Research on Graph Signal Sampling and Reconstruction Method Based on Causal Emergence

Weiping Ren

China West Normal University Nanchong Sichuan, China 1403655452@qq.com

Jialin He*

China West Normal University Nanchong Sichuan, China 1977725159@qq.com

Dongxu Cui

China West Normal University Nanchong Sichuan, China 1532664678@qq.com

Xing Fan

China West Normal University Nanchong Sichuan, China 1943806337@qq.com

Graph signal sampling and reconstruction techniques effectively reduce data dimensions while preserving the information and structure contained in networks. The distortion-free reduction of data dimensions not only aids in reducing the computational complexity of data but also facilitates the extraction of network features and attributes. This reduction in data dimensions provides convenience for the better analysis and processing of complex network data, ultimately enhancing the efficiency of complex network analysis tasks. Currently, the field of graph signal processing primarily consists of two fundamental frameworks: graph signal processing based on the graph Laplacian matrix and graph signal processing based on the graph's adjacency matrix . Although these two frameworks make full use of the topological structure of graph models, they do not consider the relationships between node information within networks, nor do they integrate individual node information with network structural information. Furthermore, they fail to approach the problem of sampling and reconstruction from the perspective of overall network information. This paper takes the network's intrinsic topological structure and the causal relationships between nodes as its starting point. The goal is to reduce network size and increase the effective information of individual nodes to make the entire network contain richer, valuable information. Combining the graph signal research method based on the graph Laplacian matrix, this paper introduces a graph signal sampling and reconstruction algorithm based on causal emergence. The algorithm initially employs a causal emergence algorithm based on spectral methods to obtain the sampled network and subsequently utilizes classical representation learning algorithms to acquire node representations of the sampled network. This approach enables us to obtain graph signal representations that are richer than those of the original network. Subsequently, we apply a reconstruction framework to reconstruct the obtained graph signals. Finally, the experimental results conducted on three real networks demonstrate that the proposed algorithm outperforms existing theories of graph signal sampling and reconstruction. This research contributes valuable insights and methods to the further development of the field of graph signal processing.

Additional Keywords and Phrases: Causal Emergence, Complex Networks, Graph Signals, Sampling and Reconstruction

* The author is corresponding author. Email: 1977725159@qq.com

1 INTRODUCTION

To describe and process the complex network structures within this high-dimensional data, researchers in graph theory, drawing from graph spectral theory [1][2], have extended classical time-frequency signal processing techniques, such as Fourier transforms, sampling and reconstruction, filtering, modulation, translation, and wavelet transforms, to the field of graph signal processing. Nowadays, we face challenges stemming from the continuously expanding data scale and the increasing diversity of data, making the curse of dimensionality an inevitable problem and challenge in data analysis. Consequently, graph signal processing is particularly crucial in the realm of graph research. The sampling and reconstruction theory within graph signal processing technology [4][5] can effectively address these issues. However, the key problem lies in comprehensively considering the characteristics of data and graph signals, identifying crucial signals within network data, and, while preserving distortion-free reconstruction of the original signal, eliminating redundant information from the original data to achieve efficient dimensionality reduction and reconstruction of high-dimensional data.

2021, Erik Hoel and colleagues introduced a novel approach in their paper titled "The Emergence of Informative Higher Scales in Complex Networks." This method involves grouping subgraphs of nodes in a network into macroscopic nodes, reducing the network's size while increasing its effective information (referred to as causal emergence) [6]. Their research revealed that high-information scales are quite prevalent in simulated and real networks across biological, social, informational, and technological domains [6].These information-theoretic metrics can quantitatively identify key nodes in the network that possess richer, effective information, and selecting these key nodes aids in better sampling.

Currently, research in the field of graph signal processing primarily revolves around two fundamental frameworks: graph signal processing based on the graph Laplacian matrix [7] and graph signal processing based on the adjacency matrix of the graph [8][9]. While both of these approaches make good use of the topological structure of graph models, they do not consider the relationships between node information in the network. They also do not combine individual node information with network structural information, nor do they approach the sampling and reconstruction problem from the perspective of overall network information.

This paper, starting from the topological structure of the network itself and the causality between nodes, aims to reduce the network's size and increase the effective information of individual nodes to make the entire network contain richer effective information. Combining the graph signal research methods based on the graph Laplacian matrix, this paper proposes a graph signal sampling and reconstruction algorithm based on causal emergence. This method first observes the entire network from both the perspective of nodes and network structure using causal emergence algorithms, identifying micro nodes that can synthesize macro nodes. It then samples these two types of nodes at different scales. This paper combines the original graph signal with the sampled graph signal features, constructs a low-pass graph signal space, and utilizes a reconstruction framework to reconstruct the graph signal, achieving distortion-free sampling and reconstruction of the original signal.

The main contributions of this paper include:

1. The proposed method addresses the previous limitations in research by considering the causal relationships between nodes and the network.

2. The graph signal sampling and reconstruction theory introduced in this paper is based on quantitative information theory, enabling the quantification of information content and measurement of node information richness.

3. This paper also discusses the analysis of different input graph signals for the same network without altering the network's topological structure, demonstrating that the quality of a graph signal representation is closely related to the graph signal mapping function, in addition to the network's topological structure.

2 BACKGROUND KNOWLEDGE

The concept of graph signals originates from traditional signal processing, but graph signal processing specializes in studying discrete signals defined on irregular topological graphs. The primary research areas include graph Fourier transforms [10][11], graph signal sampling and reconstruction [12], graph signal filtering and modulation, graph signal translation, and wavelet transforms [13], among others.

2.1 Fundamental Concepts of Graph Signals

A graph consists of two core components: node information and edge-weight information that describes the relationships between nodes. For a given graph G(V, E), where V represents the set of nodes and E represents the set of edges, a graph signal is composed of G(V, E) and a mapping function f, which maps nodes to real values. This mapping function represents the node's representation, combining the topological structure of the graph with node data information. This mapping can be represented as:

$$\mathrm{f}: \mathrm{V} \rightarrow \mathbb{R}^{\mathrm{N}*\mathrm{d}} \quad (1)$$

Here, d represents the dimension of the node attribute vector.

2.1.1 Basic Framework of Graph Signal Sampling and Reconstruction

Graph signal sampling and reconstruction are fundamental issues in graph signal processing. The goal of graph signal sampling is to obtain local information from the global signal and reconstruct the original information with either no loss or minimal distortion. The operation of graph signal sampling effectively projects the input signal into a low-frequency signal space[14] and then identifies nodes in the low-frequency space that contain valuable information, thus enabling distortion-free reconstruction of the graph signal.

In Pesenson's research [15], a sufficient condition for uniquely reconstructing band-limited signals was proposed, along with the introduction of important concepts such as unique sets and Λ-set .

A unique set signifies that for a specific node-set, the signal can be uniquely determined as long as certain conditions are met. Thus, if the sampled node set forms a unique set, the signal can be accurately reconstructed. This laid the foundation for ensuring the precise reconstruction of the original signal.

Definition 1 (Unique Set): A subset of nodes $\mathrm{S} \in \mathrm{V}$ is a uniqueness set for a space $\mathrm{PW}_{\omega}(\mathrm{G})$, $\omega > 0$,if for any two signals from $\mathrm{PW}_{\omega}(\mathrm{G})$, the fact that they coincide on S implies that they coincide on V.

The above definition implies that it is sufficient to know the value of an $\omega - \mathrm{bandlimited}$ graph signal only on the uniqueness set S. The uniqueness set is in turn found by finding its complement set S^c , which a Λ- set with $\Lambda = \frac{1}{\omega}$. A $\Lambda - \mathrm{set}$ is defined as follows:

Definition 2 (Λ-Set): A subset of nodes $\mathrm{Q} \epsilon \mathrm{V}$ is a Λ-set if any $\emptyset \in \mathrm{L}_2(\mathrm{Q})$ admits a Poincare inequality with a constant $\Lambda > 0$,i.e.,

$$\left|\left|\emptyset\right|\right| \leq \Lambda||\mathrm{L}\emptyset|| \quad (2)$$

Currently, research in graph signal processing primarily relies on two fundamental frameworks: graph signal processing based on the graph Laplacian matrix [7] and graph signal processing based on the graph's adjacency matrix [8][9]. The method proposed in this paper introduces the theory of causal emergence into these existing frameworks, providing a novel approach to graph signal sampling and reconstruction.

2.2 Causal Emergence Algorithm

Complex systems can be described at different macro and micro scales, meaning the same network can be analyzed at multiple scales[16]. When observing a network at a higher scale, more information can be obtained while reducing the network's size. This phenomenon, where macro descriptions contain more information than micro descriptions, is known as causal emergence[17].

In a network, each node has output weights represented as $W_{\mathrm{i}}^{\mathrm{out}}$, which signifies the output information of node V_i.These output weights are composed of both node V_i and node V_j. For instance, if the edge weight $w_{ij} = 0$, it indicates that there is no connection between node V_i and node V_j. The sum of output weights for each node in the network equals 1, thus, the edge weight w_{ij} in the network represents the next-time-step random transition probability from node V_i to V_j.To find the macro scale in the network with the maximum information content, Erik Hoel and others defined Effective Information (EI) based on information theory. It is used to quantify the determinism of nodes in the network and how determinism is distributed[16]. The definition of Effective Information (EI) is based on two types of uncertainty: the first one is the Shannon entropy of the average output weight vector in the network, capturing the distribution of output weights, and the second one is the average Shannon entropy of the network. Thus, the network's Effective Information can be represented as EI = H(<W>) - <H(W)>, calculated as follows:

$$\mathrm{EI} = -\sum_{\mathrm{i}=0}^{\mathrm{n}}\left(\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{N}}\mathrm{W}_{\mathrm{ij}}\log_2\right)\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{n}}\mathrm{W}_{\mathrm{ij}}\right)\right) - \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(-\sum_{\mathrm{j}=1}^{\mathrm{N}}\mathrm{W}_{\mathrm{ij}}\log_2\mathrm{W}_{\mathrm{ij}}\right) \quad (3)$$

The occurrence of causal emergence implies that the macroscopic description of the network has a higher EI value. Therefore, by quantifying the difference in EI values between the microscopic network and the macroscopic network, the degree of causal emergence can be measured.

$$\textit{casual emergence} = EI(G_M) - EI(G) \quad (4)$$

Currently, there are three main algorithms for computing causal emergence: greedy algorithm, network spectral methods, and gradient descent. Considering the basic topological structure of graph signals, this paper mainly utilizes the network spectral method to calculate the causal emergence algorithm.

2.2.1 Causal Emergence Based on Spectral Method

This study employs a causal emergence algorithm based on the spectral method[19] to extract network information from both micro and macro scales. The specific implementation of this algorithm is as follows: Given a network with N nodes, its adjacency matrix A contains the edge weight information between nodes, and we define it as the micro network (source network). Each entry in the adjacency matrix A represents the weight from node i to node j, and if there is no edge connecting two nodes, the weight is zero. The normalized

adjacency matrix is obtained by dividing each column of A by the corresponding node's degree. The normalized adjacency matrix can be viewed as a transition probability matrix for random walks on graph G. We define this transition probability matrix as P. Next, we perform eigendecomposition on P to obtain the eigenvalue matrix and the eigenvector matrix, where the i-th eigenvalue corresponds to the i-th column of the eigenvector matrix. Using this method, we can obtain a null space of the P matrix and define it as $P^{'} = \{\lambda_i(u_i)|\lambda_i \neq 0\}$[18]. Then, we use cosine similarity to measure the distance between two nodes in the null space and obtain a distance matrix (CS). Finally, we apply the OPTICS[6] algorithm to this distance matrix for clustering analysis and interpret the clustering analysis as a mapping from the micro-scale to a new macro scale, thereby obtaining a macro network (GM) containing N' nodes. The degree of this coarsening, i.e., the degree of causal emergence, depends on the distance threshold used in the clustering algorithm and also on the network's topological structure[8]. Below is the pseudocode for the algorithm:

Algorithm 1 Causal Emergence based on Spectral Analysis

Input : graph G =(V,E) and adjacency matrix A
Output: Microcosmic graph GM
1: W_out←A_norma
2: Calculate eigenvalues(λ) and corresponding eigenvector(U)
3:Used U^'={λ_i (u_i)|λ_i≠0} get U^' Space
4:Calculate U^' cosine similarity Distance CS
5:Microcosmic graph by using OPTICS clustering algorithm to calculate CS
6:return GM

Based on the causal emergence algorithm described above, we can obtain a sampled network. The sampled network is smaller than the source network and contains more valuable information[20]. This new network consists of two types of nodes: macro nodes formed by subgraphs of micro nodes, and the remaining macro nodes corresponding to the original micro nodes, as shown in Figure 2-1:

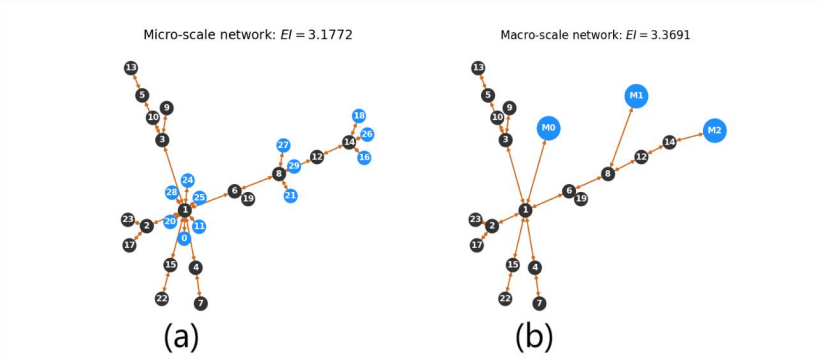


Figure 2-1: Illustration of Causal Emergence Nodes

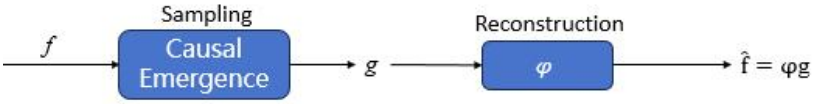
Figure 2-1(a) represents the micrograph, while Figure 2-1(b) depicts the macrograph after causal emergence. From Figure 2-1, we can observe that the effective information (EI) of the micrograph (a) is 3.1772,

while the effective information (EI) of the macrograph (b) is 3.3691, indicating a significant increase in effective information in the macrograph. It is worth noting that Node 1 maintains its original characteristics in both the micro and macro views, but Nodes 0, 11, 20, 24, 25, and 28 are represented as M0 nodes in the macro state.

Next, traditional node representation learning algorithms (such as DeepWalk, Line, Node2Vec, Hope, etc.) can serve as mapping functions for graph signals to obtain network graph signals. Through graph signal processing theory, these signals can be further analyzed in the spectral domain[21].This section provides fundamental knowledge about graph signals and the causal emergence algorithm, laying an important theoretical foundation for the subsequent research work in this paper.

3 GRAPH SIGNAL SAMPLING AND RECONSTRUCTION METHOD BASED ON CAUSAL EMERGENCE

Graph signal sampling and reconstruction are among the core tasks in the field of graph signal processing. The following outlines the basic approach to graph signal sampling and reconstruction:



Consider a network consisting of N nodes, where the node feature values obtained through traditional representation learning constitute the input graph signal f. During the sampling process, we introduce the causal emergence algorithm, which is used to obtain the sampled signal g. In the signal reconstruction phase, we use the interpolation operator φ from the reconstruction framework to transform the sampled signal g into the reconstructed graph signal, resulting in the reconstructed graph signal $\hat{f}$.

3.1 Selection of Sampling Matrix

We first employ the causal emergence algorithm based on the spectral method to sample the network, resulting in a sampled network (G_M) containing M nodes, which includes more effective information. In the sampled network, we generate two types of nodes[22]: macroscopic nodes and microscopic nodes. Macroscopic nodes are composed of several microscopic nodes, while microscopic nodes correspond to individual nodes in the original graph. For example, in Figure 2-1(b), M0 is a macroscopic node in the sampled network, composed of microscopic nodes 0, 11, 20, 24, 25, and 28, while 1 is a microscopic node in the sampled network, preserving the attributes of the original graph.

To construct the sampling matrix S, we create an N-dimensional matrix based on the number of nodes in the sampled network (M) and the number of nodes in the original network (N). In the sampling matrix, the values for microscopic nodes are set to 1, and the values for macroscopic nodes are set to -1. This matrix structure helps capture graph signal information at different scales, as shown below.

$$S[i] = \begin{cases} -1, & i \in Mic \\ 1, & i \in Mac \end{cases} \quad (5)$$

As per the significance of the causal algorithm, when causal emergence occurs in the network, the sampled network will contain more effective information than the microscopic network, encompassing more potential network information[23]. Consequently, the information of a node in the sampled network will be richer than that of a node in the microscopic network. Therefore, we sample these two types of nodes at different scales, and

from an information theory perspective, these two scales represent the positive and negative signals of the graph.

3.2 Building Low-Pass Signal Space

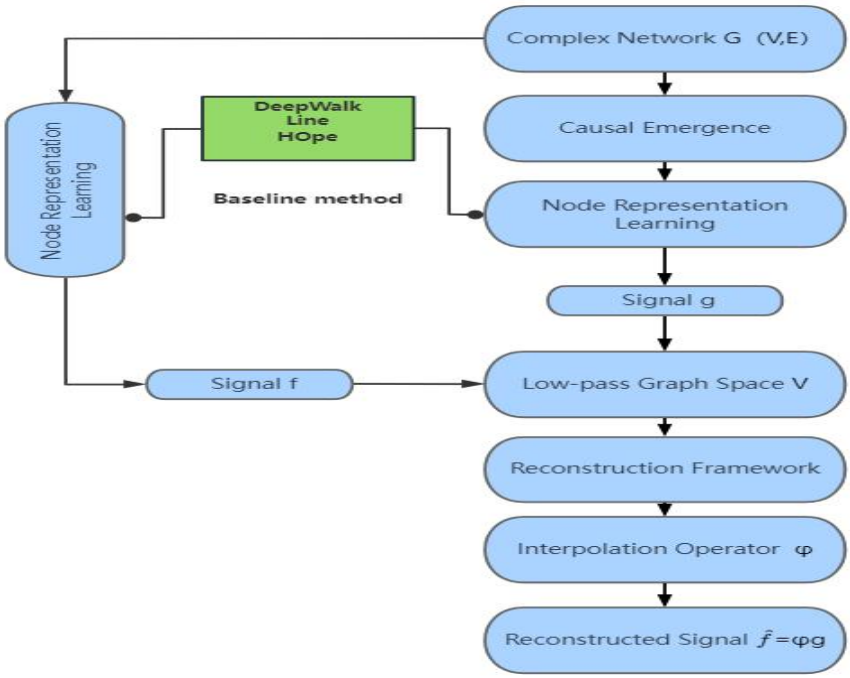
In the sampled new network, we obtain graph signal representations for M nodes, which serve as the sampled graph signal g. The graph signals from the original network are used as the graph signals for the corresponding microscopic nodes in the sampled network. The graph signals of macroscopic nodes remain unchanged. Based on the sampling, we construct the low-pass signal space V for graph signals as follows:

$$\mathrm{V[i]} = \begin{cases} \mathrm{f[i]}, \mathrm{i}\epsilon\mathrm{Mic} \\ \mathrm{g[i]}, \mathrm{i}\epsilon\mathrm{Mac} \end{cases} \quad (6)$$

3.3 Reconstruction Framework

The reconstruction process is the reverse of the sampling process, and the main purpose of the reconstruction framework is to obtain the interpolation operator. First, we need to construct an $N \times N$-dimensional identity matrix I, where N is the number of nodes in the original graph. Next, we solve the reconstruction equation $Q * S * V = I$ to obtain the low-pass operator Q in the low-pass signal space. This low-pass operator Q is used to capture the information representation of the sampled nodes in the low-pass signal space. With the known low-pass operator Q and low-pass space V, we can calculate the interpolation operator φ=VQ, thereby obtaining the interpolation operator φ. We use the interpolation operator φ to extend the low-pass signals to the entire network, achieving signal reconstruction $\hat{f} = \varphi * g$.

3.4 Flowchart of the Causal Emergence-Based Sampling and Reconstruction Algorithm



3.5 Causal Emergence-Based Sampling and Reconstruction Algorithm

Algorithm2 Sampling and Reconstruction Algorithm Based on Causal Emergence

Input: graph G =(V,E)

Output: Reconstructed Signal f

Use Traditional representation learning methods to calculate G's graph signal f

Use Algorithm1 get G_M(Sampled Graph)

Use Traditional representation learning methods to calculate G_M's graph signal g

Create a low-pass space V

Use the reconstruction framework to obtain the low-pass operator Q

Calculate the interpolation operator φ using Reconstruction Framework

Return Reconstructed Signal $\hat{f} = \varphi g$

4 EXPERIMENTAL ANALYSIS

4.1 Dataset

To evaluate the algorithm's performance, this paper selected three classical real network datasets, which differ in terms of network size and node attribute features. These datasets cover networks from various domains: the PPI-ne dataset is a lightweight protein network dataset, the Cora dataset is a standard citation network dataset constructed based on paper citation relationships, and the Wikipedia dataset is a word network extracted from Wikipedia web pages. The table below (Table 4-1) provides detailed information about these three datasets, including the number of nodes (|V|), edges (|E|), and labels (|L|).

Table 4-1. Experimental Datasets

Dataset	|V|	|E|	|L|
PPI-ne	3890	76584	50
Cora	2708	5429	7
Wikipedia	4777	184812	39

4.2 Evaluation Metrics

In this paper, We use the F1 score as the performance evaluation metric, which combines precision and recall and is calculated using the following formula:

$$Micro - F1 = 2 \cdot \frac{Precision_{micro} \cdot Recall_{micro}}{Precision_{micro} + Recall_{micro}} \quad (7)$$

In the node classification task, our goal is to maximize the F1 score because it is a comprehensive performance metric, and a higher score indicates better algorithm performance.

This paper uses the common metrics of Average Precision (AP) and the Area Under the Receiver Operating Characteristic Curve (AUC) for the link prediction task as performance evaluation standards. AP is the average precision for each class, while AUC considers the overall precision. The AUC value is defined as follows:

$$AUC = \frac{N_1 + 0.5N_2}{N} \quad (8)$$

The goal of link prediction is to maximize the AUC value to ensure the accuracy of the algorithm.

4.3 Benchmark Methods and Parameter Settings

To evaluate the performance of our method, we compared it with several classical methods:

(1) DeepWalk: DeepWalk employs a random walk strategy to sample nodes in the network and uses the SkipGram model to learn node representations. In our experiments, the parameters for DeepWalk were set as follows: number of random walk steps: 10, number of random walks per node: 5, random walk window size: 5, number of workers: 5, and iterations: 1.

(2) LINE: LINE utilizes a breadth-first search strategy to generate context nodes and leverages first-order and second-order neighbor information for learning node embeddings. The parameters for LINE in our experiments were set as follows: number of random walk steps: 10, number of random walks per node: 5, negative sampling size: 3, batch size: 20, initial learning rate for SGD: 0.025, and the use of continuous first-order approximation.

(3) HOPE: HOPE introduces higher-order proximities to describe asymmetric transitivity and uses generalized singular value decomposition for network embedding. Our experiments used the KI index as the higher-order proximity metric, with the parameter β set to 0.09.

4.4 Node Classification Task

To evaluate the performance of our algorithm in the node classification task, we selected three real network datasets from Table 4-1 and compared them with the three classical algorithms mentioned above.

As shown in Figure 4-1, the experimental results for node classification on the Cora dataset are displayed. Our proposed method demonstrates improved performance on the Cora dataset. Compared to the Hope algorithm, our method performs equally well, while the improvement is more significant when compared to the Line method. On the Cora dataset, our method achieves the highest improvement of 2.3%, 1.4%, and 6.5% compared to the DeepWalk, Hope, and Line methods, respectively.

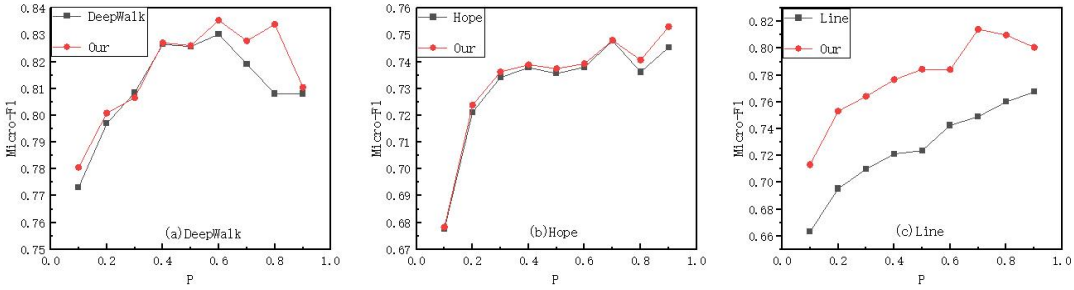


Figure 4-1. Node Classification Performance of Three Baseline Models and Our Method on the Cora Dataset

In Figure 4-2, we present the experimental results of three benchmark methods and our approach for node classification on the PPI-NE dataset. Similarly, it can be observed that our proposed method exhibits performance improvements relative to the three benchmark methods, although the improvements are relatively modest. On the PPI-NE dataset, our method achieves the highest performance improvements of up to 1.4%, 1.6%, and 5.8% compared to the DeepWalk, Hope, and Line methods, respectively.

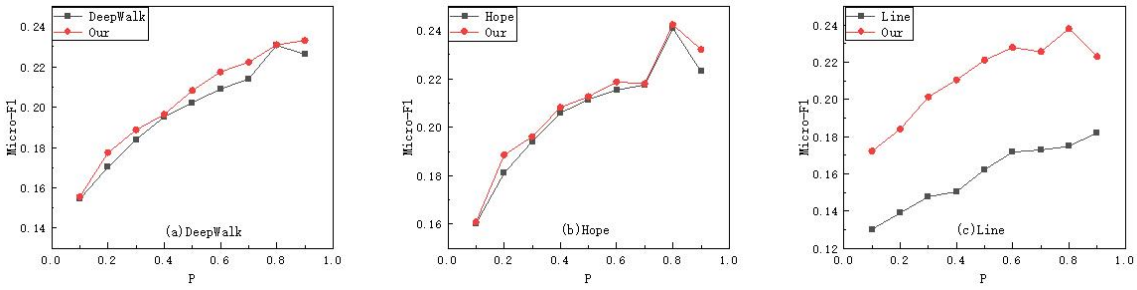


Figure 4-2. Node Classification Comparison of Three Benchmark Models and Our Method on the PPI-NE Dataset

Finally, in Figure 4-3, we present the experimental results of three benchmark methods and our approach to node classification on the Wikipedia dataset. From the graph, it is evident that our approach exhibits a significant performance improvement on the Wikipedia dataset. The Wikipedia dataset is the most outstanding example of our approach's performance. On this dataset, our method outperforms DeepWalk, Hope, and Line by 2.2%, 2.9%, and 5.6%, respectively.

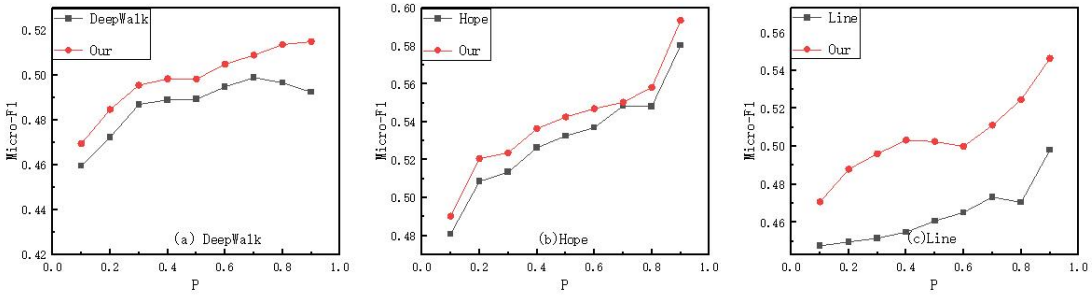


Figure 4-3 illustrates the node classification results of three benchmark models and our approach on the Wikipedia Dataset.

In summary, through experiments on node classification tasks on the three real-world network datasets mentioned above, we can conclude that our proposed method has achieved significant performance improvements across all datasets. Particularly, it shines on the Wikipedia network with a performance boost of up to 5.6%. Although the performance improvement may be relatively lower compared to methods like DeepWalk and Hope, it still represents a notable enhancement.

4.5 Link Prediction Task

To comprehensively assess the algorithm's performance, this paper not only conducted experiments and evaluations in node classification tasks but also performed link prediction tasks on three real-world networks. The Area Under the Curve (AUC) score was employed as the performance metric. The parameter settings for our method remained consistent with those mentioned for the three benchmark algorithms. In each network, 50% of the edges were used as the test set, while the remaining 50% were allocated to the training set. The experimental results are presented in Table 4-2.

Table 4-2. AUC Score Comparison of Four Benchmark Algorithms and Our Method in Link Prediction Tasks

Dataset	Benchmark Methods	Traditional Methods	Our Method
	DeepWalk	0.9171	0.9374
Cora	Hope	0.7197	0.8329
	Line	0.6981	0.7993
	DeepWalk	0.8194	0.9424
PPI-NE	Hope	0.9478	0.9549
	Line	0.5698	0.6802
	DeepWalk	0.7186	0.8946
Wikipedia	Hope	0.5857	0.7269
	Line	0.5927	0.6695
	DeepWalk	0.9171	0.9374
Cora	Hope	0.7197	0.8329
	Line	0.6981	0.7993

From Table 4-2, it is evident that the algorithm proposed in this paper achieves a significant performance improvement compared to traditional methods in link prediction tasks. It is worth noting that there are some variations in the performance improvement of our approach due to differences in the network's topology with benchmark methods. On the Cora dataset, the maximum performance improvement is 11.32%, while the minimum improvement is 2.03%. On the PPI-NE dataset, the maximum improvement is 12.3%, and the minimum is 0.71%. On the Wikipedia dataset, the maximum improvement is 17.6%, and the minimum is 7.68%. Through the analysis of the above data, it is clear that our approach has achieved significant performance improvements in link prediction tasks.

4.6 Analyzing the Performance of Different Graph Signal Representations on the Same Topology

We must take note that, whether in node classification tasks or link prediction tasks, different graph signal representations within the same network exhibit significant differences during the sampling and reconstruction processes. This implies that the quality of graph signals is influenced by multiple factors, not only related to the network's topology but also closely tied to the adopted graph signal mapping function. Therefore, to more accurately capture the intrinsic characteristics of graph signals, we must consider various aspects of information from the node level to the overall network, to find the most suitable graph signal mapping representation. This is crucial.

This paper takes into account the causal relationships between nodes and the overall network during the sampling process, a consideration that is often overlooked in existing algorithms. This comprehensive consideration of causal relationships allows us to gain a deeper understanding and capture both local and global features in the network, thereby improving the quality of graph signal representations. Hence, the approach presented in this paper not only focuses on the network's topology but also emphasizes the critical role of the graph signal mapping function.

5 CONCLUSION

The research presented in this paper demonstrates significant performance improvements in both node classification and link prediction tasks through the use of the causal emergence-based graph signal sampling and reconstruction algorithm. These improvements are attributed to the sampling process based on causal emergence, which effectively enriches the representation of graph signals from the perspectives of individual

node attributes and network structures. This method goes beyond traditional approaches, providing a novel information-theoretic view on graph signal sampling and reconstruction while emphasizing the importance of considering causal relationships between nodes and the overall network structure.

Notably, the proposed method performs exceptionally well in the link prediction task, opening up intriguing avenues for future research. Furthermore, we will continue to explore the potential of causal emergence-based graph signals in other application domains and conduct more comprehensive assessments of the algorithm's overall performance. Given that our experiments involved various types of real-world networks with potentially influential network structures, future studies will include the construction of artificial networks and further experimentation to strengthen our research findings.

REFERENCES

[1] Fan R K C. Spectral graph theory[M].Published for the Conference Board of the mathematical science by the American Mathematical Society,1997.

[2] Cvetkovic D, Rowlinson P, Simic S. An introduction to the theory of graph spectra[J]. Comparative & General Pharmacology,2009,2(6):217-224.

[3] Proakis J G, Manolakis D G. Digital signal processing: principles, algorithms, and applications [J].2006,23(4):392-394.

[4] Shuman D I, Narang S K, Frossard P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine, 2013, 30(3):83–98.

[5] Liu D, Brace C L. Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure[J].IEEE Signal Processing Magazine,2014,31(5):80-90.

[6] Brennan Klein and Erik Hoel. “The emergence of informative higher scales in complex networks”. In: Complexity (in press). url: https://arxiv.org/abs/1907.03902.

[7] Sandryhaila A, Moura J M F. Discrete Signal Processing on Graphs [J].IEEE Transactions on Signal Processing,2013,61(7):1644-1656

[8] Sandryhaila A, Moura J. Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure. IEEE Signal Processing Magazine,2014, 31(5):80–90.

[9] Sandryhaila A, Moura J M. Discrete signal processing on graphs: Frequency analysis. Signal Processing, IEEE Transactions on, 2014, 62(12):3042–3054.

[10] Sardellitti S, Barbarossa S, Lorenzo P D. On the graph Fourier transform for directed graphs[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(6):796-811.

[11] Magoarou L L, Gribonval R, Tremblay N. Approximate fast graph Fourier transforms via multilayer sparse approximations [J]. IEEE Transactions on Signal and Information Processing over Networks, 2018, 4(2): 407-420.

[12] Tanaka Y, Eldar Y C . Generalized sampling on graphs with subspace and smoothness priors[J]. IEEE Transactions on Signal Processing, 2020,68:2272-2286.

[13] Hammond D K, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory[J]. Applied and Computational Harmonic Analysis, 2011, 30(2):129-150.

[14] Pesenson I. Sampling in Paley-Wiener spaces on combinatorial graphs. Transactions of the American Mathematical Society, 2008, 360(10):5603–5627.

[15] I. Pesenson, “Sampling in Paley-Wiener spaces on combinatorial graphs,” Trans. Amer. Math. Soc, vol. 360,no.10,pp.5603–5627, 2008.

[16] OuM, CuiP,Pei J, et al. Asymmetric transitivity preserving graph embedding[C]//Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining2016: 1105-1114

[17] Erik P. Hoel. When the map is better than the territory.[J], Entropy, 2017, 19(5)

[18] Griebenow R, Klein B, Hoel E Finding the right scale of a network: Efficient identification of causal emergence through spectral clustering[J]. 2019.

[19] Finding the right scale of network:Efficient identification of causal emergence in preferential attachment networks through spectral clusting.

[20] Prabhakar Telagarapu, Gulivindala Suresh, J. Venkata Suman, and N. V. Lalitha, "Prabhakar Telagarapu, Gulivindala Suresh, J. Venkata Suman, and N. V. Lalitha," International Journal of Computer Theory and Engineering vol. 4, no. 1, pp. 81-84, 2012.

[21] Song Tiangang, Lyu Zhou, Ding Xinyang, and Wan Yi, "3D Surface Reconstruction Based on Kinect Sensor," International Journal of Computer Theory and Engineering vol. 5, no. 3, pp. 567-573, 2013.

[22] Naveed Ahmed and Imran Junejo, "Using Multiple RGB-D Cameras for 3D Video Acquisition and Spatio-Temporally Coherent 3D Animation Reconstruction," International Journal of Computer Theory and Engineering vol. 6, no. 6, pp. 447-450, 2014.

[23] Vicky Sintunata and Terumasa Aoki, "Color Segmentation Based Depth Adjustment for 3D Model Reconstruction from a Single Input Image," International Journal of Computer Theory and Engineering vol. 8, no. 2, pp. 171-176, 2016.