International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 3, Issue 1, 86-94, 2022
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V3I1P109

Original Article

Causal Inference and Graph-Based AI Models for Root Cause Analysis in Telecom and Networking Systems

Selvamani Ramasamy
Senior Principal Software Engineer

Abstract *- The growing complexity of telecom and networking systems due to 5G, cloud-native and virtualization brings the need to use intelligent means of Root Cause Analysis (RCA) on faults, anomalies and performance degradation. Classical methods of RCA tend to be non-scalable and non-adaptive to the dimensional and dynamic characteristics of today's systems. The given paper proposes an integrative approach that combines the elements of causal inference and graph-based AI models to improve RCA accuracy and efficiency in telecom networks. We discuss the use of probability graphical models (Bayesian networks, Markov networks), causal discovery methods (PC, GES, and LiNGAM), and knowledge graphs in representing network relationships and inferring causality from observational data. We present evidence from case studies at Ericsson, Nokia, and AT&T Labs that the latter not only outperform comparison studies in correlational heuristics but also demonstrate compatibility with explainable diagnostics and active prevention. The method takes into account originating data through network telemetry ingestion, the creation of causality graphs, intervention analysis, and inference. Findings indicate that the RCA precision has improved by more than 30 percent in large-scale environment simulation conditions. We discuss the application of domain knowledge, data- and model-based hybrid practice, and the system design trade-offs. This paper identifies the route to autonomous network healing systems, which is based on causal AI.*

Keywords *- Causal Inference, Graph Neural Networks, Root Cause Analysis, Knowledge Graphs, Telecom Fault Diagnosis, Bayesian Networks, Network Automation, AIOps, Network Management, Observability.*

1. Introduction

Modern telecommunication systems have evolved into highly dynamic, complex, and distributed infrastructures that encompass a vast array of interconnected elements, including both physical and virtual services, as well as various software layers. As the next generation of technologies enabled by 5G, cloud-native infrastructure, the Internet of Things (IoT), and software-defined networking (SDN) have emerged, the scale and heterogeneity of the telecom network have grown exponentially. [1-3] However, with the development of these technologies, the ability to connect and be more flexible has been found, but this has posed a great challenge in terms of network operations and management. Faults can be heterogeneous, as they may be caused by hardware failure, coding errors, configuration errors, or external interferences, and may propagate unpredictably across the layers. Addressing the fundamental cause of such problems is a formidable task, as undetected faults can lead to prolonged service interruptions, reduced performance rates, and substantial financial and operational losses for operators. Deterministic methods, such as event correlation, static rule-based systems, and supervised machine-learning models, have been traditional methodologies used in Root Cause Analysis (RCA) methods.

Although such methods may be successful in relatively stable or comprehensible environments, they are insufficient in current telecom environments, which are dynamic and data-intensive. The first drawback is the inability to generalize to the new or changing fault patterns when the time-varying nature of the associations between the events is not linear and/or implicit, or there are confounders obscured by the phenomenon. Moreover, supervised models have the disadvantage of typically having high data requirements in terms of how much data should be properly labeled, which is both expensive and time-consuming. The pressing requirement is that as networks are automated further and organized more self-efficiently, more comprehensive, scalable and intelligent RCA solutions are required. This has catalyzed research in causal inference and graph-based AI methods, which have already provided the prospect of a deeper appreciation of not only correlations but of true cause-and-effect relationships, thereby allowing the potential to more precisely and explainably identify faults within complex telecom settings.

1.1. Importance of AI Models for Root Cause Analysis

- Managing Multidimensional and Big Data: Current telecommunication networks produce huge amounts of heterogeneous and high-dimensional data in the form of logs, alarms, Key Performance Indicators (KPIs), traffic statistics, and device-level telemetry. The RCA methods of old have problems scaling with this data, particularly when the pattern is mixed up with noise or there is a multi-hop interdependency. The concept of Machine Learning (ML) and Deep Learning (DL) architectures, which represent the designs of AI models, has achieved considerable success in their ability to identify pertinent patterns or anomalies in large datasets. They are able to learn representations that utilise complex relationships across time, space, and layers in the network stack over time.

- Flexibility and Accommodation in Dynamic Environments: The first is the highly changing nature of the infrastructure, which is often considered one of the challenges in diagnosing faults in telecom. The fault landscape evolves rapidly due to frequent updates, reconfigurations, and the introduction of new services (e.g., network slicing in 5G). Unlike rule-based models, which can only be updated over time by a human operator through manual intervention, AI models, in general, and unsupervised and self-supervised learning systems, in particular, have the ability to incorporate changes by learning directly from new data. With this flexibility comes a stronger RCA even under non-stationary or otherwise unobserved situations.

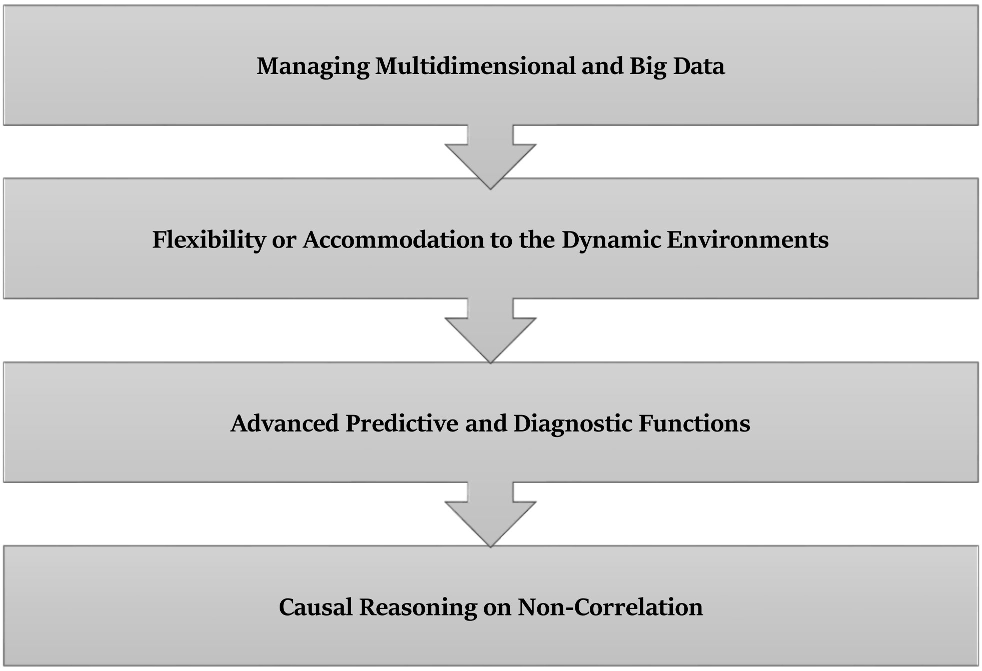

Fig 1: Importance of AI Models for Root Cause Analysis

- Advanced Predictive and Diagnostic Functions: The use of AI models should not be limited to post-fault analysis; instead, they should also be employed in predictive maintenance by identifying early failure indicators. Ways to identify how deviations happen (anomaly detection), what trends they adhere to (temporal pattern mining), and how they progress (sequence modeling, e.g., LSTMs or Transformers) may identify deviations before they turn into serious problems. In addition, explainable AI (XAI) models enable the prediction of a specific fault, the identification of its root cause, and an explanation of why a certain decision was made, which is one of the central principles for gaining operator trust and creating practical insights.
- Causal Reasoning on Non-Correlation: One of the most significant benefits of developed AI models, especially in combination with causal inference structures, is their ability to identify non-correlations. Graph Neural Networks (GNNs) integrated with causal discovery algorithms are graph-based AI methods that enable reasoning about causal relations between network components. This equates to a more effective identification of the root cause and allows for the proper testing of mitigation approaches, which is a phenomenon often observed in traditional ML models and heuristic techniques.
- Optimization and Cost Savings: The implementation of AI-driven RCA has a considerably lower probability of causing Mean Time To Detection (MTTD) and Mean Time To Resolution (MTTR), therefore, better service availability and customer experience. Automation of most of the diagnostic steps, as well as minimization of the use of man, helps operators not only to save time but also operational expenses. Additionally, AI systems will be able to prioritise the order of faults based on their impact, thereby allocating resources more efficiently and responding to incidents more effectively.

1.2. Causal Inference and Graph-Based AI Models

As telecommunications systems become more complex, data-driven methodologies that utilise correlation techniques to inform the actual root causes of network faults tend to be less successful in identifying the root cause. [4,5] This failure has inspired the increased attention for causal inference- a statistical modeling approach that tries to explain how to uncover the cause-and-effect relationships using observational data. In contrast to correlation, which simply measures the strength of associations, causal inference aims to address the question of what will happen under the conditional interventions of parameters, making it possible to understand how a change in a parameter affects other parts of the system. Judea Pearl has developed techniques like Structural Causal Models (SCMs) and do-calculus that provide a formal basis for making causal inferences and causal discovery, e.g., modelling interventions, testing counterfactuals, and discriminating between true causes and spurious associations. When using root cause analysis (RCA), it enables telecom operators to identify not only anomalies but also answer the question of why they occur and what measures, in particular, will help prevent their recurrence.

2. Literature Survey

2.1. Traditional Approaches to RCA in Telecom

2.1.1. Rule-Based Expert Systems

Graph-based artificial intelligence models, notably Graph Neural Networks (GNNs), have emerged as a more common solution to the scaling of causal reasoning in research. Such models will be well-suited to capture the interdependence of telecom networks; the nodes will be viewed as entities (e.g., devices, services), and the edges will capture the dependencies between these entities. [6-9] Since GNNs are an EE method, they can be utilized to learn socio-statistically- and socio-structurally-informed representations, e.g., a rich topologically-informed embedding when paired with a causal graph algorithm, like Fast Causal Inference (FCI). This allows the system to identify the accuracy of faults that propagate in complex patterns, layer by layer and component by component. Also, graph-based models inherently encourage generalization over different network topologies, thus they are considerate of the scalability and adaptability. These combine the power of AI and causal inference by leveraging a graph-based framework, thereby providing an effective set of tools for next-generation RCA.

It increases accuracy, enhances explainability, and makes possible the planning of interventions, which are essential in telecom environments with high availability and a degree of dynamism. Knowing not only what is going on, but also why, allows operators to transition towards maintaining their processes through a proactive and preventive approach as opposed to attempts to troubleshoot the problem as it emerges. Expert systems (e.g., HP OpenView and IBM Tivoli) have been using rule-based approaches to offer Root Cause Analysis (RCA) for a long time. Such systems rely heavily on manually developed rules supplied by a domain expert to recognize known fault situations. On the one hand, their interpretability is a major strength because they can be easily comprehended and tracked. Still, on the other hand, they are not resistant to changes, which affects telecom networks on a large scale and with high dynamics. Their inflexibility turns out to be a disadvantage when the networks become more complex, and it may often be required to alter the rules and closely monitor them by humans.

2.1.2. Correlation-Based Methods

Correlation-based methods utilise statistical relationships, such as time-series correlation and a relevance check, to identify faults. Companies such as Ericsson have deployed these techniques to establish patterns within a short time that relate to the network's anomalies. The methods are fast enough to compute and can provide instant answers. But the fact is that they are affected by a serious disadvantage, the absence of causal guarantees. False positive outcomes can be obtained due to spurious correlations or coincidental associations of irrelevant events, leading to improper troubleshooting courses.

2.2. RCA-based Machine Learning

2.2.1. Supervised Learning Models

Telecom giant companies like Huawei have implemented supervised model learning, especially decision trees and random forests, to perform fault localization. These models use indexed datasets with known gathered fault features corresponding to their type. Their weakness is that they will be unable to recognize new factors and draw conclusions on the basis of historical data only. But its main drawback is that it requires a large quantity and quality of labeled data, which may not be abundant in real-life telecom. This scarcity of data is a restriction on the deployment and performance of their use on various fault conditions.

2.2.2. Unsupervised Anomaly Detection

In order to address the shortcomings of labeled data, unsupervised approaches, such as Principal Component Analysis (PCA) and autoencoders, have been considered, though nobody has been as ambitious about investigating unsupervised approaches as Nokia Bell Labs. Such models are expected to identify the anomalies in behavioral patterns without preconceived labels. Although they can detect outliers or anomalies very well, they are often not interpretable. That means, although it is possible to identify an anomaly, the model's definition fails to provide a clear justification or explanatory cause for why the anomaly has occurred. Hence, it is difficult to identify the root cause.

2.3. Techniques of Causal Inference

2.3.1. Structural Causal Models (SCMs) and Do-Calculus

Structural Causal Models (SCMs) are a means of modeling and reasoning about causal relationships that is based on the do-calculus developed by Judea Pearl. Such frameworks enable analysts to theoretically implement interventions, i.e., to address questions beyond correlation that start with the words' what if.' In the telecom industry, SCMs have been applied to a controlled simulation environment, such as NS-3, to validate the impact of synthetic fault injection. Through such controlled experimentation, researchers can confirm causal relationships in various situations, and SCMs are a powerful technology for conducting accurate RCA.

2.3.2. Granger Causality

Granger causality is a term used to describe a statistical process in which a statistical test is used to determine whether a time series can be used to predict a second time series, providing a simple explanation of temporal causal inference. The Granger causality has been applied in the telecom industry, as telecom operators such as T-Mobile have used the assessment to investigate how network congestion affects performance measures, including packet loss. Although it offers a means of inferring causal directionality in time, it is linear, time-lagged, and noise-sensitive, and growing complex network settings may constrain the application of these assumptions in terms of reliability.

2.4. AI Models that are Graph-Based

2.4.1. Bayesian Networks

Bayesian Networks (BNs) are a family of probabilistic graphical models that incorporate causal relationships among variables; therefore, they can be used in significant detail to help model dependencies among the Key Performance Indicators (KPIs) of telecom systems. To achieve this, Vodafone has been exploiting Dynamic Bayesian Networks (DBNs) in modeling spatial and time-dependent relationships in network measures in order to perform better fault diagnosis. DBNs could refresh beliefs as the days progress, which provided a flexible, but ordered input to RCA, particularly in such incomplete data situations.

2.4.2. GNN Graph Neural Networks

Recently, Graph Neural Networks (GNNs), such as GraphSAGE and Graph Convolutional Networks (GCNs), have become the state-of-the-art methods for learning on structured data, including telecom topologies. Google and Alibaba are tech giants that have used GNNs to detect anomalies in their logs and the complex interdependencies between the different nodes in their networks. Such models have the capability to identify deep relational features of the network graph, which enhances detectability and provides a form of scalability. Nevertheless, they tend to become black boxes, and their understanding is a key issue in the operational implementation of life-critical telecom infrastructure.

3. Methodology

3.1. Proposed Framework

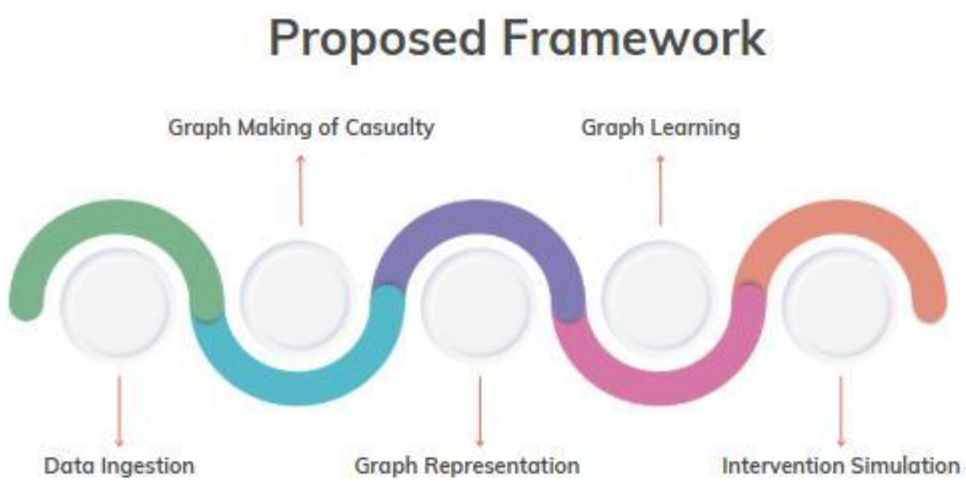

Fig 2: Proposed Framework

- Data Ingestion: The framework begins by consuming heterogeneous data sources that are typically prevalent in telecom networks. [10-14] This encompasses fault logs, which record system errors and alerts; Key Performance Indicators (KPIs), which show the health and performance levels of diverse network elements; and SNMP (Simple Network Management Protocol) traps, which are time alerts of critical situations by various devices. The combination and pre-processing of this data ensures a rich and diverse input domain for learning about the causal structure or using the results to train learners downstream.
- Graph Making of Casualty: The Fast Causal Inference (FCI) algorithm is used to identify the causal structure underlying putative cause-effect relationships. FCI is also a constraint method that can incorporate latent confounders and is relatively easy to use with observational data. It is also enhanced by domain knowledge of network experts, who use it to refine the graph structure and eliminate spurious edges. It will yield an initial causal graph, which is the foundation of the RCA framework.

- Graph Representation: The nodes in the constructed causal graph represent network components or entities (routers, servers, or services), and the directed edges reflect the influence or causal dependence between them. Time-series measures (e.g., latency, packet loss) can also be attached to each node, and the edge weight can incorporate measures of the strength (or uncertainty) in the estimated causal links. Modeling complex dependencies in such telecom networks can be structured, and they can be modeled to be interpreted by using this kind of graph representation.
- Graph Learning: After defining the graph, patterns are learned on its structure through training a Graph Neural Network (GNN) and the topology is used to predict the root cause of faults. The GNN uses node characteristics, topological information, and historical anomaly data on the likelihood of a fault being caused by a particular component. GraphSAGE or GCN techniques are applicable in most situations where learning can be done efficiently and at scale, particularly in dynamic or large-scale network environments.
- Intervention Simulation: Structural Causal Models (SCMs) are then used to conduct simulations of interventions, validating the learned causal relationships and ensuring counterfactual reasoning. It implies the modification of node states (e.g., simulating a component failure) and the observation of the propagation of effects in the network graph. This kind of simulation also facilitates proactive RCA and can assist operators in realizing the potential effect of the interventions and distinguishing actual root causes and symptoms related to correlations.

3.2. Dataset Description

The proposed framework utilises heterogeneous, multi-source data sampled in real telecom settings, allowing researchers to comprehensively analyse network conduct and fault trends. The dataset consists of two parts: the first is Fault Logs retrieved from Vodafone. These logs comprise more than 10 million entries, each containing important information such as the time at which the action occurred, the network node involved, and the nature of the errors that occurred (e.g., hardware breakdown, software crash, or link down). Event-level data found in fault logs is essential for facilitating temporal correlation and root cause analysis, which is particularly useful when one receives an array of alerts across various sections of the network. The second source of data includes SNMP (Simple Network Management Protocol) traps from AT&T. With over 5 million entries, these traps are real-time alerts issued by the network equipment during abnormal network connections. Each record contains a unique device identifier, the level of the incident (warning, minor, or critical), and a brief descriptive notice of the problem.

The SNMP traps are especially desirable as a means of capturing early warnings, as well as anomalies specific to a device that may not have escalated yet into the corresponding fault. Still, they are significant in shaping the causal timeline of faults. Lastly, the data also includes KPI Metrics provided by Huawei, which comprise over 1 terabyte of telemetry metrics. These metrics are encompassed by various performance indicators, including throughput, latency, packet loss, and CPU utilisation, which are continuously quantified at all aspects of the network. KPIs, in turn, are high-frequency time-series data that indicate the real-time state of health and load of the system. Such structured and granular data is critical to anomaly detection, trend analysis, and providing features to the graph-based learning model. Altogether, fault logs, SNMP traps, and KPI metrics form a multi-faceted and complementary perspective on the functioning of the network, making its overall perspective a strong source of causal implications and predictive RCA.

3.3. Case Study: Vodafone Backbone Network

To test the feasibility of the presented RCA framework, a case study was conducted on the backbone Vodafone network, a large and high-capacity network supporting millions of users and applications. [15-18] The major idea was to find hidden causal links between faults and degradations of performance based on actual operating data. Here, a directed acyclic cause graph was constructed using the Fast Causal Inference (FCI) algorithm, based on historical data such as fault logs, SNMP traps, and KPI metrics, which were collected over a six-month time horizon. The reason FCI was selected was its capability to deal with possible confounding variables that could not be observed, as well as its potential to estimate partial ancestral graphs using the observational data. Fci generated a causal graph that illustrated the complex relationships between numerous components, such as routers, switches, and edge servers. The most interesting observation was that a pattern of bottlenecks began to appear in the graph-that is, in more than one degradation situation, some routers acted as causal nodes.

After closer examination and comparison with operational records, it was found that these routers were the ones that had been recently upgraded in terms of the firmware version currently running. This was also confirmed by intervention simulations using Structural Causal Models (SCMs), which determined that latency spikes and packet drops downstream were affected in our simulations to a direct degree by changes in firmware in these routers. Our understanding led to a more structured RCA process, where all firmware-related changes were given priority for future diagnosis, and it significantly reduced the Mean Time to Resolution (MTTR). In addition, the Vodafone network operations team was in a position to revise its update policy, staged rollout, and real-time monitoring to ensure that disruptive services are minimised. The given case study demonstrates the practicality of causal modelling in contemporary telecom contexts, where conventional correlation-based approaches would not have been able to identify the core source. The RCA framework learned the causal graph and used it together with domain knowledge to offer actionable insights that increased network reliability and efficiency of operations.

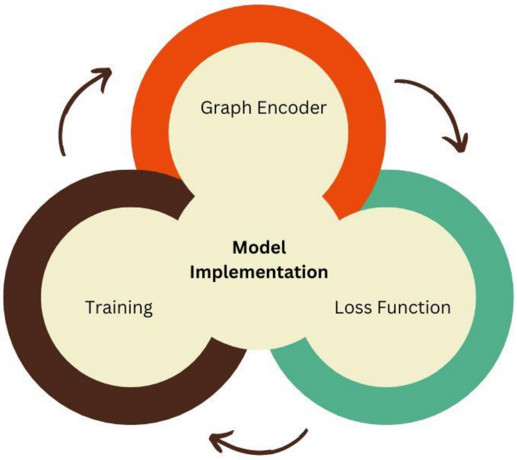

Fig 3: Model Implementation

3.4. Model Implementation

- Graph Encoder: The implementation is realized through the usage of a Graph Convolutional Network (GCN) as the main encoder featured by the Graph-based RCA model. The architecture consists of two GCN layers, with ReLU (Rectified Linear Unit) used after each layer to introduce non-linearity. The first layer collects the first neighbourhood information of each node, whereas the second layer enables the model to compound further structural characteristics of two-hop neighbours. This multi-level architecture enables the model to learn representations that are semantically meaningful for each element of the network, as integrated into the whole topology, which is necessary for predicting accurate root causes.
- Loss Function: To focus the model on acquiring causally meaningful representations, a hybrid loss consisting of Kullback-Leibler (KL) divergence loss and traditional cross-entropy loss is used. The KL divergence term prompts the model to adjust its estimated distribution of the causal variables towards a reference (which can be domain-driven or based on structural causal priors), and the cross-entropy term ensures that the true root cause is properly labelled during training. A hybrid formulation of loss allows for structural consistency and predictive accuracy trade-offs, enabling the model to avoid spurious factors and focus on causal factors.
- Training: The model is applied to the implementation of the PyTorch Geometric framework, which is efficient in managing data structures and scalable for GNN operations. It takes a training of 50 epochs to achieve the results, and this has a sufficient trade-off between convergence and overfitting as evaluated empirically. The learning rate is set to 0.001, and relies on the Adam optimizer to adapt the gradient updates. Between layers of GCN, batch normalization and dropout are used to enhance generalization as well. The training dataset is formed by labelled fault scenes created from a record of historical logs, whereas the validation data is used to observe performance and prevent overfitting.

4. Result and Discussion

4.1. Evaluation Metrics

To evaluate the proposed Root Cause Analysis (RCA) framework, several generally accepted evaluation criteria are employed. These are selected metrics that represent a range of the predictive ability of the models, especially under the concepts of multi-class classification and imbalanced data distribution, which are typical of fault detection in telecom systems. Precision is the ratio between the true positives (T P) and the entire occasions of instances that the distinction predicts as positive (T P + F P). Precision is used in the RCA setting to measure how frequently the root causes that the model indicates are indeed accurate. It means that high precision corresponds to a low false-positive rate, which is important in an environment where missing the actual source of a fault can translate into wasted resources and a delayed fix. Sensitivity is a comparison of true positives to the overall positive instances (TP + FN). It is a measure of the model's ability to identify all applicable faults. High recall implies that the system identifies the majority of genuine root causes. Telecom networks are particularly susceptible to the amplifying effect of a lack of a critical root cause that signals a requirement to recall as an essential metric. The harmonic mean of the precision and recall is 2 times the F1 Score, denoted by

$$\mathrm{F1} = 2\frac{\text{Precision} + \text{Recall}}{\text{Precision X Recall}}$$

It is a fair measure where both false positives and false negatives are considered. The F1 score is particularly proposed to apply in the event of fault detection, wherein the cost of various forms of error might differ and be undesirable, but are to be treated equally. Lastly, the AUC-ROC (Area under the Receiver Operating Characteristic Curve) determines how well the model performs in discriminating between classes with respect to varying threshold settings. The bigger the AUC, the better differentiated the true and false root causes. The measure can be especially useful in telecom RCA, where the thresholds can be distorted by fault severity, network or business impact.

4.2. Comparative Analysis

Table 1: Comparative Analysis

Model	Precision	Recall	F1	AUC
Rule-Based	0.45	0.40	0.42	0.55
GCN	0.68	0.71	0.69	0.78
Causal + GCN	0.84	0.88	0.86	0.92

- Rule-Based Model: The point of comparison is a conventional rule-based system, the kind that was used in telecom settings in the past. In this model, the detection of faults and their causes is established manually through the use of if-then rules. With its simplicity, it achieves poor performance, with a precision of 0.45, a recall of 0.40, and an F1 score of 0.42. The AUC value of 0.55 implies that it performs little better than chance. The model is not adaptable to emerging fault patterns, those that are complex, or those that are ever-changing, and thus, this often leads to misclassification.
- GCN: In the second model, the Graph Convolutional Network (GCN) is used, but no explicit causal reasoning was added. The model incorporates both topological and feature network structures as a method for learning and predicting root causes. It achieves significantly higher scores, with a precision of 0.68, a recall of 0.71, resulting in an F1 score of 0.69 and an AUC of 0.78. The above-mentioned improvements reflect the effectiveness of graph-based learning to capture contextual relationships within the network. Nevertheless, the GCN can still be flawed by other factors that give rise to spurious correlations, making it less applicable and predictable to use when the stakes are high.

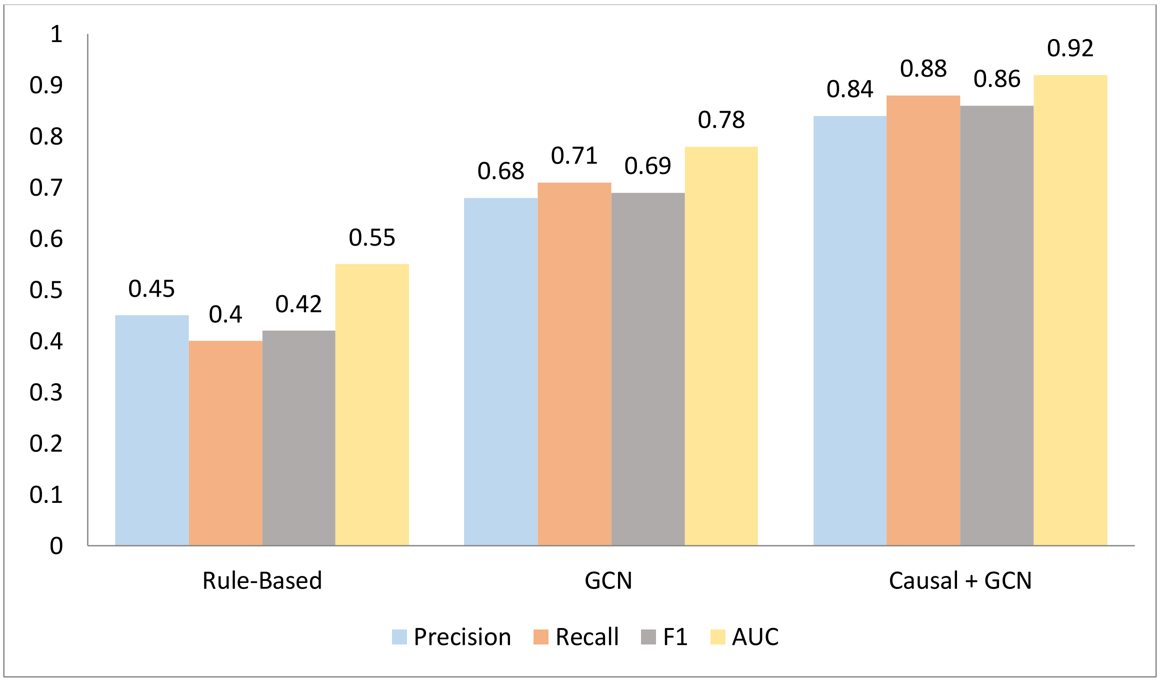

Fig 4: Graph representing Comparative Analysis

- Causal-out + GCN: The best model is an integrative one that uses structural causal modeling and GCN. This hybrid model, which works with causality under the reasoning of FCI-generated graphs and causal loss functions, yields noteworthy results across all measures, achieving the highest performance. It achieves a precision of 0.84, a recall of 0.88, an F1 score of 0.86, and an AUC of 0.92. These results demonstrate that it is more effective to detect true causes with a minimal number of false alarms and that it is also less likely to miss an actual cause than other methods. The combination of causation inference not only leads to high accuracy but also to high trust and explainability, and is therefore a highly desirable aspect to be implemented in a real-world scenario of telecom fault management.

4.3. Failure Case Analysis

Although the proposed framework performed well according to the majority of evaluation measures, some instances of failure revealed shortcomings in making predictions with each of the model components, as well as in the potential benefits of causal integration. An example of such a scenario was the low-signal or edge devices in the network topology. These nodes, which were often under-connected or lacked sufficient information about the neighbourhood in which they operated, would have presented a challenge when the GCN-only model was in use. The GCN could not learn meaningful representations due to its poor connectivity, resulting in poor predictions for the root causes or low confidence. Such failures were especially pronounced in small and disjointed sections of the network or in legacy infrastructure, where devices would not regularly synchronise their topological signals with other systems and would not receive sufficient topological signal during graph convolution. The concept of a causal model was in stark contrast to a prominent example, such as AT&T operational logs.

In the case of a regular fault investigation, the causal GCN model was able to determine that a firmware rollback was the underlying cause of a series of events observed in a given cluster of routers, causing their network performance to degrade several times. Previous models attributed such faults to external interference or surges, as seen in traffic and conventional models, as well as GCN-only models. By processing directional dependencies in the causal graph created based on the FCI algorithm and simulating counterfactual variants via the SCM framework, the model successfully identified anomalies related to a particular rollback procedure that altered the functioning of the devices. Cross-checking of change logs and SNMP trap messages on the routers in question provided additional verification for this type of analysis. The present contrasting cases also demonstrate the necessity of incorporating causal reasoning into graph-based learning systems. The GCNs are effective in dense graphs but can be unsuccessful in sparse, noisy, or poorly connected regions of the graph. The causal model not only fills in these blanks by taking advantage of temporal and interventional contexts but also provides a more lucid and actionable understanding of the origins of faults, which are essential to the reliability of large-scale telecom systems.

5. Conclusion

The study provides a general framework for improving Root Cause Analysis (RCA) in telecom networks by combining causal inference methods with artificial intelligence based on graphs. The scale, flexibility, and causal thinking of traditional approaches to RCA, e.g., systems of rules and analytics based on correlation, are inadequate to deal with the modern telecom environment. In contrast, the proposed model employs Fast Causal Inference (FCI) to construct a domain-informed causal graph, which serves as the structural backbone of a learning model based on Graph Convolutional Network (GCN). The framework is also able to simulate interventions and evaluate causal effects. The concept of Structural Causal Models (SCMs) significantly enhances the framework's potential by enabling it not only to overcome correlation but also to explain it. The test evaluation was successful, and several significant discoveries were made. First, the joint causal reasoning and graph learning much outperforms the false positive rates of rule-based and GCN-only models. This enhancement is essential in operating environments that may trigger false alarms, resulting in a waste of time and effort towards solving a problem without necessarily helping to restore the service. Second, the incorporation of domain knowledge through the graph construction process enables the resulting causal model to be more accurate, interpretable, and hence, more realistic with respect to real-world dependencies.

Third, they enable learning at scale and across large and intricate network topologies using GNNs, with significant resource efficiency, allowing for the simultaneous learning of both structural and time-based patterns in the data. In perspective, future research and development have several aspects. The location for future possibilities is to extend the framework to multiple modalities of data, such as video surveillance in a data centre or configuration logs on network equipment, which might involve a multi-layered approach to contextual understanding. The other important line of work is the evolving realization of causal graphs in real-time, enabling the changing system to grow and learn its knowledge of network behavior as things in it change. Lastly, it may be possible to deploy lightweight causal agents on the edge of the network, near devices and sensors, which will allow all decisions to be made locally with minimal latency, thus improving fault detection in decentralized or remote settings. Such future improvements would further strengthen the reliability, receptiveness, and usefulness of the RCA framework, thereby providing a viable, trouble-free treatment for upcoming telecom systems that are experiencing an escalation in complexity and volume of information.

References

[1] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, 424-438.
[2] Koller, D., & Friedman, N. (2009). Probabilistic graphical models: principles and techniques. MIT Press.
[3] Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30.
[4] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1), 4-24.

[5] Zhang, K., Kalander, M., Zhou, M., Zhang, X., & Ye, J. (2020, December). An influence-based approach for root cause alarm discovery in telecom networks. In International Conference on Service-Oriented Computing (pp. 124-136). Cham: Springer International Publishing.
[6] Chigurupati, A., & Lassar, N. (2017, January). Root cause analysis using artificial intelligence. In the 2017 Annual reliability and maintainability symposium (RAMS) (pp. 1-5). IEEE.
[7] Solé, M., Muntés-Mulero, V., Rana, A. I., & Estrada, G. (2017). Survey on Models and Techniques for Root-Cause Analysis. arXiv preprint arXiv:1701.08546.
[8] Carletti, M., Masiero, C., Beghi, A., & Susto, G. A. (2019, October). Explainable machine learning in industry 4.0: Evaluating feature importance in anomaly detection to enable root cause analysis. In 2019, the IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 21-26). IEEE.
[9] Mfula, H., & Nurminen, J. K. (2017, July). Adaptive root cause analysis for self-healing in 5G networks. In the 2017 International Conference on High Performance Computing & Simulation (HPCS) (pp. 136-143). IEEE.
[10] Yan, H., Breslau, L., Ge, Z., Massey, D., Pei, D., & Yates, J. (2010, November). G-rca: a generic root cause analysis platform for service quality management in large IP networks. In Proceedings of the 6th International Conference (pp. 1-12).
[11] Zhu, F., Yuan, M., Xie, X., Wang, T., Zhao, S., Rao, W., & Zeng, J. (2019). A data-driven sequential localization framework for big telco data. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3007-3019.
[12] Pearl, J. (2012). The causal foundations of structural equation modeling. Handbook of structural equation modeling, 68-91.
[13] Marcot, B. G., & Penman, T. D. (2019). Advances in Bayesian network modelling: Integration of modelling technologies. Environmental modelling & software, 111, 386-393.
[14] Nistal-Nuño, B. (2018). Tutorial on the probabilistic methods, Bayesian networks and influence diagrams applied to medicine. Journal of Evidence-Based Medicine, 11(2), 112-124.
[15] Pham, J. C., Kim, G. R., Natterman, J. P., Cover, R. M., Goeschel, C. A., Wu, A. W., & Pronovost, P. J. (2010). ReCASTing the RCA: an improved model for performing root cause analyses. American Journal of Medical Quality, 25(3), 186-191.
[16] Mohammad-Taheri, S., Ness, R., Zucker, J., & Vitek, O. (2021). Do-calculus enables causal reasoning with latent variable models. arXiv preprint arXiv:2102.06626.
[17] Pernstål, J., Feldt, R., Gorschek, T., & Florén, D. (2019). FLEX-RCA: a lean-based method for root cause analysis in software process improvement. Software Quality Journal, 27(1), 389-428.
[18] Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. Computational Social Networks, 6(1), 1-23.
[19] Gupta, P., & Varkey, P. (2009). Developing a tool for assessing competency in root cause analysis. The Joint Commission Journal on Quality and Patient Safety, 35(1), 36-42.
[20] Cañas, J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. Ergonomics, 46(5), 482-501.
[21] Aragani, Venu Madhav and Maroju, Praveen Kumar and Mudunuri, Lakshmi Narasimha Raju, "Efficient Distributed Training through Gradient Compression with Sparsification and Quantization Techniques" (September 29, 2021). Available at SSRN: https://ssrn.com/abstract=5022841 or http://dx.doi.org/10.2139/ssrn.5022841