International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Random Walk in the Age of GNNs: Unveiling Its Continued Relevance and Applications

Aleksandr Timashov*

Graduate Certificate in Artificial Intelligence, Stanford University, Stanford, California, United States,London, United Kingdom

Email: aleks.k.tim@gmail.com

Abstract

This article emphasizes the ongoing importance of Random Walk in improving Graph Neural Networks (GNNs). We illustrate how Random Walk enhances GNNs by offering a deeper structural understanding, better feature learning, and increased efficiency in handling large-scale graphs. The incorporation of Random Walk strategies significantly enhances performance in practical applications like drug discovery and fraud detection. Our results indicate that Random Walk continues to be an essential tool for enhancing the interpretability, scalability, and dynamic modeling of graph-based systems, highlighting its enduring significance in contemporary AI methods.

Keywords: random walk; graph neural networks; feature learning; node embeddings; scalability; drug discovery; fraud detection; node2vec; message passing; social network analysis; interpretability; sampling techniques; explainable AI; real-life applications.

1. Introduction

In recent years, Graph Neural Networks (GNNs) have emerged as a powerful tool in machine learning, revolutionizing how we process and analyze complex, interconnected data structures. These sophisticated models have found applications in various fields, from social network analysis to drug discovery, showcasing their versatility and effectiveness in handling graph-structured information. However, amidst this rapid advancement in neural network architectures, a seemingly simple concept from classical computer science continues to play a crucial role: Random Walk. Random Walk, a process that describes a path consisting of a succession of random steps, has been a fundamental tool in computational methods for decades [1] .Its simplicity belies its power in modeling complex systems and solving intricate problems.

Received: 8/16/2024
Accepted: 10/16/2024
Published: 10/26/2024

----

* Corresponding author.

As we venture into the age of GNNs, Random Walk not only retains its relevance but also finds a new purpose in enhancing and complementing these advanced neural network models.This article aims to explore the intersection between Random Walk and GNNs, unraveling how these two concepts synergize to tackle complex problems in data analysis and prediction. We will delve into their individual characteristics, examine why they work well together, and showcase real-world applications where their combination yields powerful results.

2. What are Random Walk Algorithms?

At its core, a Random Walk is a mathematical process that describes a path starting at a given node and moving to the next one at random following some strategy **R**. Imagine a drunk person stumbling home: each step they take is in a random direction with some probability, partly dependent on their previous steps. Sometimes they can take steps back, as it is written on the image. The steps of this journey are the essence of a Random Walk.

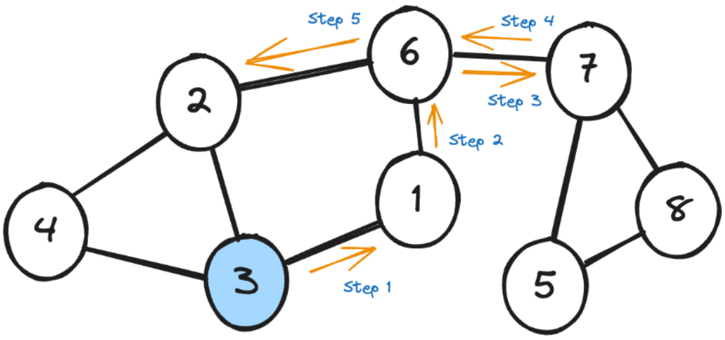

Figure 1: random walk visualization

The concept of Random Walk has a rich history in mathematics and physics, dating back to the early 20th century. It was first formalized by Karl Pearson in 1905, but its roots can be traced to the Brownian motion observed by botanist Robert Brown in 1827. Over time, Random Walk has found applications in various fields, including economics, biology, and computer science.

The Random Walk Approaches were developed to define embeddings of nodes in a graph in such a way that similar nodes would have similar embeddings, and different nodes would have different embeddings. The main idea is that if two nodes appeared on the same random walk, we would consider these nodes to be similar. There are multiple ways to define the similarity of two embeddings. One of the most straightforward ways is a dot product. We should remember to normalize it to keep it in a range between 0 and 1.

$$z_u^T z_v \approx P(u \;\&\; v \text{ both are part of a random walk})$$

Figure 2: representation of a relationship in the context of random walks on a graph

Below are the main steps of a basic Random walk approach:

1. Estimate the probability of visiting node v on a random walk starting from node u.
2. Optimize embeddings of each node that appeared during our simulations to follow the calculations of step 1.

People run some form of stochastic gradient descent to maximize the log-likelihood objective. Below are formal definitions, including the objective.

Given: Graph $G = (V, E)$

Goal: Learn a mapping $f: u \rightarrow R^d$

Objective: $\sum_{u \text{ in } V} \sum_{v \text{ in } N(u)} \log P(v \mid z_u)$

Neighbourhood of node u, following strategy R: $N_R(u)$

Figure 3: graph embedding objective and neighborhood strategy

It is worth mentioning that:

- Each note can appear multiple times on a Random Walk.
- To calculate probability people often use softmax.

A practical way of calculating probability is using softmax: we want to compare the similarity of nodes u and v with the similarity of u and other nodes in a graph.

$$P(v|z_u) = \frac{\exp(z_u^T z_v)}{\sum_{n \text{ in } V} \exp(z_u^T z_n)}$$

Figure 4: conditional probability for node embedding

The challenge with this approach is we need to sum over all nodes in the graph which is computationally expensive. The solution to this challenge is negative sampling: instead of using all nodes, we normalize the equation against some random "negative" samples. In practice, 5-20 samples are used.

In computational fields, Random Walk has been instrumental in solving a wide array of problems:

1. Graph Analysis: Random Walks help in understanding the structure of complex networks, identifying important nodes, and detecting communities.

2. Search Algorithms: PageRank, the algorithm that powered Google's initial success, uses Random Walk principles to rank web pages. This is a key application of Random Walk in modern technology [2].
3. Sampling and Monte Carlo Methods: Random Walks are used to sample from complex probability distributions and solve high-dimensional integrals [3].
4. Financial Modeling: The concept underpins many models in finance, including the famous "random walk hypothesis" in stock market analysis [4].
5. Recommendation Systems: Random Walks on user-item graphs can generate personalized recommendations.

The versatility and simplicity of Random Walk have ensured its continued relevance even as more complex algorithms have emerged.

3. What are GNNs?

Graph Neural Networks (GNNs) represent a class of deep learning methods designed to perform inference on data described by graphs. Unlike traditional neural networks that operate on regular grid-like structures (such as image pixels), GNNs can process data with irregular structures, making them ideal for analyzing complex relational data.

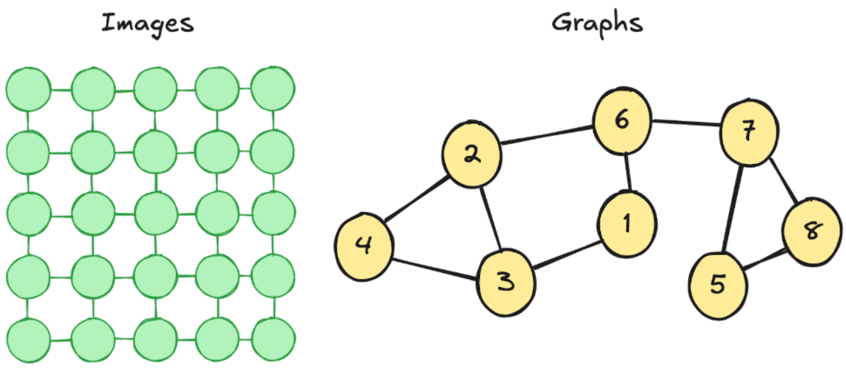

Figure 5: conditional probability for node embedding

At their core, GNNs work by propagating and transforming feature information across the nodes of a graph. Each node aggregates information from all or some of its neighbors, updates its own features, and then passes this information along. This process allows GNNs to capture both local and global structural information of the graph. The GNN layer includes two steps: Message Passing and Aggregation. Below is the example for embedding calculation for the node "1" from the image above.

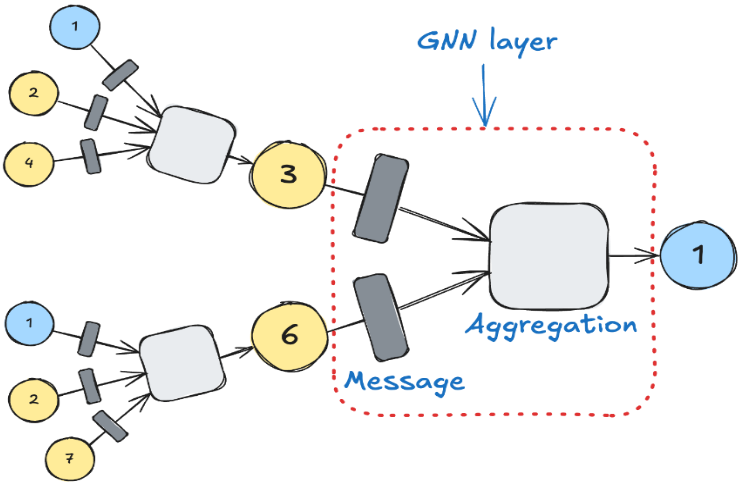

Figure 6: GNN layer with message passing and aggregation

The importance of GNNs lies in their ability to handle graph-structured data, which is ubiquitous in the real world.

The impact of GNNs has been profound in several areas:

1. Social Network Analysis: GNNs can predict links between users, detect communities, and identify influential nodes in social graphs [5].
2. Bioinformatics: In drug discovery and graph generation, GNNs predict protein-protein interactions and molecular properties [6].
3. Recommendation Systems: GNNs can model complex user-item interactions to provide more accurate and personalized recommendations [7].
4. Natural Language Processing: GNNs have been applied to tasks like text classification and machine translation, where they can capture semantic relationships between words or sentences [8].
5. Computer Vision: In tasks like scene graph generation or object tracking, GNNs help to understand the relationships between objects in an image [9].

The ability of GNNs to process and learn from graph-structured data has opened up new possibilities in machine learning, allowing us to tackle problems that were previously challenging or intractable.

4. Why is Random Walk Used Along with GNNs?

Combining Random Walk with Graph Neural Networks represents a powerful synergy between classical graph algorithms and modern deep learning techniques. This combination bridges the gap between traditional methods of graph analysis and the advanced capabilities of neural networks, resulting in more robust and effective models for graph-structured data.

Random Walk complements GNNs in several key ways:

1. Structural Understanding and Interpretability: Random Walks provide intuitive ways to explore and understand graph structure at both local and global levels. Through generated paths, they offer interpretable explanations for node relationships and information flow, complementing the often black-box nature of GNN decisions. Different walking strategies (like node2vec's p,q parameters) enable flexible control over structural exploration.
2. Improved Feature Learning: Random Walk-based embeddings capture sequential node co-occurrences that can serve as rich input features for GNNs or help initialize GNN parameters through pre-training, leading to better representations.
3. Handling of Large-Scale Graphs: For very large graphs, Random Walk-based sampling techniques enable efficient processing by creating meaningful subgraphs or node sequences that GNNs can handle more efficiently.
4. Regularization and Temporal Dynamics: The stochastic nature of Random Walks acts as a natural regularizer for GNNs while also capturing sequential patterns crucial for dynamic graphs, where traditional GNNs might focus primarily on static structure.

The synergy between Random Walk and GNNs proves particularly powerful in real-world graph learning tasks. In node classification, Random Walks enable GNNs to capture the influence of distant nodes, leading to more accurate predictions. For link prediction, Random Walk-based methods complement GNNs by providing a broader view of potential connections across the graph structure.

Moreover, this integration is especially valuable when dealing with complex, irregular graph data. Real-world networks often have heterogeneous structures, varying node degrees, and multi-scale properties. The combination of Random Walk's exploration capabilities with GNNs' learning power creates models that efficiently handle these structural complexities, making this approach particularly effective for applications from social network analysis to molecular structure prediction.

5. Integration: How Random Walk Principles are Incorporated into GNN Architectures

Random Walk principles can be integrated into GNN architectures in multiple ways, each offering distinct advantages. These integrations range from using Random Walks for initial node embeddings to incorporating walking strategies directly into message passing mechanisms. For example, Node2Vec, one of the earlier approaches, uses biased Random Walks to generate node embeddings that capture both local and global graph structures.

Node2Vec performs a series of Random Walks starting from each node in the graph. These walks are biased to balance between breadth-first search (BFS) and depth-first search (DFS) strategies. This balance is controlled by two key parameters:

- Return parameter $\boldsymbol{p}$ controls the likelihood of returning to the node visited on the previous step.
- In-out parameter $\boldsymbol{q}$ determines whether the walk explores the node's close neighborhood (like BFS) or ventures farther into the graph (like DFS).

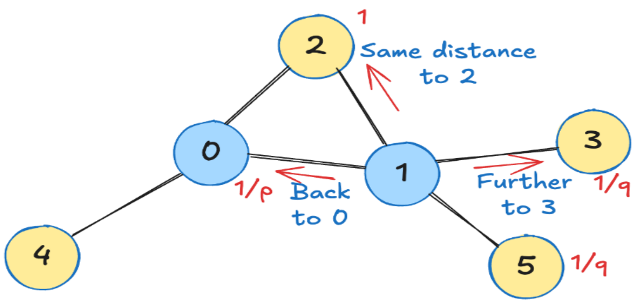

Figure 7: biased random walk strategy in graphs

These sequences serve as input to a Skip-gram model, which learns to predict context nodes within a fixed window along each walk, similar to word2vec in NLP. This optimization process generates node embeddings that preserve both local and global proximity in the graph structure.

These Random Walk-based embeddings can be integrated into GNN architectures in several ways:

- Input Features: The embeddings can serve as initial node features or be concatenated with existing features, providing GNNs with rich, structure-aware input representations.
- Alternative Message-Passing Schemes: Some specialized GNN variants incorporate Random Walk metrics into their attention mechanisms, such as using Random Walk with Restart scores to weight neighbor importance, though this remains a niche approach in GNN architectures.
- Graph Sampling: For large graphs, Random Walk-based sampling strategies enable efficient training by creating representative mini-batches of nodes or subgraphs.
- Structural Analysis: Random Walks can be used to compute various graph properties (e.g., node centrality, community structure) which can be incorporated into GNN layers or loss functions to guide the learning process.

These techniques can be used independently or in combination, adapting to specific task requirements. By incorporating Random Walk-based techniques, GNNs gain a deeper understanding of node relationships and graph structure, enabling better performance on tasks that require understanding both local and global graph properties.

6. Real-Life Applications

The combination of Random Walk and GNNs has found numerous applications across various industries, demonstrating its effectiveness in solving complex real-world problems. Here are some notable examples:

6.1. Biomedicine

In drug discovery and biomedicine, Random Walk-GNN models can be used to predict protein-protein interactions and identify potential drug targets. By representing molecules as graphs and using Random Walks to explore their structure, these models can capture complex chemical properties and interactions, leading to more accurate predictions of drug efficacy and side effects. This application in biomedicine has shown particular promise in accelerating the drug discovery process and improving our understanding of complex biological systems [10].

6.2. Fraud Detection

In financial networks, Random Walk-GNN models excel at detecting fraudulent transactions by analyzing complex transaction patterns. The Random Walk component helps trace money flow paths through the network, while GNNs learn to identify suspicious patterns in these paths. This combination has significantly improved fraud detection accuracy by capturing both local transaction patterns and broader money flow networks [11].

6.3. Social Media Analysis

Social media platforms can leverage Random Walk-GNN models to understand user behavior and information flow in vast networks. These models excel at identifying influential users, detecting coordinated inauthentic behavior, and powering recommendation systems by combining Random Walks' ability to trace information propagation with GNNs' pattern recognition capabilities [12].

6.4. Future Trends

Several trends are shaping the future use of Random Walk-GNN models:

- Explainable AI: There's a growing focus on making GNN models more interpretable, with Random Walks providing traceable paths through the graph that can help explain model decisions and predictions.
- Dynamic Graphs: As more attention is given to analyzing temporal graphs, we can expect to see new Random Walk-GNN architectures designed specifically for evolving network structures.
- Federated Learning: In scenarios where data privacy is crucial, federated Random Walk-GNN models could enable collaborative learning across distributed graph data while preserving data privacy and locality.
- Scalable Architectures: Research is moving towards more efficient Random Walk-GNN implementations that can handle increasingly large and complex graph structures.

7. Conclusion

The enduring relevance of Random Walk in the age of Graph Neural Networks underscores a fundamental truth in the field of artificial intelligence: new advancements often build upon the ideas of foundational concepts, rather than replace them. The synergy between Random Walk and GNNs demonstrates how classical algorithms can enhance modern machine learning approaches, enabling models to capture both fine-grained structures and global patterns. This integration demonstrates that the most powerful solutions often emerge from bridging time-tested principles with cutting-edge techniques.

Key takeaways from our exploration

- Random Walk remains a fundamental tool for understanding complex graph structures, despite its algorithmic simplicity.
- GNNs represent a significant advancement in processing graph-structured data, opening up new possibilities across various fields.
- The integration of Random Walk principles with GNNs creates models that excel at capturing both local and global graph properties.
- This combination drives practical improvements across industries, from accelerating drug discovery to enhancing fraud detection.
- The evolution continues toward more interpretable, dynamic, and privacy-aware Random Walk-GNN approaches.

As we continue to push the boundaries of AI and machine learning, it's clear that the interplay between foundational concepts like Random Walk and advanced techniques like GNNs will remain crucial. This synergy not only improves our ability to solve complex problems but also deepens our understanding of the underlying principles that govern networked systems.

In the ever-evolving landscape of AI, the Random Walk-GNN combination stands as a testament to the power of bridging classical and modern approaches in data science and machine learning.

References

[1]. L. Lovász, "Random Walks on Graphs: A Survey," in Combinatorics, Paul Erdős is Eighty (Volume 2), Keszthely, Hungary, 1993, pp. 1–46.

[2]. S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," Computer Science Department, Stanford University, Stanford, CA 94305, USA, sergey@cs.stanford.edu and page@cs.stanford.edu.

[3]. "Suppressing random walks in Markov chain Monte Carlo using ordered overrelaxation," 21 June 1995.

[4]. E. F. Fama, "Random walks in stock market prices," Financial Analysts Journal, vol. 21, no. 5, pp. 55-59, Sep.-Oct. 1965. Taylor & Francis, Ltd.

[5]. W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph Neural Networks for Social Recommendation," in Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA, pp. 1–11. [Online]. Available: https://doi.org/10.1145/3308558.3313488.

[6]. Z. Zhang, L. Chen, F. Zhong, D. Wang, J. Jiang, S. Zhang, H. Jiang, M. Zheng, and X. Li, "Graph neural network approaches for drug-target interactions," Curr. Opin. Struct. Biol., vol. 73, pp. 102327, Apr. 2022. [Online]. Available: https://doi.org/10.1016/j.sbi.2022.102327.

[7]. C. Cooper, S. H. Lee, Y. Siantos, and R. M. Neal, "Random walks in recommender systems: Exact computation and simulations," Department of Informatics, King's College London, U.K.,

name.surname@kcl.ac.uk; R. M. Neal, Department of Statistics and Department of Computer Science, University of Toronto, Toronto, Ontario, Canada, radford@stattoronto.edu, June 22, 1995.

[8]. G. Zhu, L. Zhang, Y. Jiang, Y. Dang, H. Hou, P. Shen, M. Feng, X. Zhao, Q. Miao, S. A. A. Shah, and M. Bennamoun, "Scene graph generation: A comprehensive survey."

[9]. J. Li, X. Gao, T. Jiang, "Graph networks for multiple object tracking," NELVT, Department of Computer Science, Peking University, China.

[10]. M. H. Al-Rabeah and A. Lakizadeh, "Prediction of drug-drug interaction events using graph neural networks based feature extraction," Scientific Reports, vol. 12, Article number: 15590, 2022.

[11]. Y. Luo and G. Wang, "Fraud detection based on graph neural network," 2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), IEEE.

[12]. A. Menzli, "Graph neural network and some of GNN applications: Everything you need to know," Neptune.ai, 11th September 2023.