frequency domain, i.e., $\mathbf{x}_k \in \mathbb{R}^N$. Here, it is noteworthy that such graph bandlimited to the nonlinear GFT operator holds for any time-varying network signals $\mathbf{u}_k \in \mathcal{U}_d$, and does not assume any signal dependency among original N vertices (i.e., the elements in $\mathbf{x}_k$ can be independently chosen from $\mathbb{R}$).

6.3.2 Sampling Theory of Nonlinear GFT

After the generalized GFT concept provided by Definition 4, we will elaborate in the following the nonlinear GFT sampling theory. We will go through (i) what the condition is for sampling vertex subset $\mathcal{C}$, and (ii) how to recover the whole time-varying network signals from the samples.

Theorem 5 *Denote a GFT operator $\mathbf{h}^{-1} : \mathcal{U}_d \to \mathcal{U}_r$, any $\mathbf{u} \in \mathcal{U}_d \subsetneq \mathbb{R}^M$, and a matrix $\mathbf{F}$ of size $L \times M$. Then, a sampling operator (matrix) $\mathbf{C}_F$ of size $S \times L$ that ensures the recovery of $\mathbf{u}$ from the sample $\mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{u}$ should keep the one-to-one mapping property of the operator $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$. The recovery of $\mathbf{u}$, denoted as $\hat{\mathbf{u}}$ is expressed as:*

$$\hat{\mathbf{u}} = \mathbf{h}\left(\left(\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}\right)^{-1}\left(\mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{u}\right)\right), \tag{6.14}$$

where $\circ$ denotes the function composition operator.

Proof 6 *We denote the graph-frequency response of $\mathbf{u}$ as $\mathbf{x} \in \mathcal{U}_r$. As such, the GFT and the inverse GFT transformations using the nonlinear GFT operator $\mathbf{h}^{-1}(\cdot)$ are:*

$$\mathbf{x} = \mathbf{h}^{-1}(\mathbf{u}), \tag{6.15}$$

$$\mathbf{u} = \mathbf{h}(\mathbf{x}), \tag{6.16}$$

given the reversible computation of the GFT operator. Then, by multiplying the sampling matrix $\mathbf{C}_F$ on both sides of $\mathbf{F} \cdot \mathbf{u} = \mathbf{F} \cdot \mathbf{h}(\mathbf{x})$, we have:

$$\mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{h}(\mathbf{x}) = \left(\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}\right)(\mathbf{x}) = \mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{u}. \tag{6.17}$$

From Eq. (6.17), it is seen that

$$\mathbf{x} = \left(\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}\right)^{-1}\left(\mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{u}\right) \tag{6.18}$$

holds if and only if $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$ has inverse operation. This is equivalent to the

one-to-one mapping property of $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$ *stated by the Theorem. Then, by taking Eq. (6.18) into Eq. (6.16), we derive the recovery of* $\mathbf{u}$ *as* $\hat{\mathbf{u}}$ *in Eq. (6.14) of the Theorem.*

After the description of Theorem 5, we here cast it for the designed Koopman observable for explanation and clarification. For the Koopman observable $\mathbf{u}_1$ with nonlinear dependency determined by the lower-sized original network signal, $\mathbf{u}_1 = \mathbf{h}(\mathbf{x}_1)$, Theorem 5 treats it as a graph bandlimited signal with nonlinear graph-frequency response $\mathbf{x}_1$ to the GFT operator $\mathbf{h}^{-1}(\cdot)$ (i.e., the inverse of the designed vector-valued observable). Then, Theorem 5 proves that the one-to-one mapping from the samples (determined by the sampling matrix) to such graph bandlimited response $\mathbf{x}_1$ can guarantee the recovery of $\mathbf{u}_1$.

Then, to determine the sampling matrix $\mathbf{C}_F$ that maintains such one-to-one mapping, we propose following Propositions.

Proposition 1 *Denote a GFT operator* $\mathbf{h}^{-1} : \mathcal{U}_d \to \mathcal{U}_r$ *where the dimension of* $\mathcal{U}_d$ *and* $\mathcal{U}_r$ *is* $dim\mathcal{U}$*. Then, one prerequisite for the recovery of any* $\mathbf{u} \in \mathcal{U}_d$ *is that the number of rows in the sampling matrix* $\mathbf{C}_F$ *is no smaller than* $dim\mathcal{U}$*.*

Proof 7 *Otherwise, let us assume the number of rows in the sampling matrix* $\mathbf{C}_F$ *is no greater than* $dim\mathcal{U} - 1$*. Given from Theorem 5,* $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$ *maintains one-to-one characteristic. This infers the set of the basic functions of* $\mathcal{U}_r$ *is spanned by at most* $dim\mathcal{U} - 1$ *element functions of* $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$*, which contradicts the dimension of* $\mathcal{U}_r$*, i.e.,* $dim\mathcal{U} \neq dim\mathcal{U} - 1$*.*

Proposition 2 *Denote a GFT operator* $\mathbf{h}^{-1} : \mathcal{U}_d \to \mathcal{U}_r$ *where the dimension of* $\mathcal{U}_d$ *and* $\mathcal{U}_r$ *is* $\dim\mathcal{U}$*. Then, one prerequisite for the recovery of any* $\mathbf{u} \in \mathcal{U}_d$ *is that, for sampling matrix* $\mathbf{C}_F$*, at least* $\dim\mathcal{U}$ *scalar-valued functions of* $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$ *are linearly independent.*

Proof 8 *Otherwise, let us assume any* $\dim\mathcal{U}$ *scalar-valued functions of* $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$ *are linearly dependent. Then, there must have* $< \dim\mathcal{U}$ *linearly independent scalar-valued functions spanning for the set of basic functions of* $\mathcal{U}_r$*. This suggests* $\dim\mathcal{U}_d = \dim\mathcal{U}_r < \dim\mathcal{U}$*, and thereby contradicts the dimension of* $\mathcal{U}_r$ *as* $\dim\mathcal{U}$*.*

As such, using Propositions 1-2, we are able to determine a sampling matrix $\mathbf{C}_F$ that maintains an approximated one-to-one characteristic. We next describe how this can be combined with the Koopman linearized time-evolution information for sampling vertex subset selection and signal recovery.

6.3.3 Nonlinear GFT Network Sampling and Recovery

Before the use of nonlinear GFT sampling theory, we first specify the Koopman linearized time-evolution information that will be combined with nonlinear GFT for the selection of sampling vertex subset $\mathcal{C}$ and the signal recovery, i.e.,

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_K \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi}^0 \\ \mathbf{\Psi}^1 \\ \vdots \\ \mathbf{\Psi}^{K-1} \end{bmatrix} \cdot \mathbf{u}_1, \tag{6.19}$$

where $k = 1, 2, \cdots, K$ are discrete times. As such, the linearly evolved information from time $k = 1, 2, \cdots, K$ can be effectively used for the recovery of $\mathbf{u}_1$ as $\hat{\mathbf{u}}_1$, and the original network signal $\hat{\mathbf{x}}_k$ is derived using the Koopman operator $\hat{\mathbf{u}}_k = \mathbf{\Psi}^{k-1} \cdot \hat{\mathbf{u}}_1$ and $\hat{\mathbf{x}}_k = \mathbf{h}^{-1}(\hat{\mathbf{u}}_k)$

Combining the linearly time-evolved information with the nonlinear GFT sampling theory, we assign:

$$\mathbf{F} = \begin{bmatrix} \mathbf{\Psi}^0 \\ \mathbf{\Psi}^1 \\ \vdots \\ \mathbf{\Psi}^{K-1} \end{bmatrix}. \tag{6.20}$$

The GFT operator is the aforementioned $\mathbf{h}^{-1}(\cdot)$, which maps the range set of the observable $\mathbf{u}_1 \in \mathcal{U}_d \subsetneq \mathbb{R}^M$ to $\mathbf{x}_1 \in \mathbb{R}^N$ (i.e., the original and initial network signal with $dim\mathcal{U} = N$ over original network $\mathcal{G}(\mathcal{N}, \mathbf{A})$). In such a manner, we convert the aim to determine the sampling vertex subset $\mathcal{C}$ for the complete recovery of the initial observable $\mathbf{u}_1 \in \mathcal{U}_d \subsetneq \mathcal{B}^M$ that is graph bandlimited to the GFT operator $\mathbf{h}^{-1}(\cdot)$ with graph-frequency response $\mathbf{x}_1 \in \mathbb{R}^N$.

Nonlinear GFT Sampling Vertex Subset Selection

We first provide the relations from the sampling vertex subset $\mathcal{C} \subset \mathcal{N}$ of the original network $\mathcal{G}(\mathcal{N}, \mathbf{A})$, to the sampling matrix $\mathbf{C}_F$ in Theorem 5. Such relation is:

$$\mathbf{C} = [1_{i,n_i \in \mathcal{C}}], \tag{6.21}$$

$$\mathcal{C}_h = \left\{ m \middle| h_m(\mathbf{x}_k) = h_m\left(\mathbf{C} \cdot \mathbf{x}_k\right) \right\}, \tag{6.22}$$

$$\mathbf{C}_h = \left[1_{j,m_j \in \mathcal{C}_h}\right], \tag{6.23}$$

$$\mathbf{C}_F = \mathbf{C}_h \otimes [\underbrace{1, \cdots, 1}_{K}]. \tag{6.24}$$

In Eq. (6.21), $1_{i,n_i \in \mathcal{C}}$ means only (i, n_i)th element of matrix $\mathbf{C}$ is 1 if $n_i \in \mathcal{C}$, and other elements are 0. In Eq. (6.23), $1_{j,m_j \in \mathcal{C}_h}$ means only (j, m_j)th element of matrix $\mathbf{C}_h$ is 1 if $m_j \in \mathcal{C}_h$, and other elements are 0. In Eq. (6.24), $\otimes$ represents the Kronecker product. To conveniently deduce further analysis and computation, we abbreviate the above equations by:

$$\mathbf{C}_F = \mathbf{\Upsilon}\left(\mathcal{C}\right). \tag{6.25}$$

Given the nonlinear GFT sampling Theorem 5, the optimal identification of the sampling vertex subset $\mathcal{C}$ should keep the one-to-one (reversible) mapping characteristic of the vector-valued function $\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h}$, which is a NP-hard challenge. To implement this, we use the Propositions 1-2 to approach a sub-optimal solution. In other words, we do so by identifying the $\dim \mathcal{U} = N$ linearly independent rows of $\mathbf{F}$, i.e.,

$$rank(\mathbf{\Upsilon}(\mathcal{C}) \cdot \mathbf{F}) = N. \tag{6.26}$$

Here, such nonlinear GFT sampling vertex subset selection is different from the previous full column-rank sampling criterion (in Chapter 2 Theorem 2). We will analyze this in Section 6.4. Then, to realize Eq. (6.26), we minimize the quotient of the 1st singular divided by the Nth singular of $\mathbf{C}_F \cdot \mathbf{F} = \mathbf{\Upsilon}(\mathcal{C}) \cdot \mathbf{F}$, i.e.,

$$\mathcal{C} = \underset{\mathcal{C} \subset \mathcal{N}}{\operatorname{argmin}} \left\{ \frac{\sigma_1\left(\mathbf{\Upsilon}\left(\mathcal{C}\right) \cdot \mathbf{F}\right)}{\sigma_N\left(\mathbf{\Upsilon}\left(\mathcal{C}\right) \cdot \mathbf{F}\right)} \right\}, \tag{6.27}$$

where $\sigma_i(\cdot)$ denotes the ith singular of the matrix.

Eq. (6.27) is implemented in a greedy algorithm framework in Algorithm 3. The inputs are the vertex set $\mathcal{N}$ of original network $\mathcal{G}(\mathcal{N}, \mathbf{A})$, and the linearly evolved matrix $\mathbf{F}$ in Eq. (6.20) that characterizes the linear mapping between initial observable $\mathbf{u}_1$ and further time-evolved observables $\mathbf{u}_{1:K}$. Step 1 is to initialize the sampling vertex subset as empty set. Steps 2-5 are to add vertex one by one with minimum quotient between 1st and Nth singulars, till a threshold γ or the required size of sampling vertex subset is approached. Finally, we output the selected sampling vertex subset $\mathcal{C}$, which will be used for sensing and collecting the samples for the recovery of time-varying network signals in the following part.

Algorithm 3 Nonlinear GFT Sampling Vertex Subset Selection

Input: $\mathcal{N}, \mathbf{F}$

1: Initialize $\mathcal{C} = \emptyset$.
2: **while** $\frac{\sigma_1(\boldsymbol{\Upsilon}(\mathcal{C})\cdot\mathbf{F})}{\sigma_N(\boldsymbol{\Upsilon}(\mathcal{C})\cdot\mathbf{F})} > \gamma$ **do**
3: $\quad n = \text{argmin}_{n\in\mathcal{N}\setminus\mathcal{C}} \left\{ \frac{\sigma_1(\boldsymbol{\Upsilon}(\mathcal{C}\cup\{n\})\cdot\mathbf{F})}{\sigma_N(\boldsymbol{\Upsilon}(\mathcal{C}\cup\{n\})\cdot\mathbf{F})} \right\}$.
4: $\quad \mathcal{C} = \mathcal{C} \cup \{n\}$.
5: **end while**

Output: Return $\mathcal{C}$.

Nonlinear GFT Signal Recovery

After the determination of the sampling vertex subset $\mathcal{C}$, we can derive the samples, and subsequently transform them into observable form, denoted as $\mathbf{y}$. This is equivalent to compute the sampling matrix $\mathbf{C}_F = \boldsymbol{\Upsilon}(\mathcal{C})$ from Eq. (6.21)-(6.25), and pursue sampling on linearly time-evolved observables $[\mathbf{u}_1, \cdots, \mathbf{u}_K]$. Such $\mathbf{y}$ is expressed as:

$$\mathbf{y} = \mathbf{C}_F \cdot \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_K \end{bmatrix}. \tag{6.28}$$

Then, we (i) take the samples $\mathbf{y}$ from Eq. (6.28) into Eq. (6.19), and (ii) transform the initial observable $\mathbf{u}_1$ into its graph frequency response using GFT operator $\mathbf{h}^{-1}(\cdot)$. The result is:

$$\mathbf{y} = \mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{u}_1 = \mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{h}(\mathbf{x}_1). \tag{6.29}$$

Noting the difficulty of the computation of $(\mathbf{C}_F \cdot \mathbf{F} \circ \mathbf{h})^{-1}$, which contains nonlinear inverse operator, we hereby compute the recovery of $\mathbf{x}_1$ using the quasi-Newton methods in following equation. The recovered $\hat{\mathbf{x}}_1$ is derived by optimizing:

$$\hat{\mathbf{x}}_1 = \underset{\mathbf{x}_1\in\mathbb{R}^N}{\text{argmin}} \left\{ \|\mathbf{y} - \mathbf{C}_F \cdot \mathbf{F} \cdot \mathbf{h}(\mathbf{x}_1)\|_{l_2}^2 \right\}, \tag{6.30}$$

with gradient:

$$\nabla = \left(\mathbf{h}(\mathbf{x}_1)^T \cdot \mathbf{F}^T \cdot \mathbf{C}_F^T - \mathbf{y}^T\right) \cdot \mathbf{C}_F \cdot \mathbf{F} \cdot \frac{\partial \mathbf{h}(\mathbf{x}_1)}{\partial \mathbf{x}_1}, \tag{6.31}$$

where $\|\cdot\|_{l_2}$ represents the l_2-norm. After the recovery as $\hat{\mathbf{x}}_1$, we derive the estimated $\hat{\mathbf{u}}_1 = \mathbf{h}(\hat{\mathbf{x}}_1)$, and $\hat{\mathbf{u}}_k = \boldsymbol{\Psi}^{k-1}\hat{\mathbf{u}}_1$. Finally, we recover the time-varying network signals

as $\hat{\mathbf{x}}_k = \mathbf{h}^{-1}(\hat{\mathbf{u}}_k)$. The schematic flow is illustrated via Fig. 6.1(c).

6.4 Novelty Compared with Other State-of-the-Arts

In this section, we distinguish our proposed Log-Koopman nonlinear GFT sampling scheme, with other two state-of-the-art schemes. The first one is referred to as the Poly-Koopman based graph observability analysis [50]. The second one is the combined linear GFT sampling with time-evolved information [46].

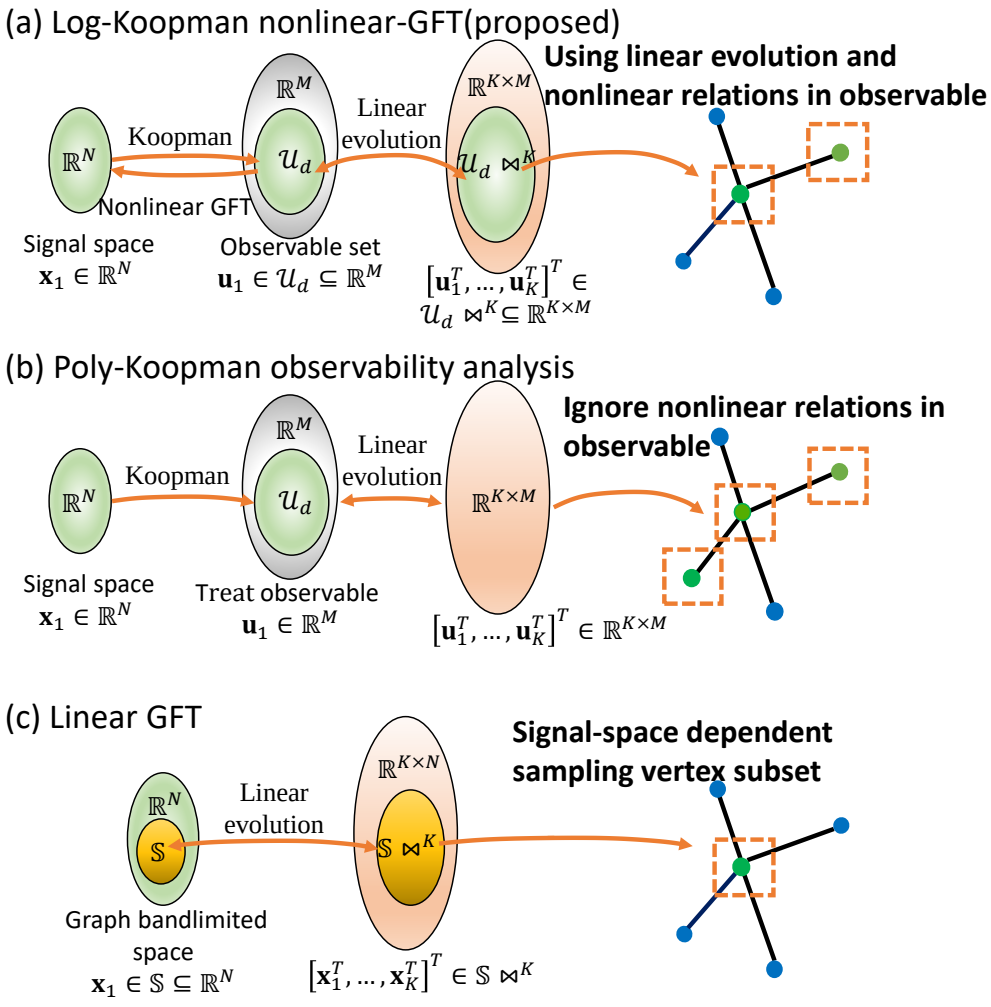

Figure 6.2: Comparison of proposed Log-Koopman nonlinear GFT (a), Poly-Koopman graph observability (b), and linear GFT (c).

6.4.1 Sampling by Poly-Koopman based Graph Observability Analysis

Poly-Koopman based graph observability analysis method has been proposed in [50], which uses the standard linear algebra directly on the poly-based Koopman linearized time-evolution model. The work briefly contains two steps. First, they designed polynomial-based observable elements (denoted as $\boldsymbol{\zeta}_k$ of size $M \times 1$ from $\mathbf{x}_k$ of size $N \times 1$), and derived the corresponding Koopman operator (denoted as $\boldsymbol{\Psi}_{\text{poly}}$ with $\boldsymbol{\zeta}_{k+1} = \boldsymbol{\Psi}_{\text{poly}} \cdot \boldsymbol{\zeta}_k$). Second, they determined the sampling vertex subset by treating the M elements in $\boldsymbol{\zeta}_k$ independent (i.e.,$\boldsymbol{\zeta}_k \in \mathbb{R}^M$), and therefore using the standard linear algebra, i.e., maximizing the energy computed by Koopman observability gramian. As such, the sampling matrix of Poly-Koopman based graph observability analysis, denoted as $\mathbf{W}_\zeta$ of size $M \times M$, is computed as [50]:

$$\mathbf{W}_\zeta = [\mathbf{I}_{l \times l}\ \mathbf{0}] \cdot \mathbf{P}_{\text{poly}}^{-1}, \tag{6.32}$$

by l largest eigenvalues of $\boldsymbol{\Psi}_{\text{poly}} = \mathbf{P}_{\text{poly}} \cdot diag(\lambda_1, \cdots, \lambda_M) \cdot \mathbf{P}_{\text{poly}}^{-1}$ to maximize the energy of the sampled observable, i.e.,

$$\begin{aligned} &\max_{\mathbf{W}_\zeta} \sum_{k=1}^{K} \boldsymbol{\zeta}_1^T \cdot (\boldsymbol{\Psi}_{\text{poly}}^k)^T \cdot \mathbf{W}_\zeta^T \cdot \mathbf{W}_\zeta \cdot \boldsymbol{\Psi}_{\text{poly}}^k \cdot \boldsymbol{\zeta}_1 \\ =&\sum_{k=1}^{K} \boldsymbol{\zeta}_1^T \cdot diag\left(\lambda_1^{2k}, \lambda_2^{2k} \cdots, \lambda_l^{2k}, 0, \cdots, 0\right) \cdot \boldsymbol{\zeta}_1. \end{aligned} \tag{6.33}$$

We distinguish from two aspects between our proposed Log-Koopman nonlinear GFT sampling method with the poly-Koopman graph observability method, as follows:

- First, the polynomial-based observable designed by [50] is only suitable and able to provide reliable linearization performance for small-scale network (e.g., the total number of vertices $N < 30$). However, as analyzed before in Eq (6.4), when it comes to address signal processing issues for the large-scale network (e.g., $N > 50$), the scheme will inevitably lead to the size explosion, by using at least $M = O(N^2)$ polynomial terms to construct the observable. This will subsequently give rise to a size of $O(N^2 \times N^2)$ Koopman operator for further network sampling and signal recovery applications, which will cost substantially huge computational and storaging resources and thereby make it less impractical. In contrast, our proposed Log-Koopman nonlinear GFT sampling method relies on the logarithm summation to reduce the number of

multi-elemental multiplicative terms, and can achieve an order of $O(N)$ number of observable elements for Koopman linearization, thereby avoiding the size explosion for further network sampling and signal recovery applications. Further illustrations of this will be provided in Figs. 6.3-6.8 from Section 6.5.1.

- Second, the direct use of the linear algebra analysis on derived Koopman linearized time-evolution model leads to an overlook of the nonlinear dependency between elements in the vector-valued observable. This is because linear analysis, i.e., the eigenvector analysis in Eq. (6.32) views the initial observable set equaling the linear space $\mathbb{R}^M$ (as is shown in Fig. 6.2). This therefore neglects the fact that the range set of the designed observable is a subset of the linear space $\mathbb{R}^M$, since the designed observable (e.g., the polynomial-based and the log-based) is fully determined by the lower-sized original network signal $\mathbf{x}_1 \in \mathbb{R}^N$. As such, the sampling vertex subset maps from $\mathbb{R}^M$ other than the range set itself $\mathcal{U}_d$ will inevitably give rise to the redundant selection of sampling vertices for the recovery of the observable that belongs to $\mathcal{U}_d \subsetneq \mathbb{R}^M$ (seen Fig. 6.2(a)-(b)). We show the comparison performance in Figs. 6.5-6.6.

6.4.2 Comparison with Linear GFT Sampling

Linear GFT sampling method is to sample and recover the time-varying network signal $\mathbf{x}_k$ that belongs to a known subspace (i.e., having signal dependency among vertices, which is also referred to as the graph bandlimited domain) of $\mathbb{R}^N$, i.e., $\forall k \in \mathbb{N}^+, \mathbf{x}_k \in span\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_r\} \subset \mathbb{R}^N$. Here, the orthogonal $r < N$ supports $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_r$ are derived either from the r-leading eigenvectors of the topology-equation based matrix (in Chapter 3), or from the prior knowledge of the signal-space (in Chapter 4). As such, the linear GFT operator $\mathbf{Q}^{-1} = \mathbf{Q}^T$ can be assigned as $\mathbf{Q}^T = [\mathbf{q}_1, \cdots, \mathbf{q}_r]^T$. And the GFT and inverse GFT processes are $\tilde{\mathbf{x}}_k = \mathbf{Q}^T \cdot \mathbf{x}_k$ and $\mathbf{x}_k = \mathbf{Q} \cdot \tilde{\mathbf{x}}_k$ respectively. In such a manner, the sampling matrix $\mathbf{C}_\Phi$ to ensure the recovery of $\mathbf{x}_k$ from $\mathbf{C}_\Phi \cdot \mathbf{\Phi} \cdot \mathbf{x}_k$ can be determined by [30, 34, 35, 38, 46]

$$rank\left(\mathbf{C}_\Phi \cdot \mathbf{\Phi} \cdot \mathbf{Q}\right) = r. \tag{6.34}$$

Here, different from the traditional graph Theorem 2 considering static graph signals, the matrix $\mathbf{\Phi}$ that specifies the linear time-evolution model is added by the work in [46], i.e., $\mathbf{\Phi} = [\mathbf{L}^0, \cdots, \mathbf{L}^{K-1}]^T$, given an exact or approximated linearized time-evolution model $\mathbf{x}_{k+1} = \mathbf{L} \cdot \mathbf{x}_k$. Then, as one derives the sampling matrix $\mathbf{C}_\Phi$ and the

samples $\mathbf{y} = \mathbf{C}_\Phi \mathbf{\Phi} \cdot \mathbf{x}_k$, the recovered signal $\hat{\mathbf{x}}_k$ can be computed by [30,34,35,38,46]:

$$\hat{\mathbf{x}}_k = \mathbf{Q} \cdot pinv(\mathbf{C}_\Phi \cdot \mathbf{\Phi} \cdot \mathbf{Q}) \cdot \mathbf{y}, \tag{6.35}$$

where $pinv(\cdot)$ represents the pseudo-inverse operator.

We provide the two main differences between our proposed Log-Koopman nonlinear GFT sampling method and the popular linear GFT sampling method as follows.

- First, as compared in Fig. 6.2(a) and Fig. 6.2(c), the linear GFT sampling method relies on the signal dependency among network vertices. As such, the selection of sampling vertex subset under such a graph bandlimited signal subspace, is not suitable for the signals that do not belong to the assumed signal-space. This thereby gives rise to the signal-space dependent sensor placement, as $\mathbf{C}_\Phi$ in Eq. (6.34) will be changed with the different assumptions of the dynamic signal-space.

- Second, when the prior knowledge of the dynamic governing equations and the signal-space are unavailable, the linear GFT sampling method will lose the compass, unable to generate an equation-driven or a data-driven GFT operator for further network sampling and signal recovery. In contrast, our proposed Log-Koopman nonlinear GFT sampling method is able to capture the nonlinear graph bandlimted property of the observable $\mathbf{u}_k = \mathbf{h}(\mathbf{x}_k)$. As such, the selection of the sampling vertex subset is achieved, which is independent with the signal-space, and is suitable for any vector $\mathbf{u}_k \in \mathcal{U}_d \subsetneq \mathbb{R}^M$ determined by any network signal $\mathbf{x}_k \in \mathbb{R}^N$. Therefore, the proposed scheme leads to a time-invariant sensor placement scheme for monitoring and recovery of the time-varying network signals.

6.5 Simulations & Experimental Results

In this section, the performances of our proposed Log-Koopman nonlinear GFT sampling method will be examined. The key performance indicators include the network domain sampling rate (i.e., the quotient of the size of sampling vertex subset divided by the total number of network vertices, $|\mathcal{C}|/N$), and the normalized root mean square error (NRMSE) of the recovered time-varying network signals,

defined in the following:

$$\text{NRMSE} = \sqrt{\frac{\sum_{k=1}^{K}(\hat{\mathbf{x}}_k - \mathbf{x}_k)^T \cdot (\hat{\mathbf{x}}_k - \mathbf{x}_k)}{\sum_{k=1}^{K} \mathbf{x}_k^T \cdot \mathbf{x}_k}}. \tag{6.36}$$

For the experimental setting, we configure the network using the Erdös–Rényi graph, in which the probability of any directed edge is 0.5 and is independent from other edges. We test 1000 different groups of data, of which the network scales (i.e., the total number of vertices N) ranges from 10 to 100, in order to evaluate the wide-range suitability for large/small network scales. The time-varying network signals are derived by two general differential equations with parameters (F, B, and R) according to [66], i.e.,

$$\frac{dx_n(t)}{dt} = F - B \cdot x_n(t) - \sum_{m=1}^{N} R \cdot x_n(t) \cdot x_m(t), \tag{6.37}$$

$$\frac{dx_n(t)}{dt} = -B \cdot x_n(t) + \sum_{m=1}^{N} R \cdot \frac{x_m(t)^2}{1 + x_m(t)^2}. \tag{6.38}$$

Eq. (6.37) is referred to as Biochemical Dynamics of protein interactions, governed by the mass-action kinetics. The corresponding parameters in Eq. (6.37) are set as $F = 10$, $B = 1$ and $R = 1$, and the initial signal for each vertex is randomly assigned as $x_i(0) \in (0, 1)$ [66, 84]. Eq. (6.38) is referred to as gene Regulatory Dynamics. The corresponding parameters in Eq. (6.38) are set as $B = 1$ and $R = 1$, and the initial signal at each vertex is randomly configured as $x_i(0) \in (0, 100)$ [66]. Here, we emphasize that although we list the governing dynamic equations in Eqs. (6.37)-(6.38), we do not rely on those exact expressions; what we require in this experiment is their generated data for performance evaluation. For the proposed logarithm observable in Eq. (6.6), we assign the scaling parameter $\eta = 500$ via a scan and selection from all η with optimal linearization accuracy.

6.5.1 Log-Koopman Linearization Performance

In this part, the linearization accuracy of our proposed log-observable based Koopman operator is provided in Figs. 6.3-6.4. For Fig. 6.3, we use $N = 50$ biochemical network dynamic in Eq. (6.37). For Fig. 6.4, an $N = 100$ gene Regulatory network Dynamic in Eq. (6.38) is adopted. To evaluate the linearization accuracy, we compare the proposed Log-observable based Koopman operator with the state-of-the-art poly-observable based Koopman operator in [47]. Here, we represent the x-axis as

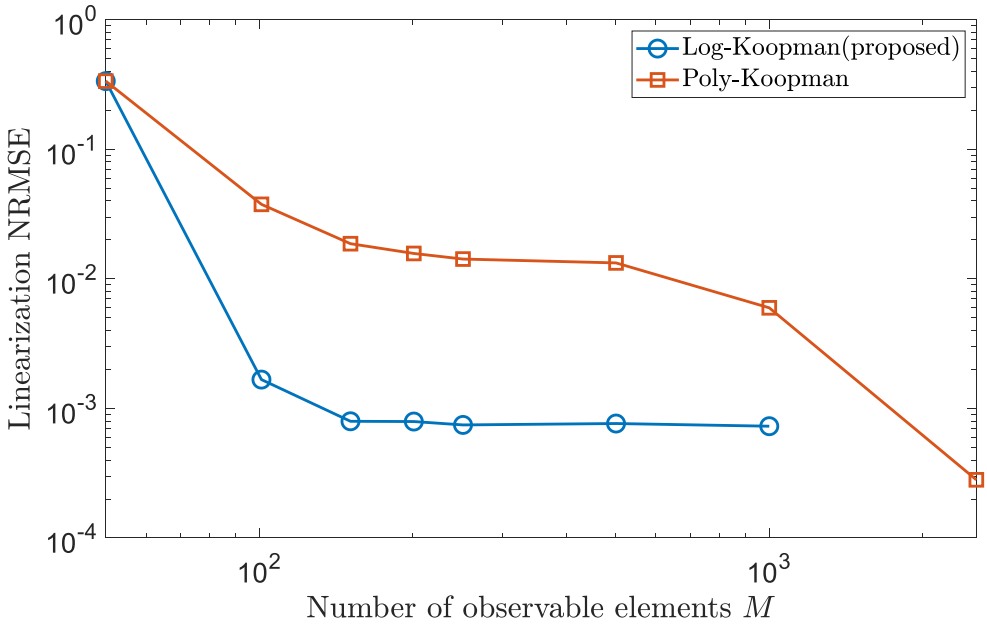

Figure 6.3: Comparison of linearization performance between different Koopman schemes, for $N = 50$ biochemical time-varying network signals of protein interactions.

the number of observable elements in $\mathbf{h}(\cdot) = [h_1(\cdot), h_2(\cdot), \cdots, h_M(\cdot)]^T$ in Eq. (6.6), i.e., M. The y-axis then presents the normalized RMSE between the time-evolved signals using Koopman operator and the original time-varying network signals.

It is firstly seen from Fig. 6.3 that, when the number of observable elements equals the total number of network vertices ($M = N$), the linearization NRMSEs of both schemes are identical (i.e., $NRMSE = 0.3$ for both schemes). This is due to that both the proposed Log-observable based Koopman and the compared Poly-observable based Koopman degenerate to the DMD, which directly use the original network signals as the Koopman observable, i.e., $\mathbf{h}(\mathbf{x}_k) = \mathbf{x}_k$. Then, it is observed from Fig. 6.3 that, as the number of observable elements M increases, the linearization NRMSEs of both schemes decrease. For example, when M increases from 50 to 200, the linearization NRMSE decreases from 0.3 to 10^{-3} for our proposed Log-Koopman, and when M increases from 50 to 2000, the linearization NRMSE gets lower from 0.3 to an order of 10^{-4} for Poly-Koopman. Within this trend, it is seen that the linearization NRMSE of the proposed Log-Koopman operator converges faster to a small value as opposed to the Poly-Koopman operator in [47]. The former uses only $M = O(N)$ (e.g., $M = 3 \times N = 150$ in Fig. 6.3) observable elements, which is much smaller than that of the Poly-Koopman operator (requiring

$M = O(N^2)$).

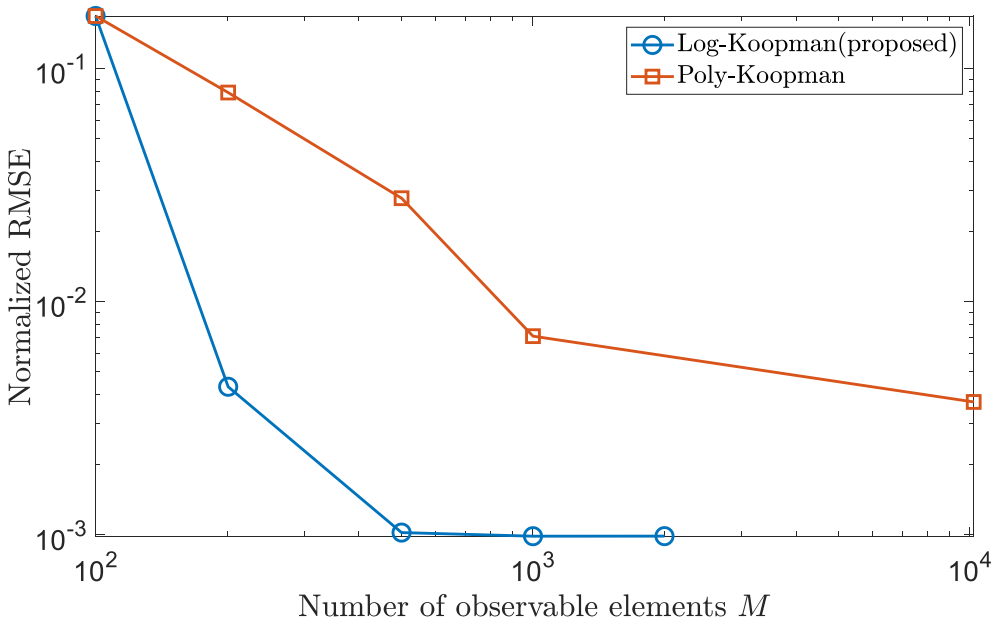

Figure 6.4: Comparison of linearization performance between different Koopman schemes, for $N = 100$ gene Regulatory Dynamics.

Similar results are also illustrated in Fig. 6.4, where $N = 100$ gene Regulatory network Dynamic in Eq. (6.38) is evaluated. First, an identical linearization RMSE (e.g., 0.2) is approached for both schemes at the point the number of observable elements equaling the total number of network vertices ($M = N$), due to the basic DMD observable selection used for both schemes. Then, we can see from Fig. 6.4 that, with an increase of the number of observable elements M, the linearization NRMSEs of both schemes get smaller. For example, when M increases from 10^2 to 10^3, the linearization NRMSE decreases from 0.2 to nearly 10^{-3} for our proposed Log-Koopman, and when M grows from 10^3 to 10^4, the linearization NRMSE decreases from 0.2 to an order of 10^{-3} for Poly-Koopman. Also, the linearization NRMSE of the proposed Log-Koopman operator converges faster when compared with the Poly-Koopman operator in [47]. The former uses only $M = O(N)$ (e.g., $M = 5 \times N = 500$ in Fig. 6.4) observable elements, much smaller than that of the Poly-Koopman operator (which is not able to reach the same linearization accuracy performance even if a group of $M = O(N^2) = 10^4$ observables are used).

We attribute such observable size reduction to the logarithm summation used by our proposed scheme to approximate the multi-elemental multiplicative terms of Taylor series in Eq. (6.3). In such a manner, only a smaller number of logarithm-

based observable elements are needed to approximate and replace the indefinite multiplicative Poly-based observable elements. This therefore indicates the ability of the proposed Log-Koopman to prevent the size explosion when addressing and linearizing the large-scale time-varying network signals (i.e., $N > 50$). Subsequently, such observable size reduction from $M = O(N^2)$ to $M = O(N)$ makes the further processes of network sampling and signal recovery more feasible and tractable, as the Koopman operator of size $M \times M$ will be used.

It is noteworthy that a major disadvantage of the proposed log-observable based Koopman operator lies in a relatively large linearization NRMSE limit (e.g., an order of 10^{-3} in Figs. 6.3-6.4), as opposed to the poly-observable based Koopman operator. This is due to the inaccuracy of the logarithm summation approximation for the multi-elemental multiplicative polynomial terms. However, in the context of the network sampling application, we address this by using the sample-awareness training data to refine the designed Log-Koopman operator. To be specific, we generate the new training data evolved by the sample-related initialization, where values from the sensor vertices are assigned as the initial samples, and others are randomly generated. By doing so, we will show in the following that, the sample-awareness Log-Koopman operator combined with the nonlinear GFT sampling theory, can provide a promising network sampling and signal recovery performances.

6.5.2 Performance of Log-Koopman Nonlinear GFT Sampling and Recovery

The performances of the proposed nonlinear GFT sampling scheme leveraging the Log-Koopman operator is evaluated in the following. Here, two state-of-the-art sampling methods are pursued for comparisons, i.e., the Poly-Koopman graph observability analysis scheme in [50], and the network time-difference smoothness signal recovery scheme in [19].

Recovery Accuracy

We illustrate the recovery NRMSE with respect to the network domain sampling rate, i.e., $|\mathcal{C}|/N$, in Figs. 6.5-6.6, where Fig. 6.5 accounts for the $N = 100$ biochemical network dynamic in Eq. (6.37), and Fig. 6.6 presents the $N = 100$ gene regulatory network dynamic in Eq. (6.38). In Fig. 6.5, we firstly observe that, as the sampling rate $|\mathcal{C}|/N$ increases from 0 to 1, the recovery NRMSEs of all schemes become lower (e.g., the recovery NRMSE decreases from 10^{-1} to an order of 10^{-3} for our proposed scheme). This is due to the fact that a larger $|\mathcal{C}|/N$ can lead to

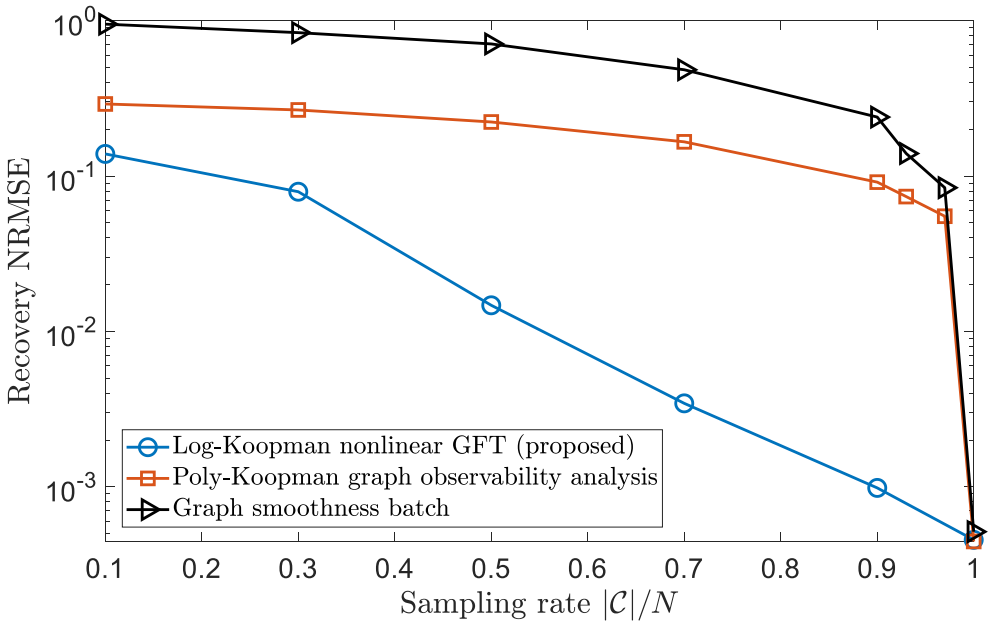

Figure 6.5: Comparison of recovery NRMSE versus sampling rate among different sampling schemes, for $N = 100$ biochemical network dynamic of protein interactions.

more samples from the network domain and therefore a better recovery performance can be achieved. Then, it is observed that the sampling rate $|\mathcal{C}|/N$ from our proposed Log-Koopman nonlinear GFT sampling scheme is much smaller as opposed to the two competitive schemes in [50] and [19]. For example, our proposed Log-Koopman nonlinear GFT sampling scheme reaches NRMSE $= 10^{-2}$ by sampling from only 50% of total network vertices, which is much smaller than that of the Poly-Koopman graph observability scheme in [50] (needing nearly 90% of vertices to guarantee the same order recovery NRMSE).

Similar results can be found in Fig. 6.6, which presents the $N = 100$ gene regulatory network dynamic in Eq. (6.38). In Fig. 6.6, it is observed that, as the sampling rate $|\mathcal{C}|/N$ increases from 0 to 1, the recovery NRMSEs of all schemes become lower, as a larger $|\mathcal{C}|/N$ can lead to more samples from the network domain and therefore a better recovery performance can be achieved. Then, we can see that the sampling rate $|\mathcal{C}|/N$ from our proposed Log-Koopman nonlinear GFT sampling scheme is much smaller as opposed to the two competitive schemes in [50] and [19]. For example, in Fig. 6.6, our proposed Log-Koopman nonlinear GFT sampling scheme reaches a NRMSE lower than 10^{-2} by sampling from only 70% of total network vertices, much smaller than that of the Poly-Koopman graph observability scheme in [50] (needing almost all of vertices to reach the same recovery accuracy).

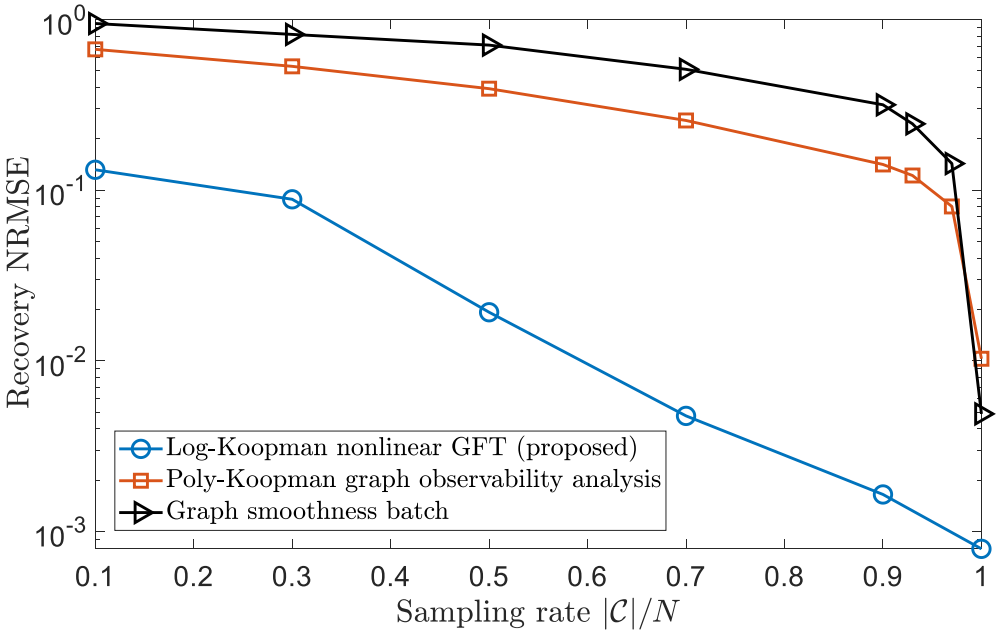

Figure 6.6: Comparison of recovery NRMSE versus sampling rate among different sampling schemes, for $N = 100$ gene regulatory network dynamics.

Recovery Robustness

We further analyze the robustness of our proposed Log-Koopman nonlinear GFT method in the context of various network scales (i.e., the total number of network vertices from $N = 10$ to $N = 100$). Figs. 6.7-6.8 provide the average recovery NRMSEs over various N, under the network domain sampling rates as $|\mathcal{C}|/N = 25\%, 50\%, 75\%$. Here, Fig. 6.7 presents the results for the biochemical network dynamic in Eq. (6.37). From Fig. 6.7, we can see that, the proposed Log-Koopman nonlinear GFT sampling scheme has smaller average recovery NRMSEs, as opposed to the Poly-Koopman graph observability scheme in [50], and the network time-difference smoothness recovery method in [19], under the same network domain sampling rates. For example, it is illustrated that when using 50% of the total network vertices for monitoring, the proposed Log-Koopman nonlinear GFT can approach an average NRMSE as an order of 10^{-2}, which is much smaller than that from the Poly-Koopman graph observability analysis (i.e., $NRMSE \approx 0.3$), and that from the graph smoothness batch (i.e., $NRMSE \approx 0.8$). The gaps become even larger when 75% of network vertices are monitored (i.e., 2×10^{-3} for proposed scheme vs. 0.2 for Poly-Koopman graph analysis and 0.4 for graph smoothness batch).

Similar results can be seen in Fig. 6.8, which gives the recovery NRMSEs at

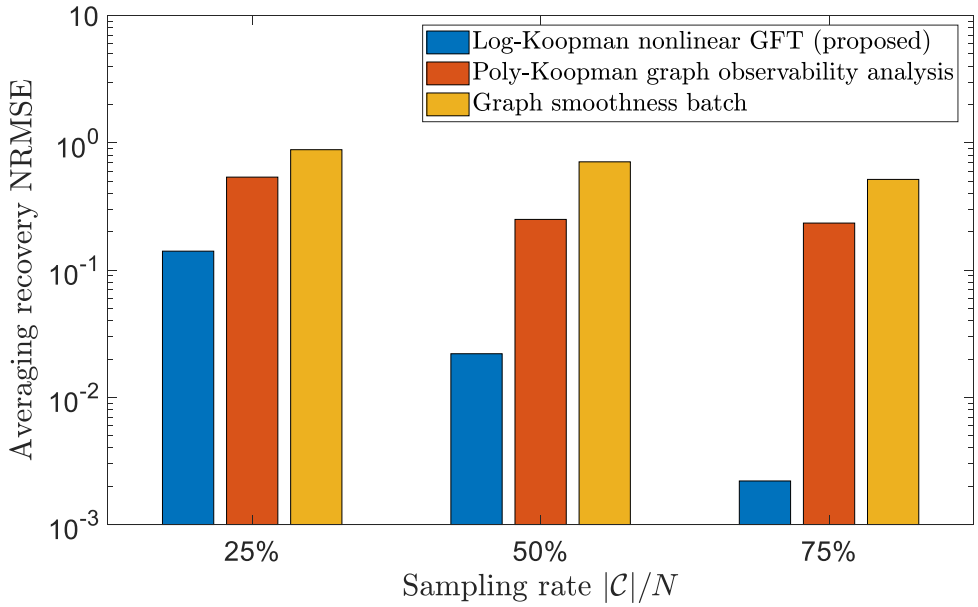

Figure 6.7: Comparison of recovery NRMSE versus sampling rate among different sampling schemes, for different network scales from $N = 10$ to $N = 100$ biochemical time-varying network signals of protein interactions.

various network scales (i.e., $N = 10$ to $N = 100$) for the gene regulatory network dynamic provided in Eq. (6.38). From Fig. 6.8, it is observed that, the proposed Log-Koopman nonlinear GFT sampling scheme outperforms the Poly-Koopman graph observability scheme in [50], and the network time-difference smoothness recovery method in [19], by providing smaller average recovery NRMSEs under the same network domain sampling rates. For example, we can see that when using 25% of the total network vertices for monitoring, the proposed Log-Koopman nonlinear GFT has an order of 10^{-1} NRMSE as opposed to that from the Poly-Koopman graph observability analysis (i.e., $NRMSE \approx 0.7$), and that from the graph smoothness batch (i.e., $NRMSE \approx 1$). Such gaps get larger when 50% and 75% of network vertices are sampled (e.g., at $|\mathcal{C}|/N = 50\%$, we have 1×10^{-2} for proposed scheme, 0.2 for Poly-Koopman graph analysis, and 0.8 for graph smoothness batch).

Combining the results from Figs. 6.7-6.8, we obtain that, the proposed Log-Koopman nonlinear GFT sampling scheme is applicable to a wide range of the network scales. Then, it is also demonstrated that, our proposed scheme requires a sampling of 50% network vertices to ensure a relative low recovery NRMSE (i.e., an order of 10^{-2}), quite smaller than those of the two competitive methods in [19, 50].

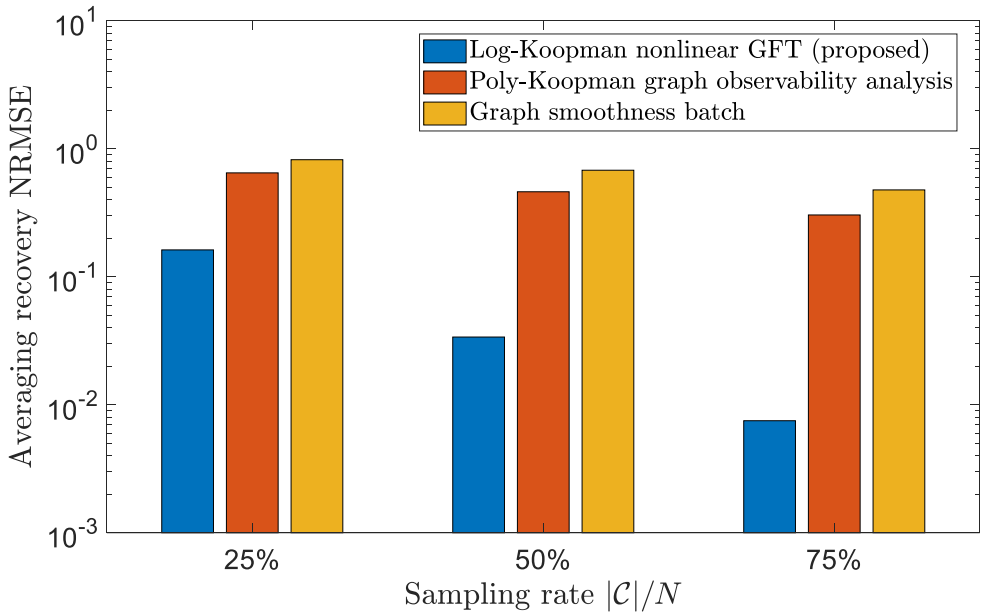

Figure 6.8: Comparison of recovery NRMSE versus sampling rate among different sampling schemes, for different network scales from $N = 10$ to $N = 100$ gene regulatory network dynamics.

Result Explanation

We explain the advantage of the sampling vertex subset reduction from our proposed Log-Koopman nonlinear GFT in the following. First, the scheme in [19], although uses the network domain smoothness (graph bandlimitedness) of the signal time-difference, overlooks the more exact time-evolution and network domain signal dependency. As such, its performance is not comparable to our proposed scheme, whereby an accurate Koopman time-evolution model and networked nonlinear GFT dependency are exploited for network sampling and signal recovery. Second, when compared with the Poly-Koopman graph observablity analysis method in [50], the sampling vertex subset reduction of the proposed Log-Koopman nonlinear GFT scheme is thanked to the exploration of the nonlinear dependency between designed observable elements. As we derive the Koopman operator, the original network signal $\mathbf{x}_k$ of size $N \times 1$ is expanded by the dependent $M > N$ observable elements, i.e., $\mathbf{u}_k = [h_1(\mathbf{x}_k), \cdots, h_M(\mathbf{x}_k)]^T$, and the purpose is then converted to search the sampling vertex subset $\mathcal{C}$ to recover the initial observable $\mathbf{u}_1$. As such, the selection of the sampling vertex subset should exploit the nonlinear dependence between the elements of $\mathbf{u}_1$. In such a view, the graph observability analysis in [50] regards the total M elements of $\mathbf{u}_1$ as independent, ignoring such nonlinear relations, and

thereby results in redundant sampling vertices. In contrast, our proposed nonlinear GFT scheme is able to transform the initial observable $\mathbf{u}_1$ to its graph bandlimited set determined by the original N elements, i.e., $\mathbf{u}_1 = \mathbf{h}(\mathbf{x}_1)$, and therefore capable of obtaining the smaller sized sampling vertex subset for network sampling and signal recovery.

6.6 Conclusions & Discussions

For this chapter, we have proposed the Log-Koopman based nonlinear GFT sampling framework for the scenarios where both the signal dependency among vertices and the dynamic time-evolution model are unknown. To use the time-evolution information for sampling vertex subset selection and network signal recovery, the Log-Koopman operator has been designed, able to derive a linearized time-evolution of the designed logarithm vector-valued observable defined on the original signal space. Compared with the state-of-the-art Poly-Koopman operator that causes a size explosion using $M = O(N^2)$ observable elements, the proposed one can reduce the observable size to $M = O(N)$ by the designed logarithm summation to approximate the multi-elemental multiplicative polynomial terms for poly-based observable. This thereby enables the further compression process that relies on the linearized time-evolution model. Then, the nonlinear GFT sampling framework has been proposed with the proven sampling theory, for the purpose to analyze and exploit the M dependent observable elements defined and determined by the N original signal. Compared to the graph observable analysis that treat the M observable elements independent, the proposed nonlinear GFT sampling can achieve a more compact sampling vertex subset by mapping the samples to the more compact range set of the designed vector-valued observable.

Chapter 7

Summary & Future Works

Network and nonlinear dynamic mechanisms underlie the coupling and complex functionalities of many engineering, ecological, social, biological systems. Optimally deploying sensors for monitoring the network's time-varying signals serves as the fundamentals for a wide-range of engineering and scientific purposes. In this thesis, we study how to place sensors on a subset of network vertices, to ensure the recovery of the time-varying network signals.

7.1 Summary of this Thesis

This thesis began by reviewing the current sampling and compression approaches categorized by relying on the signal dependency among network vertices, and the time-evolution information. The former requires an operator to uncover and characterize the dependencies of the signals on different vertices, rendering a huge difficulty for current topology-only GFT framework. The latter regards the linear/linearized time-evolution dynamic model as the prerequisite for further exploration of sequential information and sampling selection, thereby making it less practical for most of the real-world applications without explicit dynamic models.

These two challenges motivated us to develop (i) dynamic mechanism-topology combined GFT sampling, and (ii) linearized dynamic time-evolution modelling and sampling, which have been proposed and elaborated in Chapters 3-4, and Chapters 5-6 respectively.

In Chapter 3, the GFT operator with combined dynamic equation and network topology were developed in the cases with the explicit dynamic governing equations. Such an equation-topology GFT operator is able to uncover the signal dependency among different vertices. Then, the GFT sampling theory were revised

to select the time-invariant sampling vertex subset, as well as determine the discretizing rate (time-domain sampling frequency), by mapping the samples to the whole graph bandlimited frequency response. As such, the GFT framework establishes how to sample and discretize the continuous network signals under explicit governing equations.

In Chapter 4, the data-driven GFT operators were developed in order to deal with the cases where dynamic governing equations are unavailable. Learnt from the prior knowledge of the signal-space, such a data-driven GFT operator is capable of capturing the hidden dynamic mechanism from data. Then, GFT sampling theory for matrix data were developed to identify the time-invariant sampling vertex subset, by ensuring the one-to-one mapping between the samples and the whole graph bandlimited frequency response. Also, for monitoring the water-distribution network that is buried in-depth underground and is hard to be penetrated for data collection and transmission, we proposed a molecular relay mechanism for each sensor vertex to (i) encode the samples into the structure of DNA molecules, and (ii) transmit them via the pipe itself, thereby avoiding the complex penetration techniques and extra communication networks. In such a manner, we re-designed the sampling vertex subset selection algorithm to ensure the connectivity between the sensor vertices and the hub vertex for collecting reports and signal recovery. As such, the data-driven GFT sampling framework that can characterize and exploit the signal dependencies among vertex for sampling point selection has been established.

In Chapters 5-6, we considered the network sampling in the absence of the signal dependency among vertices (i.e., both the dynamic governing equations and the signal-space are unknown). We alternatively relied on the time-evolution information to achieve network sampling and signal recovery with latency. In Chapter 5, we developed the sequential data-driven GFT operator, by analyzing the principal component of the previously recovered signals. Such a sequential PCA GFT operator is able to characterize and exploit the transient signal dependency for network sampling, however, approaches the time-varying sampling vertices as the signal dependency among different vertices changes with time.

In Chapter 6, to further exploit the hidden time-evolution information, and to derive a time-invariant sampling strategy, we proposed the Log-Koopman nonlinear GFT sampling framework. Here, we designed the Log-Koopman operator to derive a linearized dynamic time-evolution model, leveraging which, the nonlinear GFT concept and sampling theory were proposed to achieve the time-invariant sampling vertex subset for complete signal recovery. As such, given its model and signal independency characteristics, our proposed Log-Koopman GFT sampling framework

is useful and can be adjusted to a wide range of network signal sampling and recovery applications.

7.2 Future Works

The future works will be mainly focused on industrial engineering and applications. This requires a pathway to a more straightforward, interpretable and trustworthy approach that the industry can easily adopt. Machine learning has hitherto been an useful and easy-to-use tool, to achieve a solution to complex and hard-to-solve problem with little expertise. As such, a combination with machine learning is demanding to expand the impact of our research.

The first step is to construct an auto-encoder and an auto-decoder using two neural networks (NN), where the former is to identify which vertices will be selected for samples, and the latter is to recover the time-varying network signals from the samples. Compared to our proposed dynamic-topology GFT framework where only the linear signal dependency among vertices can be considered, the encoder NN trained by the resulting GFT sampling points has the potential to characterize the nonlinear dependency among signals on different vertices, given the nonlinear properties of the activation functions on each neuron. This may achieve a more compact sampling vertex subset for signal recovery, as deeper nonlinear dependency governed by the hidden dynamical mechanism is taken into account.

To achieve an easy-to-use as well as trustworthy sampling framework for industrial engineering, using only NN based auto-encoder and auto-decoder are not enough, as they typically lack the transparency and are hard to be explained, due to the black-box nature of NN. This makes them vulnerable to malicious inputs and may result in unreliable outputs that are not trustworthy. As such, our second step will be spent on improving the explanability of the NN encoder and decoder. We will use the state-of-the-art metrics and methods (e.g., neurons' visualization and NN symbolic representation) introduced in [85], trying to build a reliable and explainable artificial intelligence (X-AI) framework for network sampling and signal recovery.

Bibliography

[1] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer, "Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks," *Nat. Commun.*, vol. 5, pp. 1–9, Sep. 2014.

[2] S. Dhamal, R. D. Vallam, and Y. Narahari, "Modeling spread of preferences in social networks for sampling-based preference aggregation," *IEEE Trans. Netw. Sci. Eng.*, vol. 6, pp. 46–59, Jan. 2019.

[3] A. Wilson, "Boltzmann, lotka and volterra and spatial structural evolution: An integrated methodology for some dynamical systems," *J. Roy. Soc. Interface*, vol. 5, pp. 865–871, Dec. 2007.

[4] B. Schäfer, D. Witthaut, M. Timme, and V. Latora, "Dynamically induced cascading failures in power grids," *Nat. Commun.*, vol. 9, p. 1975, May 2018.

[5] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Inferring biological networks by sparse identification of nonlinear dynamics," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, pp. 52–63, Jun. 2016.

[6] A. Yazdani and P. Jeffrey, "Complex network analysis of water distribution systems," *Chaos*, vol. 21, p. 016111, Mar. 2011.

[7] A. Annaswamy, "Ieee vision for smart grid control: 2030 and beyond roadmap," *IEEE Vision for Smart Grid Control: 2030 and Beyond Roadmap*, pp. 1–12, 2013.

[8] Z. Wei, A. Pagani, G. Fu, I. Guymer, W. Chen, J. A. McCann, and W. Guo, "Optimal sampling of water distribution network dynamics using graph Fourier transform," *IEEE Trans. Netw. Sci. Eng.*, pp. 1–1, 2019.

[9] Z. Wei, A. Pagani, B. Li, and W. Guo, "Monitoring embedded flow networks using graph Fourier transform enabled sparse molecular relays," *IEEE Commun. Lett.*, vol. 24, pp. 986–990, May 2020.

[10] F. H. Bijarbooneh, W. Du, E. C. . Ngai, X. Fu, and J. Liu, "Cloud-assisted data fusion and sensor selection for Internet of Things," *IEEE Internet Things J.*, vol. 3, pp. 257–268, Jun. 2016.

[11] J. Gao, B. Barzel, and A.-L. Barabási, "Universal resilience patterns in complex networks," *Nature*, vol. 530, p. 307, Feb. 2016.

[12] G. Moutsinas and W. Guo, "Node-level resilience loss in dynamic complex networks," *arXiv preprint arXiv:1808.05975*, 2018.

[13] A. Pagani, G. Mosquera, A. Alturki, S. Johnson, S. Jarvis, A. Wilson, W. Guo, and L. Varga, "Resilience or robustness: identifying topological vulnerabilities in rail networks," *Roy. Soc. Open Sci.*, vol. 6, p. 181301, Feb. 2019.

[14] A. Kessler, A. Ostfeld, and G. Sinai, "Detecting accidental contaminations in municipal water networks," *J. Water Resour. Plann. Manage.*, vol. 124, pp. 192–198, Jul. 1998.

[15] J. R. Chastain Jr, "Methodology for locating monitoring stations to detect contamination in potable water distribution systems," *J. Infrastruct. Syst.*, vol. 12, pp. 252–259, Dec. 2006.

[16] S. Rathi and R. Gupta, "Monitoring stations in water distribution systems to detect contamination events," *ISH J. Hydraul. Eng.*, vol. 20, pp. 142–150, May 2014.

[17] J. Berry, L. Fleischer, W. Hart, C. Phillips, and J. Watson, "Sensor placement in municipal water networks," *J. Water Resour. Plann. Manage.*, vol. 131, pp. 237–243, Jan. 2005.

[18] A. Ostfeld and E. Salomons, "Optimal layout of early warning detection stations for water distribution systems security," *J. Water Resour. Plann. Manage.*, vol. 130, pp. 377–385, Sep. 2004.

[19] K. Qiu, X. Mao, X. Shen, X. Wang, T. Li, and Y. Gu, "Time-varying graph signal reconstruction," *IEEE J. Sel. Topics Signal Process.*, vol. 11, pp. 870–883, Sep. 2017.

[20] R. Du, L. Gkatzikis, C. Fischione, and M. Xiao, "Energy efficient sensor activation for water distribution networks based on compressive sensing," *IEEE J. Sel. Areas Commun.*, vol. 33, pp. 2997–3010, Dec. 2015.

[21] S. Kartakis, G. Tzagkarakis, and J. A. McCann, "Adaptive compressive sensing in smart water networks," in *2nd International Electronic Conference on Sensors and Applications*, 2015.

[22] X. Xie, Q. Zhou, D. Hou, and H. Zhang, "Compressed sensing based optimal sensor placement for leak localization in water distribution networks," *J. Hydroinform.*, vol. 20, pp. 1286–1295, Nov. 2017.

[23] L. Xu, X. Qi, Y. Wang, and T. Moscibroda, "Efficient data gathering using compressed sparse functions," in *Proc. IEEE INFOCOM*, pp. 310–314, 2013.

[24] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3447–3461, Oct. 2012.

[25] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[26] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[27] F. R. Chung and F. C. Graham, *Spectral graph theory.* No. 92, American Mathematical Soc., 1997.

[28] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.

[29] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, pp. 3042–3054, Jun. 2014.

[30] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, pp. 6510–6523, Dec. 2015.

[31] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, pp. 80–90, Sep. 2014.

[32] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs.," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3864–3868, 2014.

[33] X. Wang, J. Chen, and Y. Gu, "Generalized graph signal sampling and reconstruction," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 567–571, Dec. 2015.

[34] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, pp. 3775–3789, Jul. 2016.

[35] S. Chen, R. Varma, A. Singh, and J. Kovacevic, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Trans. Signal Inf. Process. over Networks*, vol. 2, pp. 539–554, Dec. 2016.

[36] F. Wang, Y. Wang, and G. Cheung, "A-optimal sampling and robust reconstruction for graph signals via truncated neumann series," *IEEE Signal Process. Lett.*, vol. 25, pp. 680–684, May 2018.

[37] L. F. O. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, pp. 34–47, Jan. 2018.

[38] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, pp. 808–828, May 2018.

[39] A. Di Nardo, C. Giudicianni, R. Greco, M. Herrera, G. F. Santonastaso, and A. Scala, "Sensor placement in water distribution networks based on spectral algorithms," *EPiC Series in Engineering*, vol. 3, pp. 593–600, Sep. 2018.

[40] E. J. Candes and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, pp. 2342–2359, Apr. 2011.

[41] F. Grassi, A. Loukas, N. Perraudin, and B. Ricaud, "A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs," *IEEE Trans. Signal Process.*, vol. 66, pp. 817–829, Feb. 2018.

[42] Y. Xue, S. Pequito, J. R. Coelho, P. Bogdan, and G. J. Pappas, "Minimum number of sensors to ensure observability of physiological systems: A case study,"

in *Proc. 54th Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, USA*, pp. 1181–1188, 2016.

[43] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, vol. 78, pp. 202–210, Apr. 2017.

[44] S. Pequito, P. Bogdan, and G. J. Pappas, "Minimum number of probes for brain dynamics observability," in *Proc. 54th IEEE Conference on Decision and Control (CDC), Osaka, Japen*, pp. 306–311, Dec. 2015.

[45] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, pp. 2069–2084, Sep. 2010.

[46] E. Isufi, P. Banelli, P. Di Lorenzo, and G. Leus, "Observing and tracking bandlimited graph processes," *arXiv preprint arXiv:1712.00404*, 2017.

[47] E. Yeung, Z. Liu, and N. O. Hodas, "A koopman operator approach for computing and balancing gramians for discrete time nonlinear systems," in *Annual American Control Conference (ACC) Milwaukee, WI, USA*, pp. 337–344, Jun. 2018.

[48] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," *Nat. Commun.*, vol. 9, pp. 1–10, Nov. 2018.

[49] C. A. Johnson and E. Yeung, "A class of logistic functions for approximating state-inclusive koopman operators," in *Annual American Control Conference (ACC) Milwaukee, WI, USA*, pp. 4803–4810, Jun. 2018.

[50] A. Hasnain, N. Boddupalli, and E. Yeung, "Optimal reporter placement in sparsely measured genetic networks using the koopman operator," in *IEEE 58th Conference on Decision and Control (CDC) Nice, France*, pp. 19–24, Dec. 2019.

[51] N.-B. Chang, N. Prapinpongsanone, and A. Ernest, "Optimal sensor deployment in a large-scale complex drinking water network: Comparisons between a rule-based decision support system and optimization models," *Comput. Chem. Eng.*, vol. 43, pp. 191–199, Aug. 2012.

[52] M. Aral, J. Guan, and M. Maslia, "Optimal design of sensor placement in water distribution networks," *J. Water Resour. Plann. Manage.*, vol. 136, pp. 5–18, Mar 2010.

[53] A. Krause, J. Leskovec, C. Guestrin, and J. V. Briesen, "Efficient sensor placement optimization for securing large water distribution networks," *J. Water Resour. Plann. Manage.*, vol. 134, pp. 516–526, Nov. 2008.

[54] J. Chu, C. Zhang, G. Fu, Y. Li, and H. Zhou, "Improving multi-objective reservoir operation optimization with sensitivity-informed dimension reduction," *Hydrol. Earth Syst. Sci.*, vol. 19, pp. 3557–3570, Aug. 2015.

[55] N. Sankary and A. Ostfeld, "Incorporating operational uncertainty in early warning system design optimization for water distribution system security," *Procedia Eng.*, vol. 186, pp. 160–167, Jan. 2017.

[56] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, pp. 4363–4378, Aug. 2016.

[57] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, pp. 4845–4860, Sep. 2016.

[58] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, pp. 451–462, Feb. 2009.

[59] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, pp. 2077–2092, Apr. 2017.

[60] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

[61] J. Hart, I. Guymer, F. Sonnenwald, and V. Stovin, "Residence time distributions for turbulent, critical, and laminar pipe flow," *J. Hydraul. Eng.*, vol. 142, p. 04016024, May 2016.

[62] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition," *J. Nonlin. Sci.*, vol. 25, pp. 1307–1346, Jun. 2015.

[63] M. A. Hernández-Ortega and A. R. Messina, "Nonlinear power system analysis using koopman mode decomposition and perturbation theory," *IEEE Trans. Power Syst.*, vol. 33, pp. 5124–5134, Sep. 2018.

[64] M. Netto and L. Mili, "A robust data-driven koopman kalman filter for power systems dynamic state estimation," *IEEE Trans. Power Syst.*, vol. 33, pp. 7228–7237, Nov. 2018.

[65] A. M. Lyapunov, "The general problem of the stability of motion," *International J. Control*, vol. 55, pp. 531–534, Mar. 1992.

[66] B. Barzel and A.-L. Barabási, "Universality in network dynamics," *Nat. Phys.*, vol. 9, p. 673, Oct. 2013.

[67] L. Rossman, "Epanet 2 users manual," *U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-00/057*, 2000.

[68] S. Qiu, C. Mias, W. Guo, and X. Geng, "Hs2 railway embankment monitoring: effect of soil condition on underground signals," *SN Appl. Sci.*, vol. 1, p. 537, May 2019.

[69] Z. Sun and I. F. Akyildiz, "Magnetic induction communications for wireless underground sensor networks," *IEEE Transactions on Antennas and Propagation*, vol. 58, pp. 2426–2435, Jul. 2010.

[70] A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Information theory of molecular communication: Directions and challenges," *IEEE Trans. Mol., Biol. Multi-Scale Commun.*, vol. 2, pp. 120–142, Dec. 2016.

[71] F. Zabini, "Spatially distributed molecular communications: An asynchronous stochastic model," *IEEE Commun. Lett.*, vol. 22, pp. 1326–1329, Jul. 2018.

[72] B. A. Bilgin, E. Dinc, and O. B. Akan, "Dna-based molecular communications," *IEEE Access*, vol. 6, pp. 73119–73129, 2018.

[73] M. Pierobon and I. F. Akyildiz, "Diffusion-based noise analysis for molecular communication in nanonetworks," *IEEE Trans. Signal Process.*, vol. 59, pp. 2532–2547, Jun. 2011.

[74] Z. Wei, W. Guo, B. Li, J. Charmet, and C. Zhao, "High-dimensional metric combining for non-coherent molecular signal detection," *IEEE Trans. Commun.*, vol. 68, pp. 1479–1493, Mar. 2020.

[75] Z. Wei, B. Li, W. Hu, W. Guo, and C. Zhao, "Hamming–luby rateless codes for molecular erasure channels," *Nano Communication Networks*, vol. 23, p. 100280, 2020.

[76] Z. Luo, L. Lin, W. Guo, S. Wang, F. Liu, and H. Yan, "One symbol blind synchronization in simo molecular communication systems," *IEEE Wireless Commun. Lett.*, vol. 7, pp. 530–533, Aug. 2018.

[77] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.

[78] O. Jo, Y. Kim, and J. Kim, "Internet of Things for smart railway: Feasibility and applications," *IEEE Internet Things J.*, vol. 5, pp. 482–490, Apr. 2018.

[79] D. Romero, V. N. Ioannidis, and G. B. Giannakis, "Kernel-based reconstruction of space-time functions on dynamic graphs," *IEEE J. Sel. Topics Signal Process.*, vol. 11, pp. 856–869, Sep. 2017.

[80] P. Di Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Adaptive least mean squares estimation of graph signals," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 2, pp. 555–568, Dec. 2016.

[81] P. Di Lorenzo, P. Banelli, S. Barbarossa, and S. Sardellitti, "Distributed adaptive learning of graph signals," *IEEE Trans. Signal Process.*, vol. 65, pp. 4193–4208, Aug. 2017.

[82] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, pp. 4363–4378, Aug. 2016.

[83] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, pp. 4845–4860, Sep. 2016.

[84] E. O. Voit, *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists.* Cambridge University Press, 2000.

[85] W. Guo, "Explainable artificial intelligence for 6G: Improving trust between human and machine," *IEEE Commun. Mag.*, vol. 58, pp. 39–45, Jun. 2020.