# GRAPH NEURAL NETWORK BASED HANDOVER OPTIMIZATION FRAMEWORK

by

Yunus Umeyr Kılıç

B.S., Electrical and Electronics Engineering, Bilkent University, 2018

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Graduate Program in Computere Engineering

Boğaziçi University

2024

# ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who contributed to my thesis. First and foremost, I am immensely grateful to Prof. Tuna Tuğcu for his exceptional mentorship, unwavering support, and patience. His feedback and mentorship made the completion of this thesis possible. Special thanks are also due to the members of NETLAB for their creative insights.

Additionally, I am grateful to Prof. Ali Gökhan Yavuz and Assist. Prof. H. Birkan Yılmaz for their integral roles and contributions to my thesis committee.

Lastly, my heartfelt gratitude goes to my family, friends and colleagues whose contributions have been a cornerstone of my academic pursuit.

# ABSTRACT

## GRAPH NEURAL NETWORK BASED HANDOVER OPTIMIZATION FRAMEWORK

In the evolving landscape of mobile networks, an innovative handover optimization framework for next-generation networks in Open Radio Access Network settings is presented in this thesis. The research, focusing on embedding mobile networks including user equipments and base stations to better capture the network dynamics in handover decisioning, is facilitated through the utilization of Graph Neural Network (GNN) based framework. The core objective is to optimize critical aspects such as load balancing, handover cost, throughput gain, and coverage gain. This framework, called GNN-HOF (Graph Neural Network Based Handover Optimization Framework), is a significant departure from traditional proximity-based methods, leveraging advanced machine learning techniques to better understand and predict network dynamics. The efficacy of the approach is validated through extensive testing in simulated environments as well as real-world urban scenarios in Stuttgart and Monaco using the Simulation of Urban Mobility (SUMO) tool. The results are compelling, demonstrating that the proposed framework consistently outperforms the baseline method across all key metrics.

# ÖZET

## GRAFİK SİNİR AĞLARI TABANLI AKTARIM OPTİMİZASYON SİSTEMİ

Bu tezde, mobil ağların gelişen yapısında, Açık Radyo Erişim Ağı (O-RAN) ortamlarında sonraki nesil ağlar için yenilikçi bir el değiştirme (handover) optimizasyon çerçevesi sunulmaktadır. Grafik Sinir Ağları (GNN) tabanlı bir çerçeve kullanılarak, araştırmada, el değiştirme kararlarında ağ dinamiklerini daha iyi yakalamak için Kullanıcı Ekipmanları (UE) ve Baz İstasyonları (BS) dahil olmak üzere mobil ağları gömme üzerine odaklanılmaktadır. Temel amaç, yük dengeleme, el değiştirme maliyeti, veri akışı kazancı ve kapsama alanı kazancı gibi kritik yönlerin optimize edilmesidir. GNN-HOF (Graph Neural Network Based Handover Optimization Framework - Grafik Sinir Ağları Tabanlı Aktarım Optimizasyon Çerçevesi) ismi verilen bu çerçeve, geleneksel yakınlık tabanlı yöntemlerden önemli bir ayrılık göstermekte, ağ dinamiklerini daha iyi anlamak ve öngörmek için ileri düzey makine öğrenimi tekniklerinden faydalanmaktadır. Yaklaşımın etkinliği, Stuttgart ve Monaco'daki gerçek dünya kentsel senaryolarında yanı sıra simüle edilmiş ortamlarda kapsamlı testlerle doğrulanmıştır ve Şehirsel Hareketlilik Simülasyonu (SUMO) aracı kullanılmıştır. Önerilen çerçevenin tüm anahtar metriklerde temel yöntemi tutarlı bir şekilde aştığı görülmektedir.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SYMBOLS

| | |
|---|---|
| $b$ | Bias term |
| $c_i$ | Number of UEs connected to BS at index i |
| $E$ | Edges of a graph |
| $\hat{E}$ | Energy consumption of handover |
| $E^{\text{bs}}$ | BS to BS edges of a mobile communication graph |
| $E^{\text{ue}}$ | BS to UE edges of a mobile communication graph |
| $G$ | Graph with nodes and edges |
| $h_{u_i}^k$ | Hidden features of the neighbour node $u_i$ at iteration k |
| $J$ | Jain's index |
| $L$ | Latency of handover |
| $P$ | Packet loss of handover |
| $R$ | Resource utilization of handover |
| $T$ | Total throughput of a mobile communication network |
| $u_i$ | Neighbor node at index i |
| $V$ | Nodes of a graph |
| $V^{\text{bs}}$ | BS nodes of a mobile communication graph |
| $V^{\text{ue}}$ | UE nodes of a mobile communication graph |
| $W_{pool}$ | Trainable parameters of the network |
| $x$ | Embedding vector |
| $\mathbf{x_m}$ | Embedding vector of BSs |
| $\mathbf{x_n}$ | Embedding vector of UEs |
| | |
| $\in$ | In symbol |
| $\sigma$ | Activation function |
| $\forall$ | For all symbol |
| $\sum$ | Summation symbol |

# LIST OF ACRONYMS/ABBREVIATIONS

| | |
|---|---|
| 3GPP | 3rd Generation Partnership Project 4G 4th Generation |
| 5G | 5th Generation |
| 6G | 6th Generation |
| AI | Artificial Intelligence |
| BS | Base Station |
| GNN | Graph Neural Network |
| GNN-HOF | Graph Neural Network Based Handover Optimization Framework |
| HO | Handover |
| MoST | Monaco SUMO Traffic |
| NN | Neural Network |
| O-RAN | Open Radio Access Network |
| QoS | Quality of Service |
| RAN | Radio Access Network |
| ReLU | Rectified Linear Unit |
| RIC | RAN Intelligent Controller |
| RSRP | Received Singal Reference Power |
| RSSI | Received Signal Strength Indicator |
| SNN | Siamese Neural Network |
| STOMP | Stuttgart Open Motorway Project |
| SUMO | Simulation of Urban Mobility |
| UE | User Equipment |

# 1. INTRODUCTION

The evolution of mobile networks towards more sophisticated and high-capacity systems like 5G and beyond has necessitated advancements in network management practices [1]. Among these, handover optimization emerges as a critical area, essential for maintaining seamless communication as users move across various base station (BS) coverage areas. Traditional handover mechanisms in mobile networks, which are primarily based on signal strength and proximity, face substantial challenges in the context of modern network demands. High user mobility leads to rapid changes in network conditions, making it difficult for these traditional mechanisms to maintain stable connections, especially at high speeds or in environments with physical obstacles. Additionally, the dense deployment of BSs in urban settings complicates handover decisions due to overlapping coverage areas, leading to potential service degradation and inefficient resource utilization. As users frequently transition between multiple strong signals, traditional methods may fail to optimize network performance, often resulting in unnecessary handovers and increased signaling traffic. To overcome these limitations, modern handover algorithms need to incorporate more dynamic and comprehensive factors such as real-time traffic loads, operational capacities of BSs, and predictive analytics on user movement. These enhancements can help in making the handover process more adaptive and efficient, thereby improving network robustness and user experience in densely populated network environments.

The advent of Open Radio Access Networks (O-RAN) introduces new dimensions to network management, promoting a more open, intelligent, and adaptable approach [2]. O-RAN is an evolving architecture in mobile telecommunications that aims to increase the openness and intelligence of the radio access network. O-RAN facilitates standardization by using open interfaces, allowing multiple vendors to provide interoperable network components. The initiative promotes the use of virtualized network elements and cloud-native software solutions, incorporating advanced machine learning algorithms and data analytics to enhance network management and perfor-

mance [3]. O-RAN aims to improve the scalability of network operations, and enhance the ability of networks to adapt to changing demands dynamically. These capabilities are crucial for managing the complexities of next-generation wireless networks, which are characterized by a high density of devices and demand for data-intensive applications [4]. Thus, O-RAN represents a significant shift towards more cost-effective, flexible, and user-centric network architecture. This paradigm shift underscores the need for innovative handover optimization techniques that can dynamically adjust to the complexities of network and user behaviors. The limitations of existing handover methods, such as their reliance on simplistic criteria and their lack of consideration for the network's holistic state, motivate the exploration of new methodologies that can leverage the vast amounts of data and the dynamic nature of network environments [5].

The current handover algorithms used in mobile networks often miss important context. They fail to account for factors like how much data is currently being used at each base station or the specific needs of applications on a UE. This oversight can lead to poor decisions, like unnecessary handovers that drain battery life, increase network traffic, and potentially cause congestion. These algorithms also struggle with high-speed situations, such as users traveling on fast trains, because they can't quickly adapt or predict where the user will go next. This delay can cause frequent disruptions in service. To address these issues, there's a push for more advanced handover algorithms that take a comprehensive view of network conditions. By incorporating data analytics and machine learning, these new algorithms could adjust handover settings in real time, improving both the stability of the network and the overall user experience by ensuring smoother, more reliable connections.

This research aims to address the aforementioned challenges by proposing a novel handover optimization framework that employs Graph Neural Networks (GNNs) to model and optimize handover decisions within O-RAN architectures. GNNs are a sophisticated type of machine learning model designed specifically to work with data structured as graphs. This makes them highly effective for modeling complex relationships and interactions within networks. By representing mobile networks as graphs and

applying advanced machine learning techniques, this thesis seeks to enhance the network's understanding of its state and user movements, thereby optimizing key performance indicators such as load balancing, handover cost, throughput gain, and coverage improvement.

The integration of machine learning into network management presents a transformative opportunity to transcend traditional optimization barriers. This is particularly crucial as the complex relationships within network systems are challenging to manage using rule-based methods alone. GNNs, in particular, offer a powerful tool for modeling the complex relationships and interactions within mobile networks [6]. This research leverages GNNs to encode the topological and dynamic properties of mobile networks into a computational model, enabling the prediction and optimization of handover decisions in real-time.

The centerpiece of this thesis is the development of the Graph Neural Network Based Handover Optimization Framework (GNN-HOF), a pioneering approach that combines the theoretical underpinnings of graph theory and neural networks with practical insights from mobile network operations. The framework is built on the concept that mobile networks can be modeled as graphs where nodes represent BSs and UEs, and edges represent the connections or potential handovers between these entities. This graphical model allows the application of GNNs, which are adept at capturing the complex dependencies and interactions within the network. The GNN-HOF framework is designed to learn from network data, identify optimal handover opportunities, and adapt to changing network conditions, thereby ensuring optimal quality of service (QoS) and user experience. Moreover, this approach adapts to changing network conditions—such as varying user densities and mobility patterns—to proactively manage the network resources. This adaptability is key to maintaining optimal performance and user experience, thereby reducing the likelihood of dropped calls and service interruptions. Overall, the GNN-HOF framework encapsulates a forward-thinking approach to network management, combining theoretical advancements in machine learning and graph theory with practical network operation strategies to enhance mobile network

efficiency and user satisfaction. This integration not only improve the capabilities of current mobile networks but also sets the groundwork for future advancements in network technology.

In essence, this thesis explores the cutting-edge of network management by introducing the use of GNNs for improving handovers in O-RAN. It presents an analysis that highlights the challenges and shortcomings of current handover methods and proposes a new, data-driven solution. By combining theoretical concepts with practical research, the thesis not only contributes to academic discussions on managing mobile networks but also offers real-world solutions that could greatly improve the effectiveness, dependability, and overall quality of future mobile networks. Through detailed study and application of the GNN-HOF framework, this thesis paves the way for smarter, adaptable network management practices that can keep up with the fast pace of mobile technology advancements.

## 1.1. Contribution

The thesis introduces several significant contributions to the field of mobile communication networks, particularly in handover optimization, as detailed below:

- Machine Learning-Based Model: The thesis proposes a model grounded in machine learning principles, designed with the flexibility to be deployed in diverse geographical regions. This adaptability ensures the model's applicability across various mobile communication networks regardless of the area's characteristics.
- Graph Neural Network (GNN) Utilization: By leveraging a GNN-based architecture, the model excels at integrating the structural complexities of mobile communication networks. This method enables a more nuanced understanding and processing of network dynamics compared to traditional models.
- Holistic Handover Optimization Framework: It innovates by presenting a unified framework capable of simultaneously optimizing load balancing, handover cost, coverage, and overall throughput. This comprehensive approach addresses

multiple aspects of network performance in a cohesive manner.

- Scalability and Real-World Application: The scalability of the framework is evidenced through rigorous testing across different environments, showcasing its effectiveness in scenarios with varying BS and user equipment (UE) densities. This real-world applicability underscores the model's robustness and versatility.
- Comparative Performance Evaluation: When benchmarked against traditional RSRP-based measurement methods, the proposed framework demonstrates superior performance across all evaluated metrics. This enhancement in optimization metrics highlights the significant advancements introduced by the thesis in handover optimization techniques.

These contributions collectively mark a substantial advancement in optimizing mobile communication networks, offering a scalable, versatile, and effective solution for enhancing network performance and user experience.

# 2. LITERATURE SURVEY

## 2.1. Handover Optimization

Initial research into mobile networks focused on establishing reliable communication channels, optimizing for coverage and capacity with minimal concern for the intelligence in network management [7]. With the advent of 4G and subsequent 5G technologies, the focus shifted towards intelligent network management, emphasizing efficiency, and reliability [8].

The integration of O-RAN has transformed the telecommunications industry by introducing a network architecture that is more adaptable, intelligent, and open. This enables better interoperability and innovation in network management [9]. The literature discusses how O-RAN's architecture facilitates advanced optimization techniques, including those based on artificial intelligence (AI) and machine learning, to improve handover processes, load balancing, and overall network performance [10]. In O-RAN architectures, AI models can enhance handover management by efficiently predicting and controlling events, thereby minimizing dropped calls and enhancing the user experience through pattern analysis and movement prediction [10]. Additionally, machine learning algorithms can analyze network traffic in real time, facilitating a dynamic redistribution of network resources to optimize load and prevent bottlenecks. AI also supports continuous monitoring and proactive network performance management, dynamically adapting configurations based on real-time data analysis. However, despite the promise of AI solutions, their scalability across a complete network is a significant challenge due to the fluctuating network conditions and the inherently dynamic nature of network traffic. Moreover, these methods often lack a holistic approach, requiring separate models for each optimization metric in handover processes.

Early handover mechanisms that relied on received signal strength indicator (RSSI) based algorithms were well-suited to less complex environments where signal

strength was a reliable indicator of the best network connection. However, these algorithms are increasingly inadequate in managing the demands of modern, densely populated urban areas for several reasons [11]. RSSI-based handover decisions are primarily influenced by signal strength and proximity to a cell tower. This method is straightforward but becomes problematic in urban settings due to the high density of structures like buildings and other physical obstacles. These obstructions not only weaken signal strength unpredictably as they block and reflect signals but also create a complex propagation environment where the signal can fluctuate significantly over small areas. Moreover, RSSI-based algorithms do not account for other crucial factors such as network load and user mobility patterns. In urban areas, cellular networks can become highly congested, with numerous devices attempting to connect to the same cell tower. An RSSI-based system might direct a device to a tower that, while close and showing a strong signal, is already operating at or near capacity. This oversight can lead to poor service quality, as the network is unable to adequately handle the additional load. Additionally, the simple nature of RSSI-based handover does not effectively address the high velocity of urban users, particularly those traveling in vehicles. Fast-moving users may experience frequent, unnecessary handovers because the signal strength may rapidly vary along their route, leading to a degradation in service quality and an increase in signal dropout rates. The shortcomings of RSSI-based handover mechanisms highlight the need for more sophisticated algorithms that consider multiple variables, including signal quality, network capacity, user density, and mobility patterns. Recent studies have explored various optimization criteria, including load balancing, handover cost, and throughput, to improve the handover process, often leveraging machine learning techniques [12]. The use of machine learning, including GNNs, to understand and predict network dynamics, allowing for real-time optimization of handover decisions, marks a significant advancement over proximity-based methods [13].

GNN has emerged as a powerful tool for learning representations on graphs, capturing the relationships and interactions between nodes (which can represent UEs, BSs, etc.) in a network [14]. The application of GNNs in network settings, especially in predicting and optimizing network dynamics, has attracted considerable interest,

demonstrating the potential to significantly outperform traditional models [15]. The evolution from classic management of mobile networks to leveraging sophisticated machine learning methods has led to the creation of the GNN-HOF strategy for improving handover processes within O-RAN frameworks. By integrating GNNs, the proposed framework addresses critical challenges in mobile networks, such as load balancing, handover cost and coverage.

Most existing approaches focus on optimizing a singular aspect of the handover process, such as minimizing handover failure or enhancing signal strength [16]. The challenge intensifies when attempting to balance multiple objectives, which often have conflicting requirements. For instance, optimizing for minimal handover cost may adversely affect throughput gain or coverage [17]. The literature reveals a notable scarcity of frameworks that holistically address the multifaceted objectives of handover optimization. This includes a unified approach that simultaneously optimizes for load balancing, handover cost, throughput gain, and coverage gain in a cohesive manner [18]. A self-optimizing technique aimed at enhancing load balancing in heterogeneous wireless networks through the use of big data analytics is proposed in [19]. This method not only focuses on improving load balancing but also seeks to reduce handover costs and improve network coverage by adaptively selecting wireless access nodes. The approach promises to boost network performance and user satisfaction by intelligently redistributing network load and minimizing handover costs, which contributes to more stable and consistent connectivity. However, despite its comprehensive optimization capabilities, this technique may struggle with the complexities of rapidly changing network conditions and high user mobility in densely populated environments. Additionally, the reliance on real-time data analytics introduces significant data processing costs, potentially limiting its feasibility for large-scale deployments.

Although machine learning techniques have been applied to various aspects of network optimization, their potential to integrate and simultaneously optimize multiple handover metrics has not been fully explored or realized. This indicates a significant gap in leveraging advanced machine learning, particularly GNNs, for comprehensive

handover optimization [20]. The absence of a comprehensive method for handover optimization underscores the innovation and potential impact of the GNN-HOF framework introduced in this thesis. By leveraging GNNs, the GNN-HOF framework represents a pioneering approach in addressing the complexities of optimizing multiple handover metrics simultaneously. This not only bridges the identified research gap but also sets a new benchmark for future studies in mobile network optimization within the context of O-RAN.

Several studies, including one that proposes a comprehensive strategy for optimizing handover processes by adjusting multiple parameters, have significantly improved network performance indicators, such as the signal strength for users at the edge of cells and how evenly network traffic is distributed [20], [19]. This particular research employs sophisticated data analysis methods, like XGBoost and random forest algorithms, to tackle the complexities involved in managing handovers between different frequency bands efficiently, using techniques such as simulated annealing to address the challenges posed by the optimization problem's non-linear nature. However, the practical application and scalability of these data-driven approaches in ever-changing network conditions pose ongoing challenges. The method discussed in [19] also faces challenges in scalability as mentioned earlier.

Introducing the GNN-HOF framework for improving handovers in O-RAN networks marks a major step forward in the field. It uniquely combines the strengths of graph neural networks with advanced optimization algorithms to address the previously unmet challenge of simultaneously optimizing key handover metrics. This approach promises not only enhanced network performance but also a more intelligent and adaptive management of network resources.

The literature shows a major change in how mobile networks are managed, moving from simple signal-based methods to advanced machine learning techniques, especially with O-RAN's introduction. This change is due to the growing complexity and needs of today's networks, which require more sophisticated solutions for the best performance.

The GNN-HOF framework stands out as an innovative solution, providing a complete method for improving handovers. It uses Graph Neural Networks to effectively tackle the complex challenges of today's network environments, optimizing important aspects like load balancing, handover cost, throughput, and coverage all at once. This not only addresses a big gap in current research but also opens doors for new advancements in managing mobile networks.

# 3. BACKGROUND

## 3.1. Mobility Management

In wireless networks, mobility is defined as the capability of the network to facilitate a seamless connection of the UEs [21]. A seamless mobility requires a connection management in wireless network. Connection management is the set of processes that provide UE attachment to cell with an uninterrupted service [22]. Once the connection management is available in a wireless network, a UE has the capability to move from one cell to another without any connection interruption in the ideal case. Mobility is one of the most important fundamental features that distinguish wireless networks from wired networks [23].

The basis of the connection management is Handover. Handover in wireless communication networks is a pivotal mechanism that ensures continuous service and connectivity for UEs as they traverse different cells or network areas. This process is integral to maintaining the QoS and seamless user experience in wireless networks, adapting to the dynamic nature of user mobility and varying network conditions [24]. As the UE moves and the received signal strength varies, dropping below an acceptable level or the RSSI threshold triggers a handover process from the current serving BS to a target BS, ensuring uninterrupted connection and seamless mobility, contingent on the network's support for such functions [25]. This constitutes the fundamental approach for managing handovers, which does not account for comprehensive network conditions such as the BSs' load, the associated costs of handover, and coverage considerations.

In the evolution towards future mobile networks, the adoption of mm-wave frequencies and the proliferation of densely packed network infrastructures introduce increased complexity for mobility managements. The baseline greedy approach is not capable of handling such complexities. An efficient handover management is expected to address the challenges posed by the dense deployment, 5G mm-waves and larger

networks. Therefore, handover approaches play an essential role in future mobile communication networks [21].

## 3.2. O-RAN Architecture

The O-RAN architecture, in synergy with the 3rd Generation Partnership Project (3GPP) standards, introduces a new era in mobile communication, especially regarding handover mechanisms in 5G networks and beyond. O-RAN revolutionizes traditional Radio Access Network (RAN) systems by introducing open interfaces and modular network components, enhancing the efficiency, reliability, and intelligence of handovers [4]. This innovative approach promotes a shift from proprietary systems, fostering a more dynamic and adaptable handover process.

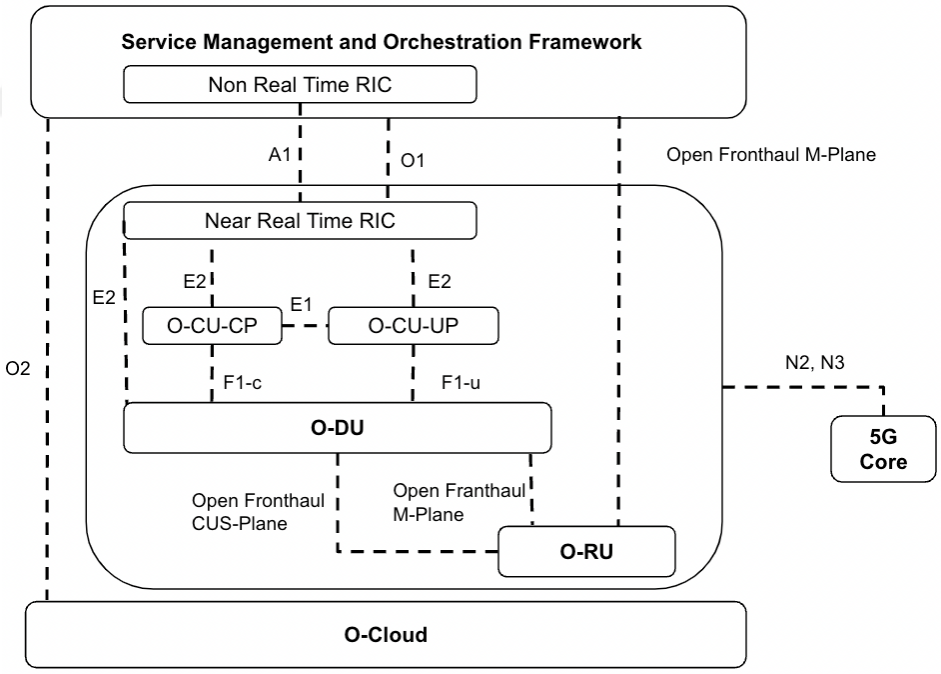



Figure 3.1. O-RAN architecture with RIC and E2 interface.

As seen in the Figure 3.1, a critical element of O-RAN is the RAN Intelligent Controller (RIC), which significantly optimizes handover procedures [2]. With its near-real-time capabilities, the RIC makes rapid decisions based on extensive data inputs, such as user mobility patterns, network load, and signal quality, thereby ensuring

efficient handovers, reducing latency, and improving overall network performance [4]. The standardization of interfaces, like the E2 interface, underscores the interoperability between diverse network elements, ensuring a seamless user experience across various network infrastructures [2].

Moreover, O-RAN leverages advanced algorithms and machine learning techniques within the RIC to predict user movements and proactively prepare the network for handovers. This predictive approach not only enhances the handover success rate but also optimizes network resources, leading to more efficient spectrum use and reduced power consumption [4]. Despite these advancements, the open nature of O-RAN necessitates rigorous security protocols to ensure safe and reliable handovers, with 3GPP standards playing a crucial role in defining these measures [4].

The integration of O-RAN within 3GPP networks marks a significant step forward in handover mechanisms, providing enhanced flexibility, improved interoperability, and intelligent mobility management. As the telecommunications industry progresses towards more open and software-driven infrastructures, O-RAN emerges as a pivotal contributor to the efficient and seamless handover processes in 5G and future mobile network generations [4], [2].

### 3.3. GNNs

In the realm of machine learning, GNNs have become a transformative force, especially in analyzing data with a graph-based structure. This innovation holds particular significance for 5G network environments, characterized by their complex topologies and the dynamic interplay among different network components. GraphSAGE, a notable variant within GNNs known for its Graph Sample and Aggregate approach, has risen to prominence due to its efficiency in producing node embeddings in graphs. These embeddings are pivotal for critical operations such as handover decisions within 5G networks, facilitating smoother transitions and enhancing network reliability.

GNNs innovate beyond conventional neural networks by tailoring their approach to data organized as graphs, capturing the essence of both individual node attributes and the overarching graph structure. This capability is invaluable in the 5G context, enabling the depiction of intricate connections and interactions among network elements like base stations, routers, and end-user devices. By employing GNNs for handover decisions, the technology can decipher the network's graph structure to anticipate issues and refine handover strategies, thereby minimizing connection failures and elevating network efficiency, particularly in densely populated zones.

GraphSAGE distinguishes itself within the GNN landscape through its method of generating node embeddings. It eschews the need for analyzing the entire graph, instead focusing on sampling and aggregating data from a node's immediate vicinity. This feature is especially beneficial for the vast and ever-evolving 5G networks, where utilizing whole-graph data is not feasible [26].

The essence of GraphSAGE lies in its ability to craft embeddings by leveraging information from a node's local environment, enabling the production of scalable and effective embeddings for extensive 5G networks. GraphSAGE is designed to generate node embeddings efficiently in large-scale graphs. Unlike earlier approaches that required information from the entire graph to generate embeddings, GraphSAGE samples a fixed-size neighborhood around each node. This allows it to scale to graphs with millions of nodes and edges by reducing the computational cost. GraphSAGE can generalize to unseen nodes. This is crucial for dynamic networks where new nodes are constantly being added. Traditional transductive methods can only generate embeddings for nodes seen during training, but GraphSAGE's inductive framework allows it to generate embeddings for new nodes based on their features and the local neighborhood structure. This is crucial for making precise, real-time decisions during handover procedures, ensuring uninterrupted service for users on the move [27, 28].

In practical applications, such as 5G handover decision-making, GraphSAGE's technique allows for the embedding of network nodes based on local topological and

traffic data. These embeddings empower handover algorithms to identify the most suitable target node, optimizing the connectivity experience for users. This is particularly useful in scenarios with rapidly changing network conditions, like urban settings or large-scale events, making the handover process more adaptable and attuned to the network's current state, thus improving service quality and user satisfaction [29].

In short, Graph Neural Networks, with GraphSAGE at the forefront, mark a significant leap forward in network embedding technology, especially for 5G handover decision-making applications. By harnessing these advancements, 5G networks are set to benefit from more intelligent and efficient handover mechanisms, ensuring superior connectivity in our increasingly digital and mobile society. As 5G technology progresses, the integration of GNNs and methodologies like GraphSAGE into network management and optimization practices is expected to play a pivotal role.

# 4. GNN BASED HANDOVER OPTIMIZATION MECHANISM

In this chapter, the system model is explained in details. Section 4.1 provides a concise overview of the entire system. Section 4.2 explores the necessary graph representations. A comprehensive explanation of the overall architecture is provided in Section 4.3. Finally, the chapter concludes by introducing a machine learning-based algorithm, presenting it as a competitive alternative to the graph approach, to validate the effectiveness of the proposed algorithm.

## 4.1. System Model

In this paper, we reframe mobility management as a graph link prediction task by transforming the O-RAN network architecture into a graph-based representation. Within this framework, UEs and BSs are conceptualized as nodes, with their interconnections represented as edges. This graph-based abstraction of the network is illustrated in Figure 4.1. The link between a UE and a BS is determined by the RSRP measurements, whereas links among BSs are established based on the Euclidean distance between them, as depicted in the figure. These links and nodes collectively form the network graph for the proposed system model.

We have developed a model aimed at predicting the emergence of new links between UEs and BSs at each time instance $t$. Detailed discussions on graph abstraction and the comprehensive system architecture are presented in the subsequent sections. Predictions of new links are based on the current state of the network. The existing connections between UEs and BSs constitute the network's current status, and thus serve as the primary input for subsequent link decisions. This approach allows for the prediction of future link formations based on historical connectivity data and real-time network conditions, enhancing the robustness and efficiency of the mobility management process.

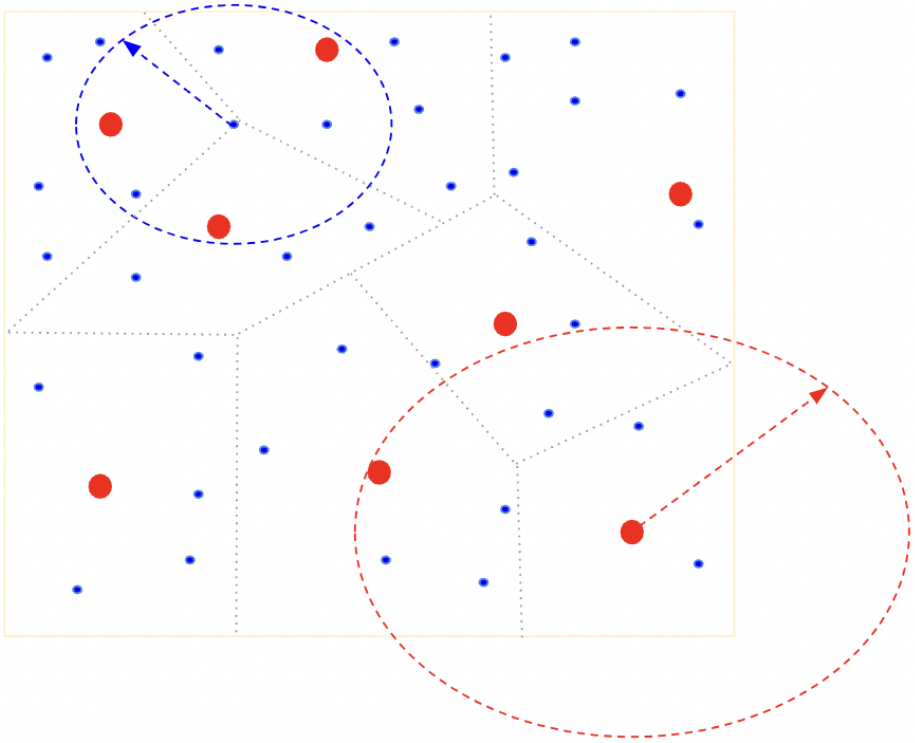


Figure 4.1. Mobile communication network graph abstraction.

## 4.2. O-RAN Network Graph Abstraction

A network consisting of $M$ BSs and $N$ UEs is abstracted as a graph $G = (V, E)$. The nodes are represented as $V = V^{\text{bs}} \cup V^{\text{ue}}$ where $V^{\text{bs}}$ are BS nodes and $V^{\text{ue}}$ are UE nodes. The edges are $E = E^{\text{bs}} \cup E^{\text{ue}}$ where $E^{\text{bs}}$ are the virtual links between the BSs and $E^{\text{ue}}$ are the connection links from the BSs to UEs. Virtual edges are defined based on the Euclidean distances between the BSs. If the distance between two BSs is below the maximum allowed distance, an undirected edge between those BSs is created.

This graph representation of the O-RAN network is illustrated in Figure 4.2, where BS nodes are connected with undirectional edges and UE nodes are connected to BS nodes with directional edges.

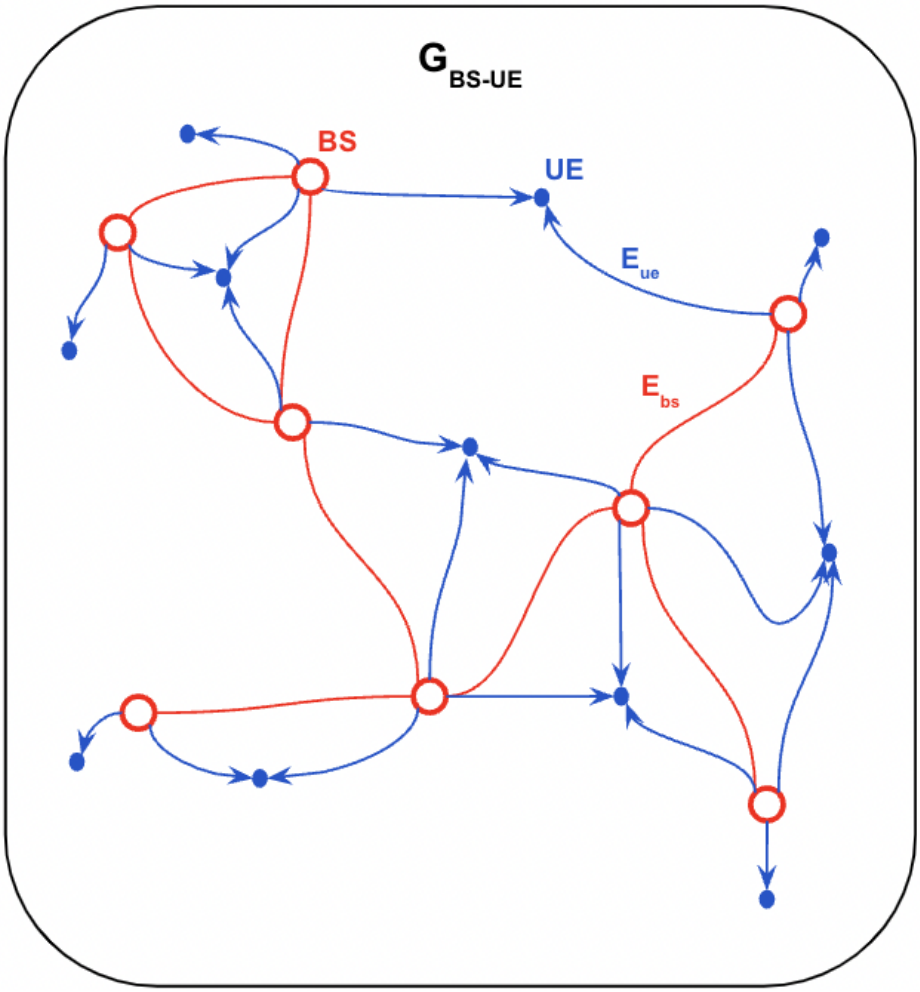



Figure 4.2. Graph representation of mobile network.

As illustrated in Figure 4.2, the diagram showcases two distinct graphs. The first graph is an undirected graph, comprising solely of BSs. In contrast, the second graph is directed and incorporates both BSs and UEs, with edges directed from BSs to UEs. This directional aspect signifies that the embeddings of the BSs are independent from those of the UEs.

## 4.3. GNN-HOG Architecture

The GNN-HOG framework comprises two principal components: the embedding process and the computation of the link confidence score. For the embedding portion, the framework utilizes GraphSAGE, a GNN-based embedding technique. The computation of the link confidence score is achieved using a a simple two layer Neural Net-

work(NN). Additionally, during the training phase, the framework incorporates several optimization metrics: handover cost, load balancing, coverage, and total throughput. The fundamental stages of the architecture are illustrated in Figure 4.3.

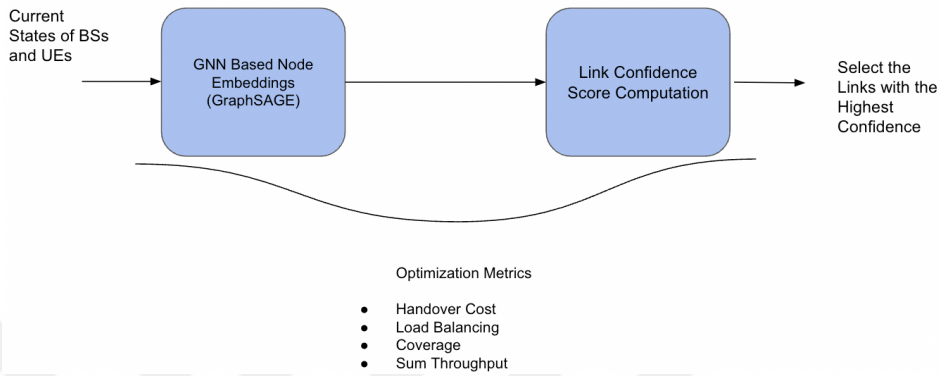



Figure 4.3. GNN-HOG framework.

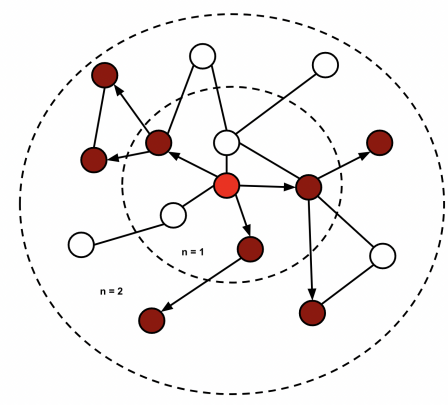
### 4.3.1. Graph Node Embedding

For the node embedding of the communication network, we utilize the inductive representation learning framework, GraphSAGE, as outlined in [26]. This approach enables the construction of two distinct node embeddings to accommodate the representations of both BSs and UEs. The embeddings of BSs are static, provided their locations do not changes. Conversely, the embeddings for UEs are dynamic, undergoing updates at each time instance $t$.

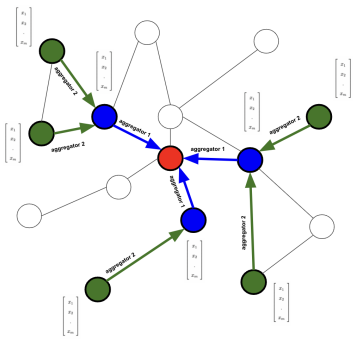
As depicted in Figure 4.4, feature aggregation is employed at every layer of the GNN utilizing an aggregation function [26]. Specifically, the max pooling aggregator, as utilized in the original GraphSAGE paper, is chosen [26]. Max pooling aggregation is defined as

$$AGGREGATE_{(k)}^{\text{max-pool}} = max(\{\sigma(W_{pool} h_{u_i}^k + b),\ \forall u_i \in N(v)\}) \tag{4.1}$$

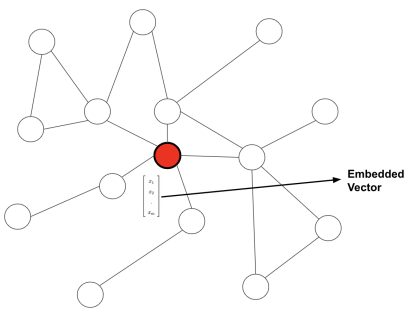
where $W_{pool}$ is the trainable parameters of the network, $\sigma$ is the activation function and $h_{u_i}^k$ is the hidden features of the neighbour node $u_i$ at iteration k.




(a)



(b)




(c)

Figure 4.4. GraphSAGE visual illustration where neighborhood sampling, feature information aggregation and graph prediction are visualized in (a), (b) and (c), respectively.

For the initial iteration, an input feature set corresponding to each node is fed into the network. The input features for BSs are generated through the RSRP measurements from each BS to the current BS. These input features form a vector of length

$M-1$, where $M$ represents the total number of BSs. Similarly, for UEs, the input features are derived from the RSRP measurements from each BS to the current UE, resulting in a vector with a length of $M$.

### 4.3.2. Link Confidence Score

A two-layer neural network architecture is designed for the task of link prediction between two node embeddings. The model initiates with the input layer, where two node embeddings are concatenated, serving as the input vector. The input vector $\mathbf{x}$ to the neural network is obtained by concatenating these two embeddings

$$x = [x_m; x_n] \tag{4.2}$$

where $\mathbf{x_m}$ and $\mathbf{x_n}$ are the final embeddings of the BS and the UE, respectively. Subsequently, the concatenated vector is processed through a hidden layer with a Rectified Linear Unit (ReLU) activation function, enhancing the model's ability to capture non-linear relationships between the node features. The final layer employs a sigmoid activation function to produce a probability score between 0 and 1, indicating the likelihood of a link existing between the two nodes. This two-layer neural network is optimized during training through a loss function appropriate for binary classification tasks, specifically binary cross-entropy, ensuring that the model learns to accurately predict the presence or absence of links between nodes. The link exhibiting the highest confidence score is identified and selected as the predicted connection between the UE and the BS.

### 4.3.3. Optimization Measures

The primary objective of the GNN-HOF framework is to enhance network performance through the optimization measures, namely handover cost, load balancing, coverage, and total throughput. These measures are defined as follows:

- Handover Cost: Computing the handover cost in a mobile communication network involves considering several factors and metrics, as handover procedures can

significantly impact the network's performance, especially in terms of latency and resource utilization. To calculate the handover cost, typically the following components are to be considered: latency, resource utilization, packet loss and energy consumption. A general formula for handover cost computation is defined as

$$Handover\ Cost = w_1 \cdot L + w_2 \cdot R + w_3 \cdot P + w_4 \cdot \hat{E} \tag{4.3}$$

where $\mathbf{L}$ is the latency, $\mathbf{R}$ is the resource utilization, $\mathbf{P}$ is the packet loss and $\hat{\mathbf{E}}$ is the energy consumption. $\mathbf{w_1}$, $\mathbf{w_2}$, $\mathbf{w_3}$, $\mathbf{w_4}$ are the weights assigned to each factor based on their contribution to the cost.

- Load Balancing: Ensuring an equal distribution of resources across BSs is pivotal for optimizing network performance and enhancing user experience. There are various metrics for equal distribution of resources [30]. Jain's index is employed to compute fairness of the load distribution over the BSs. It is defined as

$$J = \frac{(\sum_{i=1}^{m} c_i)^2}{N \cdot \sum_{i=1}^{m} c_i^2} \tag{4.4}$$

  where $\mathbf{c_i}$ is the number of UEs connected to corresponding BS.

- Coverage: Depending on the data rate of the UEs, they could be delineated as either cell-centric or cell-edge. A UE is categorized as cell-edge if its data rate falls below a predetermined threshold. Typically, this threshold is set at the 5th percentile of data rates for all UEs within the network, representing the network's coverage boundary [25]. Enhancing the data rate for cell-edge users not only improves the overall network coverage but also minimizes the occurrence of coverage gaps.

- Total Throughput: Total throughput of a mobile communication network can be defined as the total data rate provided by the network. It is calculated as

$$T = \sum_{i=o}^{M-1} \sum_{i=o}^{N-1} \frac{c(v_i, v_j)}{c_i} \tag{4.5}$$

  where $\mathbf{v_i}$ and $\mathbf{v_j}$ are the connected BS and UE respectively. $\mathbf{c_i}$ is the number of UEs connected to corresponding BS.

### 4.4. ML Based Rival Algorithm

A modified Siamese Neural Network (SNN) like ML model is employed as a machine learning-based alternative to the graph approach, serving as a competitive neural network algorithm. SNN is a specialized architecture designed to learn how to differentiate between pairs of inputs by mapping them into a feature space where the distance between comparable items reflects their similarity [31]. SNNs feature twin networks with shared weights that process two inputs in parallel, enabling the model to learn rich representations for similarity assessment. The SNN architecture has since been a cornerstone in developing models that require understanding of similarity or relationships between data points [32].

An SNN consists of two identical subnetworks, which means they have the same architecture and share the same parameters and weights. Each subnetwork processes one of the two input vectors. The outputs of these subnetworks are then combined and fed into the output layer to compute the similarity measure. Inputs are the BSs' and UEs' feature vectors that are the same features being used in GNN-HOF method. The final output is the similarity score between the BS and UE. Therefore, the primary objective of employing an SNN in our study is to highlight the distinctions between the GNN-HOF—a graph-based methodology adept at interpreting the structure of the mobility network—and a neural network model that operates independently of graph-based techniques.

# 5. EXPERIMENTS AND RESULTS

To explore the performance and reliability of the proposed methodology in wireless networks, extensive experiments are conducted within both controlled simulated environments and realistic urban contexts using the Simulation of Urban MObility (SUMO) platform. This comprehensive approach provides a dual perspective on network behavior by conducting simulations against real-world conditions, ensuring that findings are robust and applicable across various settings.

In all experiments, the performance of the proposed method is evaluated against a standard baseline, which in this case is the RSRP-based handover decision algorithm. This method, commonly utilized in mobile telecommunications, determines the optimal timing for a handover from one BS to another based on the RSRP. RSRP measures the power level of a BS's reference signals as received by a UE and is crucial for assessing the quality of the signal connection between the UE and the BS. Typically, the algorithm triggers a handover when the RSRP of the serving cell drops below a predefined threshold while the RSRP of a neighboring BS surpasses this threshold, indicating a stronger signal. The relative improvement in performance over this baseline method is quantified as our gain in each metric, demonstrating the efficacy of the proposed approach under various conditions.

Further analyses focus on scalability and adaptability of the proposed method, examining how effectively it integrates into existing network infrastructure with minimal disruption. The experiments also test the algorithm's response to sudden changes in network density and user mobility patterns, which are critical factors in urban environments. By simulating scenarios where network load and user distribution vary widely, we evaluate the resilience of our approach during peak traffic conditions and at different times of the day. These additional experiments are essential to understand the practical implications of deploying our method in live networks, providing valuable insights into its performance under stress and its potential to enhance user experience

significantly by reducing dropped connections and improving data throughput.

## 5.1. Simulations

A simulation environment that consists of BSs and UEs are created with configurable parameters. The following parameters are set for the simulation environment:

Table 5.1. Simulation parameters.

| Parameter | Value |
|---|---|
| Area | 3x3 $km^2$ |
| Average UE Speed | 30 KPH |
| Cell Transmit Power | 33 dBm |
| Channel Carrier Frequency | 30GHz |

RSRP based baseline method and the proposed GNN-HOF method are compared in the same simulation settings. BSs and UEs are located in the simulation network area. Different scenarios with varying number of BSs and UEs for both methods are conducted for comparisons. An area of 3 km x 3 km square is covered in simulation. As shown in Figure 5.1, simulation is conducted with BSs and UEs while the UEs are moving and both methods are run to decide on handover connection. There are four different movement patterns for UEs with randomized noise for each device separately. So, a random noise is added to those constant movement patterns.

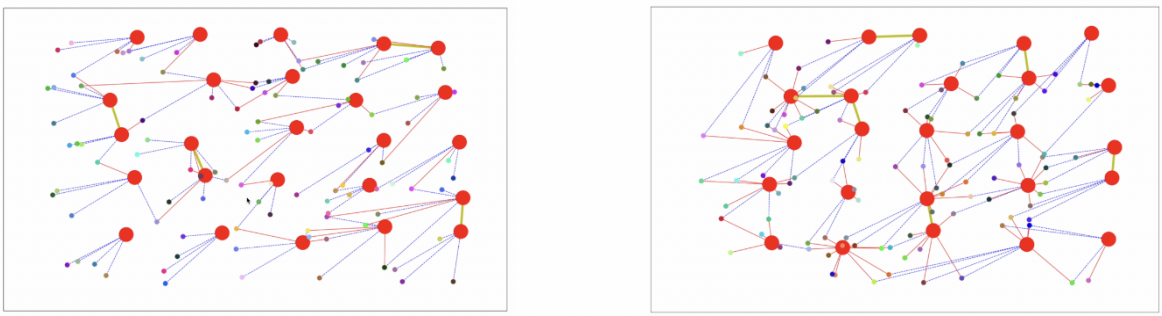


Figure 5.1. Snippets of simulation.

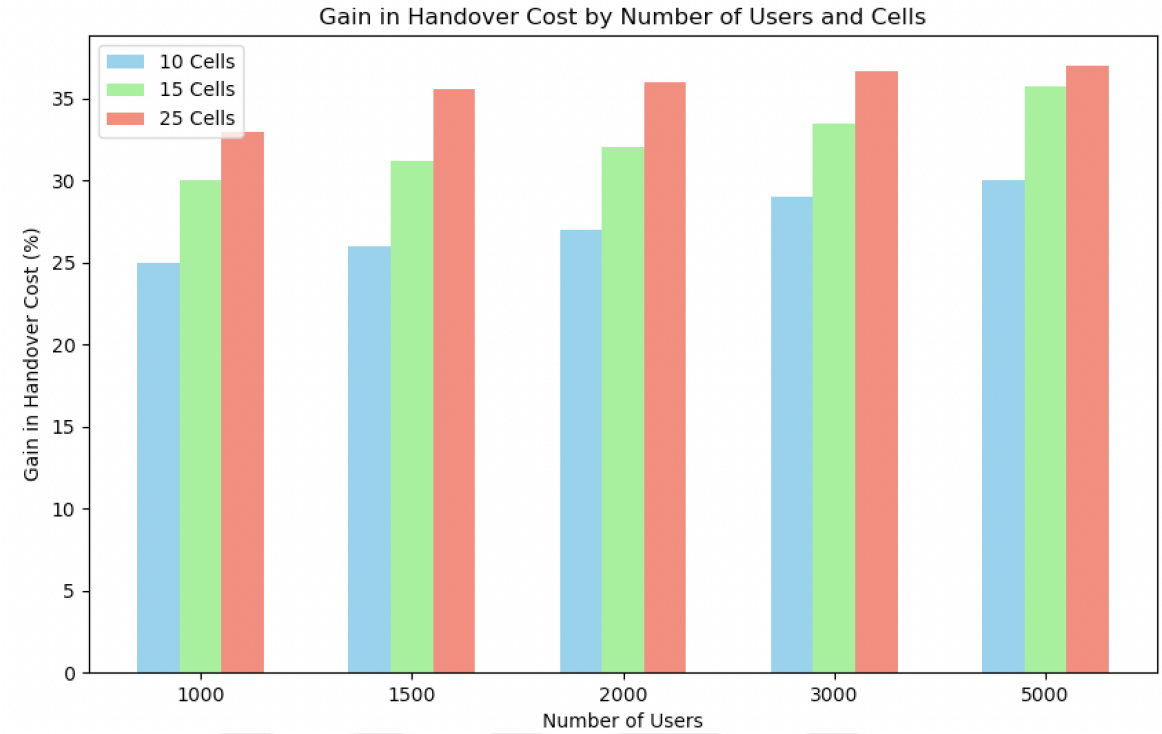



Figure 5.2. Gain in handover-cost with different network sizes.

In Figure 5.2, the reduction in handover costs compared to a baseline method across various deployment simulation scenarios is visualized. This reduction is quantified as the handover cost gain, indicating the savings achieved in handover operations. The figure reveals that as the number of UEs and BSs increases, the gain in handover cost also increases. This trend suggests a strong correlation between the UEs-BSs densities and improved handover cost efficiency. The increase in handover decisions due to greater density results in more opportunities to optimize and reduce the costs associated with these operations. This finding highlights the importance of deployment planning in densely populated network environments to maximize cost efficiency in handover processes.

In Figure 5.3, the improvements in load balancing relative to a baseline method are depicted across various deployment simulation scenarios. These improvements are strongly correlated with network density, which increases with the number of UEs and decreases with the number of BSs. The underlying reason for this pattern is the intentional design of the simulation environment to replicate realistic scenarios where load

distribution is uneven. As the network density grows, load distribution becomes more skewed, posing greater challenges to baseline distribution mechanisms. This situation offers more opportunities for the GNN-HOF method to step in and effectively optimize and balance network loads. Consequently, this shows the importance of adaptive load balancing algorithms in maintaining network efficiency in complex and densely populated network environments.

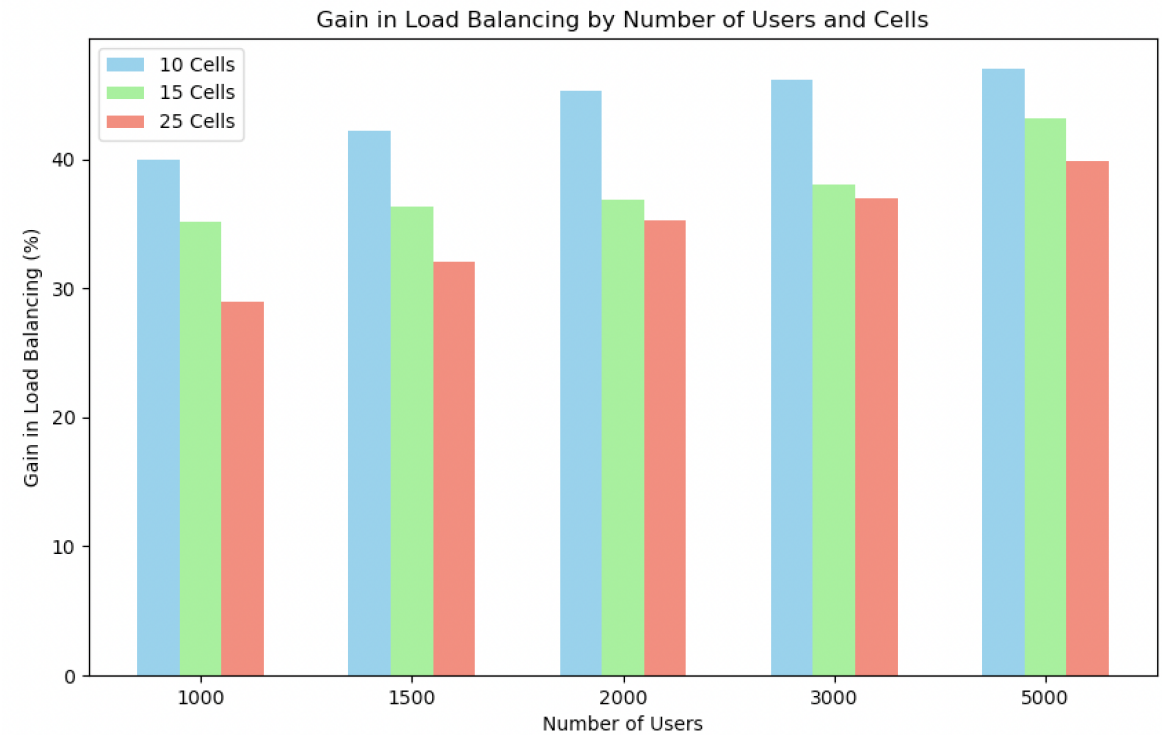



Figure 5.3. Gain in load-balancing with different network sizes.

In Figure 5.4, we examine the improvements in coverage compared to a baseline method across various deployment simulation scenarios. Due to the uneven distribution of BSs and UEs, cell-edge users often experience lower data rates. Furthermore, as the network density increases, the data rate for these cell-edge users tends to decrease further. Coverage is quantified as the 5th percentile of data rates among all UEs, leading to an expectation of reduced coverage in scenarios with denser network deployments according to the baseline method. In denser deployment scenarios, the algorithm has a better chance to enhance coverage, effectively overcoming the issues caused by uneven distribution and boosting service quality where it's needed most.

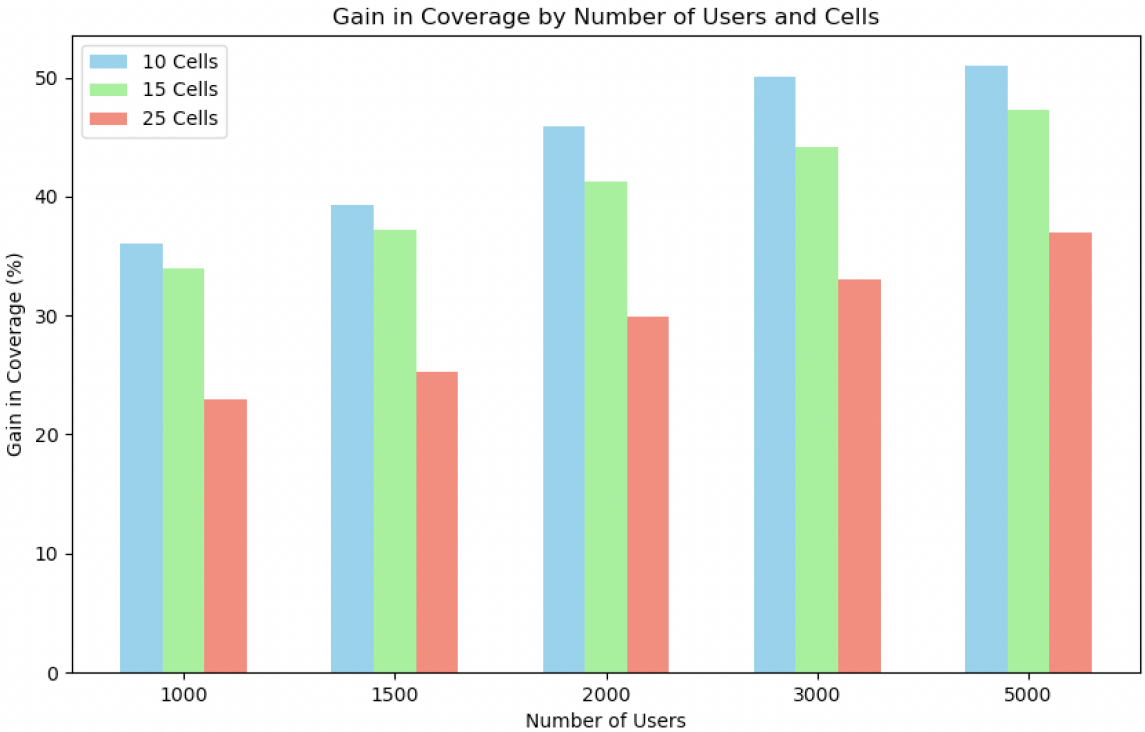



Figure 5.4. Gain in coverage with different network sizes.

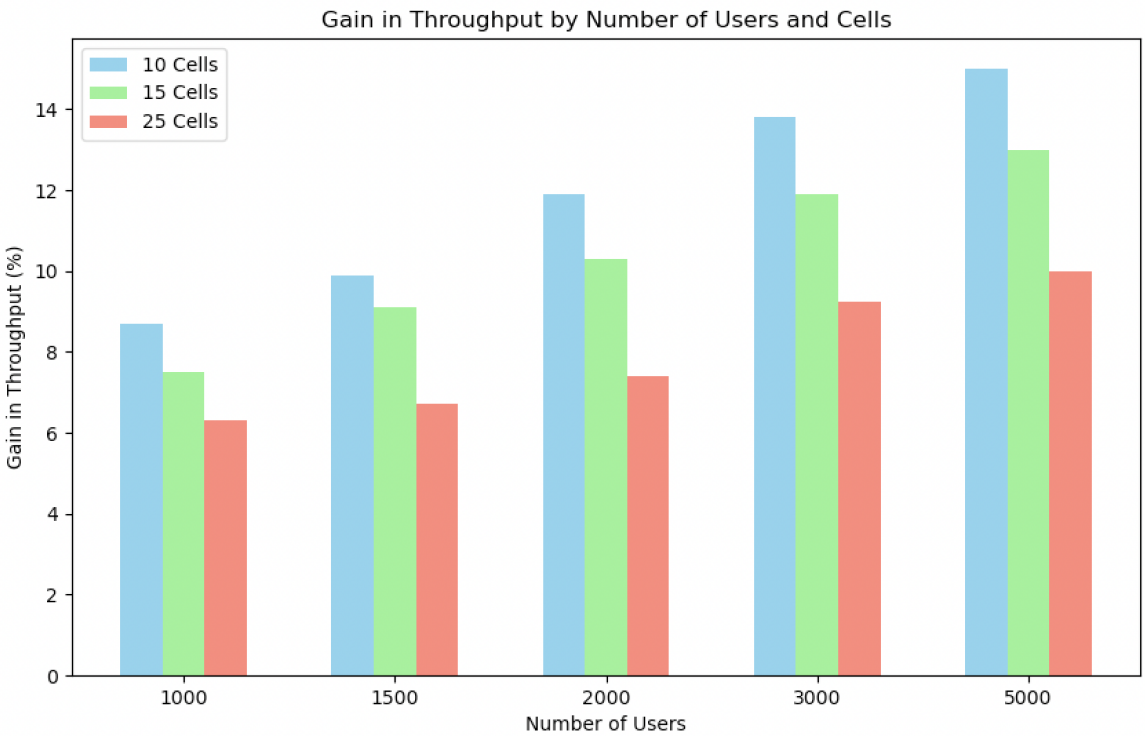



Figure 5.5. Gain in throughput with different network sizes.

In Figure 5.5, we illustrate the improvements in throughput relative to the baseline method across various simulation scenarios. As expected, the enhancement in throughput is significant in dense environments, where the impact of the GNN-HOF is more distinct. This observation aligns with previous metrics, indicating that throughput gains are closely linked to network density, highlighting the effectiveness of the GNN-HOF method in optimizing network performance under complex conditions.

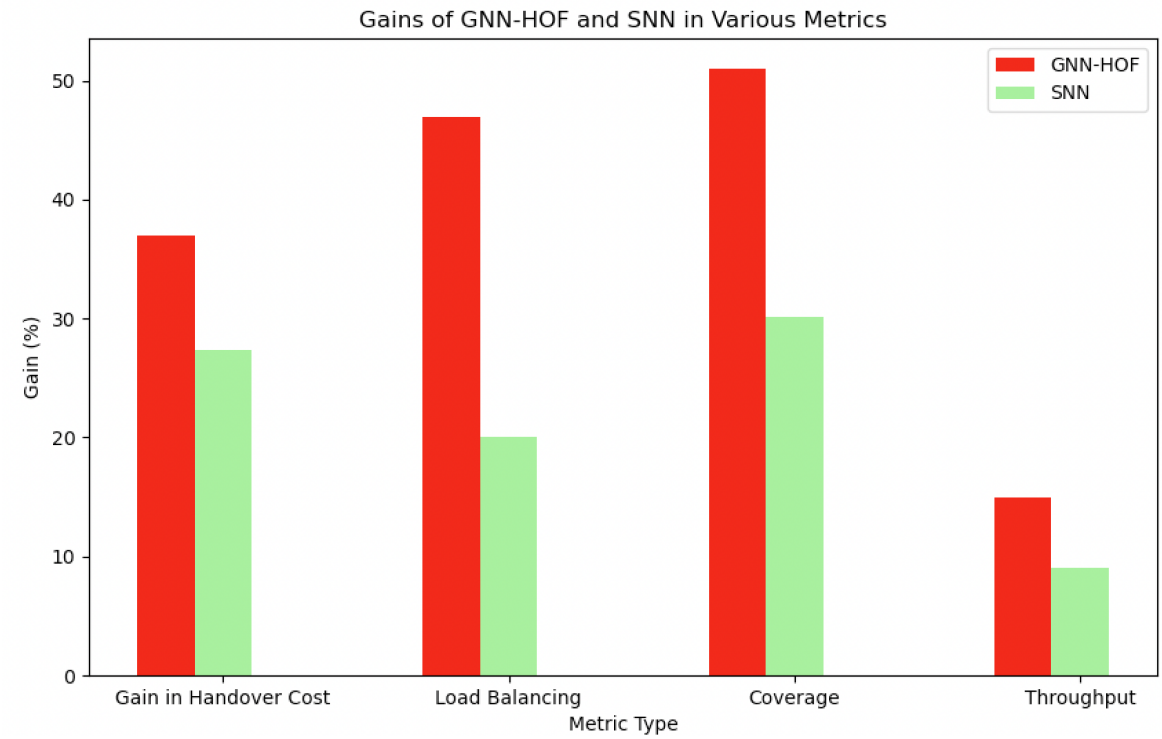



Figure 5.6. Gains of GNN-HOF and SNN in various metrics.

In Figure 5.6, we examine and compare the performance gains in handover cost, load balancing, coverage, and total throughput between the SNN and GNN-HOF methods. Consistent with expectations, the GNN-HOF method demonstrates superior performance across all these metrics due to its advanced capability to capture and utilize the structural information of the network. Furthermore, the notable improvement in load balancing achieved by the GNN-HOF method can be attributed primarily to its effectiveness in responding to the collective behavior of the network rather than just the actions of individual UEs. This shows the adaptability and efficiency of GNN-HOF in managing complex network dynamics.

## 5.2. SUMO Scenarios

To evaluate the proposed GNN-HOF framework in more realistic scenarios, the SUMO framework is employed to simulate urban mobility scenarios along with the actual BS locations in urban environments.

SUMO is an advanced, open-source traffic simulation package designed to handle large networks. It facilitates the modeling of vehicular traffic in urban environments, including the simulation of cars, public transport, and pedestrians, making it a versatile tool for researchers and planners alike. SUMO supports various traffic management and optimization scenarios, offering capabilities such as network import, demand modeling, and intermodal simulation [33]. These capabilities offer the opportunity to simulate realistic urban traffic flows. Several studies have focused on simulating the realistic urban traffic flows of specific cities including Monaco and Stuttgart [34, 35]. Those scenarios are utilized in our test cases.

The geographical positions of BSs within certain cities are accessible through online databases, notably referenced from CellMapper [36]. In the simulation environment, each BS is positioned according to its real-world location to foster the creation of highly realistic scenarios.

### 5.2.1. Monaco Traffic Scenario

The Monaco SUMO Traffic (MoST) Scenario [34] is utilized in this thesis. This scenario integrates diverse types of vehicles, pedestrian traffic reflecting the tourist influx, and public transport, offering a detailed and realistic urban mobility model. Derived from OpenStreetMap and enriched with precise geographical data, the simulation covers approximately 73.64 $km^2$, featuring a variety of road types and pedestrian pathways. Key to its design is the accurate representation of intersections and traffic management to mirror real-world dynamics of vehicle-pedestrian interactions [34]. Following parameters are used in the simulation:

Table 5.2. Simulation parameters.

| **Parameter** | **Value** |
|---|---|
| Area | 73.64 $km^2$ |
| Cell Transmit Power | 33 dBm |
| Channel Carrier Frequency | 30GHz |
| Number of BSs | 50 |

Although the number of BSs are higher in Monaco, some of the BSs are deliberately removed in order to increase the overall network density.

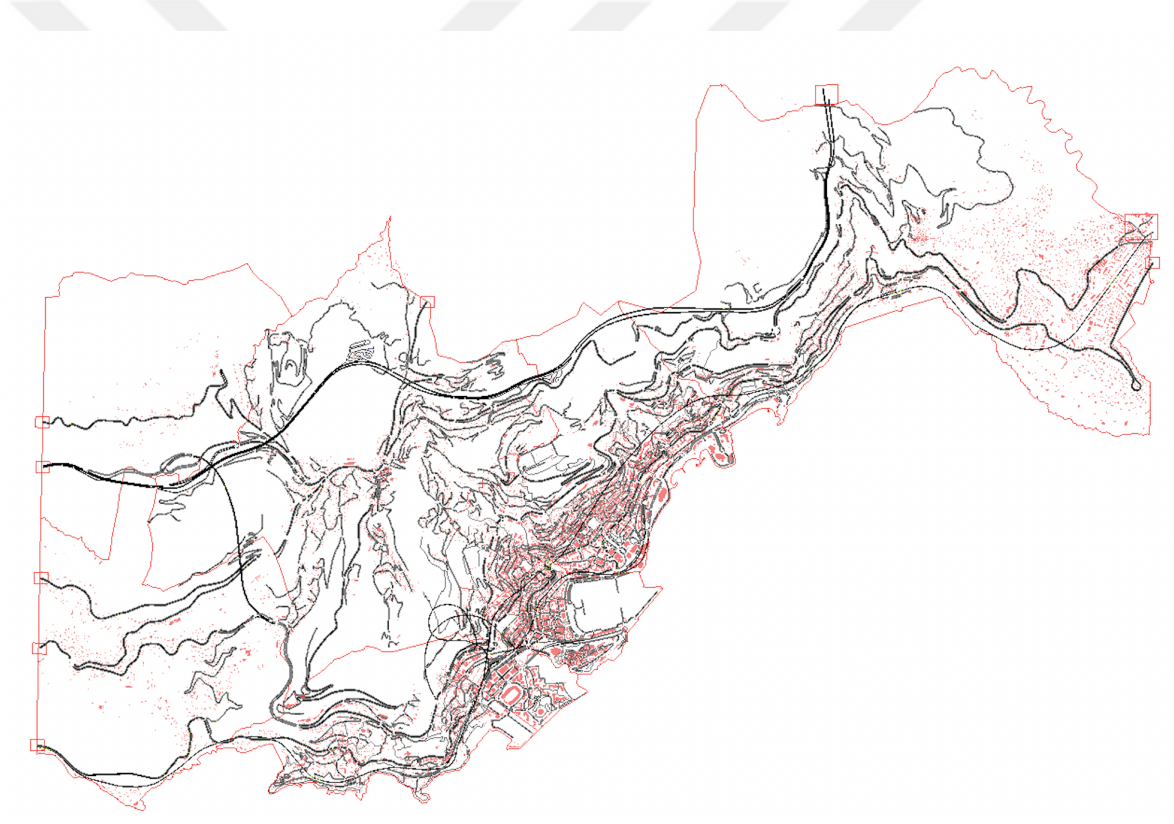


Figure 5.7. A snippet of Monaco SUMO simulation.

As depicted in Figures 5.7 and 5.8, the simulation environment extensively covers the Monaco region, simulating realistic vehicle movements.

Figure 5.8. A snippet of Monaco SUMO simulation.

There are three different regions based on traffic congestion in Monaco as follows:

- Zone 1: Heavy Traffic.
- Zone 2: Normal Urban Traffic.
- Zone 3: Little Traffic.

The simulations of the GNN-HOF and baseline methods are conducted across distinct regions to assess the impact of traffic congestion on the algorithms' performance.

In Figure 5.9, the enhancement in handover cost relative to a baseline method is illustrated across various traffic zones. The data show that as traffic congestion increases, the coverage gains achieved by the GNN-HOF method also increase, aligning with the patterns observed in simulation results. This trend shows the link between handover cost gains and network density. This correlation emphasizes the effectiveness of the GNN-HOF method in dynamically adapting to and managing varying traffic conditions, therefore optimizing network operations and reducing costs associated with handovers in congested areas.

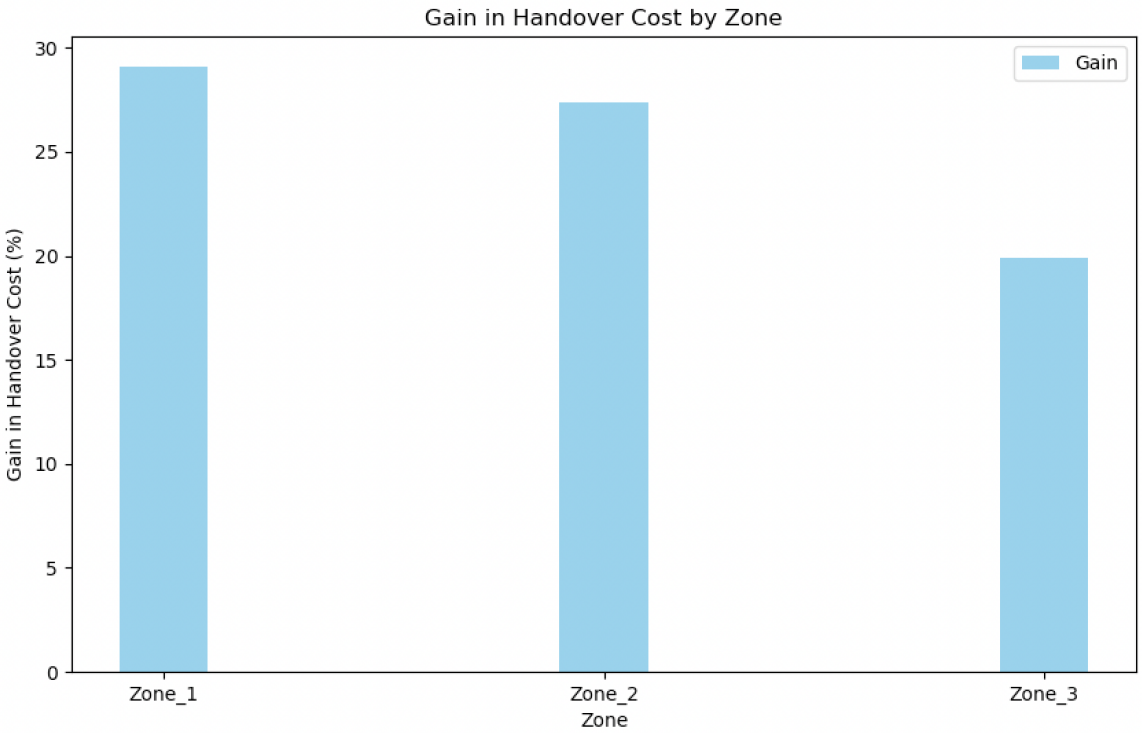



Figure 5.9. Gain in handover-cost across different zones.

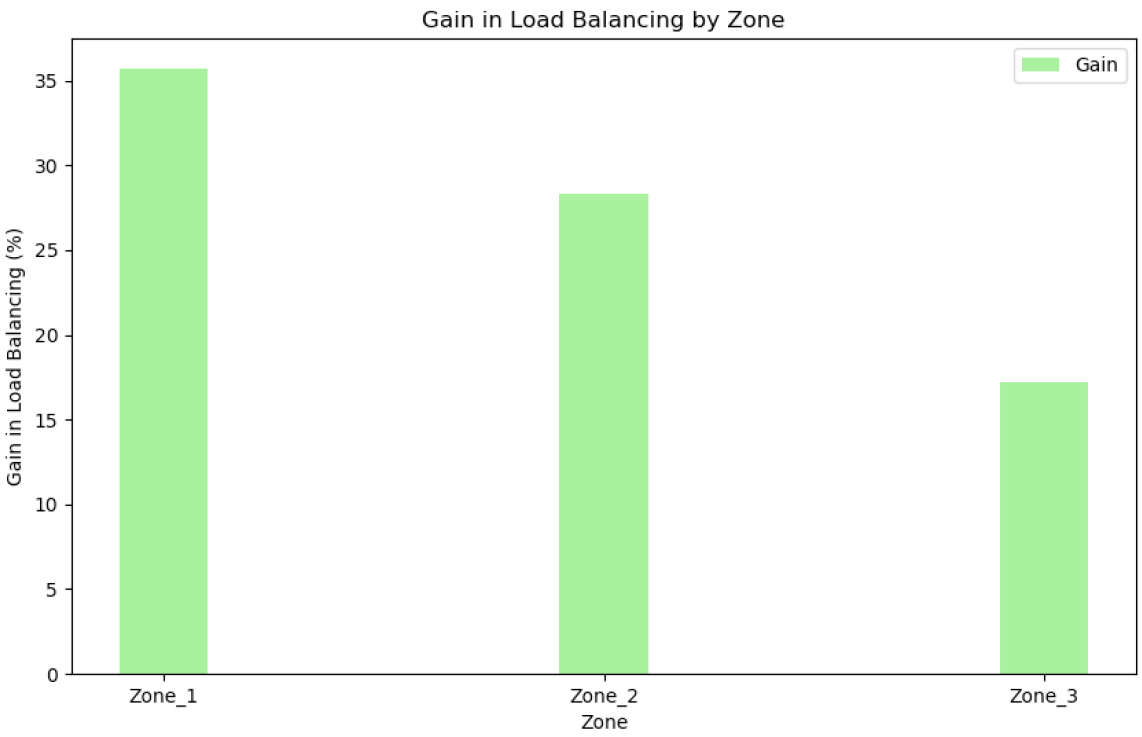



Figure 5.10. Gain in load-balancing across different zones.

In Figure 5.10, we present the improvements in load balancing relative to a baseline method across different traffic zones. As expected, there is a strong correlation between the gains in load balancing and network density. Notably, since the density of BSs does not vary significantly across the zones, the zones with a higher number of UEs exhibit increased network densities. This observation confirms the direct relationship between load balancing efficiency and network density, highlighting the importance of considering user distribution when assessing network performance and planning for resource allocation. This correlation highlights how areas with denser user populations benefit more from enhanced load balancing strategies, leading to more efficient network management.

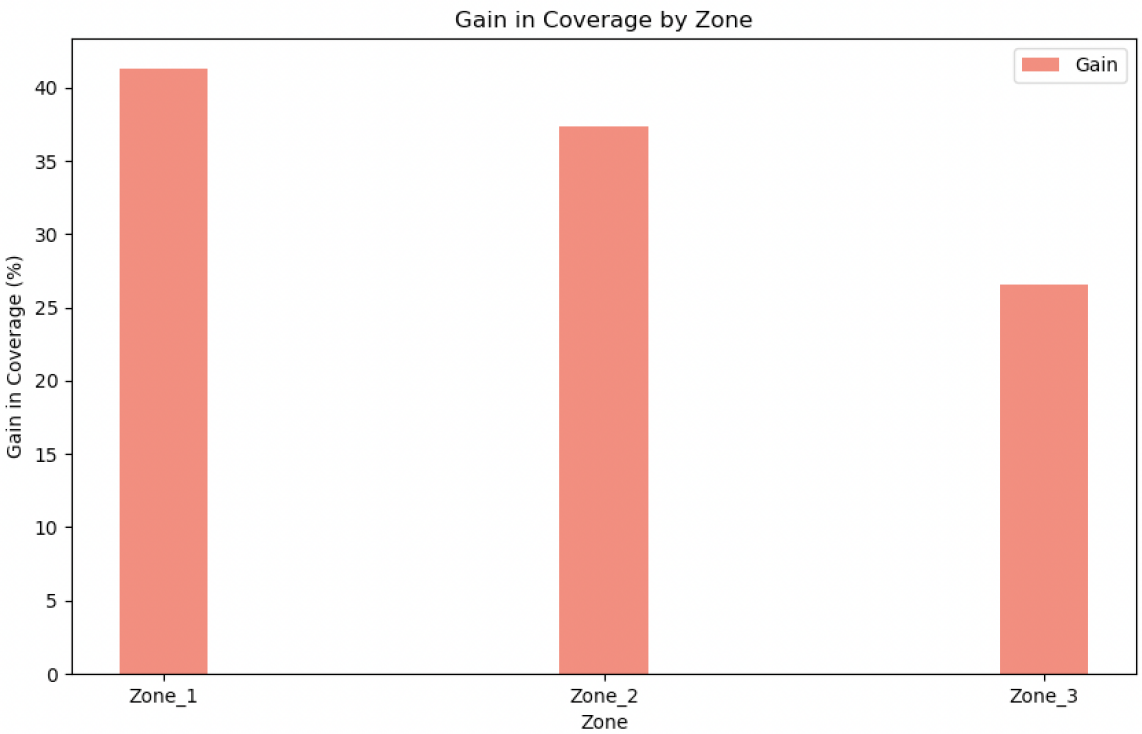



Figure 5.11. Gain in coverage with across different zones.

In Figure 5.11, we illustrate the improvements in coverage compared to a baseline method across various traffic zones. There is a clear correlation between traffic congestion and coverage gains, mirroring the trends observed in simulation studies. Notably, in the urban areas of Monaco, UEs are distributed unevenly, resulting in lower data rates for cell-edge users. The baseline method shows reduced coverage under these

conditions, as the data rate for cell-edge UEs declines with increasing network density. Therefore, the GNN-HOF method is more likely to further enhance coverage in such scenarios. This underscores the challenges of maintaining adequate coverage in densely populated areas and highlights the need for more effective distribution strategies to ensure consistent service across all zones.

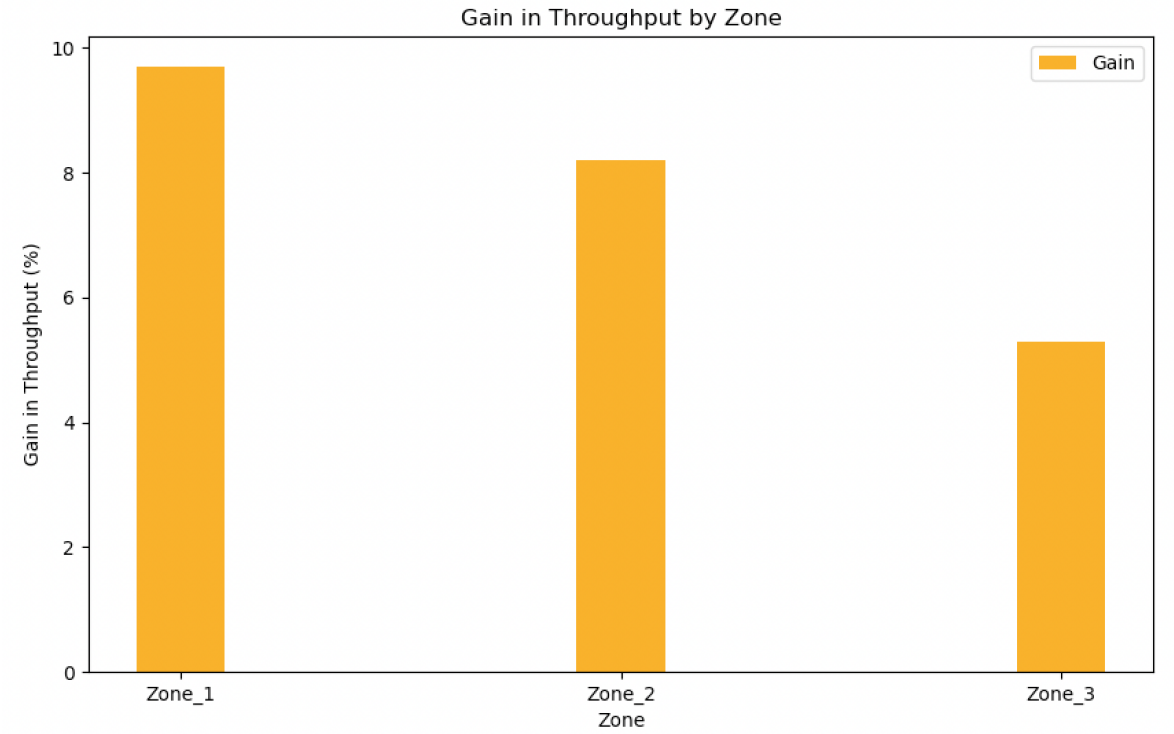



Figure 5.12. Gain in throughput across different zones.

In Figure 5.12, we illustrate the increase in throughput compared to the baseline method across various traffic zones. Similar to the pattern observed with coverage gain, the throughput gain also aligns with areas of high traffic congestion. This result confirms that the effectiveness of GNN-HOF method is more visible under the dense network deployment.

In Figure 5.13, gains in handover cost, load balancing, coverage, and total throughput between the SNN and GNN-HOF methods. As expected, the GNN-HOF method outperforms the SNN across all these metrics as in the simulations.

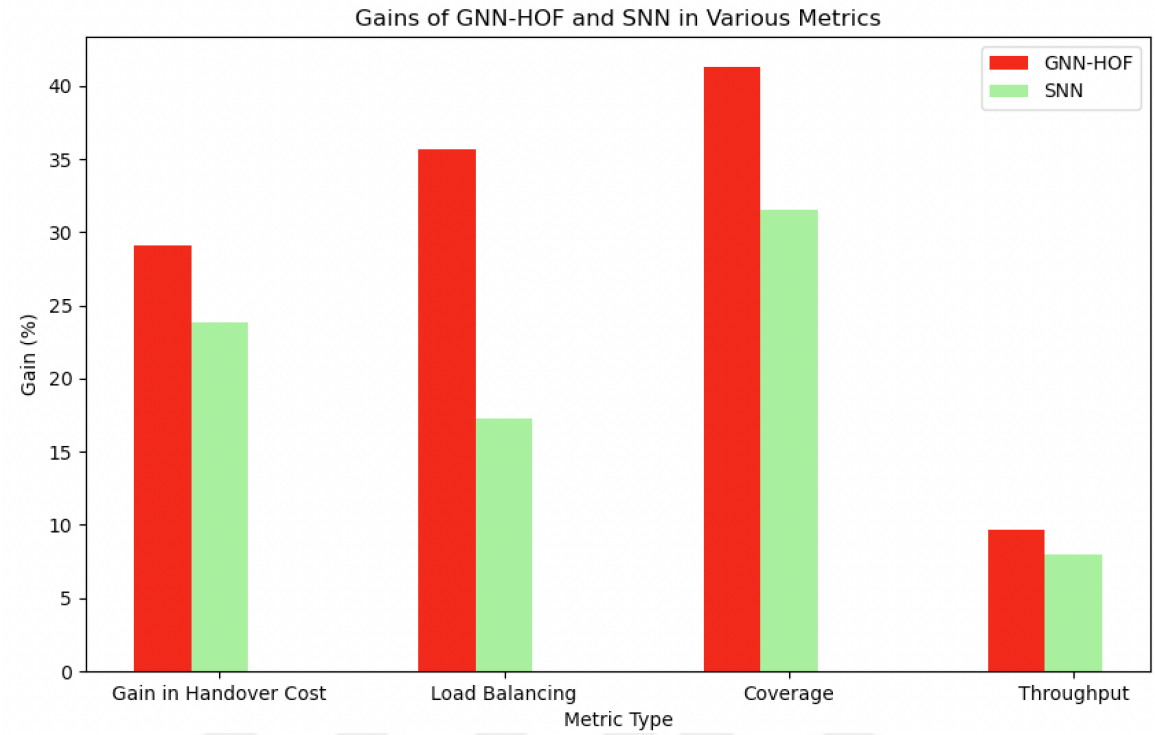



Figure 5.13. Gains of GNN-HOF and SNN in various metrics.

### 5.2.2. Stuttgart Traffic Scenario

In this simulation test, we employ the Stuttgart Open Motorway Project (STOMP) traffic scenario for Stuttgart in SUMO, as referenced in [35]. STOMP provides a synthetic 24-hour traffic simulation for a 45 km stretch of highway in Stuttgart, crafted using realistic traffic data. The scenario encompasses four distinct vehicle categories (passenger cars, motorbikes, trucks, and buses), each with authentic acceleration profiles. Following parameters are used in the simulation:

Table 5.3. Simulation parameters.

| **Parameter** | **Value** |
|---|---|
| Area | 45 km stretch of highway |
| Cell Transmit Power | 33 dBm |
| Channel Carrier Frequency | 30GHz |
| Number of BSs | 20 |

Although the number of BSs are higher in Stuttgart, some of the BSs are deliberately removed in order to increase the overall network density. Two versions of the scenario are presented: one reflecting weekday traffic patterns and another based on Sunday traffic data. This dual-scenario approach allows us to evaluate our model under varying traffic congestion levels on an identical setup. Contrary to the Monaco case, this scenario is set in a non-urban area along a highway, as illustrated in Figure 5.14.

Figure 5.14. A snippet of Stuttgart SUMO simulation.

There are three different times of day based on traffic congestion in Stuttgart case as follows:

- Morning: Rush Hour.
- Noon: Normal Highway Traffic.
- Evening: Rush Hour.

In Figure 5.15, we illustrate the improvement in handover cost relative to a baseline method at various times of the day. Consistent with earlier simulation results, the

effectiveness of the GNN-HOF method becomes more distinct as the number of UEs increases. Additionally, this simulation confirms the validity of our proposed approach by testing the same area under varying traffic conditions, thereby demonstrating the adaptability and potential benefits of the GNN-HOF method in managing handover costs effectively across different congestion levels.

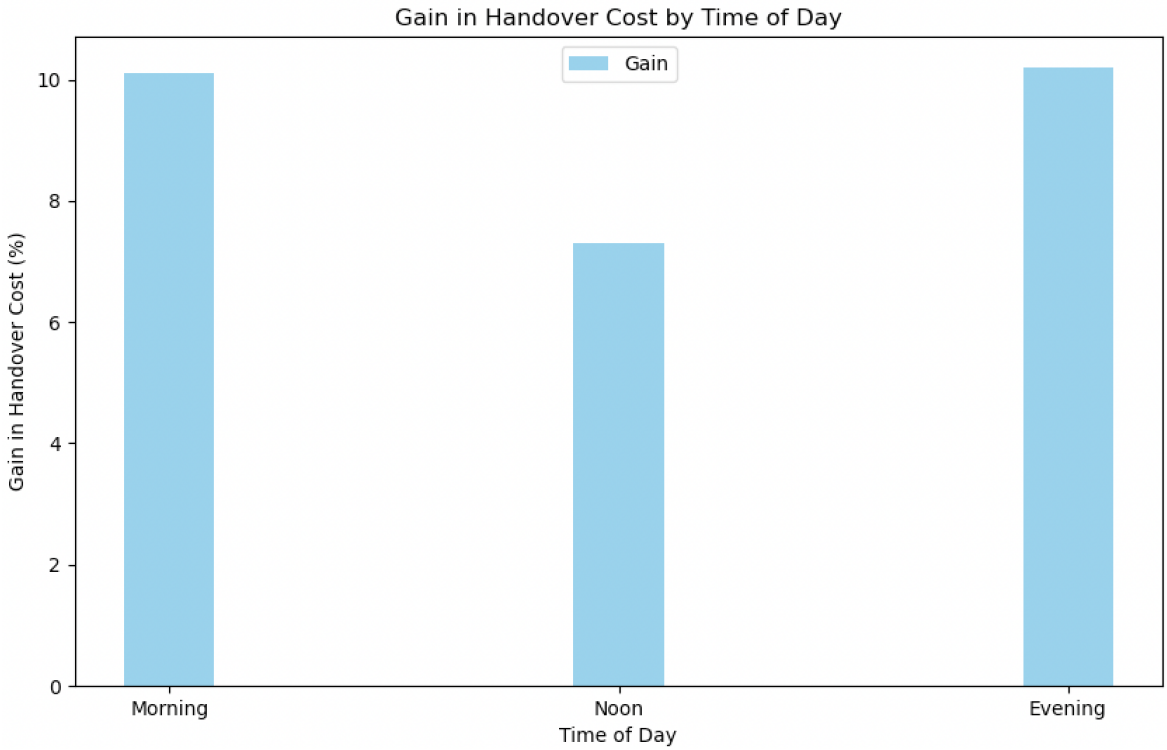



Figure 5.15. Gain in handover-cost across different times of the day.

In Figure 5.16, we demonstrate the improvements in load balancing over a baseline method at different times of the day. These enhancements are closely linked to network density, which rises with an increase in the number of UEs. A key factor in this pattern is the uneven distribution of UEs along the highway. As network density increases, the load distribution becomes more skewed, which challenges the effectiveness of traditional distribution mechanisms. This scenario provides an excellent opportunity for the GNN-HOF method to intervene and effectively optimize and balance network loads. This confirms the direct correlation between load balancing efficiency and network density, underscoring the significance of considering user distribution when evaluating network performance and planning resource allocation.

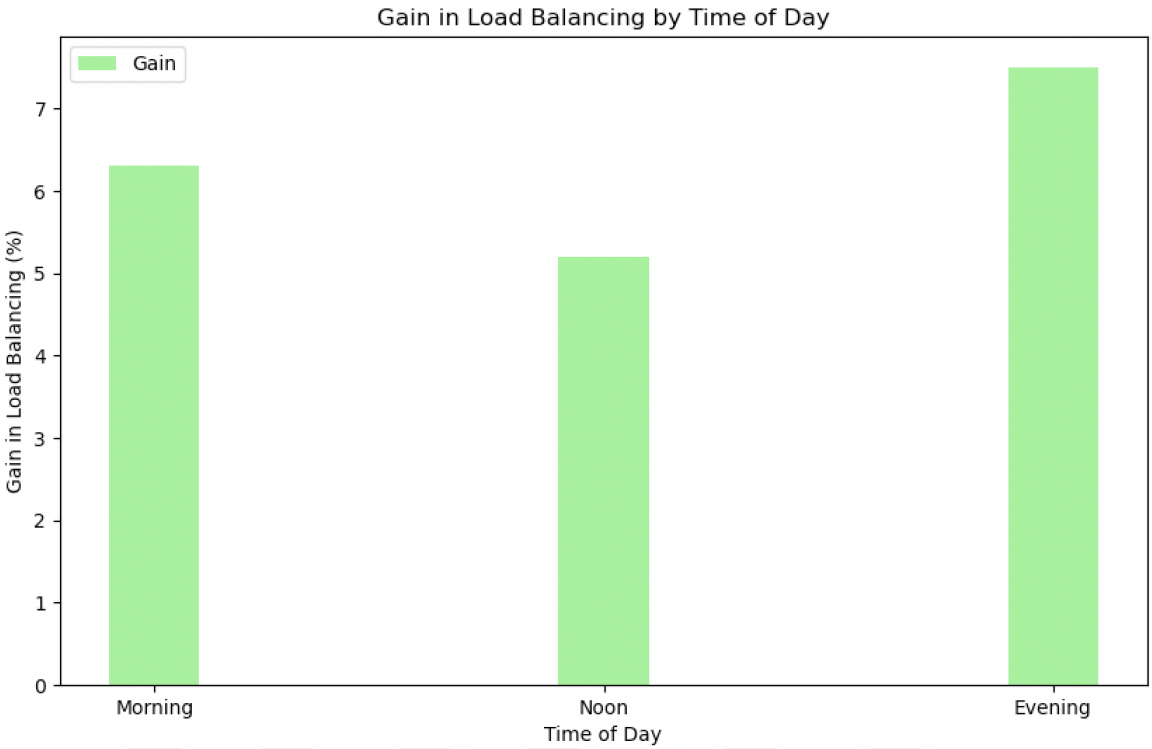



Figure 5.16. Gain in load-balancing across different times of the day.

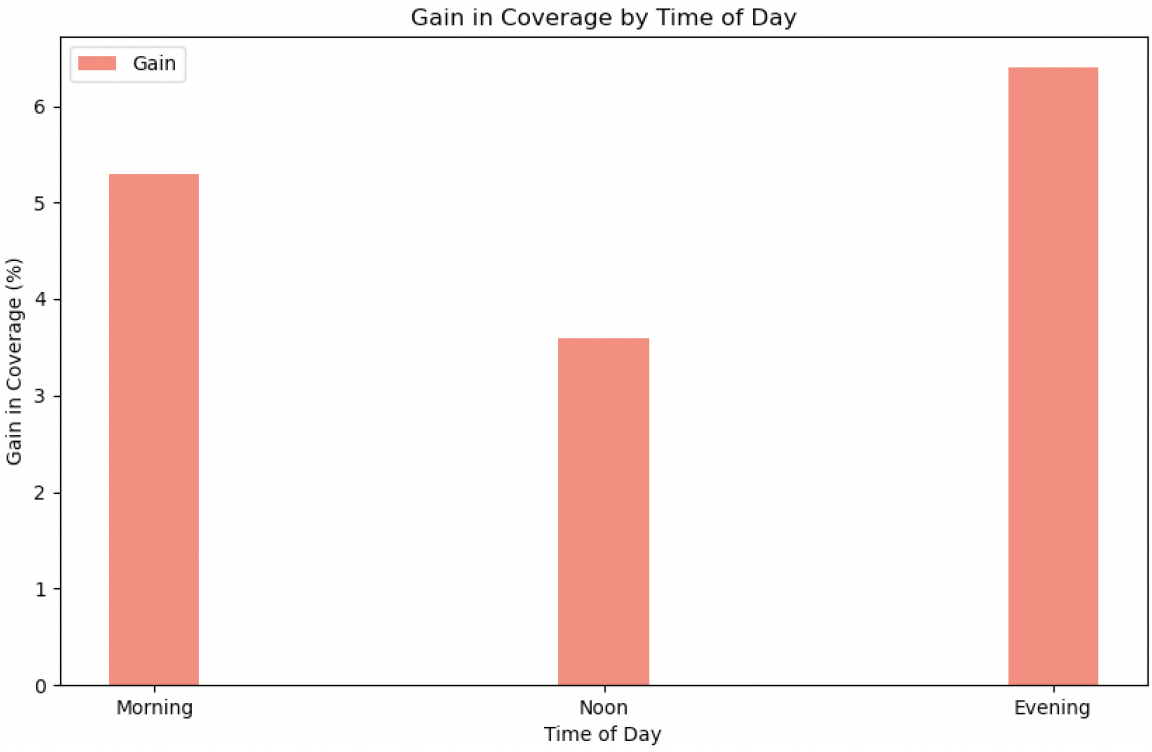



Figure 5.17. Gain in coverage with across different times of the day.

In Figure 5.17, the gain in coverage compared to the baseline method is depicted across various times of the day. Consistent with expectations, the increase in coverage is directly linked to network density. Despite using the same setup of BSs across various times, the data rates for cell-edge UEs decline as the number of UEs increases. Given that coverage is assessed based on the 5th percentile of data rates among all UEs, a decrease in coverage is expected during times of higher network density, aligning with the simulation results. This analysis highlights the challenges of maintaining consistent coverage under varying network loads and underscores the importance of adaptive network strategies to optimize coverage across different network conditions.

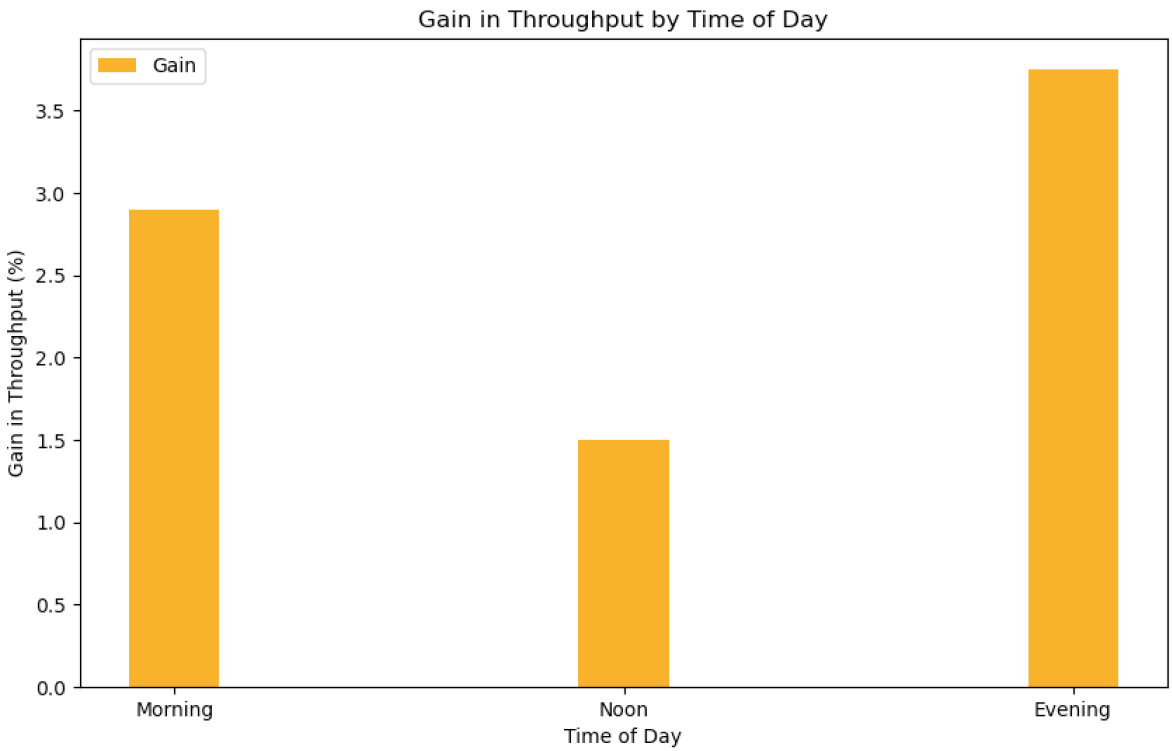



Figure 5.18. Gain in throughput across different times of the day.

In Figure 5.18, we present the enhancements in throughput relative to a baseline method throughout different times of the day. As expected, these gains in throughput are directly correlated with network density, similar to the patterns observed with coverage gains. The improvements are particularly notable in areas experiencing high traffic congestion, confirming that the GNN-HOF method exhibits greater effectiveness in densely deployed networks. This pattern is consistent with that observed for cover-

age gain, indicating that both throughput and coverage benefits from the GNN-HOF method are most pronounced under conditions of high network density. This consistency highlights the GNN-HOF method's capacity to adapt and perform optimally in challenging network environments, efficiently managing data transmission even during peak usage periods.

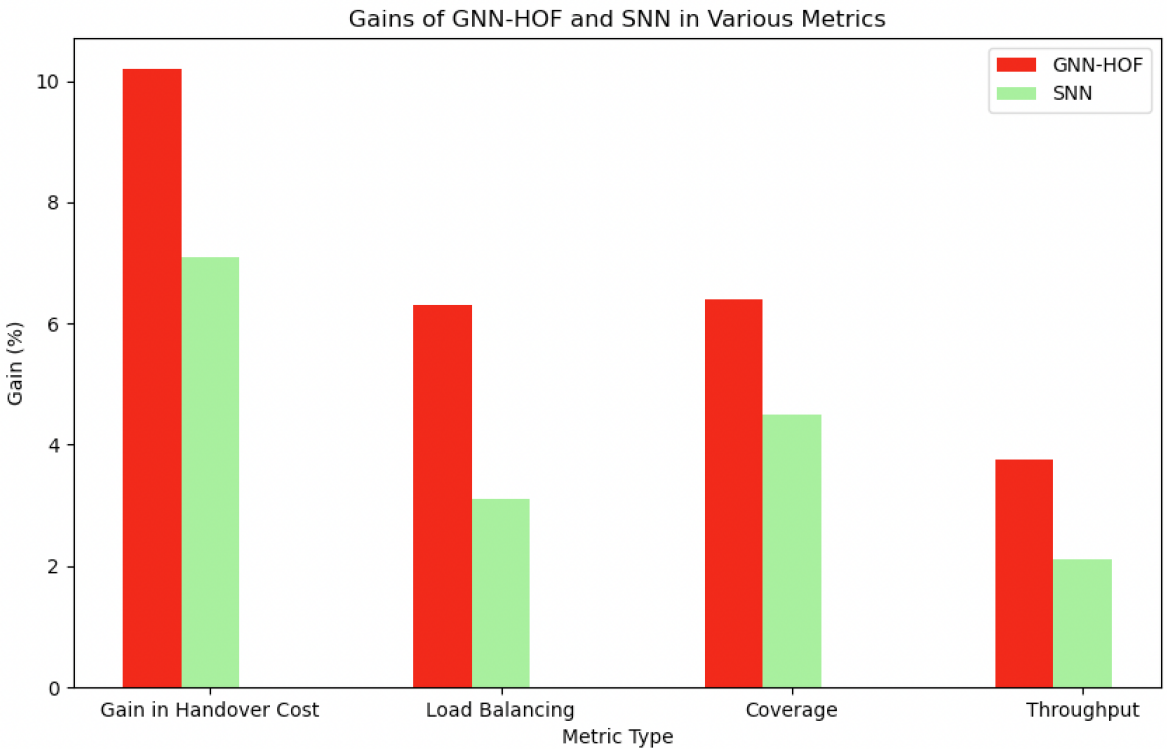



Figure 5.19. Gains of GNN-HOF and SNN in various metrics.

In Figure 5.19, we compare the gains in handover cost, load balancing, coverage, and total throughput between the SNN and the GNN-HOF methods. As anticipated, the GNN-HOF method consistently outperforms the SNN across all these metrics. This superior performance is attributed to the GNN-HOF's capability to capture and utilize the structural information of the network effectively. This comparison highlights the critical advantage of utilizing graph-based approaches in network management, particularly in environments characterized by rapidly changing network conditions.

# 6. CONCLUSION

This thesis introduces an innovative handover optimization framework for next-generation networks within O-RAN environments, showcasing the significant advancements in mobile network management through the use of a GNN based framework. This research has been pivotal in enhancing the understanding and execution of handover decisions by incorporating the dynamics of UE and BS, thus aiming to optimize critical network parameters such as load balancing, handover cost, throughput gain, and coverage gain. The developed framework, GNN-HOF, marks a substantial evolution from conventional proximity-based handover methods, employing sophisticated machine learning techniques to accurately predict and adapt to network dynamics.

The effectiveness of GNN-HOF has been rigorously tested in both simulated environments and real-world settings in Stuttgart and Monaco, using the SUMO tool. These tests have not only validated the framework's superior performance over traditional methods but have also highlighted its ability to significantly enhance network efficiency and user experience across diverse traffic scenarios and network densities.

In conclusion, the GNN-HOF represents a forward leap in the optimization of handover processes within O-RAN settings, underpinning the potential of machine learning in transforming future mobile networks. The successful application and resulting improvements from this research underscore the framework's viability and its role in paving the way for more intelligent, efficient, and user-centric mobile networks.

# REFERENCES


1. Ericsson, "5G Advanced, an Evolution Towards 6G", `https://www.ericsson.com/en/reports-and-papers/white-papers/5g-advanced-evolution-towards-6g`, 2023, accessed on March, 2024.

2. Alavirad, M., U. Hashmi, M. Mansour, A. Esswie, R. Atawia, G. Poitau and M. Repeta, "O-RAN Architecture, Interfaces, and Standardization: Study and Application to User Intelligent Admission Control", *Frontiers in Communications and Networks*, Vol. 4, No. 1, pp. 1–17, 2023.

3. Hamdan, M. Q., H. Lee, D. Triantafyllopoulou, R. Borralho, A. Kose, E. Amiri, D. Mulvey, W. Yu, R. Zitouni, R. Pozza, B. Hunt, H. Bagheri, C. H. Foh, F. Heliot, G. Chen, P. Xiao, N. Wang and R. Tafazolli, "Recent Advances in Machine Learning for Network Automation in the O-RAN", *Sensors*, Vol. 23, No. 21, pp. 1–35, 2023.

4. Polese, M., L. Bonati, S. D'Oro, S. Basagni and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges", *IEEE Communications Surveys & Tutorials*, Vol. 25, No. 2, pp. 1376–1411, 2023.

5. Yajnanarayana, V., H. Rydén and L. Hévizi, "5G Handover Using Reinforcement Learning", *2020 IEEE 3rd 5G World Forum*, pp. 349–354, Bangalore, India, 2020.

6. Shen, Y., J. Zhang, S. H. Song and K. B. Letaief, "Graph Neural Networks for Wireless Communications: From Theory to Practice", *IEEE Transactions on Wireless Communications*, Vol. 22, No. 5, pp. 3554–3569, 2023.

7. Akyildiz, I. and X. Wang, "A Survey on Wireless Mesh Networks", *IEEE Communications Magazine*, Vol. 43, No. 9, pp. S23–S30, 2005.

8. Andrews, J. G., S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong and J. C.

Zhang, "What Will 5G Be?", *IEEE Journal on Selected Areas in Communications*, Vol. 32, No. 6, pp. 1065–1082, 2014.

9. Silva, M., J. Fonseca, D. Perez Abreu, P. Martins, P. Duarte, R. Barbosa, B. Mendes, J. Silva, A. Goes, M. Araújo, B. Sousa, M. Curado and J. Santos, "O-RAN and RIC Compliant Solutions for Next Generation Networks", *INFOCOM*, New York, USA, 2023.

10. Masur, P. H., J. H. Reed and N. K. Tripathi, "Artificial Intelligence in Open-Radio Access Network", *IEEE Aerospace and Electronic Systems Magazine*, Vol. 37, No. 9, pp. 6–15, 2022.

11. Ullah, Y., M. B. Roslee, S. M. Mitani, S. A. Khan and M. H. Jusoh, "A Survey on Handover and Mobility Management in 5G HetNets: Current State, Challenges, and Future Directions", *Sensors*, Vol. 23, No. 11, p. 5081, 2023.

12. Dahouda, M. K., S. Jin and I. Joe, "Machine Learning-Based Solutions for Handover Decisions in Non-Terrestrial Networks", *Electronics*, Vol. 12, No. 8, p. 1759, 2023.

13. Jiang, C., H. Zhang, Y. Ren, Z. Han, K.-C. Chen and L. Hanzo, "Machine Learning Paradigms for Next-Generation Wireless Networks", *IEEE Wireless Communications*, Vol. 24, No. 2, pp. 98–105, 2016.

14. Wu, Z., S. Pan, F. Chen, G. Long, C. Zhang and S. Y. Philip, "A Comprehensive Survey on Graph Neural Networks", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 1, pp. 4–24, 2020.

15. Lee, M., G. Yu, H. Dai and G. Y. Li, "Graph Neural Networks Meet Wireless Communications: Motivation, Applications, and Future Directions", *IEEE Wireless Communications*, Vol. 29, No. 5, pp. 12–19, 2022.

16. Rehman, A. U., M. B. Roslee and T. Jun Jiat, "A Survey of Handover Management

in Mobile HetNets: Current Challenges and Future Directions", *Applied Sciences*, Vol. 13, No. 5, p. 3367, 2023.

17. Son, K., S. Lee, Y. Yi and S. Chong, "REFIM: A Practical Interference Management in Heterogeneous Wireless Access Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 29, No. 6, pp. 1260–1272, 2011.

18. Zhang, H., C. Jiang, X. Mao and H.-H. Chen, "Interference-Limited Resource Optimization in Cognitive Femtocells with Fairness and Imperfect Spectrum Sensing", *IEEE Transactions on Vehicular Technology*, Vol. 65, No. 3, pp. 1761–1771, 2015.

19. Beshley, M., N. Kryvinska, O. Yaremko and H. Beshley, "A Self-Optimizing Technique Based on Vertical Handover for Load Balancing in Heterogeneous Wireless Networks Using Big Data Analytics", *Applied Sciences*, Vol. 11, No. 11, pp. 1–24, 2021.

20. Farooq, M. U. B., M. Manalastas, S. M. A. Zaidi, A. Abu-Dayya and A. Imran, "Machine Learning Aided Holistic Handover Optimization for Emerging Networks", *ICC 2022 - IEEE International Conference on Communications*, pp. 710–715, Seoul, Korea, 2022.

21. Shayea, I., M. Ergen, M. Hadri Azmi, S. Aldirmaz Çolak, R. Nordin and Y. I. Daradkeh, "Key Challenges, Drivers and Solutions for Mobility Management in 5G Networks: A Survey", *IEEE Access*, Vol. 8, No. 1, pp. 172534–172552, 2020.

22. Tayyab, M., X. Gelabert and R. Jäntti, "A Survey on Handover Management: From LTE to NR", *IEEE Access*, Vol. 7, No. 1, pp. 118907–118930, 2019.

23. Sen, J., *Mobility and Handoff Management in Wireless Networks*, InTech, Rijeka, 2010.

24. Kasim, A. N., "A Survey Mobility Management in 5G Networks", *arXiv preprint arXiv:2006.15598*, 2020.

25. Orhan, O., V. N. Swamy, T. Tetzlaff, M. Nassar, H. Nikopour and S. Talwar, "Connection Management xAPP for O-RAN RIC: A Graph Neural Network and Reinforcement Learning Approach", *2021 20th IEEE International Conference on Machine Learning and Applications*, pp. 936–941, Online, 2021.

26. Hamilton, W. L., R. Ying and J. Leskovec, "Inductive Representation Learning on Large Graphs", *Proceedings of the 31st International Conference on Neural Information Processing Systems*, p. 1025–1035, California, USA, 2017.

27. Zhou, J., G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li and M. Sun, "Graph Neural Networks: A Review of Methods and Applications", *AI Open*, Vol. 1, No. 1, pp. 57–81, 2020.

28. Wang, X., H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui and P. S. Yu, "Heterogeneous Graph Attention Network", *The World Wide Web Conference*, p. 2022–2032, CA, USA, 2019.

29. Zhang, Z., P. Cui and W. Zhu, "Deep Learning on Graphs: A Survey", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 1, p. 249–270, 2022.

30. SHI, H., R. V. Prasad, E. Onur and I. Niemegeers, "Fairness in Wireless Networks: Issues, Measures and Challenges", *IEEE Communications Surveys Tutorials*, Vol. 16, No. 1, pp. 5–24, 2014.

31. Bromley, J., I. Guyon, Y. LeCun, E. Säckinger and R. Shah, "Signature Verification Using a" Siamese" Time Delay Neural Network", *Advances in Neural Information Processing Systems*, Vol. 6, No. 1, pp. 737–744.

32. Koch, G., R. Zemel, R. Salakhutdinov *et al.*, "Siamese Neural Networks for One-Shot Image Recognition", *ICML Deep Learning Workshop*, Vol. 2, Lille, 2015.

33. Lopez, P. A., M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd,

R. Hilbrich, L. Lücken, J. Rummel, P. Wagner and E. Wießner, "Microscopic Traffic Simulation using SUMO", *The 21st IEEE International Conference on Intelligent Transportation Systems*, pp. 2575–2582, Maui, HI, USA, 2018.

34. Codecà, L. and J. Härri, "Monaco SUMO Traffic MoST Scenario: A 3D Mobility Scenario for Cooperative ITS", *SUMO User Conference 2018 - Simulating Autonomous and Intermodal Transport Systems*, pp. 43–55, Berlin, Germany, 2018.

35. Förster, D., H. Löhr, A. Grätz, J. Petit and F. Kargl, "An Evaluation of Pseudonym Changes for Vehicular Networks in Large-Scale, Realistic Traffic Scenarios", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, No. 10, pp. 3400–3405, 2018.

36. CellMapper, "Cellular Tower and Signal Map", `https://www.cellmapper.net`, accessed on January, 2024.

37. Fang, R., G. Chuai and W. Gao, "Improve Quality of Experience of Users by Optimizing Handover Parameters in Mobile Networks", *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, New York, USA, 2020.

38. Alrabeiah, M., A. Hredzak and A. Alkhateeb, "Millimeter Wave Base Stations with Cameras: Vision-Aided Beam and Blockage Prediction", *IEEE 91st Vehicular Technology Conference*, Institute of Electrical and Electronics Engineers Inc., Antwerp, Belgium, 2020.

39. Wang, T., S. Wang and Z.-H. Zhou, "Machine Learning for 5G and Beyond: From Model-Based to Data-Driven Mobile Wireless Networks", *China Communications*, Vol. 16, No. 1, pp. 165–175, 2019.

40. Sana, M., A. De Domenico, E. C. Strinati and A. Clemente, "Multi-Agent Deep Reinforcement Learning For Distributed Handover Management In Dense MmWave

Networks", *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8976–8980, Barcelona, Spain, 2020.

41. Khosravi, S., H. S. Ghadikolaei and M. Petrova, "Learning-based Load Balancing Handover in Mobile Millimeter Wave Networks", *2020 IEEE Global Communications Conference*, pp. 1–7, Taipei, Taiwan, 2020.

42. Yajnanarayana, V., "Proactive Mobility Management of UEs Using Sequence-to-Sequence Modeling", *27th National Conference on Communications*, pp. 320–325, Mumbai, India, 2022.

43. Mollel, M. S., A. I. Abubakar, M. Ozturk, S. F. Kaijage, M. Kisangiri, S. Hussain, M. A. Imran and Q. H. Abbasi, "A Survey of Machine Learning Applications to Handover Management in 5G and Beyond", *IEEE Access*, Vol. 9, No. 1, pp. 45770–45802, 2021.