FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Topology Inference in Low-Voltage Distribution Networks Using Generative Models

Ana Maria Alves de Almeida



Mestrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Prof. Ricardo Jorge Gomes de Sousa Bento Bessa

October 14, 2026

© Ana Maria Alves de Almeida, 2026

Resumo

As redes de distribuição de baixa tensão enfrentam exigências crescentes de monitorização, gestão e controlo devido ao aumento dos recursos energéticos distribuídos, dos veículos elétricos e das cargas flexíveis. Um requisito essencial para estas tarefas é o conhecimento da topologia da rede, isto é, da forma como os clientes, os barramentos e os ramos se encontram efetivamente ligados num determinado instante. Na prática, os registos topológicos podem estar incompletos ou desatualizados, enquanto a informação disponível pode limitar-se a séries temporais de potência ativa e módulo da tensão recolhidas ao nível dos clientes. Embora estas medições não descrevam diretamente os ramos da rede, contêm informação estrutural indireta, uma vez que as variações de tensão dependem dos caminhos elétricos partilhados e da resposta da rede às variações de potência. Ainda assim, diferentes estruturas radiais podem originar padrões elétricos semelhantes, tornando a inferência da conectividade um problema inverso difícil.

Esta dissertação estuda a inferência de topologias em redes de baixa tensão através de modelos de aprendizagem, utilizando um conjunto de dados controlado que combina estruturas radiais geradas sinteticamente com perfis de consumo residencial derivados de medições reais. A metodologia experimental gera as redes e atribui parâmetros elétricos aos seus ramos. Aos barramentos de carga são associados perfis de potência ativa obtidos a partir de séries temporais residenciais, com resolução de 30 minutos, provenientes do *Irish Smart Metering Customer Behaviour Trial*. Para cada caso, os módulos da tensão são calculados através de simulação de fluxo de potência pelo método Forward–Backward Sweep, representando a resposta elétrica do alimentador aos perfis de consumo considerados. Este procedimento proporciona um ambiente reprodutível, com conectividade de referência conhecida e comportamento elétrico fisicamente coerente, permitindo avaliar diferentes representações das medições e arquiteturas VAE para reconstrução de redes radiais.

O trabalho começa por analisar uma arquitetura inicial de Variational Autoencoder de tamanho fixo, utilizando como entrada trajetórias brutas de potência ativa e módulo da tensão. Os resultados mostram uma diferença clara entre a reconstrução de topologias *seen* e *unseen*: o modelo reconhece adequadamente estruturas representadas durante o treino, mas generaliza de forma limitada para novas configurações radiais. Foram realizados ensaios de diagnóstico sobre a organização do espaço latente, a quantidade de dados de treino, o horizonte temporal, a regularização e a arquitetura. Estes ensaios indicam que a dificuldade de generalização não resulta apenas de uma escolha isolada de hiperparâmetros.

Em seguida, são avaliados modelos VAE convolucionais e representações estruturadas das medições, incluindo matrizes de correlação, blocos de correlação potência–tensão e matrizes de sensibilidade. Os resultados mostram que a representação da informação elétrica tem um impacto significativo na reconstrução. Nas comparações controladas, as sensibilidades fornecem a informação mais útil para a recuperação dos ramos, enquanto as trajetórias brutas de P/V apresentam um desempenho inferior. Contudo, os modelos de tamanho fixo continuam a exigir uma arquitetura distinta para cada dimensão da rede e degradam-se à medida que o número de nós aumenta.

Para ultrapassar esta limitação, é introduzida uma formulação VAE de tamanho variável. Esta

arquitetura utiliza codificação partilhada dos nós, máscaras de preenchimento, agregação ao nível do grafo, descodificação par-a-par dos ramos e projeção numa árvore radial, permitindo que um único modelo processe redes com diferentes números de nós. É também avaliada uma extensão inspirada em *Transformers*, na qual as representações dos nós são contextualizadas através de interações globais antes da agregação. Os resultados mostram uma melhor generalização para dimensões de grafo *held-out*, isto é, números de nós não representados durante o treino. A extensão do intervalo de treino para redes maiores melhora igualmente a extrapolação, embora a reconstrução de estruturas substancialmente maiores do que as utilizadas no treino permaneça desafiante.

É igualmente estudada uma formulação de completação de grafos, na qual o modelo recebe informação estrutural parcial e deve inferir os ramos em falta. A comparação entre uma entrada baseada apenas na adjacência observada e outra que a combina com sensibilidades mostra que ambas as fontes de informação são complementares. A adjacência parcial restringe o conjunto de soluções possíveis, enquanto as sensibilidades fornecem orientação elétrica sobre as relações entre os nós.

Por fim, os ensaios de robustez avaliam erros de medição aplicados à potência ativa, ao módulo da tensão e simultaneamente a ambas as grandezas. No intervalo analisado, as perturbações apenas em potência ativa apresentam um efeito comparativamente limitado, enquanto os erros de tensão provocam uma degradação mais acentuada da precisão dos ramos. O caso combinado de P+V segue essencialmente o comportamento da componente de tensão. O mesmo padrão qualitativo é observado numa implementação independente de um método analítico baseado em resistências efetivas e agrupamento recursivo. Contudo, a variabilidade dos resultados e as diferenças entre os protocolos experimentais não permitem estabelecer a superioridade geral de uma das abordagens.

De forma geral, a dissertação apresenta um estudo experimental reprodutível de modelos VAE e de modelos baseados em grafos para inferência de topologias em redes de baixa tensão. Os resultados indicam que representações fisicamente informadas, projeção radial e arquiteturas de tamanho variável contribuem para melhorar a reconstrução a partir de dados P/V. Simultaneamente, o estudo evidencia limitações relacionadas com a generalização para estruturas e dimensões não observadas, a qualidade das medições de tensão, a presença de nós não observados e a escalabilidade. Os resultados devem, por isso, ser interpretados como evidência obtida num ambiente controlado, sendo necessária validação adicional em redes reais antes de uma aplicação operacional.

Abstract

Low-voltage distribution networks face increasing monitoring, management and control requirements due to the growing penetration of distributed energy resources, electric vehicles and flexible loads. A key requirement for these tasks is knowledge of the network topology, that is, how customers, buses and branches are effectively connected at a given time. In practice, topology records may be incomplete or outdated, while the available information may be limited to active-power and voltage-magnitude time series collected at customer level. Although these measurements do not directly describe the network branches, they contain indirect structural information because voltage variations depend on shared electrical paths and on the network response to power variations. Nevertheless, different radial structures may produce similar electrical patterns, making connectivity inference a difficult inverse problem.

This dissertation studies topology inference in low-voltage networks using learning-based models and a controlled dataset that combines synthetically generated radial structures with residential consumption profiles derived from real measurements. The experimental methodology generates the networks and assigns electrical parameters to their branches. The load buses are then associated with active-power profiles obtained from 30-minute residential time series from the *Irish Smart Metering Customer Behaviour Trial*. For each case, voltage magnitudes are computed through Forward–Backward Sweep power-flow simulation, representing the electrical response of the feeder to the considered consumption profiles. This procedure provides a reproducible environment with known ground-truth connectivity and physically coherent electrical behaviour, allowing different measurement-derived representations and VAE architectures to be evaluated for radial-network reconstruction.

The work first analyses an initial fixed-size Variational Autoencoder using raw active-power and voltage-magnitude trajectories as input. The results show a clear difference between *seen* and *unseen* topology reconstruction: the model adequately recognises structures represented during training but generalises poorly to new radial configurations. Diagnostic experiments were conducted on the organisation of the latent space, the amount of training data, the temporal horizon, regularisation and architectural choices. These experiments indicate that the generalisation difficulty does not result solely from an isolated hyperparameter choice.

Convolutional VAE models and structured measurement representations are then evaluated, including correlation matrices, power–voltage correlation blocks and sensitivity matrices. The results show that the representation of the electrical information has a significant impact on reconstruction. In the controlled comparisons, sensitivity-based inputs provide the most useful information for branch recovery, whereas raw P/V trajectories achieve lower performance. However, fixed-size models still require a separate architecture for each network dimension and their performance degrades as the number of nodes increases.

To overcome this limitation, a variable-size VAE formulation is introduced. This architecture uses shared node encoding, padding masks, graph-level aggregation, pairwise branch decoding and radial-tree projection, allowing a single model to process networks with different numbers

of nodes. A *Transformer*-inspired extension is also evaluated, in which node representations are contextualised through global interactions before aggregation. The results show improved generalisation to *held-out* graph sizes, that is, numbers of nodes not represented during training. Extending the training range to larger networks also improves extrapolation, although reconstructing structures substantially larger than those used during training remains challenging.

A graph-completion formulation is also studied, in which the model receives partial structural information and must infer the missing branches. The comparison between an input based only on the observed adjacency and another that combines it with sensitivity information shows that the two sources are complementary. Partial adjacency restricts the set of possible solutions, whereas sensitivities provide electrical guidance regarding the relationships between nodes.

Finally, the robustness experiments evaluate measurement errors applied to active power, voltage magnitude and both quantities simultaneously. Within the analysed range, perturbations affecting only active power have a comparatively limited effect, whereas voltage errors cause a stronger degradation in edge accuracy. The combined P+V case largely follows the behaviour of the voltage component. The same qualitative pattern is observed in an independent implementation of an analytical method based on effective resistances and recursive grouping. However, the variability of the results and the differences between the experimental protocols do not support a general claim of superiority for either approach.

Overall, this dissertation presents a reproducible experimental study of VAE and graph-based models for topology inference in low-voltage networks. The results indicate that physically informed representations, radial projection and variable-size architectures contribute to improving reconstruction from P/V data. At the same time, the study highlights limitations related to generalisation to unseen structures and graph sizes, voltage-measurement quality, the presence of unobserved nodes and scalability. The results should therefore be interpreted as evidence obtained in a controlled environment, and further validation on real networks is required before operational deployment.

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework to achieve a better and more sustainable future for all. They include 17 goals to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

This dissertation is aligned with the need to improve the monitoring, modelling, and control of low-voltage electricity distribution networks under increasing penetration of distributed energy resources, electric vehicles, storage systems, and other flexible loads. The proposed work studies learning-based topology inference using smart-meter-like active-power and voltage-magnitude data, with emphasis on VAE-based models, structured electrical representations, variable-size architectures, robustness to measurement errors, and graph-completion settings. By enabling more reliable topology awareness, the work supports better voltage regulation, congestion management, loss reduction, validation of inferred network structures, and ultimately a more efficient and resilient operation of electricity systems.

The specific Sustainable Development Goals mentioned have the following names:

SDG 7 Ensure access to affordable, reliable, sustainable and modern energy for all

SDG 9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

SDG 11 Make cities and human settlements inclusive, safe, resilient and sustainable

SDG 13 Take urgent action to combat climate change and its impacts

SDG	Target	Contribution	Performance Indicators and Metrics
7	7.2	The dissertation supports the integration of renewable and distributed energy resources by improving topology awareness in low-voltage grids, which is essential for the reliable operation of networks with photovoltaic generation, storage, and electric vehicles.	Ability to infer radial topologies under limited observability; edge accuracy on *unseen* test topologies; robustness of topology inference under measurement uncertainty; improvement in network observability for DER-rich feeders.
	7.3	By enabling more accurate feeder models, the work supports better operational decisions related to voltage regulation, congestion management, and technical loss reduction.	Reconstruction accuracy; edge error of inferred topologies; quality of inferred models for power-flow validation; potential support for loss and voltage studies in simulation settings.
9	9.1	The proposed framework contributes to more reliable, resilient, and sustainable electricity infrastructure through data-driven topology inference and validation methods adapted to realistic smart-meter limitations.	Robustness to active-power and voltage measurement errors; scalability with network size; generalisation to *held-out* graph sizes; validation of inferred radial structures through topology metrics.
	9.5	The dissertation advances innovation in smart grid analytics by evaluating learning-based topology inference models, including fixed-size VAE, convolutional VAE, variable-size VAE, Transformer-inspired and graph-completion formulations, within a reproducible experimental framework.	Development of a reproducible benchmarking framework; comparison between fixed-size and variable-size learning models; evaluation of input representations; robustness to measurement errors and scalability with network size.
11	11.6	Smarter and more reliable low-voltage network operation contributes indirectly to more sustainable cities by supporting the electrification of urban loads and the efficient integration of distributed clean energy technologies.	Performance of topology inference in urban-style LV feeders; ability to support voltage quality and congestion-related studies; suitability for operation under EV and DER penetration scenarios.
13	13.1	Although indirect, the work strengthens the technical basis for climate action by enabling grid operation that can better accommodate low-carbon technologies and adapt to increasingly dynamic electrical conditions.	Capacity to operate under diverse operating conditions; robustness of inferred topologies under measurement uncertainty; support for resilient planning and operation under high DER penetration.

Acknowledgements

How can I express my gratitude not only for the support throughout this final chapter of my life as a student, but also throughout the last five years?

Sometimes, we take life and its circumstances for granted, as though they were ordinary and certain. Being here today, in the final days before submitting MY THESIS in Electrical Engineering at FEUP, seemed impossible to the 18-year-old Ana Maria, from a small town and a humble background, who was told so many times that she would never be capable of getting this far. The Ana Maria who thought about giving up so many times, and who genuinely believed that she would have to leave everything behind because of her love–hate relationship with programming. Thank you, AI! Since I am talking about myself, my favourite topic, I must highlight my perseverance, my ability to face obstacles and things that seem impossible, and my positivity when dealing with those obstacles. In short, believing in myself above everything and everyone else, and knowing that, when I truly want something, I can achieve it. Future Ana Maria, if you are reading this... why are you reading this? And well done, girl.

Next, I would like to thank my parents. To my mother, Natália, for always being proud of me, even for the small things, including those that I barely noticed myself, for motivating me when I needed it most and comforting me when things became difficult, for being the first person to know how far I could go, and for clipping her own wings so many times so that I could fly. To my father, António, for being my financial support throughout the last 23 years—which is also important—for teaching me to be strong and independent from the very beginning, and for always standing beside my mother, even contributing two or three words to the discussion when the subject became more serious. I do not know whether I will ever be able to repay all the love and pride you give me every day. I hope that one day I can become as good and hardworking as you are, even if only 1% of what you are. Waking up before sunrise, returning home after sunset, never stopping at weekends or during holidays... and still listening to me complain about my part-time job at the end of it all. A simple thank you is not enough.

Next, I would like to thank my grandparents, Alberto and Clarinda. The love I feel for them is different from the love I feel for my parents. While a parent loves with responsibility, a grandparent loves without thinking about what could go wrong. I saw that in them many times. Sometimes I threw myself headfirst into projects that could either go extremely well or extremely badly, but there was never any doubt in their eyes. There was always the belief that, "Se a Ana vai, a Ana vai conseguir". They were my greatest confidants throughout this journey, the people I told everything to first and who always applauded me, even when they did not fully understand what they were applauding. They are just as happy about the submission of this thesis and the end of these five years of work as they were when I got my first job in a restaurant, because both were victories and reasons to be proud.

To my godfather, Fernando, and my cousin Teresa, for giving me some of the best moments I can remember, each in their own way, for travelling all over Portugal together and for the endless memories we will share for the rest of our lives. To my godfather in particular, for creating in me

a connection with God and for teaching me to become the best version of myself every day, and to my cousin, for being my first friend, my confidante and, at times, even my arch-enemy.

To my grandfather Manuel and my grandmother Alzira. To my grandmother, for taking care of me so many times when my parents could not and for passing on her teachings to me. And to my grandfather, who watches over me from heaven every day, whose presence I feel so often, and who I know would be overflowing with happiness if he were here with us today.

A special thank you also goes to the rest of my family, who may not be as close but are equally important for always being there whenever I needed them, particularly Aunt Natália—the second one—and my uncle Rogério, for the many hours of advice and support, as well as everyone else.

Now, moving on to the group that CARRIED ME through all these years, the "Engenheiros Aspirantes". Although the group may have changed slightly in size and form over the years, it was always my main source of academic support. As my friend Diogo used to say—and he also changed his opinion over time—friends are the people who grow up with us, while university friends are merely classmates. I used to think the same way. Until it was three in the morning and we were studying together at the Faculty, or ten minutes before an exam and we were reviewing the material in the library. We suffered together, overcame challenges together and shared some of the most beautiful moments together. These friendships went beyond the academic context and became the names that will one day appear on the wedding guest list. Envy or jealousy is ultimately something intrinsic to human beings, and so many times I unintentionally felt jealous of someone close to me... I know that what I feel for them is unconditional love because I genuinely feel nothing but happiness when I see them grow more and more and become better versions of themselves. May this friendship never end, and if it does... "a gente mete mais gasóleo".

To Mara Mota, for, in a way, making me the person I am today. I was not always as determined or, as my friends call me, "desenrascada", as I am now. I owe all of that to her. Thank you for being my "mãe" and my best friend throughout all these years, and for ensuring that we were never absent when the other needed us.

To the many friendships I created at the Faculty, and to the groups in which I was involved, particularly NEEEC, which brought Zé Sousa into my life, also known as the best moustache in the course and now a lecturer. He believed in me, brought me into this project and is the reason why this thesis exists. The affection he shows everyone is clear, and the affection I have for him will last a lifetime. To FAST, where I met João Alegre, one of the most pleasant surprises of recent times, who taught me so much, believed in me so strongly and was always there during the darkest days. He truly went from being a classmate to becoming one of the most important people in my life, someone I hope to keep with me forever. A special thank you to the uncles and aunts for their advice, kind words and countless dinners when I no longer had the strength for anything else. To the other associations and to the connections I created through them, my most sincere thanks.

Finally, and turning to what brought us here, I would like to thank my supervisor, Ricardo Bessa, for his attentive guidance and for motivating me whenever I felt that the work was not good enough. To Gil Sampaio, for officially being the last person in the world willing to read a 150-page thesis word by word, for advising me, for often telling me things I did not want to hear but which made me better, for keeping me in line whenever I was about to stray from the path, and for giving me the best possible first experience in my first official job.

Calm down, will you be the last one, David? You know what they say, the last shall be first. I know you are not the greatest fan of my optimism, but here are a few words. As officially your first student, I cannot express how grateful I am to you and for all the help you gave me. Sometimes I forgot things that were so simple and ridiculous, such as Ohm's law, and you would simply breathe, count to three and remind me of the FUNDAMENTALS AND THE PHYSICS behind the law. Who would have thought that two opposites could work so well together and, "quiçá",

might even miss working with one another? To my companion during countless afternoons of work and many coffee breaks, to my scientific “sensei”, and to the owner of some of the worst handwriting I have ever seen—but that is fine because everything is done on a computer now—my most sincere thanks. During these six months, we never had lunch together. Perhaps the success of this thesis will finally change that...

May this be only the first step towards my success, but never the last moment I share with all of you. Thank you all!

Ana Maria Alves de Almeida

"Todo en esta vida tiene solución, solo no hay solución para la muerte."

Cristiano Ronaldo

Contents

1 Introduction **1**
- 1.1 Modelling Low-Voltage Distribution Networks . . . 5
- 1.2 Problem Motivation . . . 7
- 1.3 Research Gap . . . 9
- 1.4 Research Questions and Objectives . . . 10
- 1.5 Contributions . . . 11
- 1.6 Dissertation Structure . . . 12

2 Literature Review **14**
- 2.1 Distribution-Grid Topology Learning: Scope and Terminology . . . 14
- 2.2 Measurement Sources and Limited Observability in LV Grids . . . 16
- 2.3 Physics-Based Modelling, Radiality and Sensitivity Relations . . . 17
- 2.4 Smart-Meter and AMI-Based Topology Inference . . . 19
- 2.5 Hidden Nodes and Identifiability Limits . . . 22
- 2.6 Admittance and Impedance Matrix Estimation . . . 25
- 2.7 Probabilistic, Bayesian and Online Topology-Tracking Methods . . . 27
- 2.8 Generative and Attention-Based Models for Topology Inference . . . 29
  - 2.8.1 From Deterministic Estimates to Edge-Score-Based Inference . . . 29
  - 2.8.2 Variational Autoencoders . . . 30
  - 2.8.3 Convolutional Variational Autoencoders . . . 31
  - 2.8.4 Graph Generative Models . . . 31
  - 2.8.5 Transformer Encoders and Self-Attention . . . 32
  - 2.8.6 Implications for LV Topology Inference . . . 33
- 2.9 Research Gap and Positioning of the Dissertation . . . 34
- 2.10 Summary . . . 36

3 Synthetic Data Generation and Modelling Framework **37**
- 3.1 Motivation for Synthetic Data Generation . . . 38
- 3.2 Synthetic Radial Network Generation . . . 39
- 3.3 Preliminary Physical Assumptions . . . 40
- 3.4 Validation of the Modelling Assumptions . . . 42
  - 3.4.1 Empirical and Theoretical Variance . . . 42
  - 3.4.2 Acceptable Voltage-Drop Scenarios . . . 45
  - 3.4.3 Impact of Phase Angle Neglect and the X/R Ratio . . . 47
- 3.5 Residential Active-Power Profile Generation . . . 48
- 3.6 Forward-Backward Sweep Power-Flow Simulation . . . 49
- 3.7 Validation of the Synthetic Electrical Data . . . 51
- 3.8 Construction of the Synthetic P/V Dataset . . . 52

3.9 Topology Inference Modelling Pipeline . . . 53
3.10 Topology Evaluation Metrics . . . 54
3.11 Summary . . . 55

4 Exploratory Modelling and Results **57**
4.1 Initial Fixed-Size VAE Baseline . . . 58
4.2 Fixed-Size VAE Refinement and Convolutional VAE Results . . . 62
4.2.1 Effect of the Number of Training Topologies . . . 63
4.2.2 Fixed-Budget Trade-Off Between Records and Temporal Horizon . . . 64
4.2.3 Transition to a Fixed-Size Convolutional VAE . . . 65
4.3 Input Representation Study . . . 67
4.3.1 Correlation and Sensitivity Matrix Construction . . . 67
4.3.2 Exploratory Analysis of Correlation Blocks . . . 68
4.3.3 Controlled Comparison Between PV Correlations and Sensitivity Matrices 70
4.3.4 Two-Encoder Fusion of Complementary Representations . . . 71
4.3.5 Summary of the Input-Representation Study . . . 72
4.4 Variable-Size VAE Topology Inference . . . 73
4.4.1 Variable-Size Experimental Setup . . . 76
4.4.2 Variable-Size VAE Results . . . 77
4.5 Transformer-Inspired Variable-Size VAE Results . . . 78
4.6 Graph Completion with Partial Adjacency and Sensitivity Guidance . . . 82
4.7 Robustness to Measurement Errors . . . 84
4.7.1 Comparison with an Analytical Effective-Resistance Method . . . 86
4.8 Integrated Discussion . . . 89

5 Conclusions and Future Work **92**
5.1 Limitations . . . 93
5.2 Future Work . . . 93
5.3 Final Remarks . . . 94

References **96**

A Supplementary Physical Derivations **100**
A.1 Apparent Resistance Estimate for a Leaf Branch . . . 100
A.2 Extension to Non-Leaf Branches . . . 101
A.3 Voltage-Noise Propagation . . . 101
A.4 Theoretical Variance . . . 102
A.5 Experimental Validation of the Variance Expression . . . 103
A.6 Example of a Rejected Voltage-Drop Scenario . . . 103
A.7 Summary . . . 104

B Supplementary Initial Fixed-Size VAE Diagnostics **105**
B.1 Latent-Space kNN Forcing . . . 105
B.2 PCA-Based Latent-Space Forcing . . . 108
B.3 Nearest Latent-Family-Mean Forcing . . . 110
B.4 Latent-Dimension Sweep . . . 111
B.5 Summary . . . 112

C Supplementary Regularisation and Architecture Experiments **113**
C.1 KL Regularisation . . . 113
C.2 Dropout and Weight Decay . . . 115
C.3 Architecture Sweep . . . 116
C.4 Summary . . . 117

D Supplementary Fixed-Size Convolutional VAE Diagnostics **118**
D.1 Temporal-Budget Allocation . . . 118
D.2 Unseen Behaviour Under an Extended Temporal Trade-Off . . . 119
D.3 Random and Locally Perturbed Unseen Topologies . . . 120
D.4 Summary . . . 120

E Supplementary Input-Representation and Variable-Size Results **121**
E.1 Input-Representation Summary . . . 121
E.2 Examples of Full Correlation Matrices . . . 122
E.3 Variable-Size Generalisation . . . 124
E.4 Transformer-Inspired Extension . . . 124
E.5 Summary . . . 124

F Supplementary Robustness Results **125**
F.1 Correlation-Based Robustness . . . 125
F.2 Transformer-Inspired Model: Edge-Error View . . . 126
F.3 Summary . . . 126

G Supplementary Implementation Details **127**
G.1 Data Structure . . . 127
G.2 Topology Output and Radial Projection . . . 127
G.3 Evaluation Metrics . . . 127
G.4 Seen and Unseen Evaluation . . . 128
G.5 Fixed-Size and Variable-Size Training . . . 128
G.6 Summary . . . 128

List of Figures

1.1 Illustrative transition from a passive low-voltage feeder to an active low-voltage feeder with distributed energy resources, flexible loads and bidirectional power flows. The figure highlights the increasing operational complexity, variability and observability requirements associated with modern LV grids. 2

1.2 Voltage-profile comparison between a passive distribution network (PDN) and an active distribution network (ADN). In the passive case, voltage typically decreases along the feeder due to downstream loading, whereas active operation may produce local voltage rises caused by distributed generation. Extracted from [6]. . . 3

1.3 Conceptual illustration of topology ambiguity under imperfect smart-meter measurements. The same customer-level active-power and voltage-magnitude measurements may be compatible with multiple physically plausible radial topologies, motivating an inference framework that preserves topology-related uncertainty rather than relying only on a single direct reconstruction. 5

2.1 Graph-theoretic representation of a radial distribution feeder, including the root node, leaf nodes, incidence matrices and the relationship between feeder paths and matrix entries. Extracted from [7]. 15

2.2 Line infrastructure for the IEEE 37-bus feeder benchmark used in voltage-analytics-based topology verification. The figure illustrates the difference between the available line infrastructure and the operational topology to be verified. Extracted from [11]. 21

2.3 Overview of a full-scale distribution-system topology-identification framework based on smart-meter data processing, Markov Random Field structure learning and topology generation. Extracted from [28]. 21

2.4 Example of smart-meter-based radial topology reconstruction using voltage sensitivity coefficients and candidate graph generation. Extracted from [8]. 22

2.5 Illustration of a distribution network with observed nodes and hidden nodes. The figure highlights the limited-observability setting in which only part of the network is metered, while other buses remain hidden. Extracted from [17]. 23

2.6 Illustration of hidden or passive nodes and the resulting reduced network representation used in unsupervised impedance and topology-estimation settings. The figure supports the distinction between the full physical feeder and the equivalent representation available from measurements at observed nodes. Extracted from [18]. 24

2.7 Variational graph-generation architecture in which a discrete graph is encoded into a continuous latent representation and decoded into a probabilistic graph. Extracted from [50]. 31

3.1 Overall pipeline of the dissertation. The methodology starts with the synthetic generation of radial LV networks and Forward-Backward Sweep power-flow simulation to obtain P/V datasets. The subsequent stages, including feature extraction, model training, edge-score generation and radial-tree projection, are analysed in the experimental results chapter. 38
3.2 Example of a radial LV test grid used as inspiration for the synthetic radial networks considered in this dissertation. The original test-grid representation was adapted from [20] and redrawn to illustrate the type of feeder structure used in the synthetic P/V dataset generation process. 40
3.3 Comparison between the empirical and theoretical variance of the apparent-resistance estimate for terminal branch $4-14$ of a randomly generated fixed radial feeder. The base scenario considers active power only, voltage magnitudes and a purely resistive branch. 44
3.4 Comparison between the empirical and theoretical variance of the apparent-resistance estimate for terminal branch $7-10$ of the same randomly generated fixed radial feeder, after including line reactance and voltage-angle information. 45
3.5 Diagnostic associated with an acceptable voltage-drop scenario. The resistance-estimation error remains smooth as the X_{ij}/R_{ij} ratio increases, indicating that the simulated voltage behaviour remains suitable for the simplified validation setting. 46
3.6 Error introduced by neglecting voltage phase and standard deviation of the resistance estimates as a function of the X_{ij}/R_{ij} ratio. The result supports the use of voltage-magnitude-based approximations when the systematic phase-related error remains small compared with the expected measurement uncertainty. 47
3.7 Difference between voltage magnitudes obtained with and without line reactance as a function of the X_{ij}/R_{ij} ratio. The increasing trend confirms that the impact of reactance and phase becomes more relevant as the line becomes more reactive. . 48
3.8 Single-line diagram of the representative 23-bus radial feeder used in the synthetic-data validation. 51
3.9 Voltage-magnitude trajectories of the representative 23-bus radial feeder over a 96-timestamp segment. 52

4.1 Simplified evolution of the experimental path reported in this chapter. 57
4.2 Mean edge accuracy over 50 inference topologies as a function of the number of training topologies. Error bars represent one standard deviation of the edge accuracy values across the 50 inference cases. 63
4.3 *Seen* edge accuracy under a fixed data-budget trade-off between the number of training records and the temporal horizon T. 64
4.4 *Unseen* edge accuracy under a fixed data-budget trade-off between the number of training records and the temporal horizon T. 65
4.5 Gradient-based input-saliency map for the full correlation representation. Larger values indicate entries of the correlation matrix to which the model output is more sensitive. 69
4.6 Transition from fixed-size to variable-size VAE topology inference. 74
4.7 Robustness of the sensitivity-based representation under measurement errors. . . 85
4.8 Robustness of the Transformer-inspired variable-size VAE under measurement errors. 86

4.9 Comparison of mean edge accuracy under measurement perturbations for the sensitivity-based VAE and the analytical effective-resistance and recursive-grouping method. The VAE results include the training-and-inference and inference-only noise scenarios. The analytical method was independently applied to 50 noise realisations of a fixed 10-node radial network, and the error bars represent one standard deviation. In the figure legend, "single run" denotes a non-trained analytical reconstruction performed independently for each realisation, rather than a single Monte Carlo sample. 88

A.1 Example of a rejected voltage-drop scenario. The diagnostic shows the resistance-estimation error as a function of the X_{ij}/R_{ij} ratio for a case in which the voltage behaviour is not suitable for inclusion in the final synthetic P/V dataset. 104

B.1 Latent-space kNN forcing diagnostic for the original case. The PCA projection is used only to visualise the unseen sample, selected neighbours and neighbouring family. 106
B.2 Latent-space kNN forcing diagnostic for repetition 1. The unseen sample is assigned to a neighbouring family in the full latent space, but the recovered topology remains only partially correct. 107
B.3 Latent-space kNN forcing diagnostic for repetition 2. The projection illustrates that latent proximity may correspond to partial edge overlap, but not to exact topology recovery. 107
B.4 PCA-based latent-space forcing, case 1. The selected training point improves partial edge recovery, but exact topology recovery is still not achieved. 109
B.5 PCA-based latent-space forcing, case 2. The selected point is close in the PCA projection, but forcing reduces the edge accuracy relative to the original unseen reconstruction. 109
B.6 PCA-based latent-space forcing, case 3. Despite the very small PCA distance, the forced topology has low edge accuracy, showing that PCA proximity is not a reliable proxy for graph similarity. 110
B.7 Latent-dimension sweep with a fixed dataset, evaluated using edge accuracy. . . . 112

C.1 High-range sweep of the maximum KL-divergence weight. The curves report the mean *seen* and *unseen* edge accuracy, while the error bars summarise the variation between repeated runs. 114
C.2 Unseen edge accuracy for the dropout and weight-decay sweep. Rows represent dropout values, columns represent weight-decay values, and each cell reports the mean edge accuracy across three random seeds. 116
C.3 Architecture sweep within the initial fixed-size VAE regime. The bars report the mean *seen* and *unseen* edge accuracy across three random seeds, while the error bars summarise the variation between runs. 117

D.1 Fixed-budget trade-off between longer trajectories and more records per topology. The result evaluates whether performance is more affected by temporal depth or by the number of operating realisations. 119
D.2 Extended fixed-budget trade-off for unseen edge accuracy. The result shows that changing the temporal allocation affects performance, but does not fully solve unseen topology reconstruction. 119

D.3 Comparison between fully random unseen topologies and two-edge-perturbed unseen topologies. The result shows that being close to a seen topology does not necessarily make unseen reconstruction reliable. 120

E.1 Representative full correlation matrix for a 10-node network using 2000 training topologies. The matrix has dimension 20×20. 122
E.2 Representative full correlation matrix for a 20-node network using 5000 training topologies. The matrix has dimension 40×40. 123
E.3 Representative full correlation matrix for a 30-node network using 5000 training topologies. The matrix has dimension 60×60. 123

F.1 Robustness of the correlation-based representation under relative measurement errors between 0% and 1%. The six curves compare perturbations affecting P, V, and both P and V, applied either only during inference or during both training and inference. 125
F.2 Robustness of the Transformer-inspired variable-size VAE: mean edge error versus measurement error window. The curves compare errors in P, V, and both P and V, applied either only during inference or during both training and inference. . . 126

List of Tables

2.1 Measurement sources considered in distribution-grid topology inference. 17
2.2 Summary of topology-learning approaches and their limitations for LV smart-meter-based topology inference. 35

4.1 *Seen* and *unseen* edge accuracy of the initial fixed-size VAE baseline. 60
4.2 Mean edge accuracy of the latent-space diagnostics for the initial 10-node fixed-size VAE. 61
4.3 Summary of additional diagnostics within the initial fixed-size VAE regime. . . . 62
4.4 Fixed-size edge accuracy for the original VAE and convolutional VAE using raw P/V inputs. 66
4.5 Comparison between full correlation and PV-only correlation representations across fixed-size networks. 70
4.6 Controlled comparison between PV-only correlation and sensitivity representations across fixed-size networks. 71
4.7 Two-encoder fusion results across fixed-size networks. 72
4.8 Mean edge accuracy of the variable-size VAE for raw P/V, sensitivity and correlation representations across *training* and *held-out* graph sizes. 77
4.9 Transformer-inspired variable-size VAE results with sensitivity-based inputs. . . 80
4.10 Comparison between the initial variable-size VAE and the Transformer-inspired variable-size VAE on *held-out* graph sizes. 80
4.11 Transformer-inspired variable-size VAE after extending the training range up to 50 nodes. 81
4.12 Inference results of the graph-completion architecture on graph sizes *seen* during training. 83
4.13 Generalisation results of the ADJ + SENS graph-completion architecture on *held-out* graph sizes. 83

B.1 Mean results of the latent-space kNN forcing diagnostic. 106
B.2 Mean results of the PCA-based latent-space forcing diagnostic. 108
B.3 Mean results of the nearest latent-family-mean forcing diagnostic. 111

C.1 Overview of supplementary regularisation and architecture diagnostics. 113

E.1 Summary of the input representations evaluated in the topology inference experiments. 121
E.2 Graph-size split used in the variable-size VAE experiments. 124

List of Algorithms

1 Forward-Backward Sweep power-flow simulation for one topology. 50
2 Training and inference procedure of the initial fixed-size VAE. 59

List of Acronyms

AC	Alternating Current
ADJ	Adjacency Input
AMI	Advanced Metering Infrastructure
CVAE	Conditional Variational Autoencoder
DER	Distributed Energy Resource
DSO	Distribution System Operator
EA	Edge Accuracy
EE	Edge Error
ER	Exact Topology Recovery
EV	Electric Vehicle
HV	High Voltage
IEEE	Institute of Electrical and Electronics Engineers
KL	Kullback-Leibler Divergence
kNN	k-Nearest Neighbours
LV	Low Voltage
MILP	Mixed-Integer Linear Programming
MV	Medium Voltage
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PMU	Phasor Measurement Unit
PV	Photovoltaic
P/V	Active-Power and Voltage-Magnitude
RQ	Research Question
SCADA	Supervisory Control and Data Acquisition
SDG	Sustainable Development Goal
SENS	Sensitivity Input
UN	United Nations
VAE	Variational Autoencoder

Chapter 1

Introduction

Electric power systems are undergoing a structural transformation driven by decarbonisation, electrification, digitalisation and the integration of distributed energy resources. For most of their history, electricity networks were planned and operated according to a predominantly centralised paradigm, in which large generation units supplied passive consumers through transmission and distribution infrastructures designed for predictable and mostly unidirectional power flows. Within this architecture, low-voltage (LV) distribution networks were typically regarded as the final passive layer of the power system, responsible for delivering electricity from secondary substations to residential, commercial and small industrial consumers. Their operation relied on conservative planning assumptions, limited field monitoring and static network records.

Within this changing operational context, accurate knowledge of the grid topology becomes essential. In this dissertation, topology refers primarily to the connectivity of the low-voltage feeder, that is, to the way in which the transformer, electrical buses, customers and branches are connected at a given time. An inaccurate or outdated connectivity model can compromise the interpretation of measurements and the reliability of monitoring, simulation and operational decision-support tools.

This traditional paradigm is no longer sufficient. The increasing penetration of distributed generation, photovoltaic (PV) systems, electric vehicles (EVs), batteries, heat pumps and other flexible loads is changing the physical and operational behaviour of LV grids. Power flows can become bidirectional, voltage profiles can become more variable and less predictable over short time intervals, and local operating constraints may be violated by individual or aggregated customer behaviour. As a result, LV grids are progressively becoming active operational environments rather than passive delivery infrastructures. This evolution creates opportunities for flexibility, decarbonisation and local energy management, but it also increases the need for accurate observability, modelling and control of the LV network [1], [2], [3], [4], [5].

Figure 1.1 provides a conceptual illustration of this transition. In the passive case, the feeder is represented as a radial network with one-way power flow from the transformer towards passive customers. In the active case, the same type of LV feeder includes PV generation, EV charging and flexible demand, leading to bidirectional flows and more complex local interactions.

From an operational perspective, this transition also changes the type of information that Distribution System Operators (DSOs), the entities responsible for operating, maintaining and planning electricity distribution networks, need in order to manage LV networks effectively. In passive feeders, approximate planning models and conservative assumptions were often sufficient for long-term operation. In active feeders, however, local injections and flexible demand can alter voltage profiles, branch loading and the direction of power flows over short time intervals. This means that the physical model of the feeder must be sufficiently accurate and up to date for monitoring, simulation and decision-support tools. If the assumed connectivity does not match the actual network, the interpretation of voltage measurements, customer behaviour and local flexibility potential may become inaccurate.

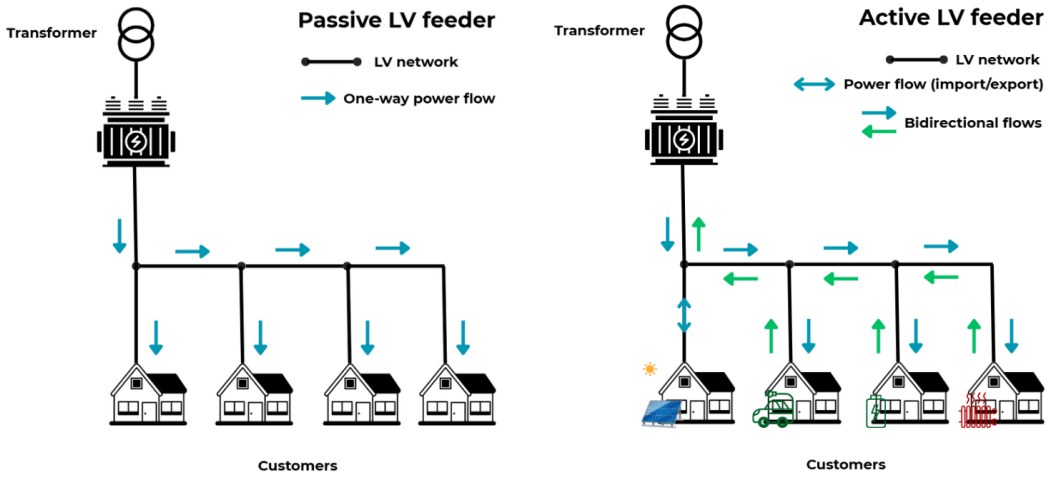


Figure 1.1: Illustrative transition from a passive low-voltage feeder to an active low-voltage feeder with distributed energy resources, flexible loads and bidirectional power flows. The figure highlights the increasing operational complexity, variability and observability requirements associated with modern LV grids.

The operational impact of this transition extends beyond the behaviour expected in traditional passive feeders. Radwan et al. [6] compare voltage profiles in passive and active distribution networks, showing how distributed generation and changing power-flow directions can modify voltage behaviour along the feeder.

In a traditional passive distribution network, power flows mainly from the upstream transformer towards downstream consumers. The voltage profile is therefore typically characterised by progressive voltage drops along the feeder, since the current supplying downstream demand produces losses and voltage reductions. Consequently, maintaining acceptable voltage levels at electrically remote customers was one of the principal operational concerns.

In active distribution networks, a monotonic voltage decrease along the feeder is no longer guaranteed. Distributed generation, particularly photovoltaic generation connected at low-voltage level, can reduce or reverse the net power flow in some branches. Local voltage rises may consequently occur during periods of high generation and low demand. Figure 1.2 illustrates this

distinction.

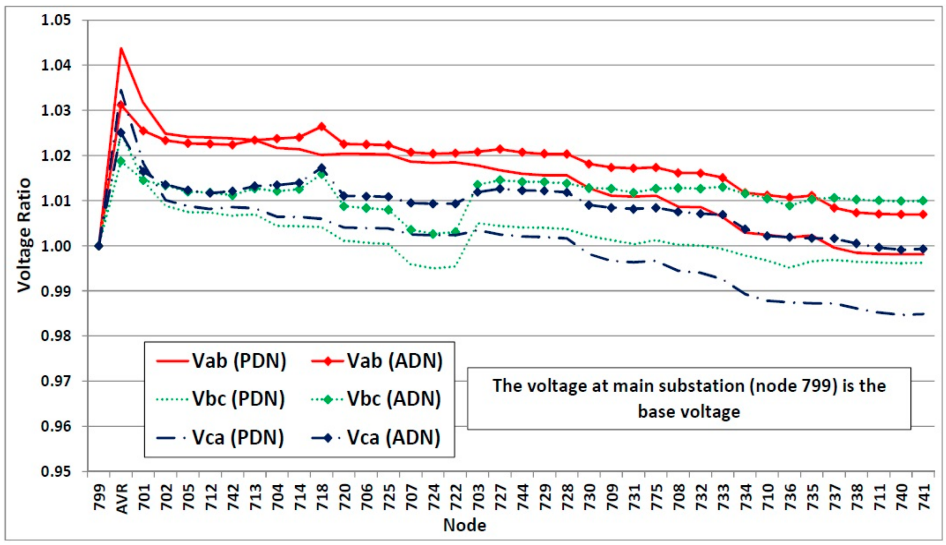


Figure 1.2: Voltage-profile comparison between a passive distribution network (PDN) and an active distribution network (ADN). In the passive case, voltage typically decreases along the feeder due to downstream loading, whereas active operation may produce local voltage rises caused by distributed generation. Extracted from [6].

Electrical parameters such as branch impedances are required for the power-flow simulations used to construct the controlled dataset, but their estimation is outside the scope of the proposed learning-based topology-inference models. These parameters are assigned during data generation and are treated as known inputs to the power-flow simulation. This connectivity information is essential for power-flow analysis, voltage estimation, loss calculation, state estimation, fault localisation, congestion management and hosting capacity assessment [7], [8], [9].

This requirement is particularly important in active LV networks, where incorrect connectivity can make voltage analysis, loss estimation and operational conclusions biased, physically inconsistent or unsuitable for decision-making, even when the underlying algorithms are mathematically sound.

In practice, LV topology information is frequently incomplete, outdated or inconsistent with reality. DSOs may possess records of the planned or documented infrastructure, but these records are not always synchronised with field interventions, incremental customer connections, maintenance actions, changes in customer phase allocation, underground cable modifications or emergency restoration procedures. This problem is particularly relevant in LV networks because they are numerous, geographically dispersed, often underground and historically less monitored than medium-voltage or transmission grids. Consequently, the digital model available to the operator may diverge from the actual topology.

This uncertainty has motivated a growing body of data-driven methods for distribution-grid topology identification, topology verification, phase identification and parameter estimation. Instead of relying exclusively on static network records, these approaches attempt to infer structural

information from measurements and operational data. Depending on the available instrumentation, the literature explores several sources of information and modelling strategies, including voltage measurements, smart-meter data, phasor measurements, micro-PMUs, probing signals, correlation measures, graphical models, sparse regression and learning-based methods [7], [8], [10], [11], [12], [13], [14].

Among these data sources, smart meters are particularly relevant for LV networks because they are already being deployed at customer connection points and provide a scalable source of measurements without requiring dedicated sensors at every internal branch or bus. For example, in Portugal, this deployment is especially relevant because the integration of all normal low-voltage customers into smart grids was established as a national objective for 2024, and E-REDES reports the installation of approximately 6.6 million smart meters across its distribution network [15], [16].

Smart meters were initially associated mainly with billing, remote meter reading and consumer-side energy management. However, their role has expanded as distribution grids become more active and data-driven. In addition to energy consumption records, smart meters can provide time series interval-average active power, voltage magnitude and, depending on the metering infrastructure, additional information such as generation, reactive power, power factor, outages or power-quality-related events. In this sense, the smart meter becomes more than a billing device: it acts as a distributed measurement point located at the edge of the LV network.

This is particularly relevant because LV networks have historically had limited monitoring when compared with transmission and medium-voltage systems. Installing high-resolution sensors, phasor measurement units or line current sensors across all LV feeders would be costly and difficult to scale. Smart meters, by contrast, are already deployed at customer premises and provide wide spatial coverage of the network. However, smart meters only observe electrical quantities at customer connection points rather than the complete internal structure of the feeder. As a result, internal electrical buses, cable cabinets and branching points may remain unobserved even when smart-meter coverage is high. The main learning-based experiments in this dissertation assume that all nodes represented in the target adjacency matrix are observed. They therefore do not attempt to reconstruct unknown intermediate physical buses. Hidden-node reconstruction is considered only through the independent analytical partial-observability comparison presented later in the dissertation.

Within this context, the present dissertation focuses on topology inference from smart-meter-like active-power and voltage-magnitude measurements. These measurements can contain indirect information about the electrical behaviour of the feeder, because voltage variations at customer connection points are affected not only by local consumption or generation, but also by shared upstream branches, voltage drops, voltage rises and network sensitivities. This motivates the use of interval-average P/V time series as the basis for reconstructing plausible LV radial topologies.

Nevertheless, the available P/V measurements may not uniquely determine the underlying feeder structure. Under realistic operating conditions, several radial topologies may be compatible with the same observations, motivating the need to preserve topology-related uncertainty rather

than relying exclusively on a single direct connectivity decision.

This ambiguity is illustrated in Figure 1.3. The same set of customer-level active-power and voltage-magnitude measurements, may be compatible with different plausible radial structures when the available data are limited, noisy or incomplete. Each structure connects the same transformer and customer nodes, but through a different radial configuration. This is consistent with recent topology-learning literature, where hidden nodes, limited monitoring, Kron-reduced representations and candidate reconstruction procedures show that the recoverable topology may be reduced, ambiguous or dependent on the measurement set [7], [8], [17], [18], [19].

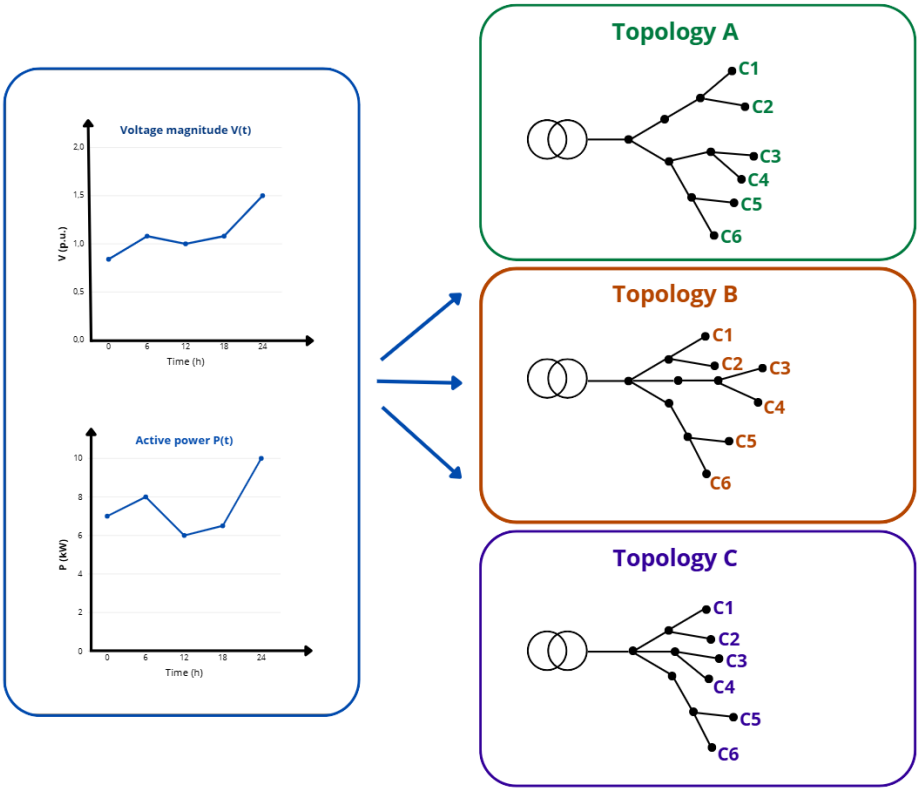


Figure 1.3: Conceptual illustration of topology ambiguity under imperfect smart-meter measurements. The same customer-level active-power and voltage-magnitude measurements may be compatible with multiple physically plausible radial topologies, motivating an inference framework that preserves topology-related uncertainty rather than relying only on a single direct reconstruction.

1.1 Modelling Low-Voltage Distribution Networks

Although practical LV networks are generally three-phase and may exhibit significant phase unbalance, this dissertation adopts a balanced single-phase equivalent representing one phase of the LV feeder. Consequently, unknown phase allocation and three-phase unbalance are outside the scope of the proposed topology-inference models. This simplification is consistent with LV topology-reconstruction approaches that represent the network as a radial single-phase system and

either assume that phase connectivity is known or address phase identification separately [20]. In a practical application, the proposed methodology would therefore be applied after assigning the monitored customers to their corresponding phases.

A voltage-sensitivity model can be used to approximate the effect of active-power variations on voltage magnitudes. Following the formulation used by Sampaio et al. [21], neglecting reactive-power variations and other exogenous voltage drivers, the new voltage magnitude at connection point i can be estimated as

$$V_i^{new} = V_i^{old} + \sum_{j=1}^{N} S_{i,j}\Delta P_j + \sum_{p=1}^{3} S_{i,p}^{SS}\Delta V_p^{SS}, \tag{1.1}$$

where V_i^{new} and V_i^{old} are the new and previous voltage magnitudes at connection point i, respectively, $S_{i,j}$ is the sensitivity of the voltage at point i with respect to the active-power variation at point j, ΔP_j is the active-power variation at point j, $S_{i,p}^{SS}$ is the sensitivity of the voltage at point i with respect to the secondary-substation voltage variation in phase p, and ΔV_p^{SS} is the voltage variation at the secondary substation in phase p.

In this dissertation, the focus is on active-power and voltage-magnitude measurements at the monitored nodes, without explicitly modelling secondary-substation voltage variations. This simplification is considered reasonable because secondary-substation voltage variations are assumed to remain relatively small and to evolve more slowly than customer-level active-power variations under the operating conditions considered. It also facilitates the demonstration of the proposed methodology by avoiding an additional source of common voltage variation. Under this simplified setting, Equation 1.1 reduces to

$$\Delta V_i \approx \sum_{j=1}^{N} S_{i,j}\Delta P_j, \tag{1.2}$$

or, in matrix form,

$$\Delta \mathbf{V} \approx \mathbf{S}\Delta \mathbf{P}, \tag{1.3}$$

where $\Delta\mathbf{V}$ is the vector of voltage variations, $\Delta\mathbf{P}$ is the vector of active-power variations, and $\mathbf{S}$ is the voltage sensitivity matrix.

Sensitivity-based formulations are widely used in distribution-system analysis because they provide a compact linear approximation of the relationship between power injections and voltage behaviour. Instead of solving the full nonlinear AC power-flow equations for every operating point, sensitivities approximate how variations in active power affect nodal voltages around a given operating condition. In this sense, they transform the local voltage-response problem from a nonlinear power-flow formulation into a linear relation between injection variations and voltage variations [8], [21], [22], [23].

These relations provide a physical link between voltage measurements and topology inference. If the feeder structure changes, the set of shared electrical paths also changes, and so does the sensitivity pattern observed in the data. Conversely, estimating the topology from P/V measurements

can be interpreted as searching for the radial structure whose electrical paths best explain the observed voltage-power relationships.

For this reason, the feeder is represented as a graph

$$\mathscr{G} = (\mathscr{V}, \mathscr{E}), \tag{1.4}$$

where $\mathscr{V}$ is the set of electrical nodes and $\mathscr{E}$ is the set of branches. If the feeder is radial and connected, then

$$|\mathscr{E}| = |\mathscr{V}| - 1. \tag{1.5}$$

The radiality constraint restricts the feasible set of topologies to trees, which is physically consistent with many LV feeders. However, it does not make the inverse problem trivial. Many different radial trees can connect the same set of nodes, and when the available measurements are noisy, incomplete or limited to voltage magnitudes and active power, several candidate structures may explain the data with similar plausibility. Therefore, topology inference remains a difficult inverse problem even under the radiality assumption.

1.2 Problem Motivation

The motivation for this dissertation arises from the mismatch between the increasing operational importance of LV grids and the limited reliability of their digital topology records. While distributed energy resources make accurate LV models more necessary, the available information often remains incomplete. In many practical settings, the operator may not know their branching structure or exact electrical connectivity within the feeder, the correct phase connection, the existence or location of intermediate nodes, or the actual line parameters.

This lack of reliable topology knowledge becomes more critical as LV networks become active. Advanced grid management tools increasingly require models that reflect the real feeder configuration. Voltage control algorithms must know which customers are electrically coupled. State-estimation tools require the correct connectivity to infer unmeasured quantities from available data. Fault-location tools must identify the likely path between the transformer and a disturbed node. Hosting-capacity assessments depend on the voltage impact of injections at different locations. Incorrect connectivity can lead to incorrect operational conclusions, including underestimated voltage violations, wrong loss estimates or inappropriate flexibility signals.

A growing body of work has addressed topology learning in distribution networks. Existing studies investigate topology detection, topology identification, phase identification, admittance estimation, parameter estimation, topology verification and reconstruction using different measurement sources and modelling assumptions [7], [11], [12], [19], [24], [25], [26], [27], [28]. These approaches demonstrate that measurement data can contain meaningful information about the structure of the distribution grid.

However, most existing methods still face at least one important limitation when applied to LV smart-meter settings. Some approaches assume that the set of candidate lines is known and

only the energised configuration must be detected. This is suitable for switching detection but less appropriate when LV records are incomplete. Other methods require measurements that are not widely available in LV networks, such as voltage angles, PMU or micro-PMU data, or other high-resolution synchronised three-phase data.

Many existing topology-inference methods also rely on assumptions about both the available measurements and the underlying network model. Some correlation-based and graphical-model approaches assume that nodal injections are sufficiently informative and, in some formulations, statistically independent or weakly correlated, so that voltage correlations can be related to electrical proximity [7], [10]. Other methods require sufficient variation in the measurements, since low variability or underexcited operating conditions can make topology recovery difficult or non-identifiable [29], [30], [31]. Several LV topology-reconstruction approaches also assume radial operation and limited or no hidden nodes, although hidden intermediate buses and unobserved branching points are known to complicate topology recovery [8], [9], [32], [33], [34]. Finally, methods that combine topology identification with electrical-parameter estimation often require accurate or estimable line parameters, such as impedances or admittances, because errors in these parameters can affect the inferred network structure [13], [14], [27].

In addition, most methods return a single estimated topology, even when the measurements are ambiguous. This last limitation is central to this dissertation. If $\mathscr{T}$ denotes the set of all feasible radial candidate topologies connecting the observed nodes, a deterministic topology reconstruction problem can be written as

$$G^{\star} = \arg\min_{G \in \mathscr{T}} \mathscr{L}(G; P, V), \tag{1.6}$$

where P and V are active-power and voltage time-series measurements, and $\mathscr{L}$ is a mismatch or fitness function. This deterministic formulation is useful, but it implicitly assumes that one topology is sufficiently better than all alternatives. In realistic conditions, several candidates may produce similar values of the objective function:

$$\mathscr{L}(G_1; P, V) \approx \mathscr{L}(G_2; P, V) \approx \cdots \approx \mathscr{L}(G_K; P, V). \tag{1.7}$$

In this case, reporting only $G^{\star}$ may hide the uncertainty inherent in the data. A more informative approach is to produce topology-related scores or alternative plausible reconstructions that can later be evaluated under radiality and physical-consistency criteria.

In the present dissertation, the learning models produce edge-score matrices from measurement-derived features such as voltage profiles, active-power time series, temporal deviations, correlations, sensitivities or learned embeddings. These edge scores are then projected into one valid radial topology for quantitative evaluation.

Therefore, the experiments reported in this dissertation do not explicitly evaluate ranked sets of candidate topologies. However, the proposed formulation provides a basis for future uncertainty-aware extensions, where topology-related scores could be used to sample, compare or validate

alternative plausible radial structures using additional physical criteria. Such extensions would require explicit calibration, sampling or ranking of multiple candidate topologies.

This viewpoint motivates the use of learning models that can produce topology-related representations or edge-score matrices from measurement-derived inputs. In the context of this dissertation, Variational Autoencoders, convolutional VAE architectures, variable-size VAE models, *Transformer*-inspired models and graph-completion formulations are explored as ways to move from purely iterative reconstruction towards offline-trained edge-score prediction.

1.3 Research Gap

The literature on distribution-grid topology learning has advanced significantly in recent years, largely due to the increasing adoption of smart metering devices and advanced metering infrastructure. The availability of customer-level voltage and power measurements has created new opportunities to infer network structure from operational data, rather than relying exclusively on static or incomplete topology records [7], [8], [9]. Existing methods provide tools for topology detection, topology identification, phase identification, parameter estimation, graphical modelling, sparse regression, active probing, state-estimation-based topology processing and learning-based inference [10], [11], [12], [13], [14], [28], [29], [30]. These contributions provide important foundations for the present work.

Nevertheless, several gaps remain when the target application is LV topology inference from realistic smart-meter data. First, many methods assume stronger observability than is commonly available in LV networks. Techniques based on phasor measurement units, micro-phasor measurement units, voltage angles or other high-resolution synchronised three-phase data can provide detailed electrical information, but such measurement systems are not deployed across LV feeders. In most practical cases, the available data are limited to active power or energy and voltage magnitude.

Second, several approaches assume that the candidate infrastructure is known. This assumption is appropriate when the task is to determine switch statuses or detect changes among known lines, but it is less suitable when the underlying LV records may be incomplete or outdated. In such cases, the problem is closer to topology reconstruction than to topology detection.

Third, hidden nodes remain a major limitation. Works on impedance estimation, Kron reduction, recursive grouping and limited observability show that the recoverable structure may be a reduced or equivalent network rather than the complete physical topology [17], [18], [27]. In LV feeders, this issue is particularly important because intermediate junctions and branching points are often not directly metered.

Fourth, many existing methods return a single deterministic estimate. While this may be sufficient when the data are highly informative, it is less appropriate when multiple radial configurations are compatible with the same measurements. In LV smart-meter settings, ambiguity can be

inherent. Therefore, methods that produce topology-related scores, uncertainty-aware representations or alternative plausible reconstructions are useful because they preserve more information than a single direct connectivity decision.

Finally, optimisation-based methods, such as Genetic Algorithms, can search large discrete topology spaces, but their iterative nature may become computationally expensive when many candidate radial structures must be explored. Learning-based models offer an alternative by learning topology-related patterns directly from measurement-derived representations and producing edge-score matrices that can be projected into radial topologies. This has the potential to shift a substantial part of the computational effort from repeated inference-time search to an offline training stage, while inference mainly requires a forward pass followed by radial projection.

The research gap addressed by this dissertation can therefore be summarised as follows: existing methods provide valuable tools for topology learning, but there is still a need to assess whether customer-level active-power and voltage-magnitude data contain enough information for radial topology reconstruction, and whether offline-trained learning-based models can produce useful topology-related edge scores from physically meaningful measurement-derived representations.

1.4 Research Questions and Objectives

This dissertation investigates how learning-based and physics-informed methods can support topology inference in LV distribution networks using customer-level active-power and voltage-magnitude data. The work is guided by the following research questions.

RQ1: Are customer-level active-power and voltage-magnitude data sufficient to support low-voltage topology reconstruction?

Hypothesis: Residential active-power profiles derived from smart-meter measurements, together with the corresponding simulated voltage-magnitude trajectories, contain structural information that can support radial LV topology reconstruction when transformed into suitable measurement-derived features, such as temporal deviations, correlations, sensitivity-based representations or learned embeddings.

Objective: To evaluate whether residential active-power profiles derived from smart-meter measurements, together with the corresponding simulated voltage-magnitude trajectories, provide sufficient information for topology reconstruction under controlled conditions, including measurement noise and uncertainty.

RQ2: Can an offline-trained learning-based framework produce useful topology-related edge scores without requiring repeated optimisation-based search for each new inference case?

Hypothesis: Optimisation-based approaches, such as genetic algorithms, typically perform most of their computational search during inference for each new case. A learning-based formulation

can instead concentrate the main computational effort during training and then produce topology-related edge scores through a forward pass at inference time.

Objective: To analyse the proposed VAE-based topology-inference framework as an offline-trained alternative to repeated optimisation-based search, focusing on its reconstruction behaviour, limitations and potential computational advantages.

RQ3: Can VAE-based models learn useful topology-related edge scores from P/V-derived representations?

Hypothesis: VAE-based models can learn latent representations of measurement-derived electrical patterns and decode them into edge-score matrices that can be projected into radial topologies.

Objective: To evaluate fixed-size, convolutional, variable-size and *Transformer*-inspired VAE models for radial topology inference from P/V-derived inputs.

RQ4: Which input representations and model architectures are most effective for topology inference across different network sizes?

Hypothesis: Physically meaningful matrix-based representations, especially sensitivity-based representations, provide more transferable information than raw P/V time series. Variable-size and *Transformer*-inspired architectures can further improve generalisation across graph sizes.

Objective: To compare raw P/V inputs, correlation matrices and sensitivity matrices across fixed-size, convolutional, variable-size, *Transformer*-inspired and graph-completion models.

RQ5: How robust are the proposed topology inference methods to measurement errors and partial topology information?

Hypothesis: Topology-related information may be affected differently by active-power errors, voltage errors and incomplete topology information.

Objective: To evaluate inferred topologies in terms of edge accuracy, edge error, radiality, robustness to active-power and voltage errors, scalability with network size and performance under partial topology information.

1.5 Contributions

The main contributions of this dissertation are the following:

- A structured formulation of LV topology inference from customer-level active-power and voltage-magnitude data.
- A reproducible controlled-data and benchmarking pipeline combining synthetically generated radial topologies, residential active-power profiles derived from real smart-meter measurements, voltage magnitudes obtained through power-flow simulation and controllable levels of measurement error.

- An empirical demonstration that P/V-derived representations contain recoverable topology-related information under controlled simulation conditions.
- An initial fixed-size VAE baseline and a fixed-size convolutional VAE formulation for topology inference from matrix-based representations.
- A variable-size topology-inference framework based on shared node encoding, masking, graph-level latent inference, pairwise edge decoding and radial-tree projection, allowing a single model to process feeders with different numbers of nodes.
- A *Transformer*-inspired variable-size model for improving global node-to-node interaction modelling and generalisation to larger *held-out* graph sizes, corresponding to node counts not represented during training.
- A comparative analysis of raw P/V, correlation and sensitivity representations, identifying which representations are most effective for LV topology inference.
- A robustness analysis of the proposed variable-size models under active-power and voltage measurement errors.
- A graph-completion formulation that combines partial topology information with sensitivity-based electrical information.
- The preparation and submission to the MEDPOWER conference of the paper *"Physics-Guided VAE-Based Topology Inference for Low-Voltage Networks"*, authored by Ana Maria A. Almeida, David Lima and Gil S. Sampaio, which is currently awaiting peer review.

The central contribution of this dissertation is the systematic evaluation of offline-trained VAE-based models for radial LV topology inference using physically informed P/V representations, pairwise edge-score prediction and radial projection across fixed and variable graph sizes.

1.6 Dissertation Structure

This dissertation is organised into five main chapters.

- **Chapter 1: Introduction.** Introduces the operational context of LV distribution networks, the challenges created by distributed energy resources and imperfect observability, and the motivation for topology inference from smart-meter-derived measurements. It defines the research gap, research questions, objectives and contributions of the dissertation.
- **Chapter 2: Literature Review.** Reviews the state of the art in distribution-grid topology learning, smart-meter-based inference, hidden-node reconstruction, phase identification, parameter estimation, graph-theoretic methods, sparse regression, active probing, optimisation methods, probabilistic approaches and graph generative modelling. The chapter identifies

the research gap addressed by the dissertation and positions the proposed learning-based approach relative to existing methods.

- **Chapter 3: Synthetic Data Generation and Modelling Framework.** Presents the controlled data-generation framework, including synthetic radial network generation, the assignment of electrical branch parameters, the use of residential active-power profiles derived from real smart-meter measurements, Forward–Backward Sweep power-flow simulation and the modelling pipeline used in the experimental chapter.

- **Chapter 4: Exploratory Modelling and Results.** Describes the experimental evaluation of fixed-size, convolutional, variable-size, *Transformer*-inspired and graph-completion models. It analyses input representations, reconstruction accuracy, generalisation to *unseen* topologies and *held-out* graph sizes, partial-topology completion and robustness to measurement errors, including an independent analytical partial-observability comparison.

- **Chapter 5: Conclusions and Future Work.** Summarises the main findings, discusses limitations and proposes future research directions, including validation with real LV data, phase-aware modelling, richer measurement representations, uncertainty-aware topology inference and deployment-oriented evaluation.

Chapter 2

Literature Review

The previous chapter introduced the motivation for learning-based topology inference in low-voltage (LV) distribution networks. The central problem is the growing need for accurate LV network models in order to face the new challenges, despite incomplete, outdated or uncertain topology records. This uncertainty has motivated a growing body of research on distribution-grid topology learning, where the objective is to infer, verify or reconstruct network connectivity from measurements, operational data and physical constraints. This chapter reviews the state of the art in this field and positions the dissertation within that literature.

The review is organised according to the progression from classical topology learning towards smart-meter-based, physics-informed and learning-based topology inference. Section 2.1 defines the scope of topology learning and clarifies the terminology used in the literature, including topology detection, topology identification, topology reconstruction, topology verification and parameter estimation. Section 2.2 discusses measurement sources and observability limitations in LV grids. Section 2.3 reviews graph representations, radial feeders and linearised power-flow relations. Section 2.4 analyses smart-meter and AMI-based methods, including voltage correlation, sensitivity-based inference and Steiner-tree reconstruction. Section 2.5 focuses on hidden nodes and identifiability limits. Section 2.6 reviews admittance and impedance matrix estimation as related tasks that can support topology inference. Optimisation, MILP and state-estimation-based approaches are discussed where relevant in relation to candidate selection, measurement assumptions and topology validation. Section 2.7 reviews probabilistic, Bayesian and online topology-tracking approaches. Section 2.8 introduces the generative-model and attention-based perspective that motivates the proposed VAE-based and *Transformer*-inspired framework. Finally, Section 2.9 summarises the research gap and positions the contribution of the dissertation.

2.1 Distribution-Grid Topology Learning: Scope and Terminology

Distribution-grid topology learning refers to the use of measurements, physical models and structural constraints to infer the operational connectivity of a distribution feeder. Deka et al. [7] provide a systematic review of topology identification and topology detection schemes for distribution

grids, emphasising methods that operate under limited measurement availability and exploit conservation laws, feeder radiality and graph-theoretic structure [7].

In the literature, several related terms are used to describe distribution-grid topology learning problems. These terms are not always used identically across papers, but they help distinguish different levels of prior knowledge and different inference objectives.

Topology detection usually assumes that the physical infrastructure, or at least a predefined set of candidate lines, switches and feasible operating configurations, is known. The task is then to determine which configuration is currently energised, or which elements are active, at a given operating time. In this sense, topology detection can be interpreted as selecting one topology from a known and limited set of possible configurations. This formulation is mainly used for switch-status estimation, topology verification and operational topology processing.

Topology identification and topology reconstruction refer to broader problems. In these cases, the objective is to infer the network connectivity itself from measurements, often when the available topology records are incomplete, outdated or uncertain. Reconstruction applies when the operator does not have a reliable feeder model and must build a plausible topology from observed customer or bus measurements. Verification, by contrast, assumes that a candidate topology already exists and evaluates whether the available data support it [7], [8], [9], [11], [24].

A further distinction concerns whether the method estimates only the graph structure or jointly estimates the graph and electrical parameters. In many operational applications, topology alone is not sufficient, because power-flow analysis, voltage estimation, loss calculation and control also require line impedances or admittances. Several works therefore address joint topology and parameter estimation [13], [14], [17], [35].

The problem addressed in this dissertation is closer to topology inference and reconstruction than to topology detection. Rather than selecting an energised configuration from a known set of switches, the objective is to infer radial LV feeder connectivity from residential active-power profiles derived from real smart-meter measurements and the corresponding voltage-magnitude trajectories obtained through power-flow simulation. The feeder is modelled as a connected radial graph, represented by an adjacency matrix or equivalent radial structure. Because measurement errors, limited observability and missing electrical quantities may prevent a unique topology from being determined, the proposed learning-based models produce topology-related edge scores that are later projected into radial connectivity structures.

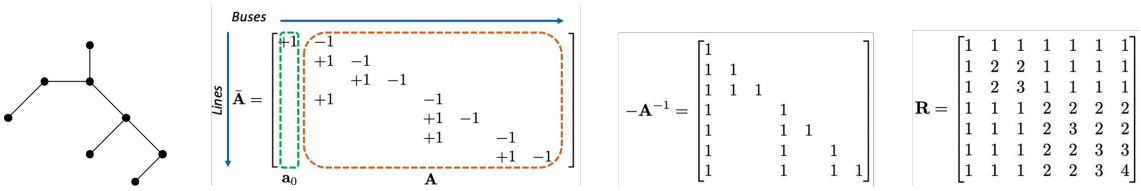


Figure 2.1: Graph-theoretic representation of a radial distribution feeder, including the root node, leaf nodes, incidence matrices and the relationship between feeder paths and matrix entries. Extracted from [7].

Figure 2.1 illustrates the graph-theoretic notation commonly used in topology-learning formulations. It highlights the representation of distribution feeders as rooted graphs and supports the distinction between nodes, branches, paths and radial structures used throughout this chapter.

This clarification is necessary because the literature addresses topology learning under different assumptions and with different levels of prior knowledge. In some cases, the task consists of identifying the active configuration within a known network model, whereas in others the goal is to infer the connectivity itself from incomplete or uncertain measurements. The present dissertation belongs to the latter setting, focusing on LV topology inference from limited customer-level active-power and voltage-magnitude data rather than on switch-state detection in a fully known grid.

2.2 Measurement Sources and Limited Observability in LV Grids

The feasibility of topology learning depends strongly on measurement availability. Transmission systems usually have better observability, including SCADA data, breaker statuses and, increasingly, phasor measurements. LV distribution networks are different. They are numerous, geographically dispersed, often underground and historically operated with limited instrumentation. As a result, topology records may be incomplete or outdated, and real-time measurements may be available only at transformer terminals, feeders, selected sensors or customer smart meters.

The main measurement sources considered in the literature include smart meters or AMI, micro-PMUs, distribution PMUs, line current sensors, switch-status data, power-flow measurements and pseudo-measurements. Each source provides different physical quantities and therefore leads to different assumptions about observability, synchronisation, measurement quality and scalability. Table 2.1 summarises these sources in terms of the physical variables they provide and their practical limitations in LV topology inference.

Smart meters are used in this dissertation because they are increasingly deployed at customer connection points and provide a scalable source of customer-level information. However, in practical LV settings, the data available for grid analysis are often limited to energy consumption or active-power-related quantities. Although many modern smart meters can record voltage magnitude, these data may not be accessible to the DSO, may be stored only as aggregated statistics or events, and may not be available at a temporal resolution suitable for topology inference. Therefore, methods based on smart-meter voltage measurements often rely on an enhanced smart-meter data setting or on additional data access beyond standard billing-oriented records.

Line current sensors can provide information about feeder loading and branch currents, but their deployment in LV networks is usually limited to specific locations, such as the outgoing feeders of secondary substations, rather than being widespread across the entire LV grid. Therefore, although they may support monitoring and topology-related analysis in some settings, they do not provide the same customer-level spatial coverage as smart meters [8], [17], [36].

The measurement source therefore constrains the class of topology-learning methods that can be applied. Methods based on voltage and current phasors can estimate richer electrical models,

Table 2.1: Measurement sources considered in distribution-grid topology inference.

Measurement source	Main physical variables	Typical LV use and limitation
Smart meters / AMI	Energy, active-power-related quantities and, when available, reactive power, power factor or voltage magnitude.	Widely deployed at customer premises, but data may be aggregated, asynchronous or incomplete. Voltage data may not always be available with sufficient resolution.
Micro-PMUs / PMUs	Synchronised voltage and current phasors, voltage angles, frequency, active and reactive power.	Rich synchronised measurements, but sparse and not commonly deployed across LV feeders.
Line current sensors	Feeder or branch current magnitude and, sometimes, current direction.	Usually installed at specific feeder points, such as secondary-substation outgoing feeders. They do not provide customer-level coverage.
Switch-status data	Open/closed status of switches or controllable network elements.	Useful when candidate infrastructure is known, but less suitable when LV records are incomplete or outdated.
Power-flow measurements	Active/reactive power flows, voltages and currents at monitored points.	More common in monitored MV/HV systems. In LV feeders, they require dedicated monitoring and sensor placement.
Pseudo-measurements	Estimated loads, injections or operating conditions from historical data, forecasts or typical profiles.	Used when direct measurements are missing, but introduce uncertainty and may bias topology or parameter estimation.

but they are less aligned with typical LV smart-meter deployments. Methods based on customer-level smart-meter data are more scalable, but they must deal with aggregation, asynchronous sampling and incomplete voltage availability. This trade-off motivates the focus of this dissertation on customer-level active-power and voltage-magnitude data and score-based topology inference under limited observability.

2.3 Physics-Based Modelling, Radiality and Sensitivity Relations

Physics-based modelling provides the foundation for distribution-grid topology inference. Using the notation introduced in Equation 1.4, a distribution feeder is represented by an undirected graph, where the vertices correspond to buses or electrical nodes and the edges correspond to electrical branches.

In this dissertation, the feeder is assumed to be connected and radial, with the corresponding radiality condition presented in Equation 1.5. Together with connectivity, this condition defines a tree structure, meaning that there is a unique path between any pair of nodes and that no closed cycles are present.

Radiality is central to many topology-learning algorithms because it reduces the feasible graph space and allows the use of tree-specific concepts such as parent-child relations, descendants, leaves, paths and recursive grouping. Nevertheless, radiality alone does not make the problem trivial. Even when the feasible set is restricted to connected trees, the number of possible radial configurations can still be very large, even for networks with a relatively small number of buses.

For example, if all labelled trees over N observed nodes were admissible, the number of possible radial structures would be given by Cayley's formula [37],

$$N_{\text{trees}} = N^{N-2}. \tag{2.1}$$

Thus, for only ten observed nodes, the unrestricted number of possible labelled radial trees is already 10^8. In practice, electrical constraints, feeder design rules and available measurements reduce this search space, but the inverse topology-inference problem remains difficult when the measurements are limited, noisy or incomplete.

In distribution-system analysis, the electrical behaviour of the network is fundamentally described by the AC power-flow equations [38]. For a node i, the complex power injection can be written as

$$S_i = P_i + jQ_i = V_i I_i^*, \tag{2.2}$$

where P_i and Q_i are the active and reactive power injections, V_i is the complex voltage at node i, I_i is the injected current, and $(\cdot)^*$ denotes the complex conjugate. The nodal currents are related to the bus voltages through the admittance matrix:

$$\mathbf{I} = \mathbf{YV}. \tag{2.3}$$

Combining these relations leads to the nonlinear AC power-flow equations, where nodal active and reactive powers depend nonlinearly on voltage magnitudes, voltage angles and network admittances. Although these equations provide a detailed physical model of the network, repeatedly solving them can be computationally demanding and may require information, such as voltage angles and complete line parameters, that is not always available in LV smart-meter settings.

For this reason, many distribution-system studies use linearised power-flow approximations. Instead of solving the full nonlinear AC power-flow equations for every operating point, these approximations express voltage deviations as a linear function of active and reactive power injections around an operating condition. For radial feeders, a commonly used approximation can be written as

$$\mathbf{v} \approx \mathbf{Rp} + \mathbf{Xq} + v_0\mathbf{1}, \tag{2.4}$$

where $\mathbf{v}$ is the vector of voltage magnitudes, $\mathbf{p}$ and $\mathbf{q}$ are the active- and reactive-power injection vectors, v_0 is the reference voltage at the root node, usually the secondary transformer, and $\mathbf{1}$ is a vector of ones that extends this reference voltage to all nodes. The matrices $\mathbf{R}$ and $\mathbf{X}$ capture the effect of network resistances and reactances on voltage behaviour [8], [22], [39].

In radial networks, the entries of $\mathbf{R}$ and $\mathbf{X}$ are not arbitrary coefficients. They depend on the feeder connectivity and on the electrical paths shared by the nodes. Under linearised radial power-flow models, these coefficients can be interpreted through the impedance accumulated along common paths between the root node and the measured buses [7], [22], [39].

Let $\mathscr{P}_i$ denote the path from the root node to node i, and let $\mathscr{P}_j$ denote the path from the root

node to node j. The common path shared by both nodes can be written as

$$\mathscr{P}_{ij} = \mathscr{P}_i \cap \mathscr{P}_j. \tag{2.5}$$

If each branch e has impedance $z_e = r_e + jx_e$, the accumulated impedance along the common path is

$$Z_{ij}^{\text{com}} = \sum_{e \in \mathscr{P}_{ij}} z_e = \sum_{e \in \mathscr{P}_{ij}} (r_e + jx_e). \tag{2.6}$$

Therefore, the resistance and reactance coefficients associated with the voltage-power relation can be interpreted as the real and imaginary parts of the common-path impedance:

$$R_{ij} = \text{Re}\left(Z_{ij}^{\text{com}}\right) = \sum_{e \in \mathscr{P}_{ij}} r_e, \qquad X_{ij} = \text{Im}\left(Z_{ij}^{\text{com}}\right) = \sum_{e \in \mathscr{P}_{ij}} x_e. \tag{2.7}$$

This path-based interpretation shows why linearised voltage-power relations are directly influenced by the underlying topology. If the feeder connectivity changes, the shared paths $\mathscr{P}_{ij}$ also change, and therefore the entries of $\mathbf{R}$ and $\mathbf{X}$ change as well. This explains why voltage and power measurements can contain information for topology inference.

The simplified active-power sensitivity relation used in this dissertation was introduced in Equation 1.3, where $\mathbf{S}$ denotes the voltage sensitivity matrix. In resistance-dominated LV feeders, active-power variations can have a strong influence on voltage magnitudes, although the relative influence of active and reactive power depends on the network parameters and operating conditions. Sensitivity-based features are therefore attractive for topology inference because they encode how changes at one node affect voltage at other nodes. This theoretical background supports the use of sensitivity-based representations as compact features for topology inference, especially when only voltage and power measurements are available [8], [22], [23].

The main limitation of physics-based and sensitivity-based formulations is that their quality depends on the quality of the measurements and on the validity of the linear approximation. If active-power variations are weak, correlated or affected by noise, the inferred sensitivities may become ambiguous. In this context, a topology may be electrically plausible without being uniquely identifiable from the available measurements, which motivates the score-based learning perspective adopted in this dissertation.

2.4 Smart-Meter and AMI-Based Topology Inference

Smart-meter and AMI-based methods exploit customer-level metering infrastructure that is increasingly deployed at scale. Their main advantage is scalability, while their principal limitations include aggregation, measurement noise, asynchronous sampling, incomplete voltage availability and restricted access to grid-relevant data. The discussion below focuses on how these measurements have been used for topology inference.

Early smart-meter-based methods often exploit voltage similarity, voltage covariance, correlation or clustering. The underlying intuition is that customers electrically close to one another tend to exhibit related voltage variations because they share upstream feeder segments and are affected by common voltage drops or voltage rises. Bolognani et al. proposed topology identification through voltage-correlation analysis, while Pappu et al. used smart-meter voltage measurements to identify LV network topology [9], [10].

A related line of work applies the same voltage-similarity principle to phase identification. In this case, the objective is not to recover the feeder topology itself, but to determine which customers are connected to the same phase. Customers on the same phase tend to exhibit more similar voltage profiles than customers on different phases, which motivates clustering and correlation-based approaches using voltage time series [12], [26], [40], [41].

A generic voltage-correlation coefficient between nodes i and j can be written as

$$\rho_{ij} = \frac{\sum_{t\in\mathscr{T}} \left(V_i(t) - \bar{V}_i\right)\left(V_j(t) - \bar{V}_j\right)}{\sqrt{\sum_{t\in\mathscr{T}} \left(V_i(t) - \bar{V}_i\right)^2}\sqrt{\sum_{t\in\mathscr{T}} \left(V_j(t) - \bar{V}_j\right)^2}}, \tag{2.8}$$

where ρ_{ij} is the Pearson correlation coefficient between the voltage-magnitude time series of nodes i and j, $\mathscr{T}$ is the set of time instants considered in the measurement window, $t \in \mathscr{T}$ denotes each individual time step, $V_i(t)$ and $V_j(t)$ are the voltage magnitudes at nodes i and j at time t, and $\bar{V}_i$ and $\bar{V}_j$ are the average voltage magnitudes of nodes i and j over the time window $\mathscr{T}$. These averages are given by

$$\bar{V}_i = \frac{1}{|\mathscr{T}|}\sum_{t\in\mathscr{T}} V_i(t), \qquad \bar{V}_j = \frac{1}{|\mathscr{T}|}\sum_{t\in\mathscr{T}} V_j(t). \tag{2.9}$$

Correlation-based methods are attractive because they are simple and can be applied to measurements collected during normal network operation. In other words, they do not require deliberate perturbations, active probing actions or controlled voltage/current excitations to reveal the feeder structure. However, they are not sufficient to uniquely recover topology in general. Voltage similarity may be caused by common upstream branches, correlated loads, PV behaviour, transformer voltage changes, phase allocation, voltage regulation actions or measurement aggregation. Thus, correlation should be interpreted as a structural feature, not as a complete topology solution.

More advanced smart-meter methods incorporate physical constraints. Cavraro, Kekatos and Veeramachaneni formulate topology verification using non-synchronised voltage data. Their method uses a statistical learning framework based on the sample voltage covariance matrix, maximum likelihood and maximum a posteriori estimation, showing that topology information can be obtained even without synchrophasors [11].

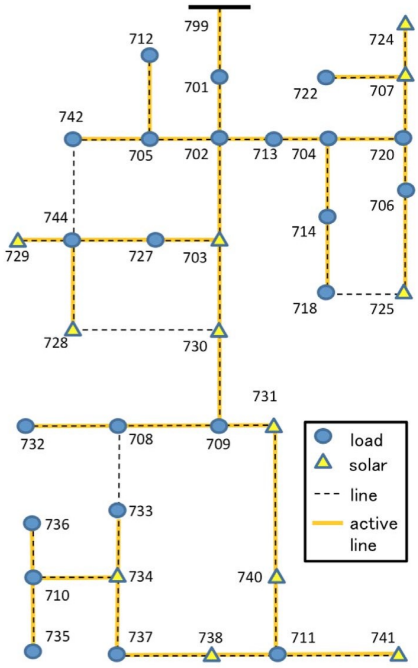


Figure 2.2: Line infrastructure for the IEEE 37-bus feeder benchmark used in voltage-analytics-based topology verification. The figure illustrates the difference between the available line infrastructure and the operational topology to be verified. Extracted from [11].

Zhao et al. propose a full-scale distribution-system topology-identification method based on Markov Random Fields. Their framework mines smart-meter and terminal data, models nodal voltage correlations through a Markov Random Field, uses pseudo-likelihood and regularisation to improve computational efficiency and avoid overfitting, and then generates topology through iterative screening. This work moves beyond pairwise correlation heuristics and treats topology identification as a probabilistic graphical-modelling problem [28].

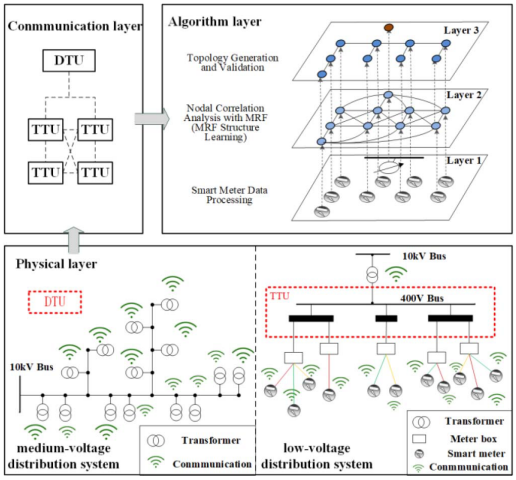


Figure 2.3: Overview of a full-scale distribution-system topology-identification framework based on smart-meter data processing, Markov Random Field structure learning and topology generation. Extracted from [28].

Pengwah et al. propose a smart-meter-based topology identification method that, in practice, addresses a topology reconstruction problem for radial LV networks. Although the work is framed as topology identification, its objective is to reconstruct the radial feeder structure from smart-meter measurements by estimating voltage sensitivity coefficients from voltage and load or power data, and then using these coefficients to infer the underlying network connectivity. The method considers the reconstruction of a Steiner tree, allowing the inferred topology to include unobserved intermediate nodes rather than only directly connecting the metered customers [8]. The method also improves the reconstruction process with backtracking, generating a cluster of candidate graphs from which the best fitting candidate is selected according to optimisation criteria.

This candidate-generation step is relevant to the motivation of this dissertation because it recognises that several radial structures may be plausible before the final selection is made. However, the approach remains essentially deterministic: the candidates are generated through an algorithmic reconstruction procedure and then filtered or selected. In contrast, the present dissertation evaluates offline-trained models that map measurement-derived representations to edge-score matrices, which are subsequently projected into radial LV topologies.

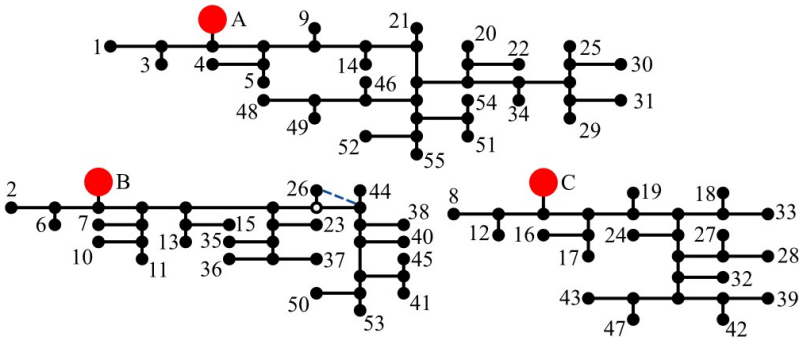


Figure 2.4: Example of smart-meter-based radial topology reconstruction using voltage sensitivity coefficients and candidate graph generation. Extracted from [8].

The reviewed smart-meter-based methods show that customer-level measurements can reveal meaningful topological information. However, their limitations are directly connected to the measurement setting. Correlation-based methods may confuse electrical proximity with correlated load behaviour or common upstream voltage variations. Sensitivity-based methods depend on sufficient excitation and on the availability of voltage measurements. Steiner-tree reconstruction can represent hidden intermediate nodes, but the final selection still follows a deterministic optimisation criterion. These limitations motivate methods that preserve topology-related uncertainty through scores, probabilities or alternative validation criteria instead of treating a single deterministic estimate as fully certain.

2.5 Hidden Nodes and Identifiability Limits

The previous sections discussed measurement availability and the practical limitations of LV grid observability. This section focuses on a related but more specific issue: the effect of hidden or

unobserved intermediate nodes on topology recovery. In LV networks, measurements are often available at customer connection points, but not at every cable cabinet, junction point, service connection point or internal branching node. Therefore, even if several customers are metered, the physical structure connecting those customers may remain only partially observed.

Hidden nodes change the nature of the topology inference problem. If all electrical nodes were measured and the line model were known, topology recovery would be closer to identifying the connectivity among observed buses. In practice, however, the measured node set may contain only a subset of the physical feeder. The inferred topology may therefore correspond either to an equivalent topology over observed nodes or to a reconstructed topology that attempts to include unobserved intermediate nodes.

Li et al. address impedance and topology estimation under limited sensing conditions, explicitly considering cases with observed and hidden nodes. Their work analyses how unobserved intermediate nodes affect both impedance estimation and topology recovery. They use OLS-based estimation under different measurement scenarios, analyse the bias introduced by hidden quantities through a factorised OLS formulation, and apply recursive grouping to infer topology in the presence of hidden nodes [17]. The main learning-based experiments in this dissertation assume that all nodes represented in the target adjacency matrix are observed and therefore do not reconstruct unknown intermediate physical buses. Hidden-node reconstruction is considered only through the independent analytical partial-observability comparison presented later in the dissertation.

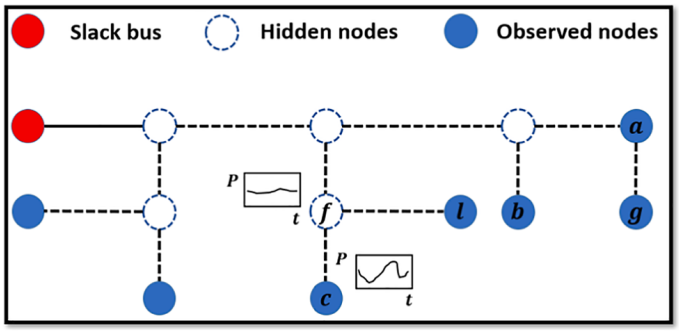


Figure 2.5: Illustration of a distribution network with observed nodes and hidden nodes. The figure highlights the limited-observability setting in which only part of the network is metered, while other buses remain hidden. Extracted from [17].

Figure 2.5 illustrates this limited-observability setting by distinguishing observed nodes, for which measurements are available, from hidden nodes, whose electrical behaviour and connectivity must be inferred indirectly.

Park, Deka and Chertkov study topology and parameter estimation in distribution grids under minimal observability assumptions. Their work analyses the conditions under which topology and line parameters can be recovered when only limited end-user measurements are available and intermediate nodes may be unobserved [34]. In this context, the term “exact” should be understood

as referring to recovery under the assumptions and identifiability conditions of the method, rather than as a general guarantee for all LV smart-meter settings.

Moffat, Bariya and von Meier analyse unsupervised impedance and topology estimation and explicitly discuss fundamental limitations caused by unobserved nodes. Their work shows that, when only a subset of active nodes is measured, the information available from measurements may correspond to an effective or reduced representation of the original network, rather than to the complete physical feeder [18]. This is closely related to Kron reduction, which is an algebraic operation applied to the network admittance matrix to eliminate unobserved or passive nodes and obtain an equivalent reduced admittance matrix over the observed nodes.

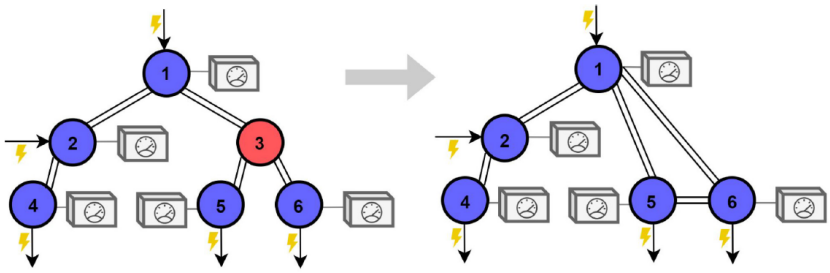


Figure 2.6: Illustration of hidden or passive nodes and the resulting reduced network representation used in unsupervised impedance and topology-estimation settings. The figure supports the distinction between the full physical feeder and the equivalent representation available from measurements at observed nodes. Extracted from [18].

It is important to clarify that Kron reduction is not simply the visual removal of nodes from a graph. It is an algebraic reduction of the admittance matrix in which unobserved nodes are eliminated, producing an equivalent model among the remaining observed nodes. This reduced model can preserve the electrical behaviour seen from the measured nodes, but it may no longer have the same sparsity pattern or direct physical interpretation as the original feeder. As a result, topology inference from reduced representations can become ambiguous: different physical networks may lead to similar equivalent behaviour at the observed nodes.

Yuan et al. formulate the inverse power-flow problem, where the goal is to infer the nodal admittance matrix from electrical measurements [27]. Their formulation assumes access to voltage and current phasor measurements, which is a stronger measurement setting than the smart-meter-like data considered in this dissertation. Under full observability, such measurements can support the identification of the admittance matrix from multiple operating points. When hidden nodes are present, the identifiable object may instead be a Kron-reduced admittance matrix over the observed nodes.

Other works further analyse how hidden nodes, partial measurements and limited excitation affect topology recovery. Doddi, Deka and Salapaka address partially observed meshed distribution grids using voltage measurements and matrix-decomposition ideas [42]. Anguluri et al. study topology identification with hidden nodes through structured norm minimisation, decomposing observed statistical relationships into sparse and low-rank components [32]. Deka et al.

study topology learning in underexcited power grids, showing that insufficient variability in the measurements can make topology recovery difficult or non-identifiable [31].

These works show that topology inference is limited not only by algorithmic design, but also by the information content of the measurements. Hidden nodes, missing measurements, reduced equivalent models and insufficient excitation can prevent the available data from uniquely determining the physical feeder. Although the main learning-based models do not reconstruct hidden intermediate nodes, these limitations show that the inferred topology is conditional on the observed-node representation and on the assumed measurement setting. Accordingly, the resulting edge scores should be interpreted as topology-related structural evidence under the stated assumptions, rather than as a general guarantee of complete physical-feeder recovery.

2.6 Admittance and Impedance Matrix Estimation

A complementary line of work estimates impedance or admittance representations of the grid. This perspective links topology inference to matrices whose structure depends directly on the connectivity and on the line parameters. In an AC network, nodal currents and voltages are related as shown in Equation 2.3, where $\mathbf{I}$ is the vector of nodal current injections, $\mathbf{V}$ is the vector of nodal voltages and $\mathbf{Y}$ is the bus admittance matrix.

The relationship between topology and electrical parameters can be made explicit using the network incidence representation. If $\mathbf{A}_{\text{inc}}$ is the branch-node incidence matrix and $\mathbf{D}_y$ is a diagonal matrix containing the branch admittances, then, neglecting shunt elements for simplicity, the bus admittance matrix can be written as

$$\mathbf{Y} = \mathbf{A}_{\text{inc}}^T \mathbf{D}_y \mathbf{A}_{\text{inc}}. \tag{2.10}$$

In this expression, $\mathbf{A}_{\text{inc}}$ encodes the physical connectivity of the network, while $\mathbf{D}_y$ contains the electrical characteristics of the branches. Therefore, the sparsity pattern of $\mathbf{Y}$ is linked to the graph structure, and its numerical values are linked to line admittances. Under appropriate modelling assumptions, estimating $\mathbf{Y}$ or related impedance representations can therefore provide information about both network topology and line parameters.

Yuan et al. formulate the inverse power-flow problem as the inference of the nodal admittance matrix from voltage and current phasor measurements [27]. This is a stronger measurement setting than the customer-level active-power and voltage-magnitude data considered in this dissertation, because it assumes access to phasor quantities rather than only active-power and voltage-magnitude measurements. Nevertheless, the work provides theoretical insight into what can be identified from electrical measurements under full and partial observability conditions.

Zhang, Wang and Zhang propose admittance-matrix estimation using linear regression and decoupled linear power-flow equations. Their method uses active-power injections, reactive-power injections and voltage-magnitude measurements, avoiding the need for voltage-angle information.

In this context, limited observability should be understood as a non-PMU setting with limited measurement types, rather than as the absence of measurements [43].

Zhang et al. further propose a numerical method for topology identification and line-parameter estimation in non-PMU distribution networks. The method is based on a two-step framework: first, a data-driven regression stage provides a preliminary estimate of the topology and line parameters; then, a joint data- and model-driven stage uses power-flow equations and a specialised Newton-Raphson iteration to refine line parameters, recover voltage angles and correct the topology. This addresses topology learning when voltage-angle measurements are not available, although it still aims at a deterministic topology and parameter estimate rather than a score-based topology-inference output [14].

Fabbiani et al. propose an online recursive least-squares method for identifying AC distribution networks. The online formulation allows the admittance-matrix estimate to be updated as new measurements arrive, avoiding the need to recompute the full estimation problem from scratch. Their approach estimates the admittance matrix, whose nonzero entries encode topological and line-parameter information, and combines recursive identification with optimal design of experiments. In this context, voltage excitations are informative voltage operating conditions selected to improve identifiability and accelerate convergence [44].

Brouillon et al. propose Bayesian methods for the identification of distribution networks. Their framework uses voltage and current measurements from micro-PMUs, accounts for measurement noise in both voltage and current data, and incorporates prior information such as sparsity patterns and known line admittances [45]. From the uncertainty-aware perspective adopted here, this introduces prior knowledge and probabilistic modelling into network identification.

Li et al. also connect impedance estimation with topology learning under limited sensing conditions. Their method estimates impedance-related quantities using OLS-based formulations under different measurement scenarios, while explicitly analysing the bias introduced by hidden nodes through a factorised OLS approach. Because the work focuses strongly on hidden-node effects and recursive grouping, it is closely related to the identifiability limitations discussed in Section 2.5 [17].

A general regularised admittance-estimation problem may be written as

$$\hat{\mathbf{Y}} = \arg\min_{\mathbf{Y}} \|\mathbf{I} - \mathbf{Y}\mathbf{V}\|_F^2 + \lambda\,\Omega(\mathbf{Y}), \tag{2.11}$$

where $\mathbf{Y}$ is the admittance matrix to be estimated, $\mathbf{I}$ and $\mathbf{V}$ are the measured current and voltage quantities, and the first term measures the mismatch between the measurements and the linear network model. The term $\Omega(\mathbf{Y})$ is a regularisation function that imposes prior structural information on the solution, such as sparsity, symmetry, Laplacian structure, prior connectivity or known line information. The parameter λ controls the relative importance of this regularisation term: larger values enforce the prior assumptions more strongly, while smaller values give more weight to the measurement-fitting term.

Such formulations combine measurement fitting with physical or structural constraints. However, they usually require measurements of sufficient quality and tend to produce a single estimated admittance matrix or a single inferred topology. This can be limiting in LV smart-meter settings, where measurements may be noisy, incomplete, asynchronous or insufficient to uniquely determine the physical feeder. In these cases, several topologies may be electrically plausible and similarly consistent with the available data. This motivates approaches that retain topology-related uncertainty through edge scores, probabilistic outputs or subsequent physical validation, rather than treating a single output as unambiguous.

The reviewed admittance and impedance estimation methods provide a bridge between electrical modelling and topology learning. Their limitation for the present dissertation is not the lack of physical grounding, but the measurement assumptions and the output structure. Several methods require phasor or current measurements that are not generally available in standard LV smart-meter settings, and most formulations still return a single estimated matrix or topology. This again motivates score-based topology inference when the available data do not uniquely identify the feeder.

2.7 Probabilistic, Bayesian and Online Topology-Tracking Methods

Probabilistic approaches provide a way to represent uncertainty, prior knowledge and incomplete information in distribution-grid topology inference. Instead of treating topology learning only as a deterministic optimisation problem, these methods model the relationship between measurements, latent network states and candidate structures in probabilistic terms. This formulation applies when the available data are noisy, incomplete, sparse or insufficient to uniquely determine one topology.

Zhao et al. propose a full-scale distribution-system topology-identification framework based on Markov Random Fields [28]. In this approach, the distribution network is represented as a probabilistic graphical model, where statistical dependencies between nodal voltage measurements are used to infer the underlying connectivity. The method uses smart-meter and terminal data to model voltage correlations between nodes, and then applies Markov Random Field structure learning to estimate the most likely network structure. Because directly computing the full likelihood of a large graphical model can be computationally expensive, the authors use pseudo-likelihood estimation and regularisation. The pseudo-likelihood reduces computational complexity by approximating the global likelihood through local conditional relationships, while the regularisation term helps avoid overfitting and promotes a more physically meaningful sparse structure. This work moves beyond simple pairwise voltage-correlation heuristics and formulates topology identification as a probabilistic structure-learning problem.

Liu et al. propose a graph-bank tracking Bayesian-network method for real-time topology estimation in active distribution systems [46]. The central idea is to construct a graph bank, that is, a predefined set of possible topology states obtained from prior network models, switching relationships and feasible operating configurations. Instead of searching over all possible graphs,

the method tracks which topology in this graph bank is most consistent with the incoming measurements. Bayesian networks are then used to model probabilistic dependencies between the topology state and real-time measurements, such as voltage and power-injection data. As new measurements become available, the method updates the probability of each candidate topology and tracks the most likely network configuration over time. This applies to operational settings in which distribution networks may change due to switching actions, faults or reconfiguration events. However, this approach assumes that a meaningful graph bank can be constructed in advance, which is different from the setting of this dissertation, where the objective is to infer plausible radial LV structures from smart-meter-like P/V data without relying on a complete predefined set of switching configurations.

Bayesian formulations also appear in admittance and parameter identification. Brouillon et al. introduce maximum likelihood and maximum a posteriori formulations for the identification of distribution networks [45]. In a maximum likelihood formulation, the objective is to find the admittance or parameter values that make the observed measurements most likely under a statistical noise model. In a maximum a posteriori formulation, prior knowledge is also included, such as expected sparsity, known line parameters or candidate connectivity information. This allows the estimation process to combine measurement fitting with prior structural assumptions. The Bayesian perspective explicitly recognises that measurements are imperfect and that prior information can guide the solution when the data alone are insufficient. However, these methods generally focus on estimating admittance or parameter values under specific measurement assumptions, often involving voltage and current measurements with higher information content than standard smart-meter records.

Another line of work studies when a distribution-network topology can be uniquely identified from a limited set of real-time measurements. Cavraro et al. develop a meter-placement strategy for topology recovery under limited monitoring [19]. Their work partitions the grid into observable islands and derives conditions under which the true topology can be uniquely recovered. For the present dissertation, this shows that topology inference depends not only on the inference algorithm, but also on where measurements are available and whether they contain enough information to distinguish between alternative configurations. This reinforces the uncertainty-aware perspective adopted here: when measurement placement or information content is insufficient, a single uniquely identifiable topology may not be guaranteed, and the available measurements may not be sufficient to support an unambiguous topology estimate.

Active probing methods provide another probabilistic or information-driven perspective on topology learning. Instead of relying only on naturally occurring load and voltage variations, probing methods deliberately perturb controllable devices, such as inverters, in order to create more informative electrical responses. Cavraro and Kekatos propose graph-based topology identification using power-grid probing and later extend this idea to inverter probing for distribution-network topology processing [29], [30]. These methods can improve excitation and make the topology more identifiable when controllable resources are available and probing actions are operationally acceptable. However, they are less aligned with the passive smart-meter-like setting considered in

this dissertation, where the objective is to infer topology from available P/V measurements rather than from deliberately injected perturbations.

Overall, probabilistic, Bayesian and online tracking methods show that topology inference can be formulated as an inference problem under uncertainty. They also show that prior information, measurement placement, candidate topology sets and measurement quality strongly influence what can be recovered. Nevertheless, many of these methods either rely on stronger measurement settings, assume a predefined set of candidate configurations, or return a most likely single topology. This dissertation builds on the same limited-observability motivation, but focuses on offline-trained learning models that produce topology-related edge scores from customer-level active-power and voltage-magnitude data.

2.8 Generative and Attention-Based Models for Topology Inference

2.8.1 From Deterministic Estimates to Edge-Score-Based Inference

The literature reviewed so far shows that many topology-learning methods return a single estimated topology, even when the available measurements are limited, noisy or ambiguous. This is suitable when the measurement information is sufficient to identify one clearly dominant solution. However, in LV smart-meter settings, this condition is not always guaranteed. Measurements may be aggregated, asynchronous, affected by noise, or limited to active power and voltage magnitude. In addition, hidden intermediate nodes, uncertain phase connections and electrically similar radial structures can make several candidate topologies plausible.

This motivates a shift from deterministic topology reconstruction towards uncertainty-aware and score-based inference. Instead of representing the problem only as the search for one optimal topology,

$$\mathscr{G}^{\star} = \arg\min_{\mathscr{G}\in\mathscr{T}} \mathscr{L}(\mathscr{G};\mathbf{X}), \tag{2.12}$$

where $\mathscr{T}$ is the set of feasible radial topologies, $\mathbf{X}$ represents measurement-derived features and $\mathscr{L}$ is a mismatch or fitness function, a learning-based formulation can first produce topology-related scores between node pairs.

In this perspective, the model does not need to output a final graph directly. It can instead learn a mapping from measurement-derived inputs to an edge-score matrix,

$$\mathbf{S} = f_{\theta}(\mathbf{X}), \tag{2.13}$$

where $\mathbf{S}_{ij}$ represents the learned plausibility of a connection between nodes i and j. The final radial topology can then be obtained by applying a radial projection operator,

$$\widehat{\mathscr{G}} = \Pi_{\text{radial}}(\mathbf{S}). \tag{2.14}$$

This formulation preserves the motivation for uncertainty-aware topology inference without requiring the explicit generation or ranking of multiple candidate topologies in the experiments. It

is also consistent with the modelling pipeline used in this dissertation, where VAE-based, convolutional, variable-size, *Transformer*-inspired and graph-completion models produce edge scores that are projected into valid radial structures for quantitative evaluation.

2.8.2 Variational Autoencoders

Variational Autoencoders (VAEs) provide a suitable generative framework for this formulation because they learn continuous latent representations and support sampling from the learned latent space. Unlike a deterministic autoencoder, which maps an input directly into a fixed latent vector and reconstructs it through a decoder, a VAE maps each input into a probability distribution in the latent space. In the standard formulation, the encoder approximates the posterior distribution over latent variables,

$$q_\phi(\mathbf{z}|\mathbf{x}), \tag{2.15}$$

where $\mathbf{x}$ is the input data, $\mathbf{z}$ is the latent variable and ϕ represents the encoder parameters.

This approximate posterior is often modelled as a Gaussian distribution with mean and variance produced by the encoder network:

$$q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}\left(\mu_\phi(\mathbf{x}), \mathrm{diag}\left(\sigma_\phi^2(\mathbf{x})\right)\right). \tag{2.16}$$

The decoder then maps samples from this latent distribution back to the data space through a likelihood model,

$$p_\theta(\mathbf{x}|\mathbf{z}), \tag{2.17}$$

where θ represents the decoder parameters.

During training, the latent variable can be sampled using the reparameterisation trick:

$$\mathbf{z} = \mu_\phi(\mathbf{x}) + \sigma_\phi(\mathbf{x}) \odot \varepsilon, \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{2.18}$$

which allows gradients to be propagated through the stochastic sampling operation.

In the standard VAE formulation, the training objective is usually written as the negative evidence lower bound. It includes a reconstruction term and a Kullback-Leibler regularisation term [47], [48]:

$$\mathscr{L}_{\mathrm{VAE}} = -E_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] + D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x}) \,\|\, p(\mathbf{z})\right). \tag{2.19}$$

The first term encourages the decoder to reconstruct the input from the latent representation. The second term regularises the approximate posterior distribution so that it remains close to the chosen prior distribution, commonly a standard normal distribution. Although several VAE variants modify, weight or extend these terms, this standard formulation captures the central idea used in this dissertation: learning a continuous latent representation from which topology-related outputs, such as edge-score matrices, can be decoded.

2.8.3 Convolutional Variational Autoencoders

Convolutional Variational Autoencoders (ConvVAEs) extend the standard VAE framework by replacing fully connected encoder and decoder blocks with convolutional layers. This allows the model to exploit local spatial patterns and structured matrix representations more efficiently than dense architectures.

In topology inference, measurement-derived representations such as correlation matrices, sensitivity matrices and edge-score matrices can be interpreted as structured two-dimensional objects. Convolutional layers can therefore exploit structured local patterns in these matrix representations before latent encoding, although the interpretation of locality depends on the adopted node ordering.

For this reason, convolutional VAEs are considered in this dissertation as an alternative to standard fixed-size VAEs.

2.8.4 Graph Generative Models

Graph generative models are also relevant because the output of topology inference is a graph. Variational Graph Autoencoders model graph structure using latent node embeddings and probabilistic edge decoding [49], while GraphVAE extends the VAE framework to the generation of discrete graph structures [50]. A generic edge decoder can be written as

$$p(A_{ij} = 1 | \mathbf{z}_i, \mathbf{z}_j) = \sigma(\mathbf{z}_i^T \mathbf{z}_j), \tag{2.20}$$

where A_{ij} is an adjacency entry, $\mathbf{z}_i$ and $\mathbf{z}_j$ are latent representations associated with nodes i and j, and σ is a sigmoid function.

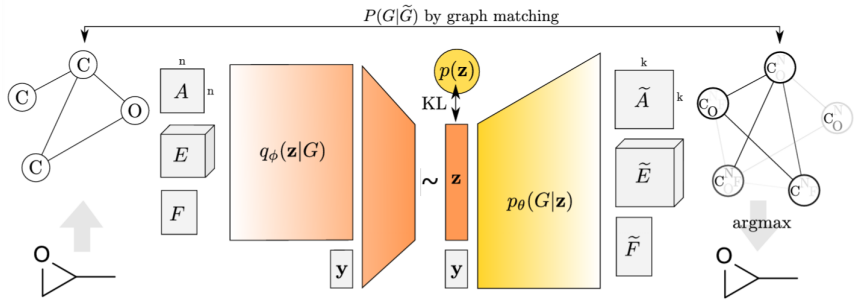


Figure 2.7: Variational graph-generation architecture in which a discrete graph is encoded into a continuous latent representation and decoded into a probabilistic graph. Extracted from [50].

Figure 2.7 illustrates the GraphVAE architecture proposed by Simonovsky and Komodakis [50]. In the original formulation, the input is a discrete graph $\mathscr{G}$ represented by structural and attribute information. The adjacency matrix $\mathbf{A}$ indicates which nodes are connected, the edge tensor $\mathbf{E}$ stores edge-type information, and the node-feature matrix $\mathbf{F}$ stores node attributes. In the molecular-generation example used in the original paper, nodes correspond to atoms, edges correspond to chemical bonds and attributes represent atom or bond types. In an LV topology-inference

context, the analogous interpretation would be that nodes represent buses or customer connection points, edges represent electrical branches, and optional attributes may describe node or branch features.

The encoder maps the input graph into a latent distribution, from which a latent vector z is sampled. The decoder then maps z back into a probabilistic graph. Rather than generating edges sequentially, GraphVAE decodes a fully connected probabilistic graph up to a predefined maximum number of nodes. This means that the decoder outputs probabilities for node existence, edge existence and graph attributes. In the context of this dissertation, the relevant idea is not the direct generation of complete graph samples, but the use of probabilistic edge decoding to obtain topology-related edge scores that can later be constrained by radial projection.

This representation is useful for graph generation because it allows a discrete graph object to be represented through differentiable probabilistic outputs during training. However, it also introduces challenges. First, graph generation is permutation-invariant: the same graph can be represented by different node orderings. For this reason, GraphVAE includes a graph-matching step to align the decoded probabilistic graph with the target graph during training. Second, unconstrained graph generation does not automatically guarantee that the generated graph satisfies domain-specific requirements. In LV topology inference, this means that additional constraints or post-processing steps are required to ensure that generated candidates are connected, radial and physically plausible.

2.8.5 Transformer Encoders and Self-Attention

Transformer architectures were originally introduced by Vaswani et al. [51] for sequence modelling, replacing recurrent processing with attention mechanisms that allow each element of an input sequence to interact directly with all other elements. The central operation is scaled dot-product attention, where input embeddings are projected into queries, keys and values. In simplified form, the attention operation can be written as

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{2.21}$$

where Q, K and V are the query, key and value matrices, respectively, and d_k is the dimension of the key vectors. The product QK^T measures the compatibility between elements, while the softmax operation converts these scores into attention weights. The output is a weighted combination of the value vectors, allowing each element to be updated using information from the remaining elements.

Although Transformers were first developed for ordered sequences, the attention mechanism is also relevant for problems involving sets of interacting elements. In graph-related applications, the input elements can be interpreted as nodes or node descriptors, and self-attention can be used to contextualise each node representation with respect to the others. This is useful when the relation between two elements depends not only on their individual properties, but also on the global configuration of the full set.

This idea is particularly relevant for LV topology inference. In a radial feeder, the existence of a branch between two nodes is not determined only by the isolated measurements of those two nodes. It also depends on how their voltage and power variations relate to the remaining nodes in the network. As the number of nodes increases, the number of possible node pairs grows rapidly, and the model must distinguish between a larger set of plausible candidate branches. A mechanism that allows node embeddings to exchange information before graph-level aggregation can therefore be useful for representing global node-to-node dependencies.

In variable-size settings, attention must also account for padding. When feeders with different numbers of nodes are grouped in the same mini-batch, shorter feeders are padded for computational alignment. A mask can be added to the attention scores so that padded positions do not contribute to the attention weights. In this masked formulation, the attention operation becomes

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V, \tag{2.22}$$

where M is an additive mask. Valid node pairs receive a neutral mask value, while padded positions receive a large negative value before the softmax operation. This prevents padded elements from influencing the contextualised node embeddings.

For the purposes of this dissertation, Transformers are not used as a general-purpose sequence model. Instead, the relevant idea is the self-attention mechanism as a way to contextualise node embeddings before graph-level pooling. This motivates the *Transformer*-inspired extension evaluated in Chapter 4, where the initial variable-size VAE is modified by inserting a masked self-attention stage between the shared node encoder and the graph-level aggregation step.

2.8.6 Implications for LV Topology Inference

The application of graph generative models to LV topology inference therefore requires additional physical and structural constraints. An inferred topology should be connected, radial, compatible with the available measurements and physically plausible in terms of voltage behaviour. Therefore, a generative or graph-learning model cannot be used only as an unconstrained graph predictor. Its outputs must be interpreted together with radiality constraints, edge-score projection, power-flow consistency and topology-validation criteria.

This dissertation is positioned within this score-based perspective. The objective is not to explicitly generate and rank multiple complete topology candidates, but to explore how VAE-based, graph-based and attention-inspired models can learn topology-related representations from smart-meter-like P/V measurements. In the experimental framework adopted here, these models produce edge-score matrices or connection probabilities, which are then projected into valid radial topologies for quantitative evaluation.

2.9 Research Gap and Positioning of the Dissertation

The reviewed literature shows that distribution-grid topology learning is an active research area, but several open challenges remain for LV topology inference from limited customer-level active-power and voltage-magnitude data.

First, many high-accuracy methods rely on measurement conditions that are not yet typical in LV grids, such as dense micro-PMU deployment, voltage angles, current phasors or line-flow measurements. Smart meters are more widely available, but they provide lower-resolution and less synchronised data.

Second, several methods assume known candidate infrastructure or switch sets. This is suitable for topology detection and verification, but less suitable when the LV topology records are incomplete or outdated and the task is closer to reconstruction.

Third, hidden nodes remain a fundamental challenge. Works on Kron reduction, recursive grouping, partial observability, underexcitation and structured norm minimisation show that hidden nodes and limited excitation restrict what can be uniquely recovered from measurements [18], [27], [31], [32], [42]. Therefore, a single inferred topology may conceal ambiguity arising from the measurement and observability conditions.

Fourth, many existing methods produce or select a single final topology. Pengwah et al. [8] are an exception because they generate multiple candidate graphs through backtracking before selecting the best-fitting candidate. Comparatively fewer studies evaluate offline-trained models that produce edge-level scores from limited customer-level P/V-type data and subsequently enforce radiality through graph projection.

Fifth, VAEs, convolutional VAEs and graph generative models are established machine-learning frameworks, but their application to radial LV topology inference remains comparatively limited. The opportunity is to use generative and graph-based learning models not simply as black-box predictors, but as mechanisms for learning topology-related representations from measurement-derived inputs and decoding them into edge-score matrices. In this dissertation, these models are used to produce topology-related edge scores and are not evaluated as explicit generators of ranked sets of complete topologies.

Table 2.2 summarises the main families of approaches reviewed in this chapter and highlights their limitations for LV topology inference from smart-meter-like P/V data.

Table 2.2: Summary of topology-learning approaches and their limitations for LV smart-meter-based topology inference.

Method family	**Data assumptions**	**Main contribution**	**Limitation for this dissertation**
Voltage correlation and clustering	Smart-meter voltage time series	Uses voltage similarity to infer electrical proximity	May confuse proximity with correlated loads, phase effects or common voltage regulation
Sensitivity-based reconstruction	Voltage and load or power data	Links voltage response to radial feeder structure	Depends on voltage availability, excitation and linear approximation quality
Hidden-node and reduced-network methods	Partial measurements with hidden nodes	Analyse unobserved internal structure	Often require assumptions stronger than smart-meter-like data
Admittance and impedance estimation	Phasors or non-PMU approximations	Relates topology and electrical parameters	Often requires richer measurements and returns one matrix or topology
MILP and state-estimation methods	Sensors, pseudo-measurements and candidate switches	Encodes constraints and switching logic	Often assumes known infrastructure or dedicated sensors
Bayesian and online tracking	Prior topology sets and real-time measurements	Tracks likely configurations under uncertainty	Often depends on predefined topology banks
Graph and generative models	Learned representations and edge-score outputs	Learns topology-related representations and pairwise connection scores	Requires radial projection and physical validation to ensure plausible LV topologies

The contribution of this dissertation is positioned at this intersection. It combines ideas from physics-based topology learning, measurement-derived electrical representations and learning-based graph reconstruction. In particular, the dissertation evaluates fixed-size VAE models, convolutional VAE variants, variable-size VAE architectures, *Transformer*-inspired extensions and graph-completion formulations. The central objective is to assess whether these models can infer radial LV topologies from P/V-derived representations under controlled simulation conditions, while enforcing radiality through graph post-processing.

2.10 Summary

This chapter reviewed the state of the art on distribution-grid topology learning, with emphasis on LV topology inference under limited observability. The literature shows that topology learning can be addressed through smart-meter correlations, sensitivity matrices, graphical models, hidden-node reconstruction, impedance/admittance estimation, MILP formulations, state-estimation approaches, active probing, Bayesian tracking and graph-learning methods. These approaches provide essential foundations, but they also reveal the limits of deterministic topology reconstruction under noisy, incomplete and partially observed measurements.

The main conclusion is that LV topology inference can be formulated as a reconstruction problem under limited observability and measurement uncertainty. Under limited customer-level measurement conditions, different radial structures may produce similar electrical patterns. This motivates the controlled evaluation developed in the following chapters, where measurement-derived representations, VAE-based models, variable-size architectures and radial graph projection are assessed for LV topology inference.

Chapter 3

Synthetic Data Generation and Modelling Framework

This chapter describes the methodology adopted to construct the controlled low-voltage (LV) grid dataset used throughout this dissertation. The dataset combines synthetically generated radial topologies, residential active-power profiles derived from real 30-minute smart-meter measurements, and voltage-magnitude trajectories obtained through power-flow simulation. This controlled construction is necessary because sufficiently large and diverse real LV datasets containing both customer-level measurements and verified ground-truth topologies are generally unavailable for training and validating supervised topology-inference algorithms. The resulting dataset is later used in Chapter 4 to evaluate different input representations and learning models.

The methodological work starts from a practical limitation: real LV grid data are difficult to obtain in a complete and reliable form. Smart-meter measurements are subject to privacy and confidentiality constraints, while LV topology records may be incomplete, outdated or inconsistent with the actual field configuration. This issue is well recognised in the topology reconstruction literature, where LV grids are often described as partially known, inaccurately documented or affected by missing electrical parameters [7], [20], [24]. In addition, supervised learning and quantitative evaluation require reliable ground-truth topologies, which are rarely available for real LV feeders.

For these reasons, a synthetic dataset was constructed. Synthetic data provide three important advantages. First, the true topology is known exactly and can therefore be used as ground truth. Second, the number of nodes, branching patterns, line parameters and operating conditions can be controlled. Third, the same procedure can be repeated many times to create diverse datasets for training and testing topology inference methods. The objective is not to replace real-world validation, but to create a physically coherent and reproducible environment in which topology inference methods can be developed and compared. Accordingly, the results should be interpreted as evidence of methodological feasibility under controlled conditions rather than as a direct demonstration of field performance.

The overall pipeline of the dissertation is shown in Figure 3.1. The present chapter focuses on

the first part of the pipeline: synthetic radial topology generation, active-power profile generation, Forward-Backward Sweep power-flow simulation and construction of the final P/V dataset. The subsequent stages, namely feature extraction, model training, edge-score generation and projection into radial topologies, are analysed in Chapter 4.

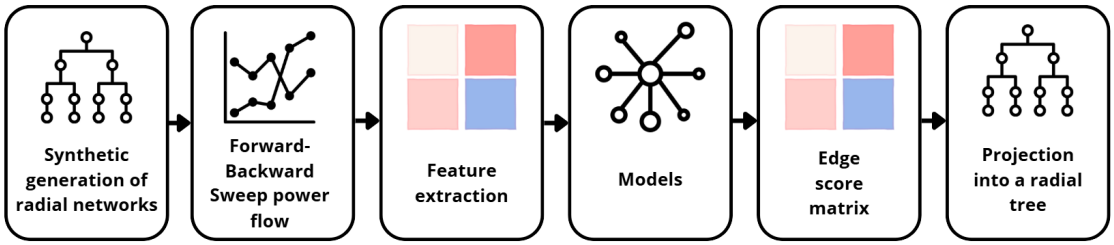


Figure 3.1: Overall pipeline of the dissertation. The methodology starts with the synthetic generation of radial LV networks and Forward-Backward Sweep power-flow simulation to obtain P/V datasets. The subsequent stages, including feature extraction, model training, edge-score generation and radial-tree projection, are analysed in the experimental results chapter.

The chapter is organised as follows. Section 3.1 explains the motivation for synthetic data generation. Section 3.2 defines the radial graph representation used for LV feeders. Section 3.3 presents the physical assumptions adopted to simplify the initial modelling problem. Section 3.4 describes the validation experiments used to support these assumptions. Section 3.5 explains the generation of residential active-power profiles. Section 3.6 presents the Forward–Backward Sweep power-flow simulation used to compute voltage profiles. Section 3.7 validates the physical coherence of the resulting synthetic electrical data. Section 3.8 summarises the construction of the synthetic P/V dataset. Section 3.9 introduces the topology-inference modelling pipeline used in the following chapter. Section 3.10 defines the topology evaluation metrics used to compare inferred and ground-truth radial networks. Finally, Section 3.11 summarises the chapter.

3.1 Motivation for Synthetic Data Generation

The methodology proposed in this dissertation relies on supervised learning and representation-learning experiments that require paired measurement–topology datasets. This requirement is specific to the controlled learning-based framework evaluated here and does not imply that all topology-learning approaches require synthetic data. In the present setting, each training or evaluation sample must contain both electrical measurements and the corresponding ground-truth network topology. In practice, obtaining this type of dataset from real LV networks is difficult because real topology records may be incomplete or outdated, measurement availability is limited, and ground-truth connectivity is not always available with the level of detail required for supervised training and controlled benchmarking.

A second difficulty concerns the topology itself. LV networks are frequently modified due to new customer connections, maintenance actions, cable replacements, network reinforcements and field interventions. These changes may not be immediately reflected in digital records. As

a consequence, the documented topology may differ from the actual energised network. This is particularly problematic for supervised learning, because training and testing require reliable labels.

The controlled dataset therefore provides an alternative in which the electrical data can be associated with exactly known connectivity. Each sample corresponds to one generated radial feeder and one complete operating window. It contains the ground-truth adjacency matrix, the electrical branch parameters assigned for the power-flow simulation, the active-power time series for all nodes and the corresponding voltage-magnitude time series obtained through power-flow simulation. The active-power profiles are derived from real 30-minute residential smart-meter data, whereas the radial topology and its branch parameters are generated synthetically. This makes it possible to evaluate topology-reconstruction errors directly by comparing the inferred and ground-truth adjacency matrices.

3.2 Synthetic Radial Network Generation

Each synthetic LV feeder is represented as an undirected graph

$$\mathscr{G} = (\mathscr{V}, \mathscr{E}), \tag{3.1}$$

where $\mathscr{V}$ is the set of nodes and $\mathscr{E}$ is the set of branches. The root node represents the secondary substation or transformer connection point. The remaining nodes represent customer connection points or equivalent LV buses.

For a feeder with N nodes, node 0 is defined as the root node. The remaining nodes are introduced sequentially. For each new node j, one parent node is selected randomly from the set of previously introduced nodes $\{0, \ldots, j-1\}$, and the branch connecting the selected parent to node j is added to the graph. Since every new node is connected through exactly one branch to an already connected graph, the resulting topology is connected, acyclic and contains exactly $N-1$ branches. Radiality is therefore guaranteed by construction rather than verified only after generation. Generated topologies are subsequently converted into canonical undirected keys so that duplicated radial structures can be identified and removed.

The networks considered in this work are radial. Therefore, for a connected network with $N = |\mathscr{V}|$ nodes, the number of branches must satisfy

$$|\mathscr{E}| = N - 1. \tag{3.2}$$

This condition is necessary but not sufficient: the graph must also be connected and acyclic. The radiality assumption is consistent with the typical operating structure of LV distribution feeders and is also widely used in LV topology identification studies based on smart-meter data [7], [20].

For each generated topology, the corresponding adjacency matrix is defined as

$$A_{ij} = \begin{cases} 1, & \text{if nodes } i \text{ and } j \text{ are electrically connected,} \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

Since the network is undirected, $A_{ij} = A_{ji}$, and since self-connections are not allowed, $A_{ii} = 0$. This adjacency matrix is later used as the ground-truth topology for evaluating the models presented in Chapter 4.

Each branch $(i, j) \in \mathscr{E}$ is associated with an electrical impedance

$$Z_{ij} = R_{ij} + jX_{ij}, \tag{3.4}$$

where R_{ij} is the branch resistance and X_{ij} is the branch reactance. In the initial simplified setting, the resistive component is assumed to dominate the line behaviour. The implications of this assumption are analysed in Section 3.4.

Figure 3.2 shows an example of a radial LV test grid. The figure was redrawn and adapted from the test-grid representation used by Lima and Sampaio [20], and it illustrates the type of feeder structure considered when constructing the synthetic P/V dataset.

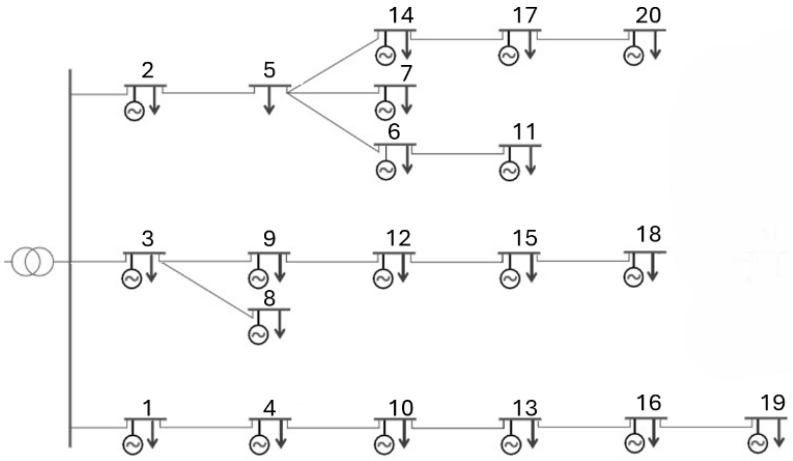


Figure 3.2: Example of a radial LV test grid used as inspiration for the synthetic radial networks considered in this dissertation. The original test-grid representation was adapted from [20] and redrawn to illustrate the type of feeder structure used in the synthetic P/V dataset generation process.

3.3 Preliminary Physical Assumptions

Before generating the final dataset, a set of physical assumptions was analysed in order to simplify the problem while maintaining electrical consistency. These assumptions are important because the dataset should not consist of arbitrary graphs and numerical values. It should consist of P/V pairs compatible with plausible LV power-flow behaviour.

The first assumption is radial operation. The feeder is represented as a tree rooted at the transformer. This allows the use of parent-child relationships and enables the Forward-Backward Sweep algorithm described in Section 3.6.

The feeder is represented using a balanced single-phase equivalent, thereby neglecting phase allocation, neutral effects, mutual coupling and three-phase voltage unbalance. This simplification provides a controlled phase-specific starting point for the topology-inference analysis and should not be interpreted as an assumption that real LV feeders are necessarily balanced.

The third assumption is that voltage drops between electrically adjacent nodes are relatively small. For a given operating point, the complex voltage-drop relation between two connected nodes i and j can be written as

$$\underline{V}_i^{(t)} - \underline{V}_j^{(t)} = Z_{ij}\underline{I}_{ij}^{(t)}, \tag{3.5}$$

where $\underline{V}_i^{(t)}$ and $\underline{V}_j^{(t)}$ are the complex voltage phasors at nodes i and j, and $\underline{I}_{ij}^{(t)}$ is the complex current phasor flowing through branch (i, j) at timestamp t. When branch voltage drops and angular differences are small relative to nominal values, first-order approximations become more accurate around the selected operating point.

The fourth assumption is that, in the initial approximation, the resistive component dominates the branch impedance. This choice is motivated by the context where low-voltage feeders commonly exhibit relatively high resistance-to-reactance ratios. Under these conditions, active-power variations have a strong influence on voltage magnitudes, making a resistance-dominated approximation a reasonable controlled starting point. This assumption is not intended to be universally valid for all LV grids. It can be expressed as

$$R_{ij} \gg X_{ij}, \tag{3.6}$$

or, more generally, by requiring the effects of branch reactance and voltage phase to remain small relative to the uncertainty introduced by measurements and operating variability. The validity and limitations of this approximation are evaluated in the following section through experiments involving the X_{ij}/R_{ij} ratio.

The fifth assumption concerns the available measurements. The initial dataset is based on active power P and voltage magnitude $|V|$. Reactive power Q and voltage phase angles are not used in the first modelling stage. This reflects the practical LV smart-meter setting considered in this dissertation: active-power or energy measurements and voltage magnitudes are more commonly available, whereas reactive-power measurements are not always recorded, stored or made accessible with sufficient temporal resolution. Voltage phase angles are even less commonly available because they require phasor-capable measurement equipment that is not typically deployed at LV customer connection points. Therefore, the omission of Q and voltage angles is an assumption about measurement availability, rather than an assumption that reactive power is physically absent from LV networks.

The complex power at node i is generally given by

$$S_i(t) = P_i(t) + jQ_i(t). \tag{3.7}$$

Under the active-power-only approximation, this becomes

$$S_i(t) \approx P_i(t). \tag{3.8}$$

This simplification reduces the amount of information required to generate and process the dataset. The following section explains how these assumptions were tested before constructing the final synthetic dataset.

3.4 Validation of the Modelling Assumptions

The validation experiments presented in this section do not estimate branch parameters for use by the topology-inference models. In the controlled dataset, branch impedances are assigned during data generation and are treated as known inputs to the power-flow simulation. Parameter estimation is therefore outside the scope of the proposed learning-based topology-inference framework.

The apparent-resistance expressions introduced below are used only as independent diagnostic tools to evaluate the internal consistency of the simplified electrical assumptions and the propagation of voltage-measurement errors. The first validation compares the theoretical and empirical variance of an apparent-resistance estimate under voltage noise. The second checks whether the generated scenarios produce admissible voltage behaviour and reliable power-flow convergence. The third evaluates the effect of neglecting voltage angles and branch reactance.

3.4.1 Empirical and Theoretical Variance

A first experiment compares the theoretical and empirical variance of an estimated branch resistance in a simplified setting. The objective is to verify whether the analytical expression for the variance of the estimated resistance is consistent with the empirical variance obtained through repeated simulations with voltage measurement noise.

The analysis focuses on a branch that supplies a leaf node, that is, one of the last branches of the network. In the notation adopted in this validation step, node i denotes the upstream node and node j denotes the downstream leaf node. This choice simplifies the interpretation of the results because the active power flowing through the selected branch corresponds directly to the active power consumed at the leaf node j. It also reduces the effect of possible inconsistencies accumulated downstream, allowing the impact of voltage measurement noise on the resistance estimate to be observed more clearly.

The initial validation is performed under a canonical simplified scenario. The branch is assumed to be purely resistive, with $X_{ij} = 0$, and the injected power is assumed to be only active, so that

$$S_j = P_j, \qquad Q_j = 0. \tag{3.9}$$

These assumptions isolate the effect of voltage measurement errors on the estimation of R_{ij}.

The theoretical expression starts from the complex power relation at the downstream leaf node j,

$$S_j = P_j + jQ_j = V_j I_{ij}^* = V_j \frac{(V_i - V_j)^*}{Z_{ij}^*}, \qquad Z_{ij} = R_{ij} + jX_{ij}. \tag{3.10}$$

Taking the real part under the assumptions used in this stage, namely $X_{ij} = 0$, $Q_j = 0$ and voltage measurements available only as magnitudes, the apparent resistance of the branch between nodes i and j reduces to

$$\hat{R}_{ij}(t) = \frac{V_j(t)\,[V_i(t) - V_j(t)]}{P_j(t)}. \tag{3.11}$$

This expression is only directly valid for the selected leaf-branch case, where node j is a leaf node and the active power in the denominator corresponds to the active power associated with that terminal node. For internal branches supplying downstream subtrees, replacing $P_j(t)$ by an aggregated downstream active power would be an additional approximation based on radial power conservation. That broader approximation is not the specific case validated in this experiment.

To evaluate the effect of measurement uncertainty, the reference voltages obtained from the power-flow simulation are perturbed with additive Gaussian noise. For each voltage magnitude,

$$\tilde{V}_i(t) = V_i(t) + \varepsilon_i(t), \tag{3.12}$$

where $\varepsilon_i(t)$ is a zero-mean Gaussian perturbation. The standard deviation used in this validation is

$$\sigma = \frac{0.005 \times 230}{3}. \tag{3.13}$$

This corresponds approximately to a $\pm 0.5\%$ interval around the nominal voltage within three standard deviations. For each noise realisation, the branch resistance is recomputed using the leaf-branch resistance expression defined above. This produces a set of resistance estimates.

For the selected leaf branch, the theoretical variance induced by the voltage measurement noise is

$$\mathrm{Var}_{\mathrm{teo}}\left(\hat{R}_{ij}\right) = \frac{\sigma^2\left(5V_j^2 + V_i^2 - 4V_iV_j\right) + 3\sigma^4}{P_j^2}. \tag{3.14}$$

The empirical variance is obtained from the simulated set of resistance estimates. If M noise realisations are generated, the empirical mean is

$$\overline{R}_{ij} = \frac{1}{M}\sum_{r=1}^{M} \hat{R}_{ij}^{(r)}, \tag{3.15}$$

and the empirical variance is computed as

$$\mathrm{Var}_{\mathrm{emp}}\left(\hat{R}_{ij}\right) = \frac{1}{M}\sum_{r=1}^{M}\left(\hat{R}_{ij}^{(r)} - \overline{R}_{ij}\right)^2. \tag{3.16}$$

This corresponds to the population variance of the accumulated resistance estimates obtained after repeated voltage-noise realisations.

Figure 3.3 compares the analytical theoretical variance with the empirical variance obtained by simulation. The purpose of this comparison is to verify whether both quantities coincide, or remain very close within the statistical uncertainty of the simulation. This validates the theoretical variance expression in the simplified scenario with $X_{ij} = 0$, $S_j = P_j$, $Q_j = 0$, and voltage measurements represented only by magnitudes.

The variance experiments are performed on a randomly generated fixed radial feeder. The base experiment shown in Figure 3.3 considers the terminal branch connecting nodes 4 and 14.

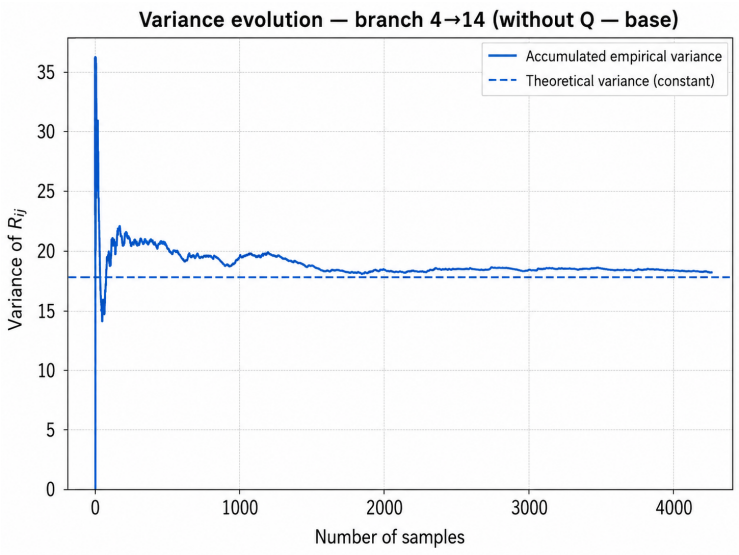


Figure 3.3: Comparison between the empirical and theoretical variance of the apparent-resistance estimate for terminal branch $4-14$ of a randomly generated fixed radial feeder. The base scenario considers active power only, voltage magnitudes and a purely resistive branch.

A second variance experiment was considered after introducing voltage phase and line reactance. In this setting, the voltage at each node is represented by a magnitude and an angle, and the branch impedance is complex. The resistance estimate therefore depends on the phase difference between adjacent nodes. A simplified expression for the phase-aware resistance estimate is

$$R_{ij}^{(\theta)}(t) = \frac{V_j(t)\left[V_i(t)\cos\left(\theta_i(t) - \theta_j(t)\right) - V_j(t)\right]}{P_j(t)}, \tag{3.17}$$

where $\theta_i(t)$ and $\theta_j(t)$ are the voltage angles at nodes i and j, respectively. This expression is used only in the preliminary diagnostic and is not part of the dataset-construction or topology-inference pipelines. It provides an apparent-resistance estimate that explicitly includes the voltage-angle difference between the adjacent nodes. Comparing this value with the magnitude-only estimate

makes it possible to quantify the approximation error introduced by neglecting voltage angles as the X_{ij}/R_{ij} ratio increases.

The second experiment uses the same randomly generated fixed radial feeder and considers the terminal branch connecting nodes 7 and 10, while including line reactance and voltage-angle information.

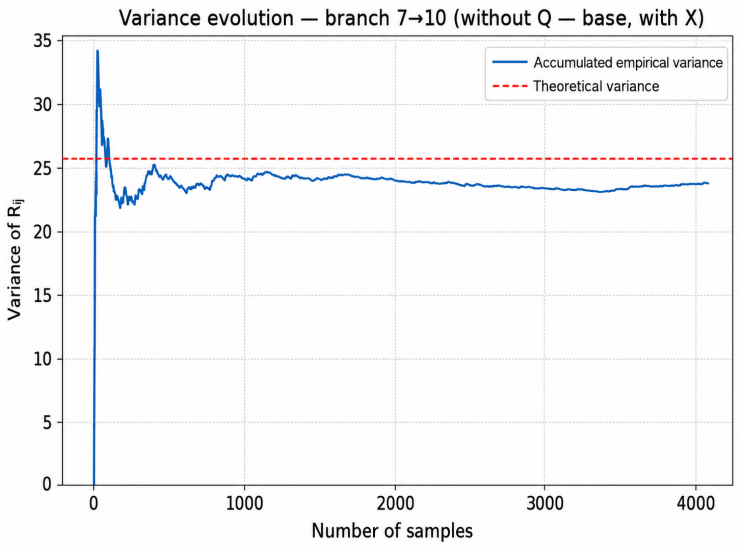


Figure 3.4: Comparison between the empirical and theoretical variance of the apparent-resistance estimate for terminal branch $7-10$ of the same randomly generated fixed radial feeder, after including line reactance and voltage-angle information.

These validations support the internal consistency of the simplified resistance-estimation model used in the preliminary analysis. Beyond validating the theoretical variance expression, they also validate, under a controlled known-topology setting, the intermediate step of estimating an electrical branch property from available P/V measurements. Because the connectivity and the selected leaf branch are known, the corresponding upstream and downstream voltage magnitudes and terminal active power can be associated with that branch, allowing its apparent resistance to be estimated.

The agreement between the empirical and theoretical variances shows that the uncertainty introduced by voltage measurement errors can also be quantified. Therefore, the procedure provides both an apparent-resistance estimate and an estimate of its measurement-induced uncertainty. This uncertainty depends on the voltage-noise level and on the active power in the denominator, meaning that very small active-power values can substantially increase the variance of the resistance estimate. This result is restricted to the controlled leaf-branch case and should not be interpreted as a general joint topology and line-parameter estimation method.

3.4.2 Acceptable Voltage-Drop Scenarios

A second validation step concerns the numerical and physical acceptability of the simulated voltage profiles. Since the synthetic topology generation process can create many possible radial

structures, not all generated cases are suitable for learning. Some combinations of topology, line parameters and load profiles may produce excessive voltage drops or unstable operating conditions.

To avoid including non-physical or numerically unstable cases in the dataset, the generated scenarios must satisfy two criteria. First, the Forward-Backward Sweep algorithm must converge. Second, the voltage magnitudes must remain within an admissible operating range:

$$V_{\min} \leq |V_i(t)| \leq V_{\max}, \qquad \forall i,t, \tag{3.18}$$

where $V_{\min}$ and $V_{\max}$ define the lower and upper voltage limits considered acceptable for the synthetic dataset.

Figure 3.5 illustrates an admissible voltage-drop scenario for the same randomly generated fixed radial feeder considered in the previous graphs. In this case, the voltage variations remain within a stable and physically plausible range, and the Forward-Backward Sweep algorithm converges reliably. This behaviour indicates that the generated topology, branch parameters and active-power profiles produce a suitable P/V sample for the topology-inference dataset.

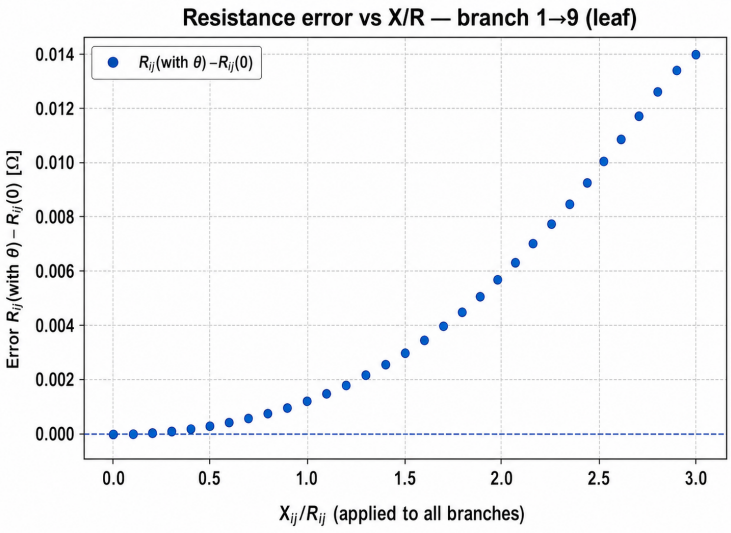


Figure 3.5: Diagnostic associated with an acceptable voltage-drop scenario. The resistance-estimation error remains smooth as the X_{ij}/R_{ij} ratio increases, indicating that the simulated voltage behaviour remains suitable for the simplified validation setting.

Cases with excessive or irregular voltage drops are not suitable for the dataset because they may indicate non-convergence, unstable operating points or physically implausible voltage behaviour. In such cases, the iterative power-flow calculation may become unstable, and the resulting voltages cannot be interpreted as reliable electrical states of the feeder. Therefore, scenarios that violate the convergence or voltage-range criteria are discarded before constructing the final P/V dataset. A representative rejected case is provided in Appendix A.6.

3.4.3 Impact of Phase Angle Neglect and the X/R Ratio

A third validation step evaluates the impact of neglecting voltage phase angles and branch reactance. In the complete branch model, the impedance is complex, as defined earlier, and the voltage drop depends on both magnitude and phase. However, the initial synthetic dataset uses voltage magnitudes and active power only. The purpose of this validation is to verify whether ignoring phase information still keeps the voltage-magnitude approximation consistent when voltage measurements are affected by errors.

The relevant ratio is

$$\rho_{ij} = \frac{X_{ij}}{R_{ij}}. \tag{3.19}$$

When ρ_{ij} is small, the branch is mainly resistive, so neglecting the phase angle introduces limited error. The voltage-magnitude approximation therefore remains reasonable when this systematic error is no greater than the uncertainty caused by voltage measurement errors. As ρ_{ij} increases, the reactive component becomes more significant and the approximation loses accuracy.

Figure 3.6 compares the phase-neglect error with the standard deviation of the resistance estimates across the X_{ij}/R_{ij} ratio. Neglecting phase is acceptable only while this systematic error remains small relative to the natural estimation uncertainty.

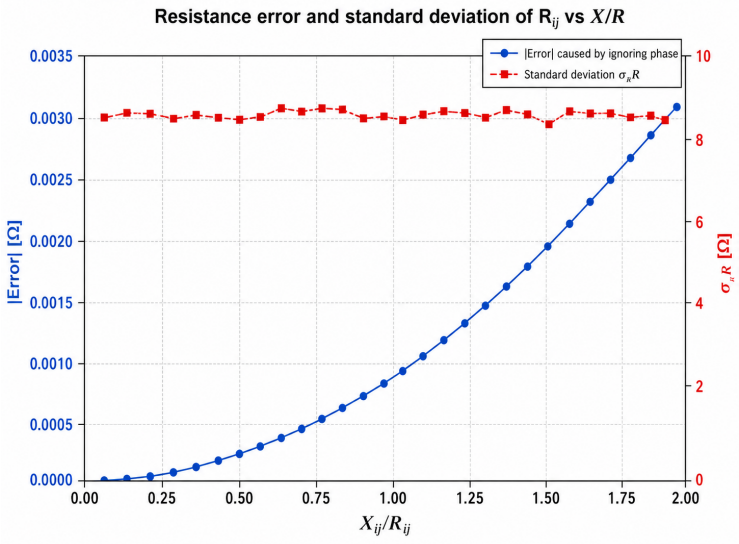


Figure 3.6: Error introduced by neglecting voltage phase and standard deviation of the resistance estimates as a function of the X_{ij}/R_{ij} ratio. The result supports the use of voltage-magnitude-based approximations when the systematic phase-related error remains small compared with the expected measurement uncertainty.

Figure 3.7 further illustrates the impact of reactance by showing how the difference between voltage magnitudes obtained with and without line reactance changes with the X_{ij}/R_{ij} ratio. The

increasing trend confirms that the effect of reactance and phase becomes more relevant as the branch becomes more reactive.

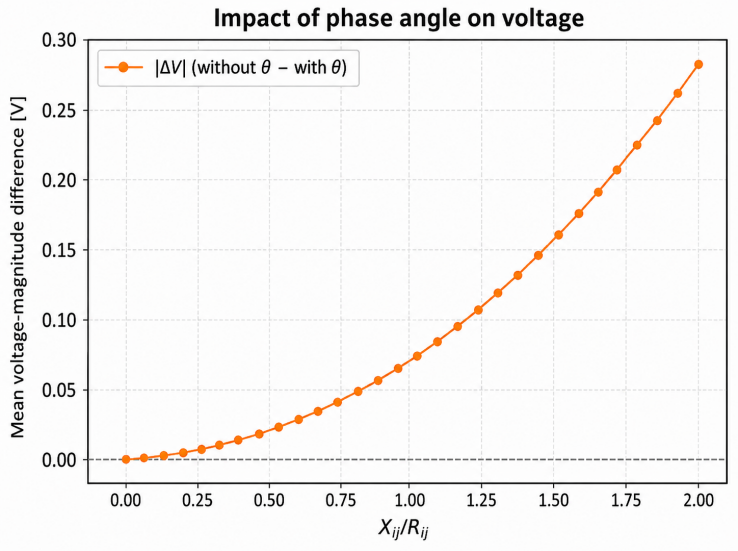


Figure 3.7: Difference between voltage magnitudes obtained with and without line reactance as a function of the X_{ij}/R_{ij} ratio. The increasing trend confirms that the impact of reactance and phase becomes more relevant as the line becomes more reactive.

These experiments support the use of a voltage-magnitude-based approximation in the initial synthetic data generation stage, while also clarifying its limits. The approximation is most appropriate for predominantly resistive LV feeders and becomes less reliable as the X/R ratio increases.

3.5 Residential Active-Power Profile Generation

After generating the radial topology and assigning line parameters, active-power profiles must be created for the customer nodes. These profiles provide the time-varying injections used by the power-flow simulation.

Instead of using arbitrary constant loads, the active-power profiles are generated to emulate residential consumption behaviour. The objective is to obtain time-varying operating conditions that reproduce the variability typically observed in low-voltage residential feeders. Each node is therefore associated with an active-power time series with residential-style temporal variation and randomly assigned consumption levels, following the simplified synthetic profile-generation approach adopted in related LV topology reconstruction work [20].

For a network with N nodes and T time steps, the active-power dataset can be represented as

$$P = [P_i(t)]_{i=1,\dots,N;\, t=1,\dots,T}\,, \tag{3.20}$$

where $P_i(t)$ is the active power at node i and time step t. These time series are generated before running the power-flow simulation and remain associated with the corresponding ground-truth topology.

The use of time-varying residential profiles is important because topology inference relies on the relationship between power variations and voltage responses. If all loads were constant or insufficiently diverse, the resulting voltage profiles would contain less structural information. By using temporally varying residential-style consumption profiles, the synthetic dataset creates operating conditions that excite the network and make topology-related voltage patterns observable.

3.6 Forward-Backward Sweep Power-Flow Simulation

The voltage profiles associated with each synthetic network are obtained using a Forward-Backward Sweep power-flow algorithm. This method is well suited for radial distribution networks because it exploits the tree structure of the feeder [20], [38]. The power-flow routine used to compute the voltage profiles is detailed in Table 1.

For each time step t, the known quantities are the active-power injections $P_i(t)$ and the network topology. The complex power at node i is generally given by

$$S_i(t) = P_i(t) + jQ_i(t). \tag{3.21}$$

In the initial simplified dataset, the reactive-power component is neglected, and therefore

$$S_i(t) \approx P_i(t). \tag{3.22}$$

The algorithm starts from a flat voltage profile, with all nodes initialised at the nominal LV voltage magnitude,

$$V_i^{(0)}(t) = V_0, \qquad \forall i, \tag{3.23}$$

where V_0 denotes the nominal voltage, such as 230 V.

At iteration k, the nodal current is computed from the complex power and the current voltage estimate:

$$I_i^{(k)}(t) = \left(\frac{S_i(t)}{V_i^{(k)}(t)} \right)^*, \tag{3.24}$$

where $(\cdot)^*$ denotes the complex conjugate. In the simplified active-power-only case, this current is mainly determined by $P_i(t)$ and the voltage magnitude estimate.

The algorithm then performs a backward sweep. Starting from the leaf nodes and moving towards the root, branch currents are computed by aggregating the currents of downstream nodes. If $\mathscr{D}(i)$ denotes the set formed by node i and all nodes downstream of i, the current in a branch can be interpreted as the sum of the currents required by the downstream subtree:

$$I_{ij}^{(k)}(t) = \sum_{m \in \mathscr{D}(i)} I_m^{(k)}(t), \tag{3.25}$$

where branch (i, j) connects node i to its parent node j.

After the backward sweep, a forward sweep is performed. Starting from the root voltage, the voltage of each downstream node is updated according to the voltage drop along the corresponding branch:

$$V_i^{(k+1)}(t) = V_j^{(k+1)}(t) - Z_{ij} I_{ij}^{(k)}(t), \tag{3.26}$$

where node j is the parent of node i.

The process is repeated until the voltage change between consecutive iterations is smaller than a predefined tolerance:

$$\max_i \left| V_i^{(k+1)}(t) - V_i^{(k)}(t) \right| < \varepsilon. \tag{3.27}$$

Algorithm 1 Forward-Backward Sweep power-flow simulation for one topology.

Input: radial topology $\mathscr{G}$, branch impedances Z_{ij}, active-power profiles $\mathbf{P}$, tolerance ε

Output: voltage profiles $\mathbf{V}$

1. For each time step $t = 1, \ldots, T$, repeat:
   1.1. Initialise all nodal voltages with a flat start $V_i^{(0)}(t) = V_0$.
   1.2. Set $k \leftarrow 0$.
   1.3. Repeat until convergence:
      1.3.1. For each node i, compute the nodal current: $I_i^{(k)}(t) = \left(S_i(t) / V_i^{(k)}(t) \right)^*$.
      1.3.2. Perform the backward sweep by aggregating downstream currents from the leaves to the root.
      1.3.3. For each branch (i, j), where j is the parent of i, compute the branch current $I_{ij}^{(k)}(t)$.
      1.3.4. Perform the forward sweep by propagating voltages from the root to the leaves.
      1.3.5. For each branch (i, j), update the downstream voltage: $V_i^{(k+1)}(t) = V_j^{(k+1)}(t) - Z_{ij} I_{ij}^{(k)}(t)$.
      1.3.6. Set $k \leftarrow k + 1$.
   1.4. Stop when $\max_i |V_i^{(k)}(t) - V_i^{(k-1)}(t)| < \varepsilon$ or when the maximum number of iterations is reached.
   1.5. Store the converged voltage magnitudes $|V_i(t)|$.
2. Return $\mathbf{V}$.

The output of the power-flow simulation is the voltage profile associated with each active-power profile and each topology. For a given topology, the final dataset contains

$$\mathscr{D}_{\mathscr{G}} = \{P_i(t), V_i(t)\}_{i=1,\ldots,N;\, t=1,\ldots,T}\,. \tag{3.28}$$

This P/V dataset forms the basis for the experiments described in Chapter 4.

3.7 Validation of the Synthetic Electrical Data

Before constructing the final P/V dataset, the generated electrical scenarios were evaluated to verify their consistency with plausible LV operation. This validation did not assess the Forward–Backward Sweep algorithm itself, which was used as a standard radial power-flow solver, but rather whether the adopted branch parameters and active-power profiles produced physically coherent voltage-magnitude behaviour.

Each feeder is synthetically generated as a connected radial graph with known ground-truth connectivity. Its branches are assigned resistance-dominated impedances based on plausible LV cable-resistance ranges and randomised branch lengths, rather than arbitrary electrical parameters. The residential active-power profiles described in Section 3.5 are then assigned to the load buses, and the corresponding voltage magnitudes are obtained as the power-flow response of the feeder. The resulting P/V samples therefore combine known radial-topology labels, resistance-consistent branch parameters and time-varying residential active-power behaviour.

Figure 3.8 shows the representative 23-bus radial feeder used in this validation example, including the slack node, the numbered load nodes and the branch resistance values R_{ij}. This allows the observed voltage behaviour to be interpreted in relation to the feeder structure and to the accumulated resistance along the paths connecting the slack node to the downstream nodes.

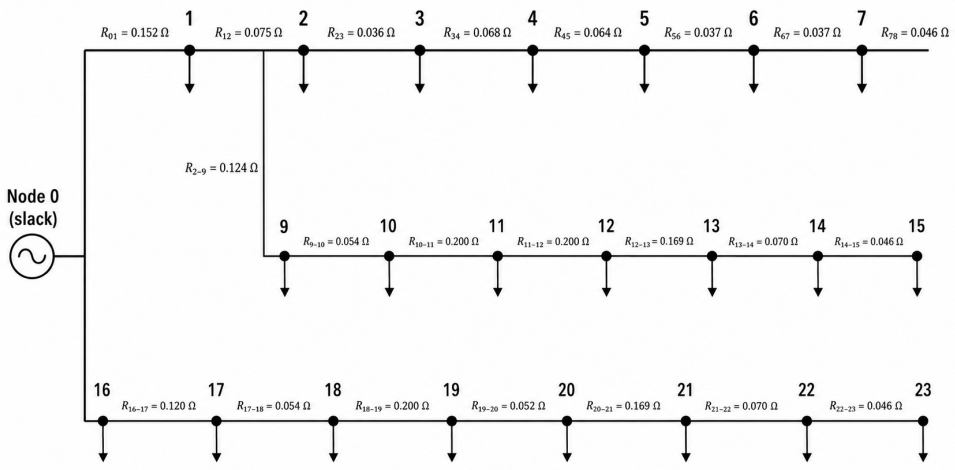


Figure 3.8: Single-line diagram of the representative 23-bus radial feeder used in the synthetic-data validation.

Figure 3.9 shows the voltage-magnitude trajectories obtained for the same feeder after applying the Forward-Backward Sweep power-flow simulation. The plot corresponds to a representative 96-timestamp segment extracted from the full 720-timestamp simulation window. Node 0 corresponds to the slack node and remains close to the nominal voltage, while the remaining trajectories reflect the voltage response of the load nodes under the assigned active-power profiles.

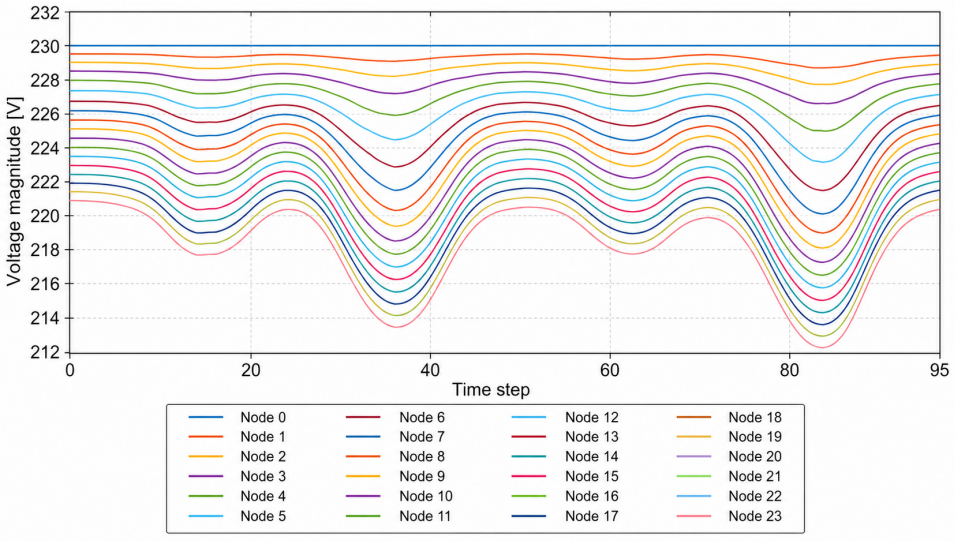


Figure 3.9: Voltage-magnitude trajectories of the representative 23-bus radial feeder over a 96-timestamp segment.

The voltage profiles remain within the admissible range considered for the synthetic validation setting. Their relative behaviour is also consistent with radial LV operation: nodes electrically farther from the slack node, or supplied through paths with larger accumulated resistance, tend to exhibit larger voltage drops during periods of higher active-power demand. Conversely, nodes located closer to the slack node remain closer to the nominal voltage.

This validation does not imply that the synthetic dataset captures all operational complexities of real LV grids, such as hidden intermediate nodes, asynchronous smart-meter readings, topology changes within the observation window or detailed customer-level modelling. However, it supports the use of the dataset as a controlled benchmark in which each sample combines known radial topology labels, resistance-consistent branch parameters, plausible residential-style active-power variation and physically coherent simulated voltage responses under the balanced single-phase equivalent adopted in this dissertation.

3.8 Construction of the Synthetic P/V Dataset

The final synthetic dataset is obtained by repeating the topology generation, active-power profile generation and power-flow simulation procedure for many radial networks. Each dataset sample contains three elements:

1. the ground-truth topology, represented by an adjacency matrix $\mathbf{A}$;

2. the active-power time series $\mathbf{P}$;

3. the voltage-magnitude time series $\mathbf{V}$ obtained through power-flow simulation.

A dataset sample can therefore be represented as

$$\mathscr{S}^{(r)} = \left(\mathbf{A}^{(r)}, \mathbf{P}^{(r)}, \mathbf{V}^{(r)}\right), \tag{3.29}$$

where r indexes the generated radial network.

The complete dataset is then

$$\mathscr{S} = \left\{\mathscr{S}^{(1)}, \mathscr{S}^{(2)}, \ldots, \mathscr{S}^{(M)}\right\}, \tag{3.30}$$

where M is the number of generated samples.

Only physically coherent samples are retained. A sample is accepted if the generated graph is radial and connected, the power-flow simulation converges, and the voltage magnitudes remain within the admissible range defined in the validation stage. Samples that fail these criteria are discarded.

The resulting dataset provides the required foundation for the remainder of the dissertation. In Chapter 4, the P/V time series are transformed into different input representations, including raw P/V signals, correlation matrices and sensitivity matrices. These representations are then used to train and evaluate fixed-size and variable-size topology inference models.

3.9 Topology Inference Modelling Pipeline

After constructing the synthetic P/V dataset, the next methodological stage is to transform the simulated measurements into representations suitable for topology inference. The objective is to learn the relationship between electrical behaviour and the underlying radial graph structure. The quantitative comparison of the models is presented in Chapter 4, but the main modelling pipeline is introduced here for completeness.

The first step is representation extraction. The raw dataset contains active-power and voltage-magnitude time series for each node. From these signals, different topology-related representations can be derived, including raw P/V inputs, voltage-correlation matrices and voltage-power sensitivity matrices. Correlation-based representations capture similarity between voltage profiles, while sensitivity-based representations approximate how active-power variations at one node are reflected in voltage variations across the network.

The second step is model-based inference. Different learning architectures can be applied to these representations, including fixed-size VAE models, convolutional VAE variants, variable-size architectures and Transformer-inspired models. These models do not directly solve a conventional power-flow problem. Instead, they learn latent representations or edge-related patterns that can be used to infer the underlying network connectivity.

The training duration was selected through a preliminary train–validation procedure rather than being fixed arbitrarily. For each model configuration, the model was first trained over a predefined maximum number of epochs while the validation loss was monitored after each epoch. The epoch corresponding to the minimum validation loss was then selected as the final training horizon. A new instance of the same model was subsequently trained for exactly this number of epochs using the same optimisation settings, and the resulting model parameters were saved for evaluation. Therefore, the reported number of training epochs corresponds to the best validation-loss behaviour observed during the preliminary run.

The third step is edge-score generation. The model output is interpreted as a matrix of edge scores or connection probabilities between pairs of nodes. Since the target topology is radial, this matrix is not used directly as the final topology. Instead, it provides ordered edge scores that guide the radial projection.

The final step is radial projection. The edge-score matrix is converted into a valid radial topology by selecting a connected set of $N-1$ edges, for example through a maximum-spanning-tree-type projection. This ensures that the final inferred topology satisfies the same structural constraints as the synthetic ground-truth networks: connectivity, no cycles and a unique path between nodes.

This modelling pipeline connects the synthetic P/V dataset produced in this chapter with the experimental evaluation presented in Chapter 4. Since the final model output is evaluated as an inferred radial topology, the next section defines the metrics used to compare the predicted topology with the known ground-truth adjacency matrix.

3.10 Topology Evaluation Metrics

The topology inference models considered in this dissertation produce either reconstructed adjacency matrices or continuous edge-score matrices. Since the target LV feeders are assumed to be radial, the final discrete prediction is evaluated after converting the model output into a valid radial topology. This post-processing step selects $N-1$ branches while enforcing connectivity and radiality, so that the inferred graph and the ground-truth graph have the same number of undirected edges.

The main quantitative metric used throughout the experiments is edge accuracy. Let $A \in 0,1^{N\times N}$ denote the ground-truth adjacency matrix and let $\widehat{A} \in 0,1^{N\times N}$ denote the inferred adjacency matrix after radial projection. Edge accuracy is defined as

$$EA = 1 - \frac{|\widehat{A} - A|_1}{4(N-1)}. \tag{3.31}$$

In Equation 3.31, N is the number of nodes and $N-1$ is the number of branches in a radial feeder. The metric is computed after radial projection, so both the predicted and ground-truth graphs contain $N-1$ undirected edges. The factor 4 accounts for the symmetric adjacency representation and for the fact that replacing one correct branch with one incorrect branch introduces

both one false negative and one false positive edge in the undirected adjacency matrix. Therefore, after radial projection, edge accuracy can be interpreted as the fraction of correctly recovered branches. An equivalent error-based metric used in some of the following experiments is the edge error. It is defined as

$$EE = \frac{\widehat{A} - A_1}{4}, \tag{3.32}$$

where EE denotes the number of incorrectly recovered undirected branches. The factor 4 accounts for the symmetric adjacency representation and for the simultaneous occurrence of one false-negative and one false-positive branch whenever a correct branch is replaced by an incorrect one. Since both the inferred and ground-truth topologies contain $N-1$ branches after radial projection, edge error and edge accuracy are related by

$$EA = 1 - \frac{EE}{N-1}. \tag{3.33}$$

For an individual topology, EE is an integer. When the metric is averaged over several inference cases, the reported mean edge error may take a non-integer value. A stricter metric, exact topology recovery, is also used in some diagnostic experiments. It is defined as an indicator variable equal to one when the inferred adjacency matrix exactly matches the ground-truth adjacency matrix, and zero otherwise:

$$ER = I\left(\widehat{A} = A\right) \tag{3.34}$$

Exact recovery is useful for identifying perfect topology reconstruction, but it is less informative when the model recovers only part of the feeder. For this reason, edge accuracy is used as the main metric in the comparative experiments, while exact recovery is mainly used as a complementary diagnostic.

3.11 Summary

This chapter presented the methodology used to generate and validate the synthetic LV dataset. The need for synthetic data was motivated by the difficulty of obtaining real LV measurements together with reliable ground-truth topologies. Synthetic radial networks were therefore generated and represented as graphs with known adjacency matrices and assigned branch impedances.

Before constructing the final dataset, several modelling assumptions were analysed. The initial formulation assumes radial operation, balanced single-phase behaviour, small voltage drops, dominant resistive effects and the use of active power and voltage magnitudes. Validation experiments were used to support these assumptions, including empirical-theoretical variance comparison, phase-aware variance comparison, voltage-drop acceptability checks and analysis of the impact of the X/R ratio on phase-related errors.

Residential active-power profiles were then generated as time-varying series with daily behaviour and randomly assigned peak values. These profiles were used as inputs to a Forward-Backward Sweep power-flow algorithm, which computed voltage profiles for each synthetic topology. The final output of the methodology is a synthetic P/V dataset with known ground-truth radial topologies. The chapter also defined the topology-inference output pipeline and the evaluation metrics used in the following results chapter, including edge accuracy, edge error and exact topology recovery after radial projection.

The next chapter uses this dataset to analyse feature representations, fixed-size and variable-size learning models, graph-completion approaches and robustness to measurement errors.

Chapter 4

Exploratory Modelling and Results

This chapter presents the experimental results obtained throughout the development of the topology inference framework. The objective is to evaluate the behaviour of the proposed learning-based approaches under different modelling choices, input representations, graph sizes and measurement errors.

The experimental path starts with an initial fixed-size VAE baseline, progresses to a fixed-size convolutional VAE, and then moves towards variable-size VAE architectures capable of handling feeders with different numbers of nodes. The final experiments analyse robustness to measurement errors and a graph-completion formulation with partial structural information.

The chapter is organised as a sequence of test cases and results. Following the synthetic-data validation presented in Section 3.7, the chapter begins with the initial fixed-size VAE experiments, with emphasis on the observed gap between *seen* and *unseen* topologies. The following sections present the fixed-size convolutional VAE, the comparison of input representations, the development of variable-size architectures, the Transformer-inspired model, the robustness analysis and the graph-completion formulation.

The main text reports the most representative result from each experimental family. When several variants of the same experiment were tested, only the configuration that best supports the main conclusion is discussed here. The remaining diagnostic results, alternative hyperparameters, secondary plots and repeated variants are reported in the appendices. This keeps the chapter focused while preserving the traceability of the full experimental process.

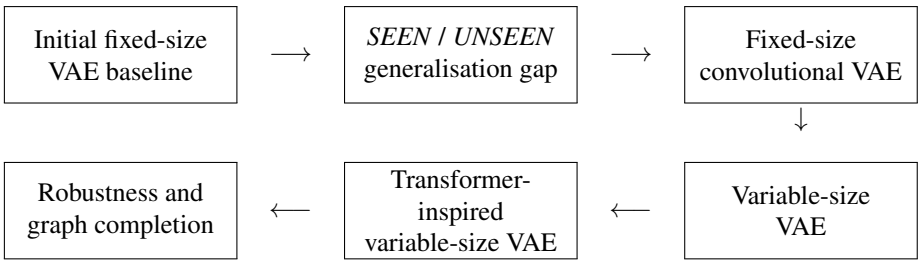


Figure 4.1: Simplified evolution of the experimental path reported in this chapter.

All experiments use the controlled dataset and topology-inference pipeline described in Chapter 3. Unless otherwise stated, the reported edge-accuracy values are computed after radial projection according to Equation 3.31.

4.1 Initial Fixed-Size VAE Baseline

The first topology-inference experiments were performed with an initial fixed-size VAE baseline using raw P/V inputs. In this regime, a separate model is associated with each network size, because the input and output dimensions depend directly on the number of nodes. Therefore, a model trained for 10-node feeders cannot be directly applied to 20-node or 30-node feeders.

The objective of this first baseline was to evaluate whether a standard fixed-size VAE could learn to reconstruct radial LV topologies from measurement-derived inputs. The evaluation uses the edge accuracy metric defined in Equation 3.31, in Section 3.10.

Two evaluation settings were considered. In the *seen* setting, the topology belongs to the set of topologies available during training, but the model is evaluated on measurement records not used for training. In the *unseen* setting, the topology itself is not present in the training set. The *unseen* setting is therefore more demanding, because the model must infer a new radial structure rather than recognise a previously observed one.

The initial fixed-size VAE was implemented as a fully connected encoder–decoder model using flattened raw P/V trajectories as input. Each sample contained $T = 720$ timestamps, with active-power and voltage-magnitude values concatenated into an input vector of dimension $2NT$, where N is the number of nodes. The encoder consisted of three dense layers with hidden dimensions 512, 256 and 128, followed by two parallel linear projections producing the latent mean and log-variance vectors. The latent dimension was set to $d_z = 64$. A latent sample was obtained through the standard reparameterisation step and passed to a mirrored decoder with hidden dimensions 128, 256 and 512. The decoder output was reshaped into an edge-score matrix, which was subsequently projected into a radial topology for evaluation.

The complete training and inference procedure of the initial fixed-size VAE is summarised in Algorithm 2. During training, the model receives a raw P/V representation and the corresponding ground-truth adjacency matrix. During inference, the decoded edge-score matrix is converted into a valid radial topology through spanning-tree projection.

Algorithm 2 Training and inference procedure of the initial fixed-size VAE.

Input: training samples $\{(\mathbf{X}^{(r)}, \mathbf{A}^{(r)})\}_{r=1}^{M}$, number of epochs E, learning rate η, and KL-divergence weight β

Output: trained encoder and decoder parameters and inferred radial topology $\widehat{\mathbf{A}}$

1. Initialise the encoder parameters ϕ and decoder parameters θ.
2. For each training epoch $e = 1, \ldots, E$, repeat:
   2.1. For each mini-batch $\{(\mathbf{X}^{(r)}, \mathbf{A}^{(r)})\}$, repeat:
      2.1.1. Flatten the raw active-power and voltage-magnitude trajectories into the fixed-size input representation $\mathbf{X}^{(r)}$.
      2.1.2. Use the encoder to compute the latent mean and log-variance: $\mu^{(r)}, \log \sigma^{2(r)} \leftarrow Encoder_{\phi}(\mathbf{X}^{(r)})$.
      2.1.3. Sample $\varepsilon^{(r)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
      2.1.4. Apply the reparameterisation step: $\mathbf{z}^{(r)} \leftarrow \mu^{(r)} + \sigma^{(r)} \odot \varepsilon^{(r)}$.
      2.1.5. Decode the latent representation into an edge-score matrix: $\mathbf{W}^{(r)} \leftarrow Decoder_{\theta}(\mathbf{z}^{(r)})$.
      2.1.6. Compute the adjacency reconstruction loss $\mathscr{L}_{\text{rec}}$ by comparing $\mathbf{W}^{(r)}$ with the ground-truth adjacency matrix $\mathbf{A}^{(r)}$.
      2.1.7. Compute the KL-divergence regularisation term $\mathscr{L}_{\text{KL}}$.
      2.1.8. Compute the total VAE loss: $\mathscr{L} \leftarrow \mathscr{L}_{\text{rec}} + \beta \mathscr{L}_{\text{KL}}$.
      2.1.9. Update ϕ and θ through gradient-based optimisation using learning rate η.
3. For a new inference sample $\mathbf{X}^{*}$, compute the latent distribution using the trained encoder.
4. Obtain the latent representation $\mathbf{z}^{*}$ and decode it into the edge-score matrix $\mathbf{W}^{*} \leftarrow Decoder_{\theta}(\mathbf{z}^{*})$.
5. Symmetrise the edge-score matrix, ignore diagonal entries and retain the scores associated with valid node pairs.
6. Apply the radial spanning-tree projection: $\widehat{\mathbf{A}} \leftarrow \Pi_{\text{radial}}(\mathbf{W}^{*})$.
7. Return the inferred radial topology $\widehat{\mathbf{A}}$.

Algorithm 2 defines the general training and inference procedure used as the starting point for the subsequent models. The following architectures preserve the same overall objective of predicting an edge-score matrix and projecting it into a radial topology, while modifying the input representation, the encoding mechanism or the treatment of variable graph sizes.

Following the epoch-selection procedure described in Section 3.9, the minimum validation loss was obtained at epoch 50. A new instance of the model was therefore trained for 50 epochs using mini-batches of 32 samples and a learning rate of 10^{-3}, and the resulting model parameters were saved for evaluation. The VAE objective combined an adjacency reconstruction term with a KL-divergence regularisation term weighted by $\beta = 10^{-3}$. The reconstruction term encouraged the decoded edge-score matrix to match the ground-truth adjacency matrix, while the KL term regularised the latent distribution towards a standard Gaussian prior.

Table 4.1 reports the *seen* and *unseen* edge accuracy of the initial fixed-size VAE baseline. The

experiments were performed for 10-node, 20-node and 30-node networks. For the 10-node case, 2000 training topologies were used. For the 20-node and 30-node cases, 5000 training topologies were used. Each reported evaluation used 50 test cases.

Table 4.1: *Seen* and *unseen* edge accuracy of the initial fixed-size VAE baseline.

Network size	**Training topologies**	**Test cases**	***Seen* EA**	***Unseen* EA**
10 nodes	2000	50	1.000	0.528
20 nodes	5000	50	0.927	0.316
30 nodes	5000	50	0.899	0.293

The results show a clear gap between *seen* and *unseen* topology reconstruction. For 10-node feeders, the model reaches perfect *seen* edge accuracy, with a value of 1.000, but only 0.528 edge accuracy on *unseen* topologies. For 20-node feeders, the *seen* edge accuracy remains high, at 0.927, while the *unseen* edge accuracy is only 0.316. For 30-node feeders, the *seen* edge accuracy is 0.899, but the *unseen* edge accuracy remains low, at 0.293.

This behaviour indicates that the initial fixed-size VAE is able to associate measurement patterns with topologies already represented in the training set. However, it does not learn a sufficiently general structural rule to reconstruct new radial configurations. The problem is therefore not only the ability to fit the training distribution, but the ability to generalise from observed topologies to *unseen* graph structures.

To investigate whether this limitation was caused by the organisation of the latent space, three latent-space diagnostics were performed for the 10-node fixed-size VAE. These diagnostics are reported in detail in Appendix B. The objective was to test whether *unseen* topologies could be improved by forcing their latent representation towards nearby training samples, nearby PCA-projected samples, or nearby latent family centroids.

In these diagnostics, a training family denotes the set of training records generated from the same ground-truth topology. A family centroid is therefore the mean latent representation of all records associated with one training topology.

The three latent-space diagnostics differ in how the reference latent point is selected. In the latent-space kNN diagnostic, the nearest neighbours are selected in the full latent space, while PCA is used only for visualisation. In the PCA-based diagnostic, the closest training point is selected directly in the two-dimensional PCA projection. In the nearest latent-family-mean diagnostic, the *unseen* latent vector is compared with the centroid of each training family in the full latent space.

Table 4.2 summarises the mean edge accuracy obtained in each diagnostic. The table reports the number of *unseen* cases tested, the *seen* edge accuracy, and the mean *unseen* edge accuracy before and after the latent-space intervention. The value in parentheses indicates the mean variation caused by the intervention.

Table 4.2: Mean edge accuracy of the latent-space diagnostics for the initial 10-node fixed-size VAE.

Diagnostic	**Cases**	***Seen* EA**	***Unseen* EA before/after**
Latent-space kNN forcing	3	1.000	0.552 / 0.556 (+0.004)
PCA-based latent forcing	3	1.000	0.481 / 0.444 (−0.037)
Nearest latent-family-mean forcing	20	1.000	0.528 / 0.506 (−0.022)

The results show that the *seen* topologies were reconstructed with edge accuracy equal to 1.000 in all three diagnostics. This confirms that the initial fixed-size VAE was able to recognise topologies already represented in the training set. However, the *unseen* results remained substantially weaker, even after applying the latent-space interventions.

In the latent-space kNN diagnostic, the mean *unseen* edge accuracy changed only from 0.552 to 0.556, corresponding to an average improvement of 0.004. This improvement is negligible and does not indicate a meaningful correction of the *unseen* reconstruction problem. Therefore, selecting nearby samples in the full latent space does not provide a reliable mechanism for recovering *unseen* radial topologies.

In the PCA-based latent forcing diagnostic, the mean *unseen* edge accuracy decreased from 0.481 before forcing to 0.444 after forcing. This shows that visual proximity in the two-dimensional PCA projection is not a reliable proxy for topological similarity. Two points may appear close in the projected latent space while corresponding to substantially different edge sets.

In the nearest latent-family-mean diagnostic, the *unseen* latent vector was replaced by the centroid of the closest training family in the full latent space. This experiment was performed over 20 *unseen* cases. The mean *unseen* edge accuracy decreased from 0.528 to 0.506, and exact topology recovery remained zero in all 20 cases. Although some individual cases improved, the average result shows that forcing the latent vector towards the nearest family centroid does not reliably improve *unseen* topology reconstruction.

These three diagnostics lead to the same conclusion. The latent space learned by the initial fixed-size VAE is not organised according to exact graph similarity. Nearby points may share some branches, but latent proximity does not reliably imply topological equivalence. Therefore, the weak *unseen* performance of the initial VAE cannot be solved simply by manipulating the learned latent representation.

Additional fixed-size diagnostics were also performed to test whether the *seen–unseen* gap could be reduced by changing the latent capacity, the regularisation strategy or the architecture while remaining within the same initial fixed-size VAE regime. The latent-dimension sweep is reported in Appendix B, together with the latent-space diagnostics, while the regularisation and architecture experiments are reported in Appendix C. Table 4.3 summarises the main conclusion of each diagnostic.

Table 4.3: Summary of additional diagnostics within the initial fixed-size VAE regime.

Diagnostic	**Main conclusion**
Latent-dimension sweep	Increasing the latent dimension did not produce a systematic improvement in *unseen* topology reconstruction.
KL regularisation sweep	Stronger variational regularisation degraded reconstruction quality and did not solve *unseen* generalisation.
Dropout and weight-decay sweep	Classical regularisation changed performance locally, but did not produce consistent strong *unseen* reconstruction.
Architecture sweep	Changing width, depth, activation functions, normalisation and decoder structure within the fixed-size VAE formulation did not remove the *seen–unseen* gap.

These diagnostics show that the weak *unseen* performance was not caused by a single hyperparameter choice. Increasing the latent dimension did not solve the problem, which indicates that the bottleneck was not simply insufficient latent capacity. Stronger KL regularisation also failed to improve *unseen* reconstruction, suggesting that the latent space was not made topology-aware simply by imposing a stronger prior. Similarly, dropout and weight decay did not remove the gap, meaning that the limitation cannot be interpreted only as classical overfitting.

The architecture sweep leads to the same conclusion. Although different fixed-size VAE variants were tested, changing the width, depth or decoder details did not provide the relational inductive bias required for *unseen* radial topology reconstruction. Overall, the initial fixed-size VAE baseline reveals a structural limitation: it can reconstruct or recognise topologies represented during training, but it does not generalise reliably to *unseen* radial configurations.

These results motivate the subsequent transition to convolutional processing, structured input representations and variable-size architectures, which are evaluated in the following sections.

4.2 Fixed-Size VAE Refinement and Convolutional VAE Results

After the initial fixed-size VAE diagnostics, the next step was not immediately to change the input representation. The first objective was to understand whether the weak *unseen*-topology performance could be improved by changing the amount and organisation of the available training data while keeping the same general fixed-size raw P/V setting. In particular, three aspects were analysed: the number of training topologies, the number of timestamps used to describe each electrical trajectory, and the trade-off between temporal depth and the number of available operating records.

This sequence is important because it separates two possible explanations for the initial limitation. If the weak *unseen* performance were mainly caused by insufficient data coverage, then increasing structural diversity or temporal information should produce a clear improvement. If, instead, the limitation were mainly related to how the model processed the raw P/V information, then these data-related changes would only produce local or limited gains.

4.2.1 Effect of the Number of Training Topologies

The first refinement consisted of increasing the number of training topologies. The motivation was that *unseen* topology inference may require the model to observe a sufficiently diverse set of radial structures during training. A larger training database exposes the model to more branching patterns and could therefore improve its ability to reconstruct new configurations.

Figure 4.2 reports the mean edge accuracy as the number of training topologies increases.

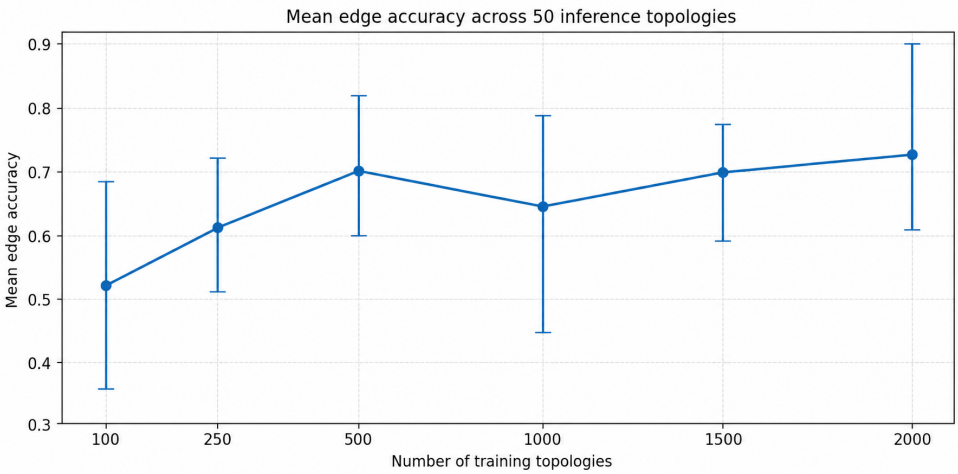


Figure 4.2: Mean edge accuracy over 50 inference topologies as a function of the number of training topologies. Error bars represent one standard deviation of the edge accuracy values across the 50 inference cases.

The topology-count sweep shows a general improvement in mean edge accuracy as the number of training topologies increases. For each training-set size, the marker represents the mean edge accuracy computed over 50 inference topologies. The error bars represent one standard deviation of the edge accuracy values across those 50 inference cases, rather than the standard error of the mean. Therefore, the relatively large error bars indicate that reconstruction difficulty varies substantially from one inferred topology to another.

This behaviour suggests that structural diversity in the training set is beneficial for partial edge recovery. However, increasing the number of training topologies does not make the fixed-size model uniformly reliable across *unseen* cases, nor does it remove the *seen–unseen* generalisation gap. This result indicates that the limitation is not only caused by the number of available training topologies; the model also requires an architecture capable of learning transferable relational patterns from the raw P/V trajectories.

4.2.2 Fixed-Budget Trade-Off Between Records and Temporal Horizon

A complementary experiment was performed under a fixed training-data budget. In this setting, the objective was to analyse whether it is more useful to train the model with many short operating records or with fewer longer records. The term fixed data budget means that the total number of timestamped P/V observations was kept approximately constant, while the allocation between the number of records and the temporal horizon was varied. In this experiment, the budget can be interpreted as

$$B = n_{\text{records}} \times T \approx 16000,$$

where n_{records} is the number of training records and T is the number of timestamps in each record. Therefore, a configuration such as $16000/T1$ corresponds to many short records with one timestamp each, whereas $25/T640$ corresponds to a much smaller number of records with a longer temporal horizon. The total amount of timestamped data is kept approximately fixed, but the balance between sample diversity and temporal context changes.

Figures 4.3 and 4.4 show this trade-off for *seen* and *unseen* topologies, respectively. Each point corresponds to a different allocation of the same approximate data budget between the number of available training records and the temporal horizon T.

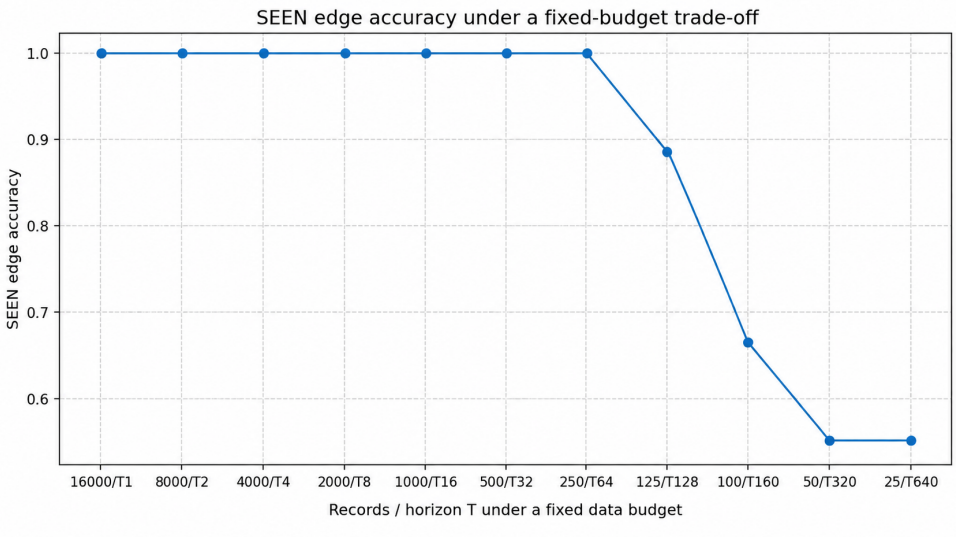


Figure 4.3: *Seen* edge accuracy under a fixed data-budget trade-off between the number of training records and the temporal horizon T.

For *seen* topologies, the model maintains perfect edge accuracy over a broad range of short and intermediate horizons. A clear degradation only appears when the temporal horizon becomes very long and the number of available records becomes too small. This indicates that, for structures already represented during training, preserving a sufficient number of training examples is more important than using very long individual records.

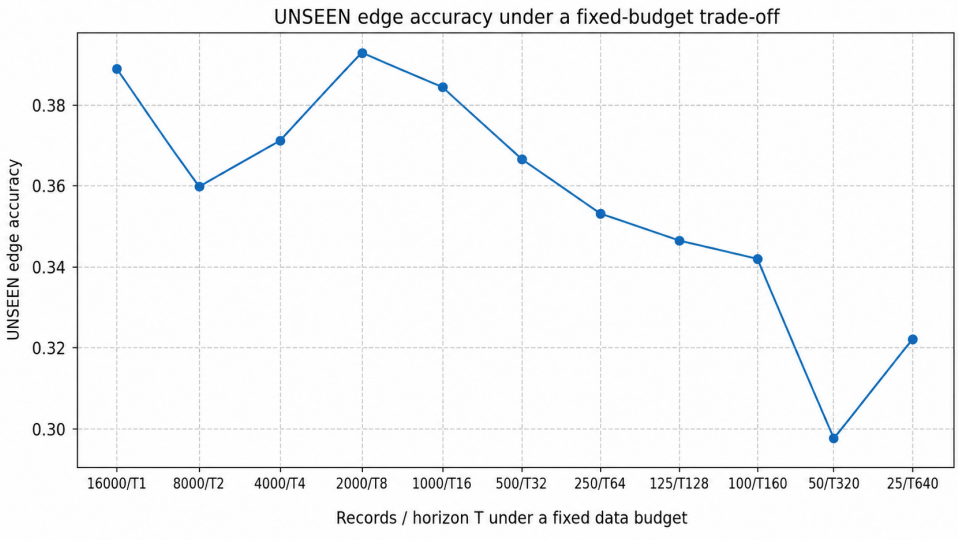


Figure 4.4: *Unseen* edge accuracy under a fixed data-budget trade-off between the number of training records and the temporal horizon T.

For *unseen* topologies, the best behaviour is obtained in an intermediate regime. Very short records provide limited temporal context, while very long records reduce the number of available training examples and therefore reduce the structural diversity seen during training. This shows that *unseen* topology inference requires a balance between temporal information and dataset diversity. Under a fixed data budget, increasing the temporal horizon is not always beneficial if it substantially reduces the number of training records.

Overall, these preliminary refinements showed that increasing data quantity or changing the allocation of a fixed data budget could improve performance locally, but did not remove the main limitation of the initial fixed-size VAE. This motivated a change in the model architecture while still using raw P/V trajectories as input.

4.2.3 Transition to a Fixed-Size Convolutional VAE

The next stage replaced the initial fully connected fixed-size VAE with a fixed-size convolutional VAE. At this point, the objective was not yet to introduce correlation or sensitivity matrices, but to test whether a convolutional architecture could extract topology-related patterns more effectively from raw P/V inputs.

The raw P/V input can be arranged as a two-channel structure, where one channel contains active-power trajectories and the other contains voltage-magnitude trajectories. A convolutional architecture can process this structured input while preserving local patterns across the input dimensions more effectively than a fully connected model applied to a flattened vector.

Relative to Algorithm 2, the convolutional VAE replaces the fully connected processing of flattened inputs with convolutional layers applied to structured input representations. The latent-variable formulation, adjacency reconstruction objective and radial projection stage remain unchanged.

Since the spatial dimensions of the input and output representations depend on the number of nodes, a separate model is still required for each network size.

The fixed-size convolutional VAE used the same temporal horizon and latent-space configuration as the previous fixed-size VAE. Each input sample contained $T = 720$ timestamps and was represented as a two-channel tensor of size $2 \times N \times T$, where N is the fixed number of nodes for the corresponding experiment. The encoder consisted of convolutional feature-extraction layers applied to the two-channel P/V input, followed by a fully connected projection with hidden dimensions 512, 256 and 128. As in the initial VAE, two parallel linear layers produced the latent mean and log-variance vectors, with latent dimension $d_z = 64$. The decoder used the corresponding mirrored structure: a latent sample obtained through the reparameterisation step was passed through dense layers with hidden dimensions 128, 256 and 512, followed by the output mapping to an edge-score matrix.

The convolutional VAE followed the same optimisation and epoch-selection procedure used for the initial fixed-size VAE. The selected training horizon was 50 epochs, therefore, a new model instance was trained for this duration using mini-batches of 32 samples and a learning rate of 10^{-3}, and the resulting model parameters were saved for evaluation. The loss function combined the adjacency reconstruction term with a KL-divergence regularisation term weighted by $\beta = 10^{-3}$. The decoded edge-score matrix was then projected into a radial topology before computing edge accuracy.

The results in this subsection are reported using the edge-accuracy metric defined in Equation 3.31. Table 4.4 compares the original fixed-size VAE and the fixed-size convolutional VAE using raw P/V inputs only. All values are rounded to three decimal places.

Table 4.4: Fixed-size edge accuracy for the original VAE and convolutional VAE using raw P/V inputs.

Network size	**Topologies**	**Original VAE P/V**	**Conv. VAE P/V**	**Improvement**
10 nodes	2000	0.528	0.682	+0.154
20 nodes	5000	0.316	0.551	+0.235
30 nodes	5000	0.293	0.446	+0.153

The convolutional VAE improves edge accuracy for all tested network sizes when compared with the original VAE using the same raw P/V input family. For 10-node networks, edge accuracy increases from 0.528 to 0.682. For 20-node and 30-node networks, the improvement is smaller but still clear, increasing from 0.316 to 0.551 and from 0.293 to 0.446, respectively.

These results show that the architecture itself matters. Even before introducing matrix-based electrical representations, convolutional processing extracts more useful information from the raw P/V input than the original fully connected VAE. However, the resulting performance is still not sufficient for robust topology inference, especially as the network size increases. The drop from 10-node to 30-node networks indicates that larger radial feeders remain more difficult to reconstruct, due to the larger number of branches and the larger space of possible configurations.

Therefore, the fixed-size convolutional VAE is interpreted as an important intermediate step. It shows that improving the architecture helps, but it also reveals that raw P/V trajectories do not expose node-to-node electrical relationships strongly enough. This motivates the next stage of the work, where correlation and sensitivity representations are analysed as more explicit matrix-based inputs for topology inference.

4.3 Input Representation Study

After confirming that the convolutional architecture improved the fixed-size raw P/V setting, the next question was which measurement-derived representation carried the most useful information for topology inference. Raw P/V trajectories contain the original electrical time series, but the relationships between nodes are only implicit. Since topology is fundamentally relational, this section studies representations that make node-to-node dependencies more explicit.

Three main representation families were analysed: raw P/V trajectories, correlation-based matrices and sensitivity-based matrices. Additional hybrid combinations were also tested to determine whether concatenating several sources of information improved reconstruction accuracy.

4.3.1 Correlation and Sensitivity Matrix Construction

The correlation and sensitivity representations are both computed from temporal variations rather than from the raw absolute values of active power and voltage. For each node (i), temporal deviations are defined as

$$\Delta P_i(t) = P_i(t) - P_i(t-1) \tag{4.1}$$

$$\Delta V_i(t) = V_i(t) - V_i(t-1) \tag{4.2}$$

For a feeder with N nodes and an operating window with T timestamps, these deviations are assembled into the matrices

$$\Delta P \in R^{N\times(T-1)}, \qquad \Delta V \in R^{N\times(T-1)}. \tag{4.3}$$

This means that each row corresponds to one node and each column corresponds to one temporal difference. Making this matrix structure explicit is important because the correlation and sensitivity representations are not computed from scalar quantities, but from the full temporal-deviation matrices.

The correlation representation is constructed by computing pairwise correlations between voltage and active-power deviations. In its full form, the resulting block matrix is

$$C = \begin{bmatrix} C_{VV} & C_{VP} \\ C_{PV} & C_{PP} \end{bmatrix} \tag{4.4}$$

where C_{VV} and C_{PP} describe voltage-voltage and power-power correlations, while C_{VP} and C_{PV} describe cross-correlations between voltage and active-power deviations. A reduced PV-only correlation variant was also tested, using only the power-voltage coupling block instead of the full four-block correlation matrix.

The sensitivity representation estimates how voltage deviations respond to active-power deviations. The sensitivity matrix is computed as

$$S = \Delta V \Delta P^T \left(\Delta P \Delta P^T + \lambda I_N\right)^{-1}. \tag{4.5}$$

Here, $S \in R^{N\times N}$, I_N is the $N \times N$ identity matrix, and λ is a ridge regularisation term used to improve numerical stability. The ridge term avoids numerical singularity or ill-conditioning in $\Delta P \Delta P^T$, which can occur when active-power variations across nodes are correlated or insufficiently excited. Following the configuration adopted throughout the experiments, the value used was $\lambda = 10^{-6}$, interpreted as a small numerical stabilisation term rather than as a physical parameter.

The term $\Delta V \Delta P^T$ has dimension $N \times N$, and $\Delta P \Delta P^T + \lambda I_N$ is also an $N \times N$ matrix. Therefore, the resulting sensitivity matrix assigns one value to each ordered pair of nodes, approximating how active-power variations at one node are associated with voltage variations at another node. This makes the representation physically meaningful for topology inference, because voltage-response patterns are constrained by the electrical structure of the feeder.

4.3.2 Exploratory Analysis of Correlation Blocks

Before comparing correlation-based inputs with sensitivity-based inputs, an intermediate diagnostic was performed to understand which parts of the correlation representation were most informative for topology inference. The full correlation matrix contains voltage-voltage, power-power, and power-voltage correlation blocks. However, not all of these blocks are expected to contribute equally to the reconstruction of the electrical topology.

From a physical perspective, the relation between active-power variations and voltage-magnitude variations is the most relevant component for topology inference, since the network structure constrains how changes in nodal injections propagate into voltage responses. Therefore, the objective of this diagnostic was to assess whether the power-voltage correlation block preserved most of the useful information contained in the full correlation matrix.

All results reported in the fixed-size input-representation study were computed using the same database configuration adopted in the previous fixed-size experiments. The 10-node experiments used 2000 training topologies, while the 20-node and 30-node experiments used 5000 training topologies. This ensures that the comparison between full correlation, PV-only correlation, sensitivity matrices and two-encoder combinations is performed under the same data-generation conditions for each network size.

Representative examples of the full correlation matrices for 10-node, 20-node and 30-node networks are provided in Appendix E.2. These examples are included as visual diagnostics to

illustrate the block structure of the full correlation representation before reducing the input to the power-voltage block.

To interpret the contribution of the different correlation blocks, a gradient-based input-saliency diagnostic was applied to the trained correlation-based model. For each test case, the gradient of a scalar model-output score was computed with respect to each entry of the input correlation matrix. The resulting saliency map therefore measures how sensitive the model output is to small local perturbations in each correlation coefficient. The saliency maps were accumulated over test topologies that were not present in the training set, producing an aggregate indication of which regions of the full correlation matrix most influenced the model output.

This diagnostic should be interpreted as a differential sensitivity analysis rather than as a causal or statistical feature-importance measure. Larger values indicate entries of the correlation matrix for which small variations would produce larger changes in the model output. In this sense, the map provides a model-based indication of which correlation blocks are most used by the trained network.

Figure 4.5 shows the resulting gradient-based input-saliency map for the full correlation representation. The most salient region is concentrated around the cross-variable power-voltage interaction blocks, while the voltage-voltage and power-power regions appear less influential. This supports the hypothesis that cross-variable correlations between active power and voltage are more relevant for topology reconstruction than correlations computed within the same variable type.

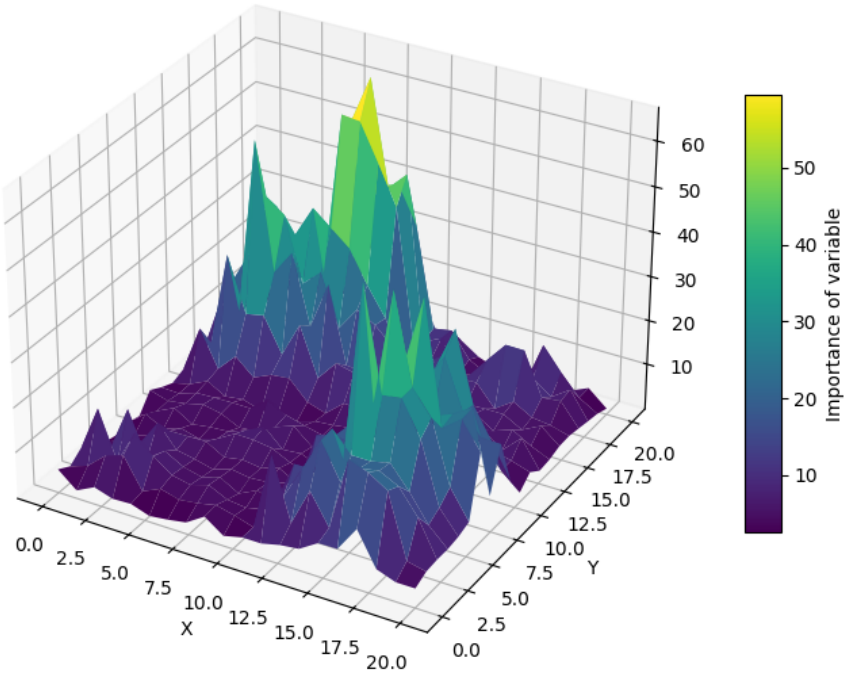


Figure 4.5: Gradient-based input-saliency map for the full correlation representation. Larger values indicate entries of the correlation matrix to which the model output is more sensitive.

To quantify this observation, the full correlation representation was compared with a reduced representation based only on the power-voltage correlation block. The results are shown in Table 4.5. All values correspond to edge accuracy and are rounded to three decimal places.

Table 4.5: Comparison between full correlation and PV-only correlation representations across fixed-size networks.

Input representation	**Metric**	**10 nodes**	**20 nodes**	**30 nodes**
Full correlation	*Seen*	0.951	0.508	0.308
	Unseen	0.740	0.393	0.279
PV-only correlation	*Seen*	0.884	0.536	0.330
	Unseen	0.704	0.402	0.286

The results show that the full correlation matrix performs best for the 10-node case, reaching an *unseen* edge accuracy of 0.740, compared with 0.704 for the PV-only correlation block. This indicates that, for smaller networks, the additional voltage-voltage and power-power correlation blocks may still provide useful information.

However, this advantage does not persist as the network size increases. For 20-node networks, the PV-only correlation block achieves an *unseen* edge accuracy of 0.402, slightly above the 0.393 obtained with the full correlation matrix. For 30-node networks, the same trend is observed, with PV-only correlation reaching 0.286, compared with 0.279 for the full correlation matrix.

These results suggest that the cross-variable power-voltage block retains most of the useful information contained in the full correlation matrix, especially as the topology inference problem becomes larger. Although the full correlation representation is stronger for the 10-node case, the PV-only representation is more compact and remains competitive or slightly superior for 20-node and 30-node feeders. This supports its use as a reduced correlation-based representation in the following comparisons.

4.3.3 Controlled Comparison Between PV Correlations and Sensitivity Matrices

After identifying that the power-voltage correlation block retained most of the useful information contained in the full correlation matrix, the next step was to compare this compact correlation-based representation with the sensitivity matrix. Both representations encode relationships between active-power variations and voltage responses, but they do so in different ways. The PV-only correlation block captures statistical association between the two variables, whereas the sensitivity matrix provides a more physically structured approximation of how active-power perturbations affect voltage magnitudes across the feeder.

The objective of this comparison was to verify whether the compact PV correlation representation remained competitive across different network sizes, and whether the sensitivity matrix provided an additional advantage due to its closer physical interpretation. The results are reported in Table 4.6. All values correspond to edge accuracy and are rounded to three decimal places.

Table 4.6: Controlled comparison between PV-only correlation and sensitivity representations across fixed-size networks.

Input representation	Metric	10 nodes	20 nodes	30 nodes
PV-only correlation	Seen	0.884	0.536	0.330
	Unseen	0.704	0.402	0.286
Sensitivity matrix	Seen	0.936	0.505	0.319
	Unseen	0.680	0.371	0.283

The comparison shows that both PV-only correlation and sensitivity matrices contain relevant topology-related information, but their behaviour differs across network sizes. For 10-node networks, the sensitivity matrix achieves a higher *seen* edge accuracy, with 0.936 compared with 0.884 for PV-only correlation. However, the *unseen* edge accuracy is higher for PV-only correlation, with 0.704 compared with 0.680 for the sensitivity matrix.

The same tendency is observed for 20-node networks. PV-only correlation reaches an *unseen* edge accuracy of 0.402, while the sensitivity matrix reaches 0.371. For 30-node networks, the two representations are very close, with *unseen* edge accuracies of 0.286 for PV-only correlation and 0.283 for the sensitivity matrix.

These results indicate that the sensitivity matrix is physically meaningful and remains competitive, but it does not dominate the fixed-size representation study. In this set of experiments, the PV-only correlation block is the most consistent compact representation in terms of *unseen* edge accuracy. This suggests that the statistical relation between active-power variations and voltage-magnitude variations captures a large part of the information needed for topology inference, without requiring the full four-block correlation matrix.

4.3.4 Two-Encoder Fusion of Complementary Representations

After comparing PV-only correlations and sensitivity matrices individually, an additional experiment was performed to evaluate whether complementary representations could be combined more effectively. Instead of concatenating all input channels into a single representation, this experiment used a two-encoder architecture. Each encoder processed one input representation independently, and the resulting latent features were then combined before the adjacency reconstruction stage.

The motivation for this experiment was to test whether different representations captured complementary information. In particular, sensitivity matrices provide a physically structured approximation of voltage response to active-power perturbations, while PV-only correlations capture statistical dependence between active-power and voltage variations. Raw P/V trajectories were also tested in combination with structured representations to assess whether the original temporal measurements added useful information when processed through a separate encoder.

Table 4.7 reports the two-encoder fusion results for 10-node, 20-node and 30-node networks. All values correspond to edge accuracy and are rounded to three decimal places.

Table 4.7: Two-encoder fusion results across fixed-size networks.

Input representation	**Metric**	**10 nodes**	**20 nodes**	**30 nodes**
Sensitivity + PV-only correlation	Seen	0.971	0.522	0.323
	Unseen	0.724	0.393	0.281
Raw P/V + sensitivity matrix	Seen	0.858	0.474	0.307
	Unseen	0.676	0.395	0.273
Raw P/V + PV-only correlation	Seen	0.920	0.485	0.317
	Unseen	0.724	0.392	0.271

The two-encoder fusion results show that combining representations does not produce a consistent improvement across network sizes. For 10-node networks, the best two-encoder configurations reach an *unseen* edge accuracy of 0.724, obtained by both sensitivity plus PV-only correlation and raw P/V plus PV-only correlation. This is slightly above the PV-only correlation result of 0.704, but still below the full correlation result of 0.740.

For 20-node networks, the two-encoder configurations are very close to each other and to the single-representation baselines. The best fusion result is obtained by raw P/V plus sensitivity, with an *unseen* edge accuracy of 0.395, but this remains slightly below the PV-only correlation baseline, which reaches 0.402. For 30-node networks, none of the fusion configurations improves over PV-only correlation or sensitivity alone.

These results indicate that adding more input channels or using separate encoders does not automatically improve topology inference. In particular, combining raw P/V trajectories with structured representations does not provide a clear benefit. This suggests that the raw temporal trajectories may introduce redundant or less organised information when a compact structured representation is already available.

Overall, the two-encoder experiments reinforce the conclusion that representation quality is more important than representation quantity. The PV-only correlation block remains the most consistent compact representation in the fixed-size input-representation study, while full correlation is strongest only in the 10-node case. Nevertheless, the decrease in edge accuracy from 10-node to 20-node and 30-node networks shows that improving the input representation alone is not sufficient to solve the scalability problem. This motivates the transition to variable-size architectures in the following section.

4.3.5 Summary of the Input-Representation Study

The input-representation study shows that the way the electrical measurements are represented has a strong impact on fixed-size topology inference. Raw P/V trajectories require the model to learn node-to-node electrical relationships implicitly, while correlation and sensitivity matrices expose these relationships more directly.

The comparison between the full correlation matrix and the PV-only correlation block shows that the full correlation representation performs best for the 10-node case. However, this advantage does not persist for larger networks. For 20-node and 30-node feeders, the PV-only correlation block is slightly better in *unseen* edge accuracy. This indicates that the cross-variable relation between active-power variations and voltage-magnitude variations retains most of the useful topological information contained in the full correlation matrix, while providing a more compact representation.

The sensitivity matrix remains a physically meaningful representation, because it approximates how voltage variations respond to active-power perturbations across the feeder. Nevertheless, in the fixed-size experiments reported in this section, the sensitivity matrix does not consistently outperform the PV-only correlation block. The two representations are close for some network sizes, especially for 30-node networks, but the PV-only correlation block is the most consistent compact representation in terms of *unseen* edge accuracy.

The two-encoder experiments further show that combining representations does not guarantee better reconstruction. Although some fusion configurations improve over individual representations in the 10-node case, this improvement is not consistent for 20-node and 30-node networks. In particular, adding raw P/V trajectories to structured representations does not provide a clear advantage and may introduce redundant or less organised information.

Overall, these results suggest that compact cross-variable representations are preferable to larger or more complex input combinations in the fixed-size setting. However, the strong degradation observed when moving from 10-node to 20-node and 30-node networks shows that input representation alone is not sufficient to solve the scalability problem. This motivates the transition to variable-size architectures in the following section, where the objective is no longer to train one fixed-dimensional model per network size, but to learn a topology-inference rule that can be applied across different graph dimensions.

4.4 Variable-Size VAE Topology Inference

The previous sections showed that topology inference improves substantially when the input representation exposes node-to-node electrical relationships. In particular, sensitivity matrices and PV-only correlations provided a more informative representation than raw P/V trajectories, and convolutional encoders were better suited to exploit the resulting matrix structure. However, all experiments up to this point were still based on a fixed-size formulation. Since the previous experiments showed that *seen* performance is consistently higher and less informative about generalisation, the remaining analysis focuses on *held-out* graph sizes and *unseen* topology reconstruction as distinct generalisation settings.

In the fixed-size setting, each model is tied to a specific number of nodes. A 10-node feeder, a 20-node feeder and a 30-node feeder require different input and output dimensions, and therefore separate models. This is a significant limitation for practical topology inference, since real LV feeders do not all contain the same number of monitored nodes. A scalable model should be

able to process feeders with different sizes without being retrained from scratch for each network dimension.

The variable-size VAE changes the architecture more substantially. Instead of flattening the complete input matrix, it applies a shared encoder to the representation associated with each node. Padding masks distinguish valid nodes from padded positions, and the valid node embeddings are aggregated to infer a graph-level latent distribution. After latent sampling, pairwise combinations of node representations are decoded into edge scores. The resulting score matrix is then projected into a valid radial tree.

The optimisation procedure remains conceptually consistent with Algorithm 2, but the encoder and decoder are reformulated so that the same trained model can process feeders with different numbers of nodes.

Figure 4.6 illustrates this transition. In the fixed-size VAE, the complete feeder-level input is encoded as a whole, mapped into a latent representation and decoded into a topology estimate with a fixed number of nodes. In the variable-size VAE, the input is reorganised into node descriptors. These descriptors are processed by a shared node encoder, producing one embedding per valid node. The node embeddings are then aggregated into a graph-level representation and used, together with pairwise node information, to infer the final topology.

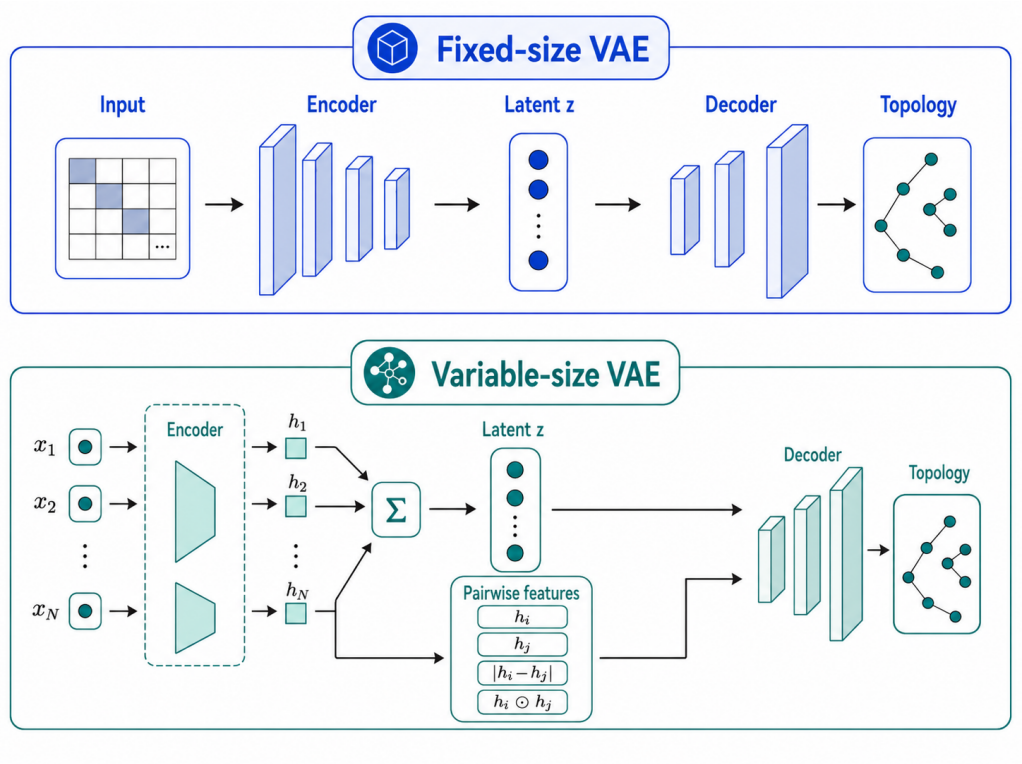


Figure 4.6: Transition from fixed-size to variable-size VAE topology inference.

For a feeder with N nodes, the model represents the input as a set of node descriptors,

$$X = \{x_1, x_2, \ldots, x_N\}, \tag{4.6}$$

where x_i contains the measurement-derived information associated with node i. The exact form of x_i depends on the input representation. For raw P/V inputs, each descriptor corresponds to the active-power and voltage-magnitude trajectories of one node over the full time window,

$$x_i = [P_i(1:T), V_i(1:T)]. \tag{4.7}$$

For matrix-based inputs, such as sensitivity or correlation matrices, each node descriptor is built from the row and column associated with node i. In this case, the descriptor represents how node i is related to the remaining nodes in the feeder. This is the key difference from the fixed-size model: the matrix is no longer processed only as one complete feeder-level image, but is reorganised into node-level relational descriptors.

Each descriptor is passed through the same node encoder,

$$h_i = f_\theta(x_i), \qquad i = 1,\dots,N, \tag{4.8}$$

where h_i is the embedding of node i and f_θ denotes the shared encoder. Since the encoder parameters θ are shared across all nodes and all graph sizes, the model is not tied to a fixed number of buses. For example, a 10-node feeder produces 10 node embeddings, while a 30-node feeder produces 30 node embeddings, but both are processed by the same encoder.

To allow mini-batch training with feeders of different sizes, padding and masking are used. Feeders with fewer nodes are padded only for computational alignment. A binary node mask identifies which positions correspond to real nodes and which positions correspond to padding. This mask is essential because it prevents padded positions from contributing to graph aggregation, pairwise edge decoding and loss computation. Therefore, padded entries do not represent artificial buses and cannot create artificial branches in the inferred topology.

The valid node embeddings are aggregated into a fixed-dimensional graph-level representation using masked pooling,

$$h_G = [\mathrm{mean}_{i\in\mathscr{V}}(h_i) \,\|\, \max_{i\in\mathscr{V}}(h_i)], \tag{4.9}$$

where $\mathscr{V}$ is the set of valid nodes and $\|$ denotes concatenation. This operation produces a graph-level representation with fixed dimension, independently of the number of nodes in the feeder. The graph-level representation is then used to compute the latent posterior,

$$q_\phi(z \mid G) = \mathscr{N}\left(\mu_\phi(h_G), \mathrm{diag}\left(\sigma^2_\phi(h_G)\right)\right). \tag{4.10}$$

The latent variable z captures feeder-level information, while the node embeddings h_i retain node-specific information. The decoder then evaluates all valid node pairs. For each pair (i,j), the decoder receives the latent graph variable, the two node embeddings and explicit pairwise interaction features,

$$\ell_{ij} = g_\psi(z, h_i, h_j, |h_i - h_j|, h_i \odot h_j), \tag{4.11}$$

where g_ψ is a shared pairwise decoder, $|h_i - h_j|$ represents the absolute difference between node embeddings, and $h_i \odot h_j$ represents their element-wise product. These pairwise features help the decoder compare nodes directly when estimating whether an electrical branch should exist between them.

The resulting edge-score matrix is symmetrised, diagonal entries are ignored, and all pairs involving padded nodes are masked out. During training, the model is compared with the ground-truth adjacency matrix using only valid node pairs. During inference, the predicted edge scores are projected onto a valid radial topology using a maximum spanning tree procedure. This guarantees that the final inferred graph is connected and contains exactly $N-1$ branches, as required for radial LV feeders.

This variable-size formulation therefore addresses the main limitation of the fixed-size models. Instead of training one model per network size, a single model can be trained on feeders with different numbers of nodes and evaluated on both *training* and *held-out* graph sizes. The following experiments evaluate whether this formulation preserves the advantages of structured input representations while improving scalability across network dimensions.

4.4.1 Variable-Size Experimental Setup

The variable-size experiments were designed to evaluate whether a single model could process feeders with different numbers of nodes. In this setting, *training sizes* denote node counts represented in the mixed-size training set, whereas *held-out sizes* denote node counts excluded from training and used only for evaluation. This terminology differs from the *seen/unseen* topology split used in the fixed-size experiments: an *unseen* topology has a connectivity pattern not represented during training, although its number of nodes may still correspond to a *training size*.

The training set included graph sizes

$$N \in \{5, 10, 15, 20, 25, 30\}. \tag{4.12}$$

Generalisation was evaluated on *held-out* graph sizes,

$$N \in \{8, 13, 23, 50, 80\}. \tag{4.13}$$

The intermediate *held-out* sizes, namely 8, 13 and 23 nodes, evaluate whether the model can interpolate between or near the graph dimensions observed during training. The larger *held-out* sizes, namely 50 and 80 nodes, evaluate extrapolation beyond the maximum graph size used for training.

All variable-size models were trained using the same operating window adopted in the previous experiments, with 720 timestamps per sample. The objective was to evaluate whether a single model could process feeders with different numbers of nodes while preserving the advantages of the structured input representations studied in Section 4.3.

4.4.2 Variable-Size VAE Results

Table 4.8 reports the mean edge accuracy of the initial variable-size VAE for raw P/V, sensitivity and correlation representations.

Table 4.8: Mean edge accuracy of the variable-size VAE for raw P/V, sensitivity and correlation representations across *training* and *held-out* graph sizes.

N	Size group	Raw P/V	Sensitivity	Correlation
5	*Training size*	1.000	0.750	0.750
10	*Training size*	0.853	0.978	0.933
15	*Training size*	0.793	0.983	0.960
20	*Training size*	0.776	0.981	0.948
25	*Training size*	0.748	0.972	0.928
30	*Training size*	0.763	0.957	0.827
8	*Held-out size*	0.846	0.957	0.857
13	*Held-out size*	0.833	0.977	0.962
23	*Held-out size*	0.780	0.977	0.943
50	*Held-out size*	0.693	0.861	0.392
80	*Held-out size*	0.589	0.642	0.187

The results show that the variable-size formulation can process multiple graph dimensions with a single model. This is the main difference relative to the fixed-size models, where each network size required a separate architecture and training procedure. The variable-size model preserves high performance on several graph sizes represented during training and also generalises to *held-out* graph sizes.

The sensitivity representation is the strongest overall input representation in the variable-size setting. The only exception is the 5-node *training size* case, where raw P/V reaches perfect edge accuracy while the sensitivity and correlation inputs reach 0.750. This isolated result should be interpreted with caution because very small radial graphs contain only a few branches, so a small number of edge mistakes has a comparatively large effect on the final metric. From 10 to 30 nodes, the sensitivity representation remains close to or above 0.95 edge accuracy, reaching 0.978 for 10 nodes, 0.983 for 15 nodes, 0.981 for 20 nodes, 0.972 for 25 nodes and 0.957 for 30 nodes. This confirms that sensitivity matrices remain effective when moving from fixed-size inference to the shared-node variable-size formulation.

The intermediate *held-out* graph sizes also show strong generalisation. For 8, 13 and 23 nodes, the sensitivity-based model reaches edge accuracies of 0.957, 0.977 and 0.977, respectively. These values indicate that the model is able to interpolate between or near the graph sizes observed during training. This is an important result because it shows that the shared encoder, masking and pairwise decoding structure can transfer the learned topology-inference rule to graph dimensions that were not directly used during training.

The largest *held-out* graph sizes are more challenging. For 50 nodes, the sensitivity representation still reaches an edge accuracy of 0.861, showing meaningful extrapolation beyond the

training range. However, for 80 nodes, the edge accuracy decreases to 0.642. This drop indicates that extrapolation far beyond the largest training graph remains difficult, even when using a physically meaningful input representation.

The comparison with raw P/V and correlation inputs reinforces the importance of representation design. Raw P/V inputs remain below the sensitivity representation for most graph sizes, because the model must infer node-to-node electrical relationships implicitly from temporal trajectories. Correlation inputs perform well for several *training* and intermediate *held-out* sizes, but degrade strongly for larger *held-out* graphs, reaching only 0.392 for 50 nodes and 0.187 for 80 nodes. This suggests that correlation matrices capture useful statistical dependencies, but these dependencies do not extrapolate as reliably as sensitivity-based voltage-response patterns.

Overall, the results show that the proposed variable-size formulation generalises across both training and *held-out* graph sizes, with sensitivity matrices providing the most effective representation. The degradation observed for the 80-node case nevertheless indicates limited extrapolation to much larger feeders, motivating the Transformer-inspired extension presented next.

4.5 Transformer-Inspired Variable-Size VAE Results

Although the initial variable-size VAE generalises across graph sizes, node embeddings are still aggregated before explicit global interaction. This motivates introducing a Transformer-inspired self-attention stage before graph-level pooling.

However, the initial variable-size architecture still has an important limitation. As described in Equation 4.8, each node descriptor is first encoded independently. The resulting node embeddings are then aggregated through the masked pooling operation introduced in Equation 4.9. Therefore, in the initial formulation, graph-level aggregation occurs before the model explicitly allows each node embedding to exchange information with all other node embeddings.

This can be restrictive for topology inference. The existence of a branch between two nodes is not determined only by the isolated properties of each node, but by the relation between nodes and by their position within the complete feeder. This limitation becomes more relevant as the number of nodes increases, because the number of possible node pairs grows rapidly with network size. In larger feeders, the model must distinguish between many more candidate branches, and this requires a more expressive mechanism for modelling global node-to-node interactions.

The Transformer-inspired extension preserves the variable-size pipeline and introduces a masked self-attention stage between the shared node encoder and the graph-level aggregation operation. This stage allows each node embedding to be contextualised using information from all other valid nodes before the graph-level latent representation is constructed.

Let H denote the matrix containing the node embeddings produced by the shared encoder,

$$H = h_1^T h_2^T \vdots h_N^T \in R^{N \times d_h}, \tag{4.14}$$

where each row corresponds to the embedding of one valid node, N is the number of valid nodes in the feeder and d_h is the node-embedding dimension. Each embedding h_i is obtained using the shared encoder defined in Equation 4.8. The Transformer-inspired block computes query, key and value projections from these embeddings,

$$Q = HW_Q, \qquad K = HW_K, \qquad V = HW_V, \tag{4.15}$$

where W_Q, W_K and W_V are learned projection matrices. The contextualisation step is then performed using masked scaled dot-product attention,

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + \widetilde{M}\right)V. \tag{4.16}$$

In Equation 4.16, d_k is the key dimension and $\widetilde{M}$ is an additive mask derived from the node mask. Valid node pairs receive a zero mask value, while padded positions receive a large negative value before the softmax operation. This prevents padded nodes from contributing to the attention weights. Therefore, padding is still used only for mini-batch alignment and does not introduce artificial buses or artificial branches.

The output of the attention block is a contextualised embedding matrix,

$$\widetilde{H} = \text{Attention}(Q,K,V), \tag{4.17}$$

where each contextualised embedding $\widetilde{h}_i$ contains information not only about node i, but also about its relation to the remaining valid nodes in the feeder. This is the key architectural difference relative to the initial variable-size VAE. In the initial model, node embeddings are pooled before explicit global interaction. In the Transformer-inspired model, node embeddings first interact through self-attention and are only then aggregated.

After this contextualisation stage, the rest of the pipeline follows the variable-size formulation described in Section 4.4. The contextualised embeddings are aggregated using the masked pooling operation of Equation 4.9, replacing h_i by $\widetilde{h}_i$. The graph-level representation is then used to compute the latent posterior as in Equation 4.10. Finally, the pairwise decoder introduced in Equation 4.11 is applied to valid node pairs, again using the contextualised node embeddings. The score matrix is symmetrised, diagonal entries are ignored, padded pairs are masked out, and the final prediction is projected onto a valid radial topology using maximum spanning tree projection. The padding masks, latent-variable formulation, pairwise edge decoder and radial-tree projection remain the same as in the initial variable-size model. Therefore, the principal architectural difference is the introduction of global inter-node interactions before pooling.

This architecture is expected to be particularly useful for larger feeders. By allowing node embeddings to exchange information before graph-level compression, the model can represent global node-to-node dependencies more explicitly. This should help the decoder distinguish between plausible and implausible branches when the number of candidate edges increases.

The Transformer-inspired model was evaluated using sensitivity-based inputs, since the previous sections showed that sensitivity matrices were the most consistent representation across graph sizes. The first training configuration used the same graph-size range as the initial variable-size VAE, namely $N \in \{5, 10, 15, 20, 25, 30\}$. The model was then evaluated on graph sizes represented during training and on *held-out* graph sizes, including $N \in \{8, 13, 23, 50, 80\}$.

Table 4.9 reports the results obtained with this first configuration. The mean edge error, defined in Equation 3.32, represents the average number of incorrectly recovered branches across the evaluated topologies. All values were computed after radial projection.

Table 4.9: Transformer-inspired variable-size VAE results with sensitivity-based inputs.

N	Size group	Mean edge error	Edge accuracy
8	*Held-out size*	0.100	0.986
10	*Training size*	0.200	0.978
13	*Held-out size*	0.160	0.987
15	*Training size*	0.260	0.981
20	*Training size*	0.420	0.978
23	*Held-out size*	0.160	0.987
30	*Training size*	1.200	0.959
50	*Held-out size*	5.660	0.885
80	*Held-out size*	20.000	0.747

The results show that the Transformer-inspired model preserves strong performance for graph sizes within or close to the training range. For the intermediate *held-out* sizes, the model achieves edge accuracies of 0.986 for 8 nodes, 0.987 for 13 nodes and 0.987 for 23 nodes. These values indicate that the attention-based contextualisation does not harm interpolation between graph sizes and can maintain high reconstruction accuracy on dimensions near the training range.

The most relevant difference appears for larger *held-out* networks. In the initial variable-size VAE, the sensitivity-based model achieved edge accuracies of 0.861 and 0.642 for the 50-node and 80-node cases, respectively. With the Transformer-inspired architecture, these values increased to 0.885 and 0.747. This comparison is summarised in Table 4.10.

Table 4.10: Comparison between the initial variable-size VAE and the Transformer-inspired variable-size VAE on *held-out* graph sizes.

N	Initial variable-size VAE	Transformer-inspired VAE	Improvement
8	0.957	0.986	+0.029
13	0.977	0.987	+0.010
23	0.977	0.987	+0.010
50	0.861	0.885	+0.024
80	0.642	0.747	+0.105

This comparison shows that the Transformer-inspired architecture is especially useful when extrapolating to larger graph sizes. The improvements for 8, 13 and 23 nodes are modest because the initial variable-size model was already very strong for intermediate *held-out* sizes. In contrast, the gain for 80 nodes is substantially larger, increasing edge accuracy by 0.105. This supports the interpretation that global node-to-node interactions become increasingly important as the number of candidate branches grows.

However, the 80-node case also shows that extrapolation remains challenging. Although edge accuracy improved from 0.642 to 0.747, the mean edge error remained high. This means that the Transformer-inspired architecture improves large-graph generalisation, but does not completely solve the problem of extrapolating far beyond the largest graph size included in training.

To further evaluate this limitation, a second training configuration was tested. In this configuration, the Transformer-inspired model was retrained using a broader range of graph sizes, extending the training set up to 50 nodes with $N \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$. The objective was to determine whether exposing the model to larger feeders during training would improve extrapolation to the 80-node case.

The results for the extended training configuration are shown in Table 4.11.

Table 4.11: Transformer-inspired variable-size VAE after extending the training range up to 50 nodes.

N	Size group	Mean edge error	Edge accuracy
5	*Training size*	0.667	0.833
10	*Training size*	0.060	0.993
15	*Training size*	0.160	0.989
20	*Training size*	0.100	0.995
25	*Training size*	0.420	0.983
30	*Training size*	0.800	0.972
35	*Training size*	0.820	0.976
40	*Training size*	1.120	0.971
45	*Training size*	1.980	0.955
50	*Training size*	1.980	0.960
8	*Held-out size*	0.100	0.986
13	*Held-out size*	0.120	0.990
23	*Held-out size*	0.080	0.996
80	*Held-out size*	13.760	0.826

The extended training configuration confirms that the model benefits from seeing larger graph structures during training. For graph sizes between 10 and 50 nodes, the edge accuracy remains consistently high, mostly above 0.955. This indicates that the Transformer-inspired variable-size

model can scale across a broader training range while preserving strong reconstruction performance.

The lower value obtained for the 5-node case should be interpreted with caution. Very small radial graphs contain only a few branches, so even a small number of incorrect edge decisions can have a stronger effect on the final metric. Therefore, this result does not necessarily indicate a failure of the architecture, but rather reflects the sensitivity of the metric for very small graphs.

The most important result is the improvement on the 80-node *held-out* case. With training limited to graph sizes up to 30 nodes, the 80-node edge accuracy was 0.747. After extending the training range up to 50 nodes, this value increased to 0.826. This suggests that the attention-based architecture extrapolates more effectively when larger intermediate graph sizes are included in the training set.

Overall, the Transformer-inspired variable-size VAE improves the initial variable-size formulation by introducing global node-to-node interactions before graph-level aggregation. The results show strong interpolation to *held-out* sizes close to the training range and improved extrapolation to larger feeders, especially when the model is trained on a broader range of graph sizes. Nevertheless, the remaining error on the 80-node case indicates that large-scale extrapolation is still not fully solved. This limitation is important for practical deployment and motivates the following robustness analysis, where the sensitivity of the proposed models to measurement noise is evaluated.

4.6 Graph Completion with Partial Adjacency and Sensitivity Guidance

The final experimental family reformulates the problem as graph completion. Instead of inferring the entire topology from electrical measurements alone, the model receives a partially observed topology and must infer the missing structure. This formulation is practically relevant because DSOs may have incomplete network records rather than no structural information at all.

The graph-completion formulation retains the variable-size processing and pairwise edge decoding, but changes the information supplied to the model. Depending on the considered variant, the model receives either a partially observed adjacency matrix alone or the same structural information combined with the sensitivity-based electrical representation. The decoder therefore predicts the missing connectivity while preserving the structural information already provided in the input.

The partial observation contains three types of information: known existing edges, known absent edges and unknown entries. In the implementation, this is encoded through three channels: known positive entries, known negative entries and an observation mask. Each true branch was kept visible with probability $p_{\text{keep_edge}} = 0.50$, meaning that approximately 50% of the existing edges were provided as known positive information. In addition, each non-edge was explicitly marked as a known absence with probability $p_{\text{observe_nonedge}} = 0.05$. All remaining entries were

treated as unknown. During inference, known positive edges must be preserved, known negative edges must not be introduced, and the final graph must remain a connected radial tree.

The numerical results for graph sizes *seen* during training are reported in Table 4.12. The difference between the two variants is substantial.

Table 4.12: Inference results of the graph-completion architecture on graph sizes *seen* during training.

N	ADJ ONLY error	ADJ ONLY accuracy	ADJ + SENS error	ADJ + SENS accuracy
5	0.667	0.833	0.000	1.000
10	2.920	0.676	0.000	1.000
15	4.240	0.697	0.060	0.996
20	5.240	0.724	0.040	0.998
25	7.480	0.688	0.240	0.990
30	10.640	0.633	0.540	0.981

Partial adjacency alone is not sufficient to resolve the missing structure accurately. However, when sensitivity information is added, performance improves sharply. The ADJ + SENS model achieves perfect reconstruction for 5-node and 10-node networks and remains above 0.98 edge accuracy up to 30 nodes.

The *held-out*-size generalisation behaviour of the ADJ + SENS model is reported in Table 4.13.

Table 4.13: Generalisation results of the ADJ + SENS graph-completion architecture on *held-out* graph sizes.

N	Type	Mean edge accuracy
8	*Held-out size*	0.969
13	*Held-out size*	0.987
23	*Held-out size*	0.994
50	*Held-out size*	0.751
80	*Held-out size*	0.616

The model generalises very well to nearby *held-out* sizes, especially 8, 13 and 23 nodes. Performance decreases for 50-node and 80-node feeders, which correspond to a stronger extrapolation regime. Even so, the results remain meaningful and show that sensitivity-guided completion is a strong formulation when partial structural knowledge is available.

The main interpretation is that partial topology information and electrical sensitivity information are complementary. The partial adjacency constrains the feasible structure, while the sensitivity matrix provides physical guidance about how nodes are electrically related. Their combination

reduces ambiguity more effectively than either source alone.

4.7 Robustness to Measurement Errors

The previous sections evaluated topology inference under clean synthetic measurements. However, practical deployment in LV networks would depend on measurements collected from smart meters or monitoring devices, which may be affected by acquisition errors, calibration uncertainty, synchronisation issues and communication-related data quality limitations. For this reason, robustness to measurement errors is a key requirement for assessing the practical relevance of the proposed topology inference models.

The robustness studies were performed with the variable-size formulation, using 10-node networks, 2000 training topologies and 50 *unseen* test cases. The test topologies were different from the training topologies, so the experiment evaluates both *unseen*-topology generalisation and sensitivity to measurement perturbations while keeping the graph size fixed. Measurement errors between 0% and 1% were applied to active power, voltage, or both active power and voltage.

For each node i and timestamp t, the perturbed active-power and voltage-magnitude measurements can be written as

$$\widetilde{P}_i(t) = P_i(t)\left(1+\varepsilon_i^P(t)\right), \tag{4.18}$$

$$\widetilde{V}_i(t) = V_i(t)\left(1+\varepsilon_i^V(t)\right), \tag{4.19}$$

where $\widetilde{P}_i(t)$ and $\widetilde{V}_i(t)$ denote the corrupted measurements. The relative errors are sampled from a zero-mean Gaussian distribution,

$$\varepsilon_i^P(t), \varepsilon_i^V(t) \sim \mathcal{N}(0,\sigma^2), \qquad \sigma \in [0, 0.01]. \tag{4.20}$$

The parameter σ defines the standard deviation of the relative measurement error. Therefore, $\sigma = 0$ corresponds to the clean-data case, while $\sigma = 0.01$ corresponds to the maximum tested error level of 1%. The error range from 0% to 1% was selected as a representative robustness window, motivated by a preliminary review of commercial smart-meter specifications and related studies, where active-power and voltage measurement accuracies were commonly reported or evaluated within approximately $\pm 0.5\%$ to $\pm 1\%$.

Two scenarios were considered. In the inference-only scenario, the model was trained on clean data and noise was introduced only at inference time. This evaluates how sensitive a clean-trained model is to measurement degradation. In the database-and-inference scenario, noise was introduced both when constructing the training database and during inference. This evaluates whether exposing the model to noisy data during training improves robustness.

Three perturbation modes were tested in each scenario. In the first mode, noise was applied only to active-power measurements. In the second mode, noise was applied only to voltage measurements. In the third mode, active power and voltage were perturbed simultaneously. These cases are referred to as P, V and $P+V$ measurement errors, respectively.

Figure 4.7 shows the representative robustness result for the sensitivity-based representation in the variable-size VAE.

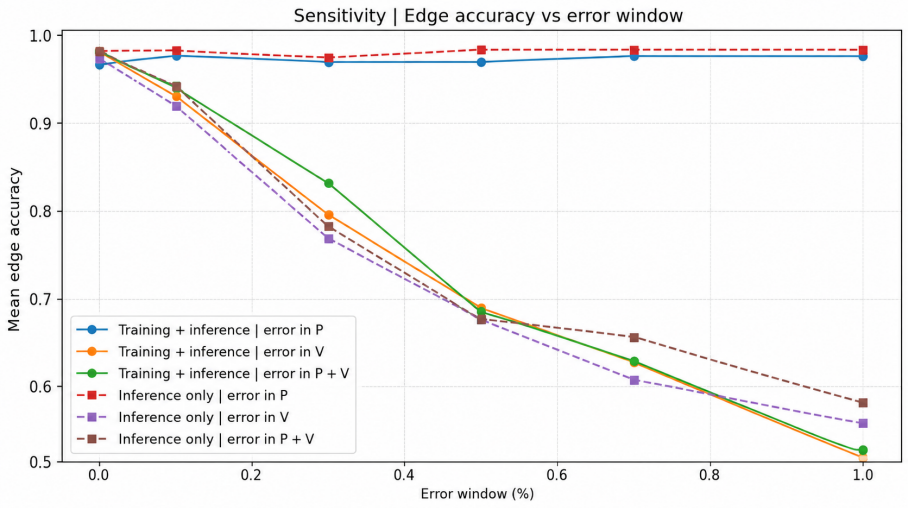


Figure 4.7: Robustness of the sensitivity-based representation under measurement errors.

The sensitivity-based model remains highly stable when the perturbation affects only active power. The corresponding curves are almost flat across the tested error range, indicating that active-power errors up to 1% have limited impact on topology reconstruction. This suggests that small perturbations in the active-power measurements do not substantially alter the relational information used by the model.

By contrast, voltage perturbations produce a much stronger degradation. This behaviour is visible both when voltage alone is perturbed and when active power and voltage are perturbed simultaneously. The $P+V$ curves tend to follow the voltage-only behaviour, showing that the combined-error case is dominated by the voltage component.

This result is physically coherent with the sensitivity-based representation. Sensitivity matrices are constructed to approximate how variations in active power are reflected in voltage responses. Therefore, voltage measurements carry the response signal that encodes much of the topology-related information. When voltage measurements are corrupted, the inferred voltage-response patterns become distorted, which directly affects the quality of the sensitivity matrix and, consequently, the topology reconstruction.

The inference-only and database-and-inference scenarios show that including noise during training can provide some robustness improvement, especially in the voltage and $P+V$ cases. However, this improvement is only partial. Training with noisy data reduces the degradation in some cases, but it does not recover the clean-data performance when voltage measurements are

affected. This indicates that noise-aware training helps, but voltage quality remains a critical factor for sensitivity-based topology inference.

Figure 4.8 shows the corresponding robustness trend for the Transformer-inspired variable-size VAE.

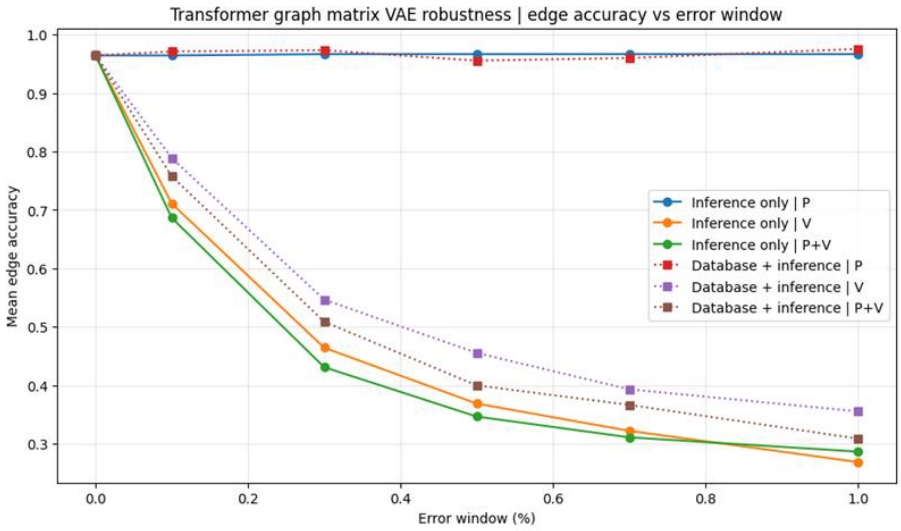


Figure 4.8: Robustness of the Transformer-inspired variable-size VAE under measurement errors.

The qualitative trend is consistent with the sensitivity-based variable-size VAE. Active-power perturbations have a limited effect on the reconstructed topology, while voltage perturbations lead to a much stronger performance degradation. When active power and voltage are perturbed simultaneously, the resulting behaviour again follows the voltage-dominated pattern.

This result highlights an important trade-off of the Transformer-inspired architecture. As shown in Section 4.5, the attention-based model improves clean-data performance and generalisation across graph sizes by exploiting global node-to-node relationships before graph-level aggregation. However, this also means that the model relies strongly on the consistency of fine-grained relational patterns in the input representation. When voltage noise distorts these patterns, the model can no longer exploit them as reliably.

Therefore, the Transformer-inspired architecture improves the use of clean relational information, but it does not remove the dependence on measurement quality. In particular, the model remains sensitive to errors that affect voltage magnitudes, because these errors alter the inter-node response patterns encoded in the sensitivity matrix.

4.7.1 Comparison with an Analytical Effective-Resistance Method

To determine whether the robustness behaviour observed for the sensitivity-based VAE was specific to the proposed learning architecture, an additional comparison was performed with an analytical radial topology-identification method based on effective electrical resistances and recursive grouping. Effective-resistance-based topology-learning approaches are reviewed by Deka et al. [7].

This comparison provides a direct connection to the sensitivity-based representation adopted in the proposed VAE. However, the reconstruction mechanisms are fundamentally different. The VAE learns an offline mapping from the sensitivity matrix to topology-related edge scores, which are subsequently projected into a radial tree. By contrast, the analytical method derives pairwise electrical distances directly from the sensitivity matrix and uses their additive properties to reconstruct the radial connectivity.

The analytical method first estimates the active-power-to-voltage sensitivity matrix $\hat{\mathbf{S}}$ from temporal variations in the active-power and voltage-magnitude data using the relationship given in Equation 1.3.

Pairwise effective resistances are then computed as

$$\widehat{R}_{\text{eff}}(m,n) = \widehat{S}(m,m) + \widehat{S}(n,n) - 2\widehat{S}(m,n).$$

In a radial network, these effective resistances behave as additive electrical distances along the unique paths connecting the nodes. Consequently, the effective resistance between two nodes corresponds to the sum of the electrical distances associated with the branches belonging to their connecting path.

The recursive-grouping procedure exploits this additive property by comparing the effective distances of pairs of nodes relative to the remaining nodes. For each pair of nodes i and j, the procedure evaluates

$$\Phi_{ij}(k) = \widehat{R}_{\text{eff}}(i,k) - \widehat{R}_{\text{eff}}(j,k),$$

for every remaining node k. If $\Phi_{ij}(k)$ is approximately equal to $\widehat{R}_{\text{eff}}(i,j)$ across the admissible values of k, node j is identified as the parent of node i. Conversely, if $\Phi_{ij}(k)$ is approximately equal to $-\widehat{R}_{\text{eff}}(i,j)$, node i is identified as the parent of node j.

When $\Phi_{ij}(k)$ remains approximately constant across the remaining nodes but differs from these two limiting values, nodes i and j are identified as belonging to the same sibling group and are associated with a common parent. After a parent–child relation or sibling group is identified, the corresponding family is contracted into its parent representation. The electrical distances between this representation and the remaining nodes are then updated using the same additive tree relations. This process is repeated recursively until the radial connectivity is reconstructed.

To evaluate robustness under measurement uncertainty, perturbations are first added to the active-power and/or voltage measurements. Because these perturbations prevent the tree-metric relations from holding exactly, the equality conditions in the analytical topology inference algorithm are replaced by a tolerance criterion. The tolerance is estimated during a calibration stage using training data and then kept fixed throughout the robustness analysis. A globally calibrated tolerance of 0.175 was used for all noise levels and for the P, V, and P+V perturbation modes.

The experiment used a fixed 10-node radial network. The analytical results were obtained over 50 Monte Carlo realisations. Across these realisations the true topology remained fixed,

while the active-power profiles and the measurement perturbations varied randomly. The same measurement-error range used in the VAE experiments, from 0% to 1%, was applied, considering errors in active power only, errors in voltage only, and simultaneous errors in active power and voltage.

For comparability, both the VAE-based approach and the analytical method were evaluated on 10-node radial networks and over the same measurement-error range. Nevertheless, the two experimental populations are not identical. The VAE results aggregate 50 *unseen* test topologies, whereas the analytical results aggregate 50 noise realisations of one fixed topology. Consequently, the comparison is intended to assess qualitative robustness patterns rather than to establish a direct performance ranking.

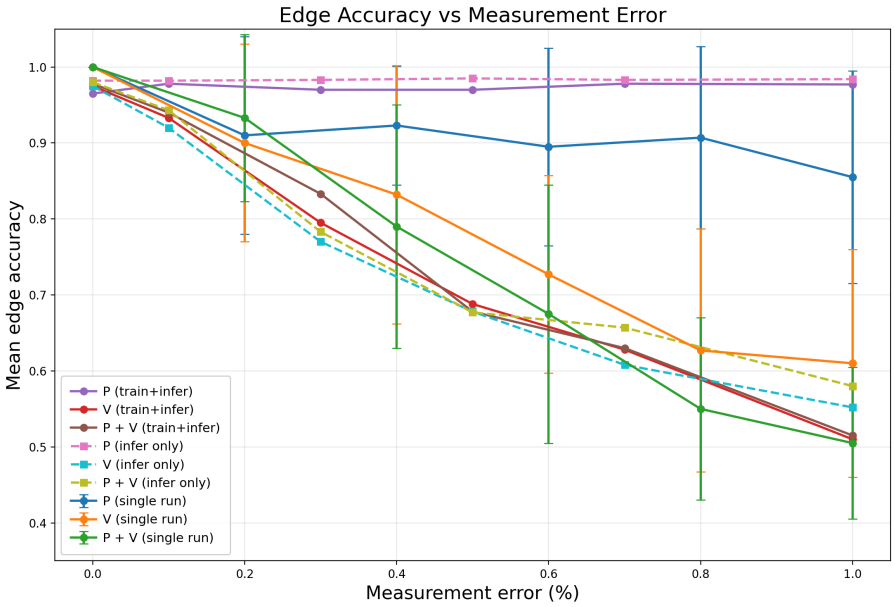


Figure 4.9: Comparison of mean edge accuracy under measurement perturbations for the sensitivity-based VAE and the analytical effective-resistance and recursive-grouping method. The VAE results include the training-and-inference and inference-only noise scenarios. The analytical method was independently applied to 50 noise realisations of a fixed 10-node radial network, and the error bars represent one standard deviation. In the figure legend, "single run" denotes a non-trained analytical reconstruction performed independently for each realisation, rather than a single Monte Carlo sample.

Figure 4.9 shows that the two approaches exhibit the same main qualitative robustness pattern. Errors applied only to active-power measurements have a comparatively limited effect on edge accuracy. By contrast, voltage perturbations lead to a much stronger performance degradation, and the combined $P+V$ case follows a similar voltage-dominated trend. This agreement is relevant because the VAE and recursive-grouping methods reconstruct the topology through substantially different mechanisms, yet both indicate that voltage quality is the main limiting factor when sensitivity-based information is used.

At some error levels, the VAE-based approach attains a higher mean edge accuracy than the analytical method, whereas at other levels the analytical method performs similarly or better. However, the analytical results exhibit considerable variability, as shown by the relatively large error bars. Together with the difference between the respective experimental populations, this variability prevents a conclusive claim that either method is generally superior. The main conclusion drawn from this comparison is therefore not a performance ranking, but the consistency of the robustness pattern observed for the proposed VAE with that of an established analytical topology-identification approach.

Taken together, the VAE experiments and the analytical comparison lead to a consistent robustness conclusion. Active-power measurement errors within the tested range have limited impact, whereas voltage measurement errors are the dominant source of performance degradation. This confirms that the strongest topology-inference representations developed in this work depend primarily on reliable voltage-response information.

From a practical perspective, this result is important. It suggests that future deployment of sensitivity-based topology inference would require careful attention to voltage measurement quality, including calibration, filtering and uncertainty-aware preprocessing. The results also indicate that model improvements alone are not sufficient to guarantee robustness if the voltage measurements used to build the input representation are significantly corrupted.

The robustness study therefore complements the clean-data results presented in Sections 4.4 and 4.5. The proposed models can achieve strong topology-reconstruction performance when the voltage-response patterns are reliable, but their performance degrades when those patterns are distorted by voltage measurement noise. This limitation should be considered when assessing the practical deployment of the proposed framework in real LV networks.

4.8 Integrated Discussion

The experiments presented in this chapter followed a progressive path from simple fixed-size topology inference models to variable-size architectures and robustness analysis. The objective of this final discussion is not to repeat the numerical results reported in the previous sections, but to consolidate the main interpretation of the experimental progression.

The first important observation is that raw P/V trajectories are not, by themselves, the most effective representation for topology inference. In the initial fixed-size experiments, the models were able to learn partial structural information, but their generalisation to *unseen* topologies remained limited. This showed that simply providing the temporal active-power and voltage trajectories to a VAE does not guarantee that the model can extract the electrical relationships required to reconstruct the feeder topology.

The subsequent fixed-size refinements clarified this limitation. Increasing the number of training topologies, changing the temporal horizon and modifying the data allocation strategy produced some improvements, but did not fundamentally solve the problem. The transition to convolutional

architectures improved performance because matrix-like structures were processed more appropriately, but the main improvement came from changing the input representation itself. This indicates that topology inference depends strongly on how the electrical information is presented to the model.

The input representation study showed that the most relevant information is concentrated in the relation between active-power variations and voltage responses. The exploratory analysis of the correlation matrix showed that the power-voltage block retains most of the useful information of the complete correlation representation, while avoiding the computational effort of constructing and storing the full four-block matrix. This justified the use of PV-only correlations as a compact correlation-based input. However, the controlled comparison also showed that sensitivity matrices provide a more physically structured representation of the same underlying relationship. While PV-only correlations describe statistical association, sensitivity matrices approximate the directional response of voltage magnitudes to active-power perturbations. This explains their strong and consistent performance across the following experiments.

The two-encoder experiments further reinforced the importance of representation design. Combining sensitivity matrices with PV-only correlations produced the strongest result among the tested fusion strategies, suggesting that the two structured representations contain partially complementary information. By contrast, combining raw P/V trajectories with structured representations did not improve performance. This shows that adding more input channels is not sufficient by itself. The additional information must expose relationships that are useful for edge reconstruction.

The transition to variable-size models addressed a different limitation of the fixed-size setting. Fixed-size VAEs require a separate model for each number of nodes, which limits their practical applicability to LV feeders with different dimensions. The variable-size VAE replaced this formulation with shared node encoders, masking, graph-level pooling and pairwise edge decoding. This allowed a single model to process feeders with different numbers of nodes. The results showed that variable-size topology inference is feasible, particularly when sensitivity matrices are used as input. The model achieved strong performance for graph sizes represented during training and for intermediate *held-out* sizes, confirming that the learned inference rule can transfer across nearby network dimensions.

Nevertheless, the initial variable-size model still struggled with stronger extrapolation to much larger graphs. This motivated the Transformer-inspired extension. By inserting a masked self-attention stage before graph-level aggregation, the Transformer-inspired architecture allowed node embeddings to exchange information before being compressed into a graph-level latent representation. This improved performance especially for larger *held-out* graph sizes, supporting the idea that global node-to-node interactions become more important as the number of possible branches increases. Extending the training range up to 50 nodes further improved the 80-node result, showing that large-graph extrapolation benefits both from a more expressive architecture and from exposure to larger intermediate graph sizes during training.

The robustness analysis introduced an important practical qualification to these results. Across

the tested architectures, active-power measurement errors had limited impact in the considered range, whereas voltage errors caused much stronger degradation. This behaviour is consistent with the role of voltage-response information in the proposed representations. Sensitivity matrices and PV-only correlations both rely on the relation between active-power variations and voltage variations; therefore, when voltage measurements are corrupted, the structural signal used for topology inference is directly affected. The Transformer-inspired model improves the use of clean relational information, but it does not eliminate the dependence on voltage quality.

Finally, the graph-completion experiments complement the purely measurement-based results by considering the case where partial structural information is available. This is relevant in practical LV networks, where the operator may not know the complete topology but may have access to partial connectivity information from GIS records, historical documentation or field knowledge. In this context, graph completion can be interpreted as an intermediate deployment scenario: instead of inferring the entire topology from measurements alone, the model uses available structural information to guide the reconstruction of the missing branches.

Overall, the experiments show that topology inference benefits from physics-informed representations and scalable variable-size architectures, while voltage quality remains the main practical limitation. The broader implications of these findings are discussed in Chapter 5.

These conclusions define the final position of the proposed framework. Sensitivity-based variable-size VAEs, especially when enhanced with Transformer-inspired contextualisation, provide a scalable and physically informed approach for LV topology inference. However, the method should be understood as a promising data-driven framework under controlled assumptions, rather than as a complete replacement for all existing topology identification procedures. Its practical use would benefit from reliable voltage measurements, uncertainty-aware preprocessing and, when available, partial structural information to reduce ambiguity in the inferred topology.

Chapter 5

Conclusions and Future Work

This dissertation addressed the problem of topology inference in low-voltage distribution networks under limited smart-meter observability. The work was motivated by the increasing operational complexity of LV grids and by the fact that accurate and up-to-date topology information is often unavailable, incomplete or uncertain in practical distribution networks.

The proposed approach formulated topology inference as a graph reconstruction problem, using synthetic radial feeders with known ground-truth adjacency matrices to enable controlled supervised learning. The experimental progression showed that topology inference benefits from two main ingredients: measurement-derived representations that expose physically meaningful relationships between nodes, and learning architectures that can operate across different graph sizes.

The results showed that raw P/V trajectories are limited when used directly, since the model must infer node-to-node electrical relationships implicitly. Correlation-based representations improved this by capturing statistical dependencies between active-power and voltage variations. However, sensitivity matrices provided the most consistent representation, because they approximate how active-power perturbations are reflected in voltage responses across the feeder.

Architecturally, the work progressed from fixed-size VAEs to convolutional VAEs, variable-size VAEs and Transformer-inspired variable-size VAEs. The fixed-size experiments revealed the limitations of models tied to a single network dimension. The variable-size formulation addressed this limitation through shared node encoders, masking, graph-level aggregation, pairwise edge decoding and radial projection. The Transformer-inspired extension further improved generalisation by introducing global node-to-node interactions before graph-level aggregation.

Overall, the results support the central conclusion that LV topology inference is improved by combining physically meaningful electrical representations with variable-size graph learning architectures. At the same time, the robustness experiments showed that voltage measurement quality remains a critical practical constraint, since voltage errors dominate performance degradation.

5.1 Limitations

Notwithstanding the promising results obtained, the work presented in this dissertation is subject to several limitations that should be acknowledged.

First, the experiments were based on synthetic radial LV feeders. This choice was necessary to generate labelled datasets with known ground-truth topologies and to perform controlled comparisons between architectures, input representations and graph sizes. However, synthetic data cannot fully reproduce the complexity of real LV networks, including incomplete asset records, customer diversity, measurement irregularities, switching events and undocumented field changes.

Second, the modelling assumptions were simplified. The main experiments considered active-power and voltage-magnitude measurements and assumed a known phase-specific representation. Although these assumptions are useful for isolating the topology inference problem, real LV feeders may require explicit consideration of reactive power, unknown phase allocation, unbalanced three-phase coupling, neutral effects and non-metered intermediate junctions.

Third, the robustness analysis showed that the proposed models remain sensitive to voltage measurement errors. This limitation is particularly important because the strongest input representations, namely sensitivity matrices and PV-only correlations, depend directly on voltage-response information. When voltage measurements are corrupted, the relational structure used for topology inference is degraded.

Fourth, the models infer radial topologies and rely on radial projection during post-processing. This is appropriate for many LV distribution feeders, but it may not directly represent networks with meshed operation, normally open points or temporary switching configurations. The projection step also selects a single final topology from a dense edge-score matrix, which may hide uncertainty when several candidate branches have similar scores.

Finally, the evaluation focused mainly on structural reconstruction metrics such as edge accuracy and edge error. Although these metrics are appropriate for measuring topology recovery, they do not fully quantify the operational value of the inferred topology. Additional evaluation would be needed to assess the impact of topology inference on power-flow estimation, voltage prediction, state estimation, flexibility management or fault localisation.

5.2 Future Work

Future research directions stemming from this work are multiple and diverse. A key priority lies in validating the proposed topology inference framework using real LV feeder data. This would require access to smart-meter voltage and active-power measurements, transformer-level measurements, network asset records and, ideally, field-validated connectivity information. Such validation would make it possible to assess how the proposed models behave under practical measurement irregularities, incomplete GIS records, unknown phase allocation, missing nodes and real customer behaviour.

Another promising direction involves improving robustness and uncertainty awareness. The robustness analysis showed that voltage measurement errors are the dominant source of performance degradation, while active-power errors have a more limited effect in the tested range. Future work should therefore explore voltage filtering, calibration procedures, denoising strategies, uncertainty-aware sensitivity estimation and noise-aware training. In addition, the current single-output radial projection could be extended to candidate-based evaluation, where several plausible radial topologies are sampled or compared using confidence scores and physical-consistency criteria.

A further research direction is a systematic comparison with other topology-discovery methods discussed in the literature review, particularly genetic-algorithm-based approaches. The same networks, measurement windows and error conditions should be used to compare the proposed VAE-based models with genetic algorithms in terms of edge accuracy, edge error, exact recovery, computational effort, inference time, scalability and robustness to measurement errors. Offline VAE training costs and online inference times should be reported separately to enable a fair comparison with the iterative search required by genetic algorithms. Hybrid approaches could also be investigated, for example by using the VAE predictions to initialise the genetic-algorithm population or to reduce the candidate-edge search space, followed by evolutionary refinement of the inferred topology.

In terms of architectural development, future work should further improve scalability to larger and more heterogeneous feeders. The Transformer-inspired variable-size VAE improved generalisation to larger *held-out* graph sizes, especially when the training range was extended up to 50 nodes. Nevertheless, extrapolation to much larger graphs remained challenging. Future work could explore curriculum learning over graph sizes, larger and more diverse training datasets, graph-specific positional encodings, attention mechanisms guided by electrical distance, and alternative graph neural architectures designed explicitly for variable-size radial networks.

Lastly, further research is warranted to investigate the integration of partial DSO information and downstream operational validation. In practice, operators may not know the complete topology, but they often have access to partial information from GIS records, historical documentation or field inspections. Future work should therefore extend the graph-completion formulation to account for uncertain, missing or incorrect prior topology information. The inferred topologies should also be evaluated in operational tasks such as voltage estimation, sensitivity estimation, hosting-capacity assessment, flexibility management and fault localisation, in order to quantify their practical value beyond structural accuracy alone.

5.3 Final Remarks

This dissertation demonstrated that topology inference in LV distribution networks can benefit significantly from combining measurement-derived electrical representations with variable-size learning architectures. Sensitivity matrices were shown to be particularly effective because

they encode physically meaningful relationships between active-power variations and voltage responses. Variable-size VAEs then made it possible to apply the same model across feeders with different numbers of nodes, reducing the dependence on fixed network dimensions.

The progression from fixed-size VAEs to convolutional, variable-size and Transformer-inspired architectures showed that both representation and architecture are important. Better results were obtained not simply by increasing model complexity, but by structuring the input and the model around the electrical and graph-based nature of the problem. The Transformer-inspired extension further showed that global node-to-node interactions are useful for improving generalisation to larger feeders.

At the same time, the work identified important practical challenges. Robustness to voltage measurement errors, extrapolation to much larger networks and integration with incomplete real-world DSO records remain open issues. These limitations do not invalidate the proposed framework, but they define the conditions under which it should be further developed and tested.

Overall, the dissertation provides a structured experimental path for LV topology inference using physics-guided representations and variable-size generative models. It shows that data-driven topology inference is feasible under controlled assumptions, and that its practical deployment will depend on combining learning-based reconstruction with reliable measurements, physical validation and available network knowledge.

References

[1] Thomas Ackermann, Göran Andersson, and Lennart Söder. "Distributed Generation: A Definition". In: *Electric Power Systems Research* 57.3 (2001), pp. 195–204. DOI: 10.1016/S0378-7796(01)00101-8.

[2] Hassan Farhangi. "The Path of the Smart Grid". In: *IEEE Power and Energy Magazine* 8.1 (2010), pp. 18–28. DOI: 10.1109/MPE.2009.934876.

[3] Kristien Clement-Nyns, Edwin Haesen, and Johan Driesen. "The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid". In: *IEEE Transactions on Power Systems* 25.1 (2010), pp. 371–380. DOI: 10.1109/TPWRS.2009.2036481.

[4] Reinaldo Tonkoski, Dave Turcotte, and Tarek H. M. El-Fouly. "Impact of High PV Penetration on Voltage Profiles in Residential Neighborhoods". In: *IEEE Transactions on Sustainable Energy* 3.3 (2012), pp. 518–527. DOI: 10.1109/TSTE.2012.2191425.

[5] Vinoth Kumar Ponnusamy et al. "A Comprehensive Review on Sustainable Aspects of Big Data Analytics for the Smart Grid". In: *Sustainability* 13.23 (2021), p. 13322. DOI: 10.3390/su132313322.

[6] Ali A. Radwan et al. "Active Distribution Network Modeling for Enhancing Sustainable Power System Performance; a Case Study in Egypt". In: *Sustainability* 12.21 (2020), p. 8991. DOI: 10.3390/su12218991.

[7] Deepjyoti Deka, Vassilis Kekatos, and Guido Cavraro. *Learning Distribution Grid Topologies: A Tutorial*. arXiv:2206.10837. 2023.

[8] Abu Bakr Pengwah et al. "Topology Identification of Radial Distribution Networks Using Smart Meter Data". In: *IEEE Systems Journal* 16.4 (2022), pp. 5708–5719. DOI: 10.1109/JSYST.2021.3128175.

[9] S. J. Pappu et al. "Identifying Topology of Low Voltage Distribution Networks Based on Smart Meter Data". In: *IEEE Transactions on Smart Grid* 9.5 (2018), pp. 5113–5122. DOI: 10.1109/TSG.2017.2680542.

[10] Saverio Bolognani et al. "Identification of Power Distribution Network Topology via Voltage Correlation Analysis". In: *52nd IEEE Conference on Decision and Control*. 2013, pp. 1659–1664. DOI: 10.1109/CDC.2013.6760120.

[11] Guido Cavraro, Vassilis Kekatos, and Sriharsha Veeramachaneni. "Voltage Analytics for Power Distribution Network Topology Verification". In: *IEEE Transactions on Smart Grid* 10.1 (2019), pp. 1058–1067. DOI: 10.1109/TSG.2017.2758600.

[12] Yinjiao Liao et al. "Unbalanced Multi-Phase Distribution Grid Topology Estimation and Bus Phase Identification". In: *IET Smart Grid* 2.4 (2019), pp. 557–570. DOI: 10.1049/iet-stg.2019.0052.

[13] J. Yu, Y. Weng, and R. Rajagopal. "PaToPaEM: A Data-Driven Parameter and Topology Joint Estimation Framework for Time-Varying System in Distribution Grids". In: *IEEE Transactions on Power Systems* 34.3 (2019), pp. 1682–1692. DOI: 10.1109/TPWRS.2018.2888619.

[14] Jun Zhang et al. "Topology Identification and Line Parameter Estimation for Non-PMU Distribution Network: A Numerical Method". In: *IEEE Transactions on Smart Grid* 11.5 (2020), pp. 4440–4453. DOI: 10.1109/TSG.2020.2980659.

[15] ERSE. *Redes inteligentes*. Available online. 2024. URL: https://www.erse.pt/atividade/regulamentos-eletricidade/redes-inteligentes/ (visited on 05/19/2026).

[16] E-REDES. *Contadores Inteligentes*. Available online. 2026. URL: https://www.e-redes.pt/pt-pt/contadores-inteligentes (visited on 05/19/2026).

[17] Haoran Li et al. "Distribution Grid Impedance and Topology Estimation with Limited or No Micro-PMUs". In: *International Journal of Electrical Power & Energy Systems* 129 (2021), p. 106794. DOI: 10.1016/j.ijepes.2021.106794.

[18] Keith Moffat, Mohini Bariya, and Alexandra von Meier. "Unsupervised Impedance and Topology Estimation of Distribution Networks: Limitations and Tools". In: *IEEE Transactions on Smart Grid* 11.1 (2020), pp. 846–856. DOI: 10.1109/TSG.2019.2956706.

[19] Guido Cavraro et al. "Real-Time Identifiability of Power Distribution Network Topologies with Limited Monitoring". In: *IEEE Control Systems Letters* 4.2 (2020), pp. 325–330. DOI: 10.1109/LCSYS.2019.2926101.

[20] D. Lima and G. Sampaio. "Topology Reconstruction of Low Voltage Grids Using Genetic Algorithms". In: *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2025, pp. 7029–7034. DOI: 10.1109/SMC58881.2025.11342963.

[21] Gil Sampaio et al. "Conditional Parametric Model for Sensitivity Factors in LV Grids: A Privacy-Preserving Approach". In: *Electric Power Systems Research* 211 (2022), p. 108316. DOI: 10.1016/j.epsr.2022.108316.

[22] Saverio Bolognani and Sandro Zampieri. "On the Existence and Linear Approximation of the Power Flow Solution in Power Distribution Networks". In: *IEEE Transactions on Power Systems* 31.1 (2016), pp. 163–172. DOI: 10.1109/TPWRS.2015.2395452.

[23] Manish K. Singh et al. "Learning to Optimize Power Distribution Grids using Sensitivity-Informed Deep Neural Networks". In: *Proceedings of the 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2020, pp. 1–6. DOI: 10.1109/SmartGridComm47815.2020.9302942.

[24] F. Dalavi, M. E. H. Golshan, and N. D. Hatziargyriou. "A Review on Topology Identification Methods and Applications in Distribution Networks". In: *Electric Power Systems Research* 234 (2024), p. 110538.

[25] Sourabh Katoch, Sumit Singh Chauhan, and Vijay Kumar. "A Review on Genetic Algorithm: Past, Present, and Future". In: *Multimedia Tools and Applications* 80.5 (2021), pp. 8091–8126. DOI: 10.1007/s11042-020-10139-6.

[26] F. Therrien, L. Blakely, and M. J. Reno. "Assessment of Measurement-Based Phase Identification Methods". In: *IEEE Open Access Journal of Power and Energy* 8 (2021), pp. 128–137. DOI: 10.1109/OAJPE.2021.3068798.

[27] Ye Yuan et al. “Inverse Power Flow Problem”. In: *IEEE Transactions on Control of Network Systems* 10.1 (2023), pp. 261–273. DOI: 10.1109/TCNS.2022.3199084.

[28] Jian Zhao et al. “Full-Scale Distribution System Topology Identification Using Markov Random Field”. In: *IEEE Transactions on Smart Grid* 11.6 (2020), pp. 4714–4726. DOI: 10.1109/TSG.2020.2995164.

[29] Guido Cavraro and Vassilis Kekatos. “Graph Algorithms for Topology Identification Using Power Grid Probing”. In: *IEEE Control Systems Letters* 2.4 (2018), pp. 689–694. DOI: 10.1109/LCSYS.2018.2846801.

[30] Guido Cavraro and Vassilis Kekatos. “Inverter Probing for Power Distribution Network Topology Processing”. In: *IEEE Transactions on Control of Network Systems* 6.3 (2019), pp. 980–992. DOI: 10.1109/TCNS.2019.2901714.

[31] Deepjyoti Deka et al. “Tractable Learning in Underexcited Power Grids”. In: *IEEE Transactions on Control of Network Systems* 9.2 (2022), pp. 763–774. DOI: 10.1109/TCNS.2021.3124299.

[32] Rajasekhar Anguluri et al. “Grid Topology Identification with Hidden Nodes via Structured Norm Minimization”. In: *IEEE Control Systems Letters* 6 (2022), pp. 1244–1249. DOI: 10.1109/LCSYS.2021.3089993.

[33] Yinjiao Liao et al. “Urban MV and LV Distribution Grid Topology Estimation via Group Lasso”. In: *IEEE Transactions on Power Systems* 34.1 (2019), pp. 12–27. DOI: 10.1109/TPWRS.2018.2868877.

[34] Sejun Park, Deepjyoti Deka, and Michael Chertkov. “Exact Topology and Parameter Estimation in Distribution Grids with Minimal Observability”. In: *2018 Power Systems Computation Conference (PSCC)*. 2018, pp. 1–6.

[35] Vedantham Lakshmi Srinivas and Jianzhong Wu. “Topology and Parameter Identification of Distribution Network Using Smart Meter and μPMU Measurements”. In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–13. DOI: 10.1109/TIM.2022.3179464.

[36] Mohammad Farajollahi, Alireza Shahsavari, and Hamed Mohsenian-Rad. “Topology Identification in Distribution Systems Using Line Current Sensors: An MILP Approach”. In: *IEEE Transactions on Smart Grid* 11.2 (2020), pp. 1159–1170. DOI: 10.1109/TSG.2019.2933006.

[37] Chris Godsil and Gordon F. Royle. *Algebraic Graph Theory*. Vol. 207. Graduate Texts in Mathematics. New York: Springer, 2001.

[38] William H. Kersting. *Distribution System Modeling and Analysis*. 4th ed. Boca Raton, FL: CRC Press, 2018.

[39] Lingwen Gan and Steven H. Low. “Convex Relaxations and Linear Approximation for Optimal Power Flow in Multiphase Radial Networks”. In: *2014 Power Systems Computation Conference*. 2014, pp. 1–9. DOI: 10.1109/PSCC.2014.7038399.

[40] W. Wang et al. “Phase Identification in Electric Power Distribution Systems by Clustering of Smart Meter Data”. In: *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2016, pp. 259–265. DOI: 10.1109/ICMLA.2016.0046.

[41] A. Hoogsteyn et al. “Low Voltage Customer Phase Identification Methods Based on Smart Meter Data”. In: *Electric Power Systems Research* 212 (2022), p. 108524. DOI: 10.1016/j.epsr.2022.108524.

[42] Harish Doddi, Deepjyoti Deka, and Murti Salapaka. "Learning Partially Observed Meshed Distribution Grids". In: *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. 2020, pp. 1–6.

[43] Jiawei Zhang, Peng Wang, and Ning Zhang. "Distribution Network Admittance Matrix Estimation with Linear Regression". In: *IEEE Transactions on Power Systems* 36.5 (2021), pp. 4896–4899. DOI: 10.1109/TPWRS.2021.3090250.

[44] Emanuele Fabbiani et al. "Identification of AC Distribution Networks with Recursive Least Squares and Optimal Design of Experiment". In: *IEEE Transactions on Control Systems Technology* 30.4 (2022), pp. 1750–1757. DOI: 10.1109/TCST.2021.3116856.

[45] Jean-Sebastien Brouillon et al. "Bayesian Methods for the Identification of Distribution Networks". In: *2021 60th IEEE Conference on Decision and Control (CDC)*. 2021, pp. 3646–3651. DOI: 10.1109/CDC45484.2021.9683503.

[46] Youbo Liu et al. "Real-Time Topology Estimation for Active Distribution System Using Graph-Bank Tracking Bayesian Networks". In: *IEEE Transactions on Industrial Informatics* 19.4 (2023), pp. 6127–6136. DOI: 10.1109/TII.2022.3197311.

[47] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *International Conference on Learning Representations (ICLR)*. 2014.

[48] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: *International Conference on Machine Learning (ICML)*. 2014.

[49] Thomas N. Kipf and Max Welling. *Variational Graph Auto-Encoders*. arXiv:1611.07308. 2016.

[50] Martin Simonovsky and Nikos Komodakis. "GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders". In: *International Conference on Artificial Neural Networks*. 2018, pp. 412–422. DOI: 10.1007/978-3-030-01418-6_41.

[51] Ashish Vaswani et al. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems*. 2017, pp. 5998–6008.

Appendix A

Supplementary Physical Derivations

This appendix complements the physical assumptions introduced in Chapter 3. It only includes intermediate derivation steps that are too detailed for the main text, namely the simplified apparent-resistance estimate and the associated voltage-noise variance expression. Equations already introduced and discussed in the main chapters are not repeated unless they are required for the derivation.

A.1 Apparent Resistance Estimate for a Leaf Branch

The simplified resistance estimate used in the initial physical validation is applied to a branch feeding a leaf node. This condition is relevant because, for a leaf branch, the branch current is associated with the load at the terminal node. Therefore, the active power in the denominator can be interpreted directly as the active power absorbed downstream of that branch.

Considering a branch between an upstream node i and a downstream leaf node j, the complex voltage drop is

$$V_i - V_j = Z_{ij} I_{ij}. \tag{A.1}$$

The complex power at the downstream node is

$$S_j = P_j + jQ_j = V_j I_{ij}^*. \tag{A.2}$$

Substituting the branch-current expression into the complex-power equation gives

$$S_j = V_j \left(\frac{V_i - V_j}{Z_{ij}} \right)^*. \tag{A.3}$$

In the initial validation stage, three simplifying assumptions are imposed: voltage phase angles are not used explicitly, the branch is considered purely resistive, and reactive power is neglected. Therefore,

$$X_{ij} = 0, \qquad Q_j = 0. \tag{A.4}$$

Under these assumptions, and writing voltage magnitudes simply as V_i and V_j, the active-power relation reduces to

$$P_j \approx \frac{V_j(V_i - V_j)}{R_{ij}}. \tag{A.5}$$

Solving for the apparent resistance gives

$$\widehat{R}_{ij} = \frac{V_j(V_i - V_j)}{P_j}. \tag{A.6}$$

This expression is not a general impedance-estimation formula. It is a simplified apparent-resistance estimate valid under the initial assumptions used to isolate the effect of voltage-magnitude errors.

A.2 Extension to Non-Leaf Branches

For a branch that feeds a downstream subtree, the current flowing through the branch is not associated with a single nodal load. In that case, the active power in the denominator should be replaced by the aggregate downstream active power,

$$P_{\mathscr{D}(j)}(t) = \sum_{k \in \mathscr{D}(j)} P_k(t), \tag{A.7}$$

where $\mathscr{D}(j)$ is the set of nodes supplied downstream of node j. The corresponding approximation is

$$\widehat{R}_{ij}(t) \approx \frac{V_j(t)\,[V_i(t) - V_j(t)]}{P_{\mathscr{D}(j)}(t)}. \tag{A.8}$$

This extension is only approximate. It neglects downstream losses and assumes that the voltage drop is mainly explained by the aggregate downstream active power. For this reason, the variance validation is performed on a branch feeding a leaf node, where this ambiguity is minimised.

A.3 Voltage-Noise Propagation

The variance validation isolates the effect of voltage measurement noise on the apparent resistance estimate. The measured voltages are written as

$$\widetilde{V}_i = V_i + \varepsilon_i, \qquad \widetilde{V}_j = V_j + \varepsilon_j, \tag{A.9}$$

with independent zero-mean Gaussian errors,

$$\varepsilon_i, \varepsilon_j \sim \mathscr{N}(0, \sigma^2). \tag{A.10}$$

The active power is kept fixed in this validation step. Substituting the noisy voltages in the apparent-resistance estimate gives

$$\widehat{R}_{ij} = \frac{(V_j + \varepsilon_j)\left[(V_i + \varepsilon_i) - (V_j + \varepsilon_j)\right]}{P_j}. \tag{A.11}$$

Expanding the numerator gives

$$\widehat{R}_{ij} = \frac{1}{P_j}\left[V_j(V_i - V_j) + V_j\varepsilon_i - V_j\varepsilon_j + \varepsilon_j(V_i - V_j) + \varepsilon_i\varepsilon_j - \varepsilon_j^2\right]. \tag{A.12}$$

The deterministic component is the apparent resistance obtained without voltage noise. The random perturbation is therefore

$$\delta R_{ij} = \frac{V_j\varepsilon_i + (V_i - 2V_j)\varepsilon_j + \varepsilon_i\varepsilon_j - \varepsilon_j^2}{P_j}. \tag{A.13}$$

This expansion is the key intermediate step omitted from the main text.

A.4 Theoretical Variance

The perturbation term can be separated into a linear part and a quadratic part,

$$L = V_j\varepsilon_i + (V_i - 2V_j)\varepsilon_j, \qquad Q = \varepsilon_i\varepsilon_j - \varepsilon_j^2. \tag{A.14}$$

Using the independence of the voltage errors, the variance of the linear part is

$$Var(L) = V_j^2\sigma^2 + (V_i - 2V_j)^2\sigma^2 = \sigma^2\left(5V_j^2 + V_i^2 - 4V_iV_j\right). \tag{A.15}$$

For the quadratic part, the Gaussian moments give

$$Var(\varepsilon_i\varepsilon_j) = \sigma^4, \qquad Var(\varepsilon_j^2) = 2\sigma^4. \tag{A.16}$$

The covariance terms vanish because the errors are independent and zero mean. Hence,

$$Var(Q) = 3\sigma^4. \tag{A.17}$$

Combining the linear and quadratic terms yields the theoretical variance of the apparent-resistance estimate,

$$Var_{\text{theo}}\left(\widehat{R}_{ij}\right) = \frac{\sigma^2\left(5V_j^2 + V_i^2 - 4V_iV_j\right) + 3\sigma^4}{P_j^2}. \tag{A.18}$$

This expression is used only for the controlled validation scenario with a leaf branch, $X = 0$, $Q = 0$, fixed active power and voltage-magnitude noise.

A.5 Experimental Validation of the Variance Expression

The empirical validation compares the theoretical variance expression derived in Appendix A.4 with the variance obtained by Monte Carlo simulation. Starting from reference voltages computed by the power-flow algorithm, independent Gaussian noise is added to the voltage magnitudes and the apparent resistance is recalculated for each realisation.

The noise standard deviation is

$$\sigma = \frac{0.005 \times 230}{3} \text{ V}. \tag{A.19}$$

This corresponds to an interval of approximately $\pm 0.5\%$ around the nominal voltage when interpreted as a three-standard-deviation range.

For each Monte Carlo realisation m, the resistance estimate is computed from the noisy voltages. The empirical variance is then obtained from the resulting set of resistance values. Agreement between the empirical and theoretical variances supports the consistency of the simplified noise-propagation model under the initial assumptions.

A.6 Example of a Rejected Voltage-Drop Scenario

During the synthetic data generation process, some randomly generated combinations of topology, branch parameters and active-power profiles produced voltage behaviour that was not considered suitable for the final P/V dataset. These cases were not used for model training because their voltage drops were too large to represent stable and physically plausible low-voltage operating conditions.

This diagnostic was introduced during the preliminary validation of the power-flow pipeline. The objective was to distinguish cases in which the generated voltages remained compatible with the assumptions of the model from cases in which the voltage variations became excessive. In the admissible case discussed in Section 3.4.2, the voltage drops remain within a stable range and the Forward-Backward Sweep algorithm converges reliably. By contrast, in the rejected case, the voltage differences between nodes become too large and the iterative calculation may become numerically unstable.

The rejected scenario is associated with large voltage variations across the feeder. In the preliminary analysis, one such case produced a voltage of 230.0000 V at the reference node and 83.5992 V at another node. The corresponding voltage difference is

$$\Delta V = |230.0000 - 83.5992| = 146.4008 \text{ V}. \tag{A.20}$$

This corresponds to approximately 63.7% of the nominal voltage. Such a voltage drop is far above the admissible variation assumed for stable operation in the simplified validation setting. Under these conditions, the power-flow equations may become difficult to satisfy iteratively,

and the resulting numerical values should not be treated as physically meaningful samples for topology-inference training.

Figure A.1 shows the diagnostic plot associated with this rejected case. The plot represents the error in the estimated resistance when phase effects are considered, as a function of the ratio X_{ij}/R_{ij}, for the analysed leaf branch. Unlike the admissible case, the behaviour is irregular and exhibits comparatively large errors for several values of the X/R ratio. This is consistent with an operating point in which the voltage behaviour is no longer suitable for the simplified assumptions used in the preliminary validation.



Figure A.1: Example of a rejected voltage-drop scenario. The diagnostic shows the resistance-estimation error as a function of the X_{ij}/R_{ij} ratio for a case in which the voltage behaviour is not suitable for inclusion in the final synthetic P/V dataset.

This rejected case illustrates why the synthetic data generation procedure cannot retain every randomly generated topology and operating condition. Cases with excessive voltage drops may violate the assumptions of the simplified model, lead to numerical instability in the Forward-Backward Sweep calculation, and produce values with limited physical meaning. For this reason, only scenarios with convergent power-flow simulations and physically plausible voltage magnitudes are retained in the final dataset.

A.7 Summary

This appendix clarifies the intermediate mathematical steps behind the physical validation used in the methodology. The derivation shows that the apparent-resistance estimate is valid only under a controlled set of assumptions: leaf-branch analysis, voltage magnitudes only, purely resistive branch, neglected reactive power and fixed active power during the voltage-noise test.

The appendix also explains why the theoretical variance expression contains both a σ^2 term and a σ^4 term. The first comes from the linear propagation of voltage noise, while the second comes from the quadratic noise terms introduced by multiplying noisy voltage quantities.

Appendix B

Supplementary Initial Fixed-Size VAE Diagnostics

This appendix presents supplementary diagnostics associated with the initial fixed-size VAE baseline discussed in Section 4.1. The main chapter reports the central conclusion: the initial fixed-size VAE achieves high seen-topology reconstruction, but its generalisation to unseen radial topologies is weak.

The purpose of this appendix is to analyse whether this limitation could be explained by the organisation of the latent space. Three diagnostics were performed. The first used nearest neighbours in the full latent space. The second selected the nearest training point in a two-dimensional PCA projection. The third forced the unseen case towards the closest latent family centroid. Finally, a latent-dimension sweep is included to verify whether the limitation was caused by insufficient latent capacity.

B.1 Latent-Space kNN Forcing

The first diagnostic tested whether nearest neighbours in the original latent space corresponded to structurally similar topologies. An unseen case was encoded by the VAE encoder, producing a latent mean vector μ_{unseen}. The nearest training samples were then selected using Euclidean distance in the full latent space. The PCA projection was used only for visualisation and was not used to select the neighbours.

The objective was to test whether latent-space proximity was aligned with edge-level similarity. If the latent representation were organised according to topology, the nearest neighbours of an unseen sample would be expected to share most of its branches.

The diagnostic was analysed over three cases. In the original case, the selected kNN family was family 46 and the unseen edge accuracy remained 0.556 before and after the diagnostic. In the first repetition, the selected family was family 53 and the unseen edge accuracy remained 0.667. In the second repetition, the selected family was family 172 and the unseen edge accuracy changed from 0.434 to 0.444.

Table B.1 summarises the mean results of the three cases.

Table B.1: Mean results of the latent-space kNN forcing diagnostic.

Quantity	**Value**
Number of unseen cases	3
Seen edge accuracy	1.000
Mean unseen edge accuracy before forcing	0.552
Mean unseen edge accuracy after forcing	0.556
Mean variation in edge accuracy	+0.004
Exact recovery after forcing	0 out of 3

The mean unseen edge accuracy changed only from 0.552 to 0.556, corresponding to an average improvement of 0.004. This improvement is negligible and does not indicate that kNN forcing provides a reliable correction mechanism for unseen topology reconstruction. The selected neighbours sometimes share part of the edge set with the unseen topology, but this partial overlap is not sufficient to recover the complete radial structure.

Figures B.1, B.2 and B.3 show the PCA projections used to visualise the three kNN forcing cases. In these figures, the unseen sample, its selected neighbours and the associated training families are represented in two dimensions. These projections should be interpreted only as visual diagnostics, because the actual neighbour search was performed in the full latent space.

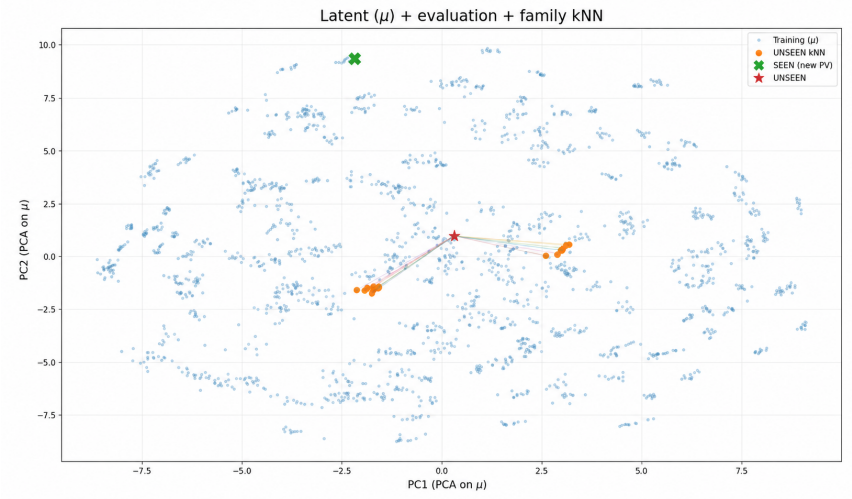


Figure B.1: Latent-space kNN forcing diagnostic for the original case. The PCA projection is used only to visualise the unseen sample, selected neighbours and neighbouring family.

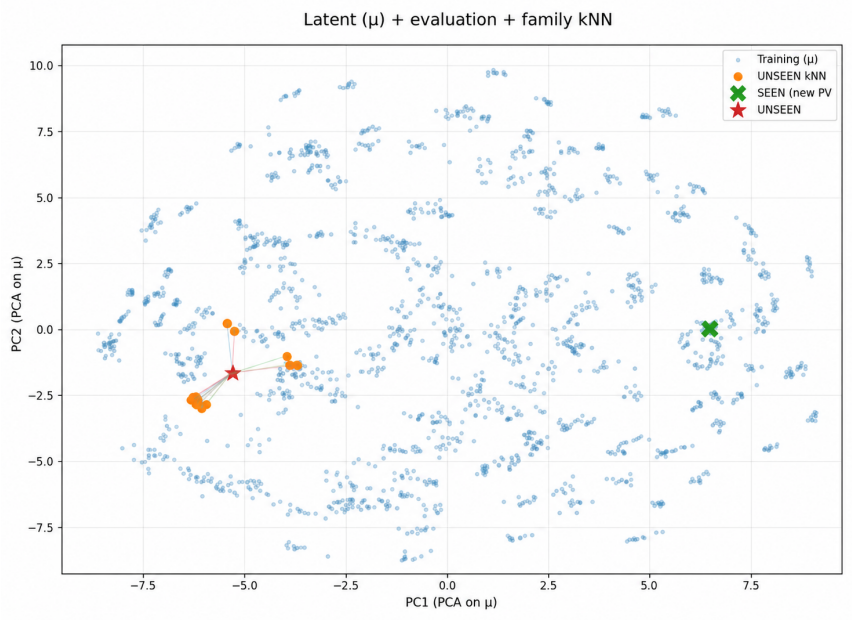


Figure B.2: Latent-space kNN forcing diagnostic for repetition 1. The unseen sample is assigned to a neighbouring family in the full latent space, but the recovered topology remains only partially correct.

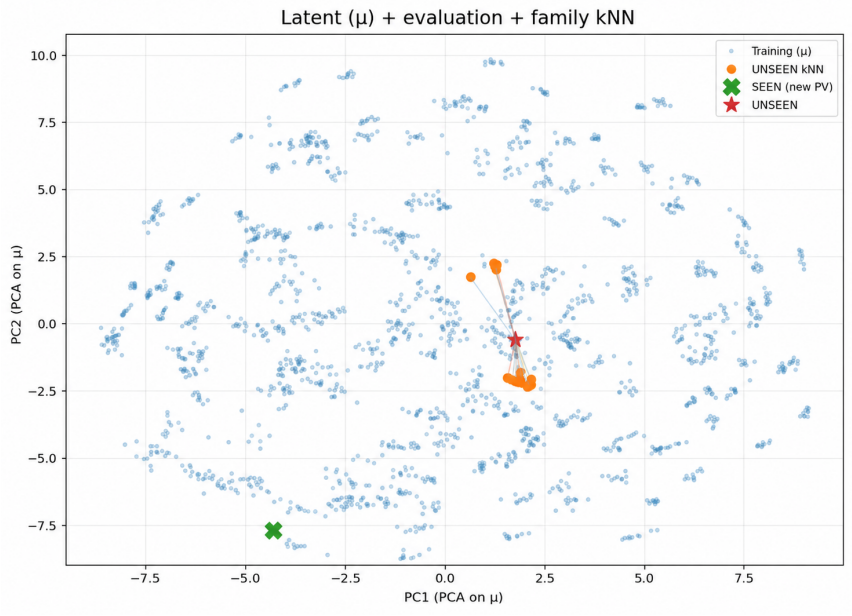


Figure B.3: Latent-space kNN forcing diagnostic for repetition 2. The projection illustrates that latent proximity may correspond to partial edge overlap, but not to exact topology recovery.

Overall, the kNN diagnostic shows that the latent space of the initial fixed-size VAE contains some local structure, but this structure is not sufficiently aligned with exact topological similarity. Nearby latent samples may share some branches with the unseen case, but they do not provide a reliable basis for reconstructing the full radial topology.

B.2 PCA-Based Latent-Space Forcing

The second diagnostic tested a more direct use of the PCA projection. In this experiment, the latent mean vectors of the training records were first projected into a two-dimensional PCA space. The unseen latent vector was projected using the same transformation. The closest training record was then selected directly in this two-dimensional PCA projection.

The unseen case was first reconstructed normally, producing the metrics before forcing. Then, the unseen latent vector was replaced by the latent vector of the selected training record. This forced vector was passed through the decoder, and maximum spanning tree projection was applied to obtain a valid radial topology.

The diagnostic was repeated for three unseen cases. In Case 1, the PCA distance to the selected point was 0.549, and forcing improved edge accuracy from 0.333 to 0.556. In Case 2, the PCA distance was 0.258, but forcing reduced edge accuracy from 0.778 to 0.556. In Case 3, the PCA distance was only 0.039, but forcing reduced edge accuracy from 0.333 to 0.222.

Table B.2 summarises the mean effect of PCA-based forcing.

Table B.2: Mean results of the PCA-based latent-space forcing diagnostic.

Quantity	**Value**
Number of unseen cases	3
Seen edge accuracy	1.000
Mean unseen edge accuracy before forcing	0.481
Mean unseen edge accuracy after forcing	0.444
Mean variation in edge accuracy	-0.037
Mean edge error before forcing	4.667
Mean edge error after forcing	5.000
Mean variation in edge error	+0.333
Exact recovery after forcing	0 out of 3

The average edge accuracy decreased from 0.481 before forcing to 0.444 after forcing. The mean edge error also increased from 4.667 to 5.000. Therefore, selecting the nearest point in the PCA projection did not improve unseen reconstruction on average.

The individual cases show why this result is important. PCA forcing improved Case 1, but worsened Cases 2 and 3. Case 3 is particularly relevant because the PCA distance was very small, only 0.039, yet the forced topology achieved only 0.222 edge accuracy. This shows that visual

proximity in the PCA projection can be misleading: two points may be close in the projected latent space while corresponding to substantially different radial structures.

Figures B.4, B.5 and B.6 show the three PCA-based forcing cases.

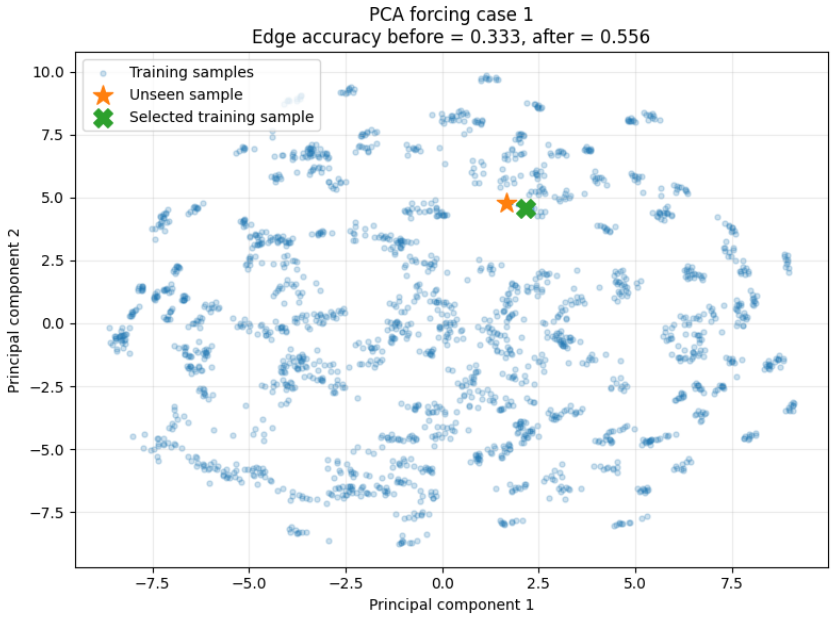


Figure B.4: PCA-based latent-space forcing, case 1. The selected training point improves partial edge recovery, but exact topology recovery is still not achieved.



Figure B.5: PCA-based latent-space forcing, case 2. The selected point is close in the PCA projection, but forcing reduces the edge accuracy relative to the original unseen reconstruction.

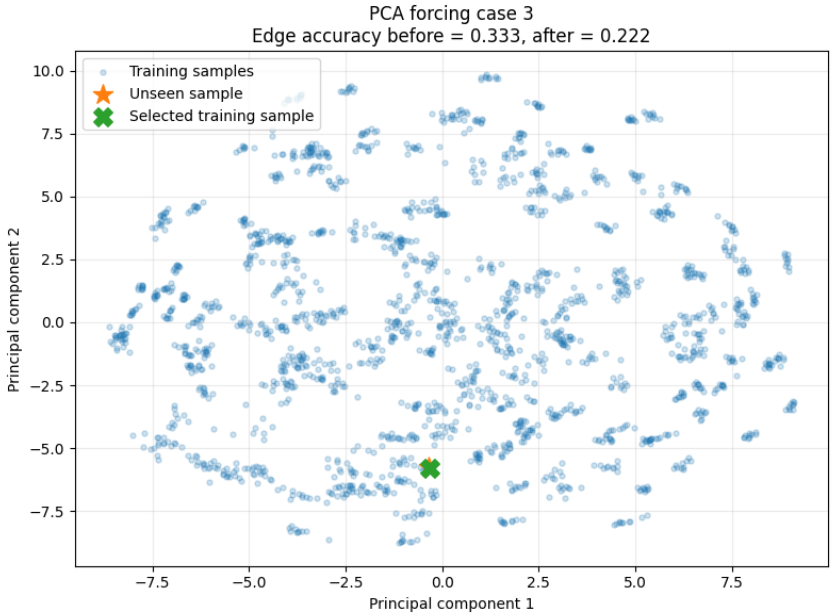


Figure B.6: PCA-based latent-space forcing, case 3. Despite the very small PCA distance, the forced topology has low edge accuracy, showing that PCA proximity is not a reliable proxy for graph similarity.

The PCA-based diagnostic confirms that two-dimensional visual proximity should not be interpreted as evidence of topological similarity. The PCA projection is useful for inspection, but forcing an unseen case towards the nearest projected training point does not provide a reliable reconstruction mechanism.

B.3 Nearest Latent-Family-Mean Forcing

The third diagnostic tested whether a more stable family-level latent reference could improve unseen reconstruction. Instead of selecting a single nearest neighbour or a nearest PCA-projected point, this experiment computed the mean latent vector of each training family.

For a training family g, the latent centroid is defined as

$$\bar{\mu}_g = \frac{1}{N_g} \sum_{i \in g} \mu_i, \tag{B.1}$$

where N_g is the number of records in family g and μ_i is the latent mean vector of record i.

For each unseen case, the latent mean vector μ_{unseen} was compared with all family centroids in the full latent space. The closest family was selected according to

$$g^* = \arg\min_g \|\mu_{unseen} - \bar{\mu}_g\|_2 . \tag{B.2}$$

The unseen topology was first reconstructed normally using μ_{unseen}. Then, the unseen latent vector was replaced by the selected family centroid,

$$\mu_{forced} = \bar{\mu}_{g^*}. \tag{B.3}$$

The forced latent vector was passed through the decoder and the resulting edge-score matrix was projected into a radial topology using maximum spanning tree projection.

This experiment was performed over 20 unseen cases. Instead of reporting all individual cases, Table B.3 summarises the mean behaviour of the diagnostic.

Table B.3: Mean results of the nearest latent-family-mean forcing diagnostic.

Quantity	**Value**
Number of unseen cases	20
Seen edge accuracy	1.000
Mean centroid distance	4.867
Mean edge accuracy before forcing	0.528
Mean edge accuracy after forcing	0.506
Mean edge accuracy variation	-0.022
Mean edge error before forcing	4.250
Mean edge error after forcing	4.450
Mean edge error variation	+0.200
Exact recovery before forcing	0 out of 20
Exact recovery after forcing	0 out of 20

The centroid-based diagnostic reinforces the previous conclusions. On average, edge accuracy decreased from 0.528 to 0.506, while edge error increased from 4.250 to 4.450. Exact topology recovery remained zero both before and after forcing.

The detailed case count also supports this interpretation: 4 out of 20 cases improved, 9 remained unchanged, and 7 worsened after forcing. Therefore, although the closest family centroid can help in some individual examples, it does not provide a reliable correction mechanism on average.

This result shows that the limitation is not only caused by selecting an individual neighbour or by information loss in the PCA projection. Even when each family is represented by its mean latent vector, the closest latent family does not consistently correspond to the correct unseen topology.

B.4 Latent-Dimension Sweep

A latent-dimension sweep was performed to test whether the initial fixed-size VAE was limited mainly by insufficient latent capacity. If the latent bottleneck were the dominant problem, increasing the latent dimension would be expected to improve unseen reconstruction.

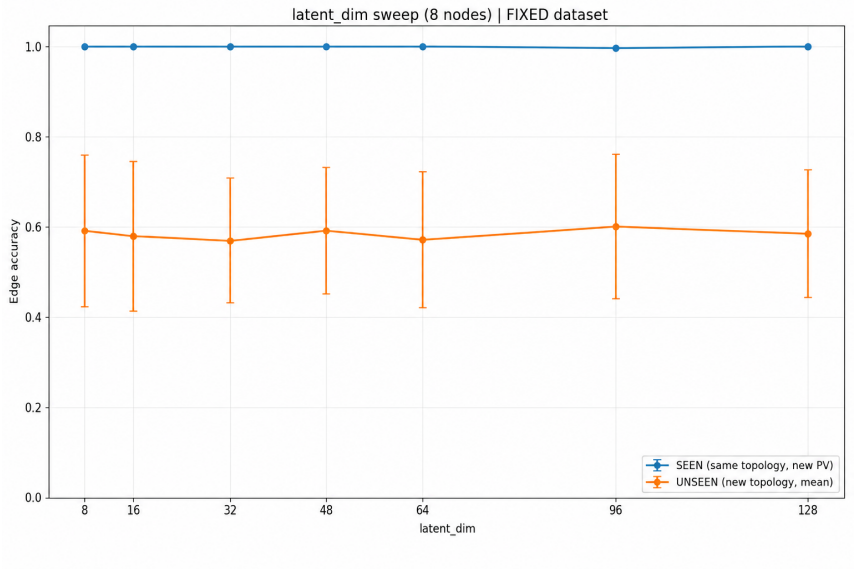


Figure B.7: Latent-dimension sweep with a fixed dataset, evaluated using edge accuracy.

Figure B.7 shows that increasing the latent dimension does not produce a systematic improvement in unseen topology inference. This indicates that the limitation is not simply caused by a latent space that is too small. Instead, the issue is related to how the initial fixed-size VAE organises and decodes graph structure.

B.5 Summary

The experiments in this appendix support the interpretation presented in Section 4.1. The initial fixed-size VAE can reconstruct seen topologies with high accuracy, but its latent space is not organised according to reliable topological similarity.

The kNN forcing diagnostic produced only a negligible mean improvement, from 0.552 to 0.556. The PCA-based forcing diagnostic reduced mean edge accuracy from 0.481 to 0.444. The nearest latent-family-mean forcing diagnostic also reduced mean edge accuracy, from 0.528 to 0.506, over 20 unseen cases. In all diagnostics, exact topology recovery remained zero after forcing.

The latent-dimension sweep further shows that the problem is not solved by increasing latent capacity. Overall, these results indicate that the limitation of the initial fixed-size VAE is structural. This motivated the transition to convolutional processing, structured input representations and variable-size models in the main results chapter.

Appendix C

Supplementary Regularisation and Architecture Experiments

This appendix reports supplementary regularisation and architecture experiments associated with the initial fixed-size VAE regime discussed in Section 4.1. These tests are kept in the appendix because they do not change the main results, but they are important for interpreting why the work moved beyond the original fixed-size VAE formulation.

The objective of these diagnostics was to test whether the seen–unseen generalisation gap could be solved by tuning the original model family. Three types of modifications were considered: stronger KL regularisation, classical regularisation through dropout and weight decay, and architecture changes within the same fixed-size VAE regime.

Table C.1 summarises the purpose of each diagnostic.

Table C.1: Overview of supplementary regularisation and architecture diagnostics.

Diagnostic	What was tested	Purpose
KL regularisation	Sweep of the maximum KL-divergence weight in the VAE objective	Test whether stronger latent regularisation could organise the latent space more effectively.
Dropout and weight decay	Sweep of dropout and weight-decay combinations	Test whether the *seen–unseen* gap was mainly caused by classical overfitting.
Architecture sweep	Changes in width, depth, activation functions, normalisation and decoder structure	Test whether the limitation could be solved by modifying the fixed-size VAE architecture.

C.1 KL Regularisation

The first diagnostic evaluated whether stronger variational regularisation could improve generalisation to *unseen* topologies. In the VAE objective, the KL-divergence term constrains the approxi-

mate latent posterior towards the prior distribution. Its contribution to the total loss can be written as

$$\mathscr{L} = \mathscr{L}_{\text{rec}} + \beta \, \mathscr{L}_{\text{KL}}, \tag{C.1}$$

where $\mathscr{L}_{\text{rec}}$ is the adjacency reconstruction loss and $\mathscr{L}_{\text{KL}}$ is the KL-divergence regularisation term. Increasing β can promote a smoother and more organised latent space, but excessive regularisation may suppress topology-related information and degrade reconstruction.

To isolate this effect, the dataset, architecture and remaining training settings were kept fixed while only the maximum KL weight was varied. The experiment used 8-node networks, $T = 800$ timestamps, 200 training topologies, 20 records per topology and a latent dimension of 48. The model was trained for 200 epochs, with the KL weight progressively increased during a 150-epoch warm-up period. The high-range sweep considered

$$\beta_{\max} \in \{0.01, 0.1, 1, 3, 10\}. \tag{C.2}$$

In Figure C.1, the horizontal axis represents the maximum KL weight, while the two curves show the mean *seen* and *unseen* edge accuracy. The error bars summarise the variation between repeated runs.

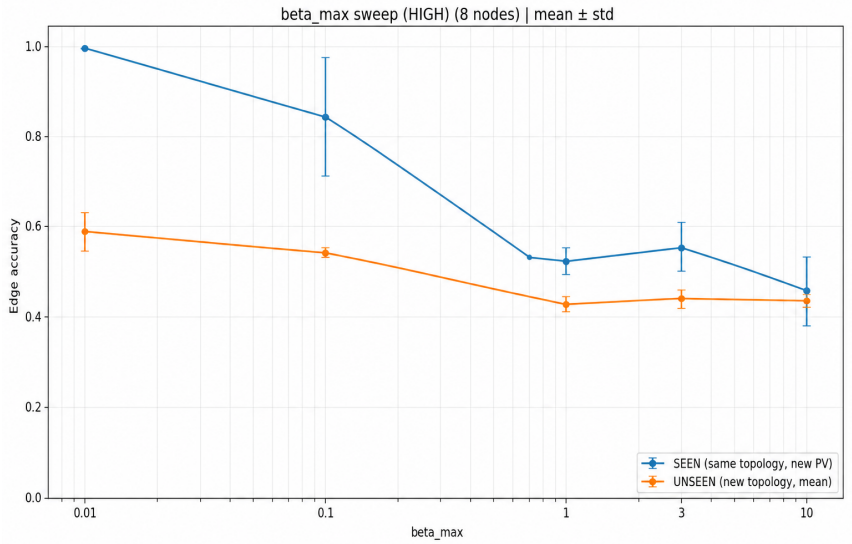


Figure C.1: High-range sweep of the maximum KL-divergence weight. The curves report the mean *seen* and *unseen* edge accuracy, while the error bars summarise the variation between repeated runs.

Figure C.1 shows that increasing the KL weight does not improve reconstruction of *unseen* topologies. At low values within the tested range, the model preserves high *seen* performance, while the *unseen* accuracy remains substantially lower. As $\beta_{\max}$ increases, the *seen* performance progressively degrades, but no corresponding improvement is observed for *unseen* cases.

This behaviour is consistent with over-regularisation: a very strong KL term forces the latent posterior closer to the prior and reduces the information available to the decoder. The experiment therefore indicates that the generalisation limitation is not caused simply by insufficient latent regularisation. Stronger KL regularisation does not organise the latent space into a reliably topology-aware representation and instead reduces reconstruction quality.

C.2 Dropout and Weight Decay

The second diagnostic tested classical regularisation through dropout and weight decay. The purpose was to evaluate whether the weak *unseen* performance could be explained mainly by overfitting to the training topologies. If this were the case, an appropriate level of regularisation would be expected to reduce the *seen–unseen* generalisation gap.

To isolate the effect of regularisation, the dataset, architecture and remaining training settings were kept fixed while only dropout and weight decay were varied. The experiment used 8-node networks, $T = 800$ timestamps, 200 training topologies and 20 records per topology. The latent dimension was 48, and the model was trained for 200 epochs using a batch size of 64, a learning rate of 7×10^{-4}, and $\beta_{\max} = 10^{-4}$ with a warm-up period of 150 epochs. Each regularisation configuration was evaluated using three random seeds.

The tested dropout values were

$$p_{\text{dropout}} \in \{0.00, 0.05, 0.10, 0.20, 0.35\},$$

and the tested weight-decay values were

$$\lambda_{\text{wd}} \in \{0, 10^{-5}, 10^{-4}, 3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}\}.$$

In Figure C.2, the rows correspond to the dropout values and the columns correspond to the weight-decay values. Each cell reports the mean *unseen* edge accuracy obtained across the three random seeds for the corresponding combination. Larger values indicate better partial reconstruction of the ground-truth topology.

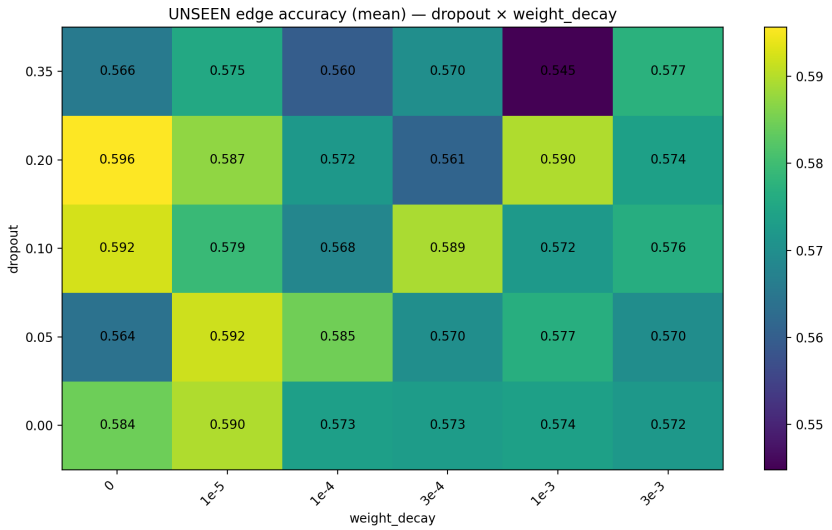


Figure C.2: Unseen edge accuracy for the dropout and weight-decay sweep. Rows represent dropout values, columns represent weight-decay values, and each cell reports the mean edge accuracy across three random seeds.

Figure C.2 shows moderate local variations across the tested configurations. The mean *unseen* edge accuracy generally remains within a similar range, although some combinations produce small improvements or degradations. No configuration produces a consistent transition to strong *unseen* topology reconstruction.

These results indicate that the limitation of the initial fixed-size VAE cannot be explained only as conventional overfitting. Dropout and weight decay affect performance locally, but they do not provide the relational structure required to map new measurement patterns reliably to new radial topologies.

C.3 Architecture Sweep

The third diagnostic evaluated whether the *seen–unseen* generalisation gap was mainly caused by the specific architecture selected for the initial fixed-size VAE. The dataset, raw P/V input representation, loss function, training procedure and evaluation metric were kept fixed, while different components of the encoder and decoder were modified. Each configuration was evaluated using three random seeds to reduce the influence of parameter initialisation, data shuffling and stochastic latent sampling.

The configuration labels in Figure C.3 are organised into four groups. The A configurations change the hidden-layer widths while retaining the same general depth. The B configurations change the number of hidden layers and their corresponding dimensions. The C configurations evaluate different normalisation and activation choices, including GELU, ReLU and SiLU. Finally, the D configurations modify the decoder capacity and structure, including compact repeated-width

decoders and a decoder that mirrors the encoder. The numbers included in each label indicate the dimensions of the corresponding hidden layers.

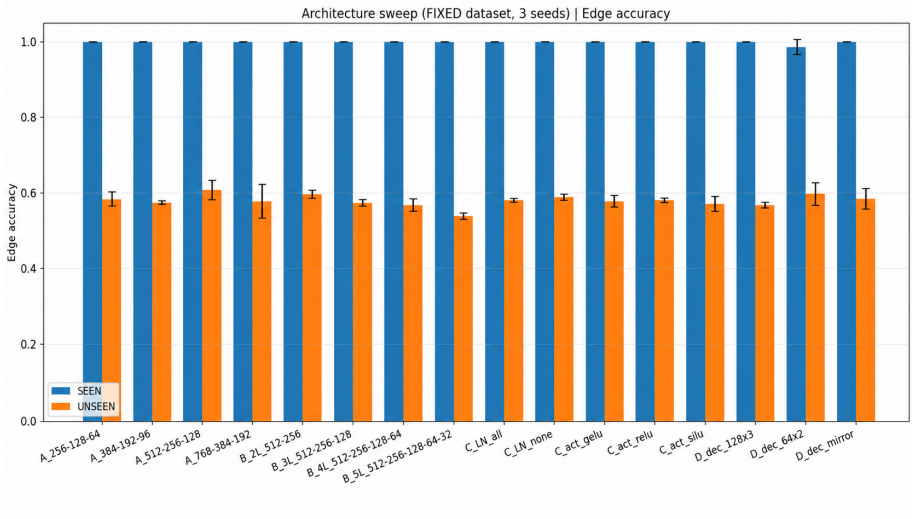


Figure C.3: Architecture sweep within the initial fixed-size VAE regime. The bars report the mean *seen* and *unseen* edge accuracy across three random seeds, while the error bars summarise the variation between runs.

Figure C.3 shows that the *seen* edge accuracy remains close to one for almost all configurations. By contrast, the *unseen* edge accuracy remains substantially lower, generally varying only within a limited range across the tested architectures. Some changes produce small local improvements or degradations, but no configuration consistently removes the generalisation gap.

These results indicate that the limitation was not caused only by the width, depth, activation function, normalisation strategy or decoder capacity of the original model. Architectural tuning within the same fixed-size VAE formulation did not provide the relational inductive bias required for reliable *unseen* topology inference. This observation motivated the subsequent transition to convolutional processing, structured matrix-based representations and variable-size graph architectures in Chapter 4.

C.4 Summary

The results in this appendix show that KL tuning, dropout, weight decay and architecture changes inside the initial fixed-size VAE family do not eliminate the seen–unseen generalisation gap. These diagnostics support the methodological transition made in Chapter 4: rather than continuing to tune the same fixed-size formulation, the work moved towards convolutional encoders, physically structured input representations and variable-size architectures.

Together with the latent-space diagnostics reported in Appendix B, these experiments show that the limitation of the initial VAE was not caused by a single hyperparameter choice. The problem was more fundamental: the original fixed-size VAE did not organise the latent representation in a way that reliably supported unseen topology reconstruction.

Appendix D

Supplementary Fixed-Size Convolutional VAE Diagnostics

This appendix provides supplementary diagnostics for the fixed-size convolutional VAE experiments discussed in Chapter 4. The main chapter reports the central results: the convolutional architecture improves the fixed-size setting, and the effect of increasing timestamps and training topologies is analysed directly through the main edge-accuracy plots. The additional results shown here are kept only as supporting evidence.

D.1 Temporal-Budget Allocation

A fixed-budget temporal experiment was performed to evaluate how the available information per topology should be distributed between longer trajectories and a larger number of shorter operating realisations. This test complements the timestamp sweep reported in Chapter 4. In this experiment, the total temporal budget assigned to each topology was kept constant at approximately $B = n_{\text{records}} T = 8000$ timestamped P/V observations. The tested configurations therefore changed only how this budget was distributed. For example, the $10/T800$ configuration uses 10 long records with 800 timestamps each, whereas $320/T25$ uses 320 shorter records with 25 timestamps each. This comparison separates the effect of temporal depth from the effect of observing a larger number of operating realisations.

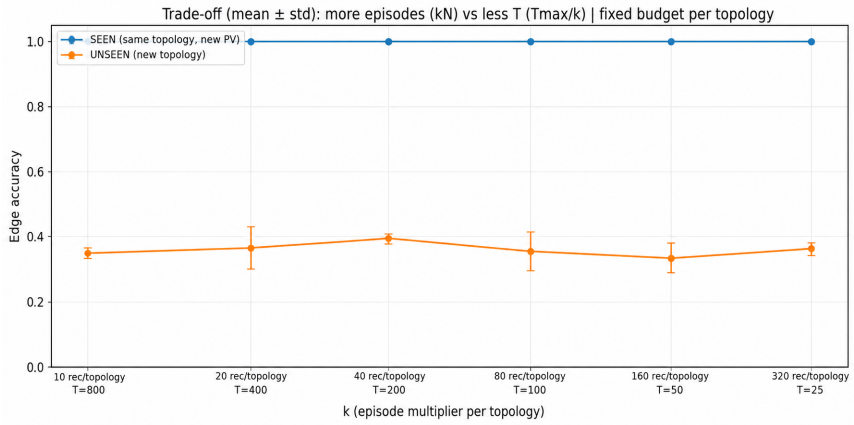


Figure D.1: Fixed-budget trade-off between longer trajectories and more records per topology. The result evaluates whether performance is more affected by temporal depth or by the number of operating realisations.

The result indicates that temporal information is useful, but the allocation of the temporal budget does not by itself remove the limitations of the fixed-size setting. Very short windows may reduce useful operating variability, while too few records per topology reduce the diversity of observed conditions.

D.2 Unseen Behaviour Under an Extended Temporal Trade-Off

The extended temporal-budget experiment focuses on the unseen setting, which is the most relevant case for topology generalisation. This result is shown separately because the seen setting can remain strong even when unseen reconstruction remains limited.

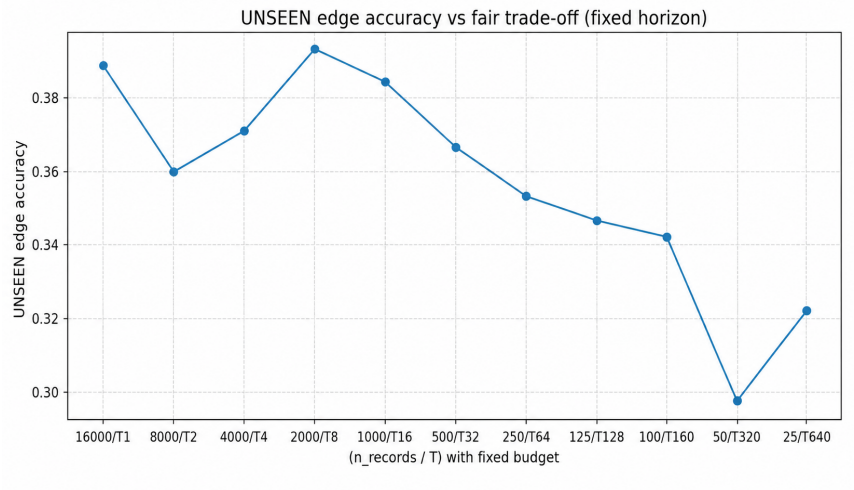


Figure D.2: Extended fixed-budget trade-off for unseen edge accuracy. The result shows that changing the temporal allocation affects performance, but does not fully solve unseen topology reconstruction.

This result supports the interpretation used in Chapter 4: increasing temporal information helps only up to a point. The model also requires sufficient structural diversity and an architecture capable of exploiting topology-related patterns.

D.3 Random and Locally Perturbed Unseen Topologies

A complementary structural diagnostic compared fully random *unseen* topologies with locally perturbed *unseen* topologies. Each locally perturbed case was derived from a topology represented during training and differed from it in two undirected branches, while preserving the same number of nodes and a valid connected radial structure. The purpose was to test whether a new topology that remains structurally close to a *seen* topology is easier to infer than a fully random *unseen* topology.

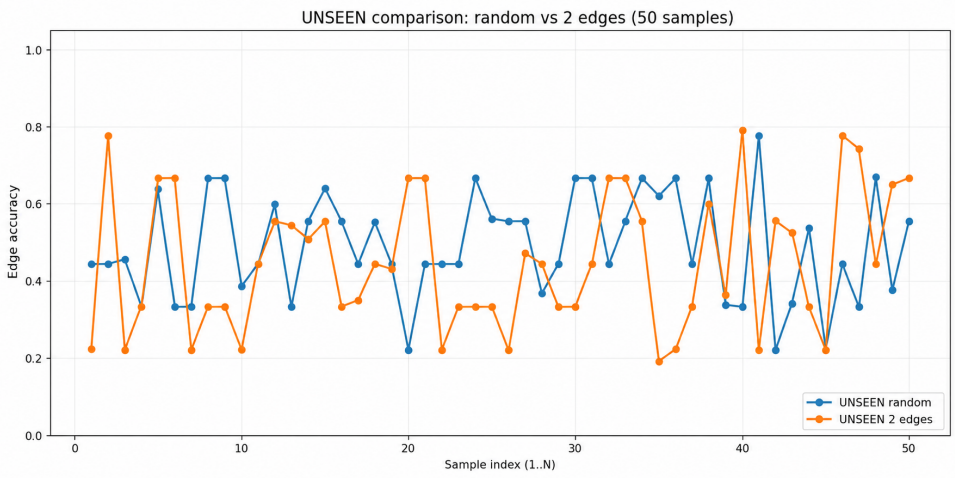


Figure D.3: Comparison between fully random unseen topologies and two-edge-perturbed unseen topologies. The result shows that being close to a seen topology does not necessarily make unseen reconstruction reliable.

The absence of a clear advantage for locally perturbed topologies suggests that unseen inference is not simply a local interpolation problem around the training database. This reinforces the motivation for moving beyond fixed-size diagnostics towards representation studies and variable-size models.

D.4 Summary

The supplementary diagnostics are consistent with the interpretation presented in Chapter 4 and further support the transition towards matrix-based representations and variable-size architectures.

Appendix E

Supplementary Input-Representation and Variable-Size Results

This appendix complements the input-representation and variable-size results presented in Chapter 4. The main chapter reports the numerical performance of the raw P/V, correlation and sensitivity representations. This appendix summarises the role of each representation, illustrates the block structure of the full correlation matrices, and provides additional context for the variable-size and Transformer-inspired results.

E.1 Input-Representation Summary

The experiments considered three main input families: raw P/V trajectories, correlation matrices and sensitivity matrices. Their roles are summarised in Table E.1.

Table E.1: Summary of the input representations evaluated in the topology inference experiments.

Representation	Main advantage	Main limitation
Raw P/V	Uses the original active-power and voltage trajectories with minimal preprocessing.	The model must learn node-to-node relationships implicitly from temporal data.
Correlation	Encodes statistical dependencies between electrical quantities.	Correlations can be affected by common load patterns and do not necessarily reflect topology alone.
Sensitivity	Encodes the approximate response of voltage variations to active-power variations.	Depends strongly on voltage quality and can degrade under voltage measurement noise.

The results in Chapter 4 show that the sensitivity representation is the most consistent option across *training* and *held-out* graph sizes. This is because it provides a physically meaningful matrix where each entry is related to how active-power variations at one node are reflected in voltage variations elsewhere in the feeder.

Correlation matrices also provide useful information, especially through the cross-variable relation between active power and voltage. However, the full correlation matrix contains several blocks with different physical interpretations. For this reason, representative examples of these matrices are shown in the next section.

E.2 Examples of Full Correlation Matrices

This section provides representative examples of the full correlation matrices used in the input-representation study. These examples are included as visual diagnostics to illustrate the block structure of the correlation representation for different network sizes.

As defined in Equation 4.4, the full correlation matrix is organised into voltage-voltage, voltage-power, power-voltage and power-power blocks. For a network with N nodes, the full correlation representation has dimension $2N \times 2N$, because it combines voltage-magnitude and active-power variables. The first N rows and columns correspond to voltage variables, while the last N rows and columns correspond to active-power variables.

The examples shown below correspond to the same training-topology quantities used in the fixed-size experiments: 2000 training topologies for 10-node networks, and 5000 training topologies for both 20-node and 30-node networks.

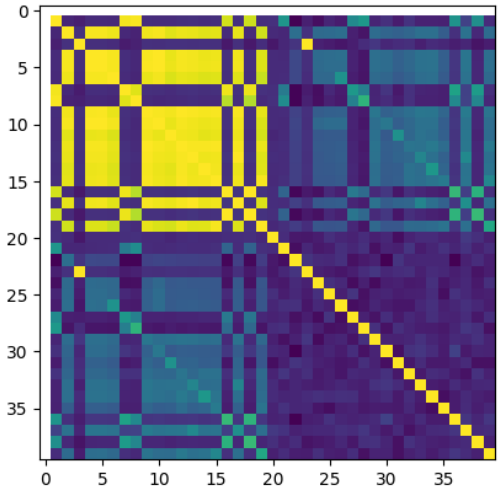


Figure E.1: Representative full correlation matrix for a 10-node network using 2000 training topologies. The matrix has dimension 20×20.

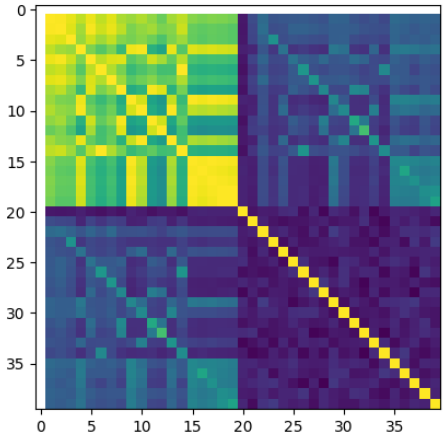


Figure E.2: Representative full correlation matrix for a 20-node network using 5000 training topologies. The matrix has dimension 40×40.

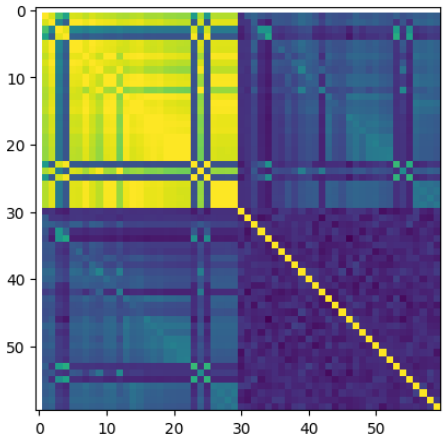


Figure E.3: Representative full correlation matrix for a 30-node network using 5000 training topologies. The matrix has dimension 60×60.

These examples illustrate the structured nature of the full correlation representation. The voltage-voltage block tends to show strong common correlations, reflecting the shared electrical behaviour of nodes within the same feeder. The active-power block is generally less structured, since residential active-power profiles can vary more independently. The cross-variable blocks are particularly relevant for topology inference because they relate active-power variations to voltage responses, which is the physical relation exploited by both PV-only correlation and sensitivity-based representations.

The visual block structure also helps explain why the full correlation matrix is informative but not always the most compact representation. Some information in the voltage-voltage block may be redundant because all voltage magnitudes are affected by common feeder-level behaviour. By contrast, the power-voltage interaction blocks are more directly related to the topology-inference task, since they encode how active-power variations are associated with voltage changes across the network.

E.3 Variable-Size Generalisation

The transition to variable-size models was necessary because fixed-size architectures require one model for each network dimension. The variable-size VAE avoids this limitation by using shared node encoding and masking-aware aggregation.

Table E.2 summarises the graph-size split used in the variable-size experiments.

Table E.2: Graph-size split used in the variable-size VAE experiments.

Set	**Graph sizes**
Training sizes	$N \in \{5, 10, 15, 20, 25, 30\}$
Held-out inference sizes	$N \in \{8, 13, 23, 50, 80\}$

The intermediate *held-out* sizes, namely 8, 13 and 23 nodes, evaluate interpolation between or near training dimensions. The larger *held-out* sizes, namely 50 and 80 nodes, evaluate extrapolation beyond the training range.

The main results show that sensitivity inputs generalise well to intermediate *held-out* sizes and remain the strongest representation for larger *held-out* networks. However, the drop observed at 80 nodes indicates that extrapolation to much larger feeders remains challenging.

E.4 Transformer-Inspired Extension

The Transformer-inspired variable-size VAE was introduced to improve the modelling of global node-to-node interactions. Instead of compressing node embeddings immediately through pooling, the Transformer encoder allows each node representation to be contextualised by the other nodes before graph-level aggregation.

This extension improved the behaviour on larger *held-out* graphs. The improvement was especially relevant for 50-node and 80-node networks, where global interactions are more difficult to capture with simple masked pooling.

A further training configuration extended the training range up to 50 nodes. This improved extrapolation to 80-node networks, showing that the model benefits from being exposed to larger graph structures during training.

E.5 Summary

The supplementary experiments are consistent with the conclusions of Chapter 4, confirming the advantages of sensitivity-based representations and the improved scalability of the Transformer-inspired variable-size architecture.

Appendix F

Supplementary Robustness Results

This appendix presents supplementary robustness results that complement the robustness analysis in Chapter 4. The main chapter reports the sensitivity-based robustness curves and the Transformer-inspired edge-accuracy curves. Here, only two additional views are retained: the correlation-based robustness result and the Transformer-inspired edge-error result.

F.1 Correlation-Based Robustness

The correlation-based representation was evaluated under measurement errors affecting active power, voltage, or both. Two settings were considered: inference-only, where the model was trained on clean data and errors were introduced only during evaluation; and database-and-inference, where the same perturbations were applied during both training and evaluation. The relative error window ranged from 0% to 1%. This experiment complements the sensitivity-based robustness analysis reported in Chapter 4.

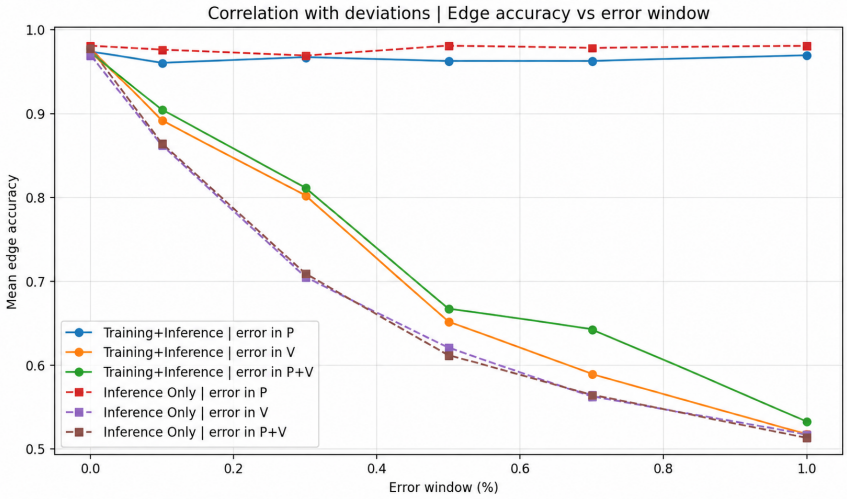


Figure F.1: Robustness of the correlation-based representation under relative measurement errors between 0% and 1%. The six curves compare perturbations affecting P, V, and both P and V, applied either only during inference or during both training and inference.

The result follows the same qualitative trend observed for the sensitivity representation. Active-power perturbations have limited impact in the tested range, whereas voltage perturbations cause a stronger degradation. This confirms that voltage quality is a central factor for topology inference when the input representation depends on voltage-derived relationships.

F.2 Transformer-Inspired Model: Edge-Error View

The main chapter reports the robustness of the Transformer-inspired model using mean edge accuracy. The complementary mean edge-error view is shown in Figure F.2. The mean edge error, defined in Equation 3.32, represents the average number of incorrectly recovered branches across the evaluated topologies; therefore, larger values indicate poorer reconstruction. The curves use the same inference-only and database-and-inference perturbation settings described in Section F.1.

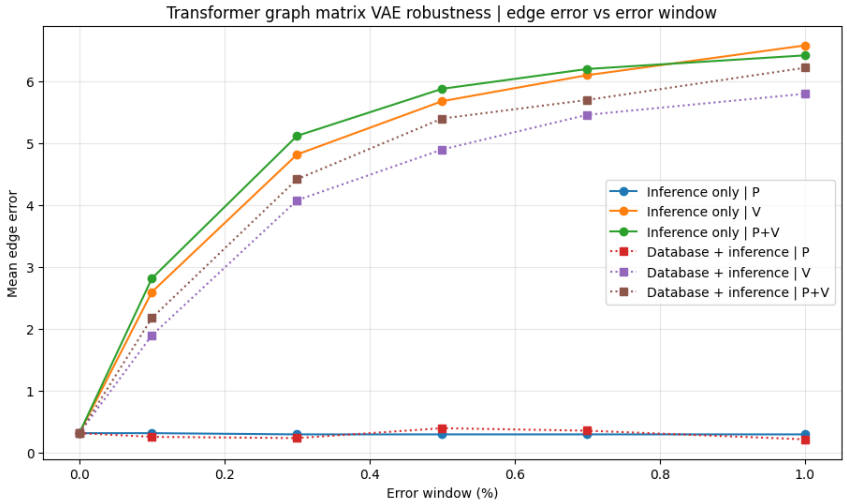


Figure F.2: Robustness of the Transformer-inspired variable-size VAE: mean edge error versus measurement error window. The curves compare errors in P, V, and both P and V, applied either only during inference or during both training and inference.

This view confirms the conclusion obtained from the edge-accuracy curves. Errors in voltage measurements increase the number of incorrectly inferred edges more strongly than active-power errors. Therefore, the Transformer-inspired model improves clean-data performance and large-graph generalisation, but it remains sensitive to voltage perturbations.

F.3 Summary

These supplementary robustness experiments are fully consistent with the interpretation presented in Chapter 4 and reinforce the importance of voltage-measurement quality for practical deployment.

Appendix G

Supplementary Implementation Details

This appendix summarises implementation details that support the experimental results reported in Chapter 4. The objective is not to repeat the methodology, but to clarify the main data-processing, training and evaluation choices used across the experiments.

G.1 Data Structure

Each synthetic sample contains a radial topology, active-power profiles and voltage-magnitude profiles. The topology is represented by an adjacency matrix, while the electrical data are represented either as raw P/V trajectories or transformed into matrix-based representations such as correlations and sensitivities.

For fixed-size models, each experiment is associated with a single network size. For variable-size models, graph sizes are processed in the same pipeline by using padding and masks. The mask identifies which nodes are real and which positions correspond to padding.

G.2 Topology Output and Radial Projection

The learning models produce edge scores or reconstructed adjacency-like outputs. Since LV feeders are assumed to be radial, the final inferred topology must satisfy the tree structure. Therefore, the predicted scores are converted into a connected radial graph during post-processing.

This step ensures that the final prediction contains the correct number of branches for a radial network and avoids disconnected or cyclic outputs. The projection does not replace the learning model; it imposes the physical radiality constraint on the model output.

G.3 Evaluation Metrics

The main metric used in the experiments is edge accuracy. It measures the fraction of correctly recovered edges in the inferred topology. This metric is useful because it captures partial reconstruction quality, even when the predicted topology is not exactly equal to the ground truth.

Exact topology recovery is stricter. It is equal to one only when all predicted edges match the ground-truth edge set. For this reason, exact recovery can remain low even when edge accuracy indicates that a substantial part of the topology was correctly inferred.

G.4 Seen and Unseen Evaluation

Two evaluation settings are considered. In the seen setting, the topology appears in the training database, but the electrical realisation used for testing is different. In the unseen setting, the topology itself is absent from the training database.

This distinction is important because strong seen performance does not necessarily imply generalisation to new radial structures. The unseen setting is therefore the more relevant test for topology inference when the objective is to reconstruct feeders not previously observed during training.

G.5 Fixed-Size and Variable-Size Training

The fixed-size experiments are useful for controlled analysis because all samples have the same number of nodes. However, this setting requires a different model for each network size.

The variable-size experiments address this limitation by allowing a single model to process several graph dimensions. Shared node encoders, masking-aware aggregation and Transformer-based node interactions are used to support this transition.

G.6 Summary

This appendix clarifies the implementation choices behind the reported experiments. The most relevant points are the use of radial post-processing, the distinction between edge accuracy and exact recovery, and the separation between seen and unseen evaluation. These details are important for interpreting the results in Chapter 4.