

TOPOLOGICAL BLINDNESS IN DEEP LEARNING

ISOLATING THE RELATIONAL INDUCTIVE BIAS OF GRAPH CONVOLUTIONAL NETWORKS

NICHOLAS J. HAMMER¹ N. J. HAMMER@STUDENT.TUDELFT.NL
 SUPERVISORS: DAVID M. J. TAX¹, CASPER C. J. VAN ENGELENBURG²
¹EEMCS, ²ARCHITECTURE AND THE BUILT ENVIRONMENT | DELFT UNIVERSITY OF TECHNOLOGY

1. Background & Research Question

The Motivation:

- Deep learning architectures have inherent assumptions about the data they process → inductive biases.
- Which machine learning method is the best for a given problem?

Graph Convolutional Networks (GCNs)

- **Relational Inductive Bias:** Process data in graph domains (e.g., financial and molecular networks).
- **Tasks:** Graph classification, node classification, link prediction, graph generation.
- **Message Passing:** Iteratively update node representations by aggregating local neighborhood features.
- **Multi-Hop Routing:** Aggregate features from distant nodes by stacking network layers.
- **Permutation Invariance:** Ignore node indexing, in contrast to most deep learning architectures.

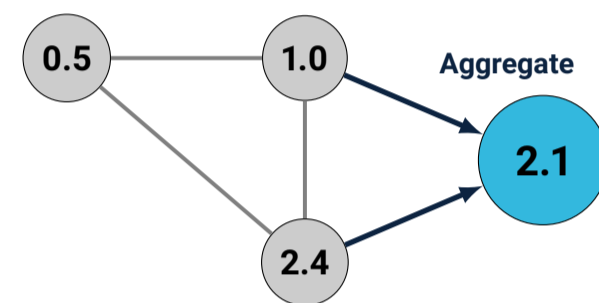


Fig. 1: Message Passing: Iteratively updating node representations by aggregating local neighborhood features.

Primary Research Question:

Can a classification task be procedurally generated such that a GCN achieves a lower error than architectures with alternative inductive biases across varying noise levels and sample sizes (Multilayer Perceptrons (MLPs), Convolutional Neural Networks (CNNs), Recurrent NNs (RNNs), and Transformers)?

Framework Goals:

1. **Generation:** Embed multi-hop dependencies into a vector.
2. **Translation:** Map the vector deterministically into native architecture formats.
3. **Evaluation:** Assess sample complexity and robustness to topological noise.

2. The Universal Vector Space

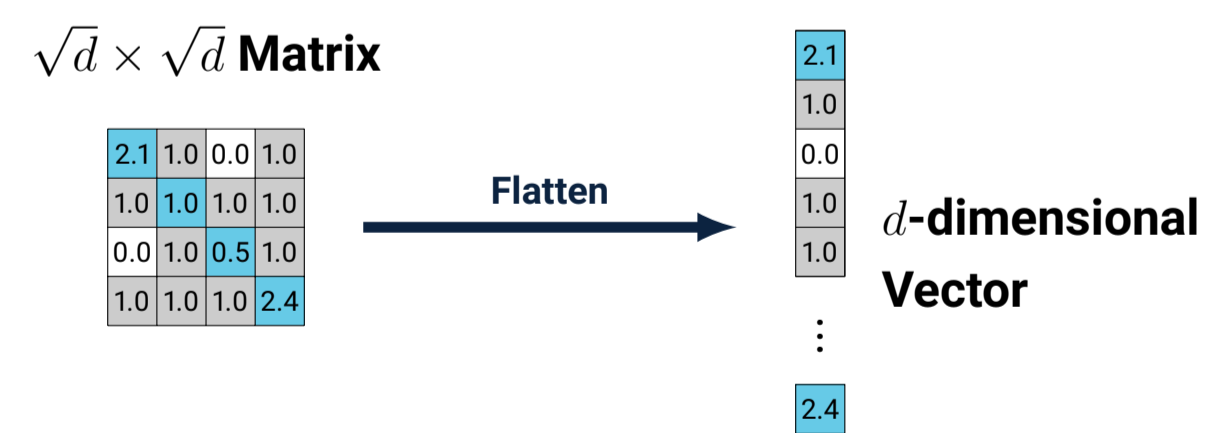


Fig. 2: All data originates as a vector generated by flattening a Unified Adjacency-Feature Matrix

Vector Serialization Scheme:

- **Information Budget:** Data is restricted to a 64-dimensional vector ($d = 64$).
- **Lossless Representation:** Matrix formulation avoids data compression or distortion.
- **Node Features:** Embedded directly on the primary matrix diagonal ($i = j$).
- **Edge Weights:** Dictate the off-diagonal relational matrix values ($i \neq j$).
- **Permutation Chaos:** Shuffled node indices introduce permutation variance.
- **Anti-Pre-Solving:** Avoid computing statistical summaries or running feature extraction prior to model ingestion.

Conclusion: The universal format standardizes information budgets so no model gains an unfair data structural advantage at ingestion.

3. The Translation Pipeline

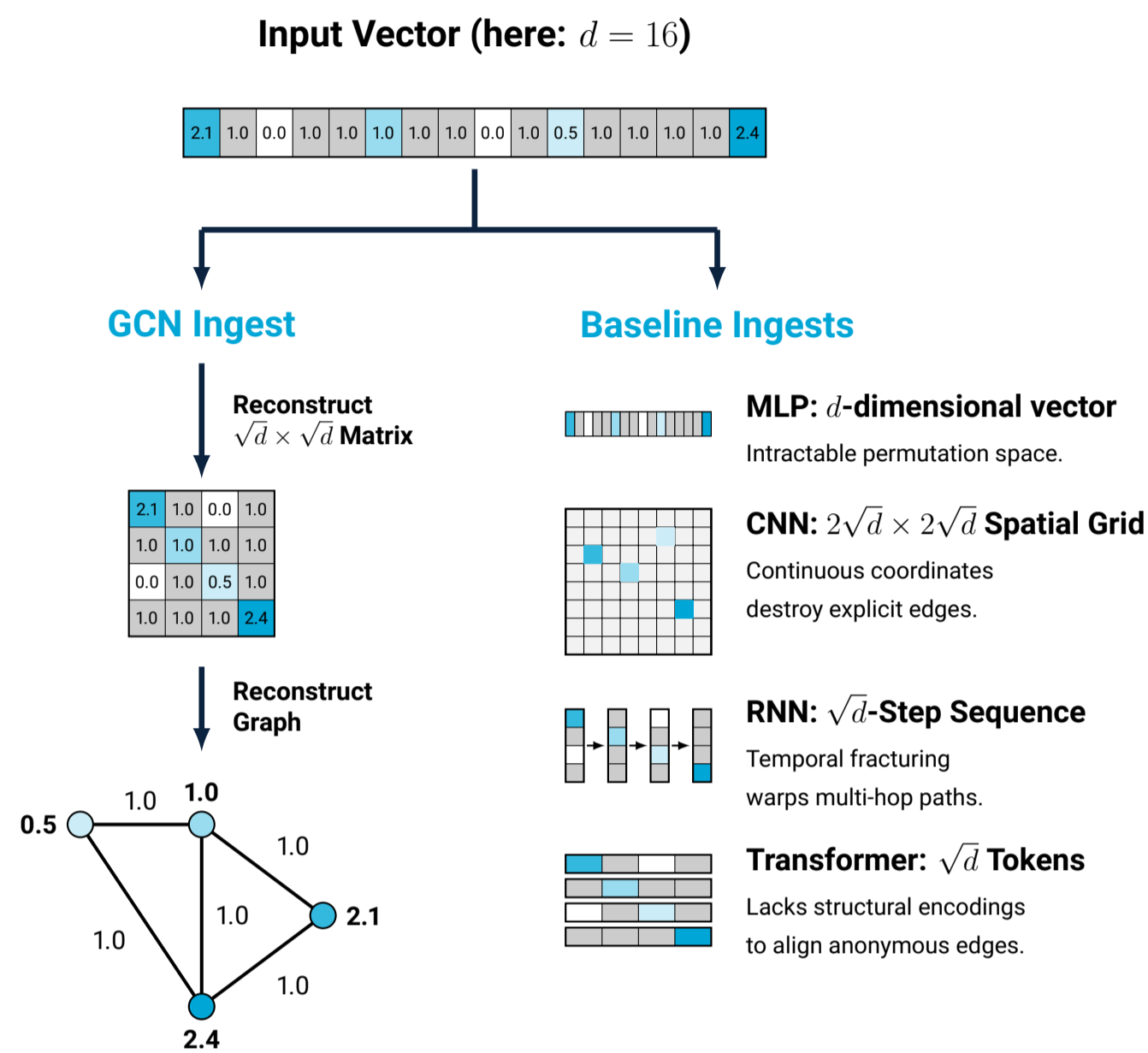


Fig. 3: A 16-D abstraction of the deterministic translation pipeline. Only the GCN retains explicit edges.

Conclusion: Translating graphs into other formats blinds models with misaligned inductive biases.

4. The Stochastic Smurfing Archetype

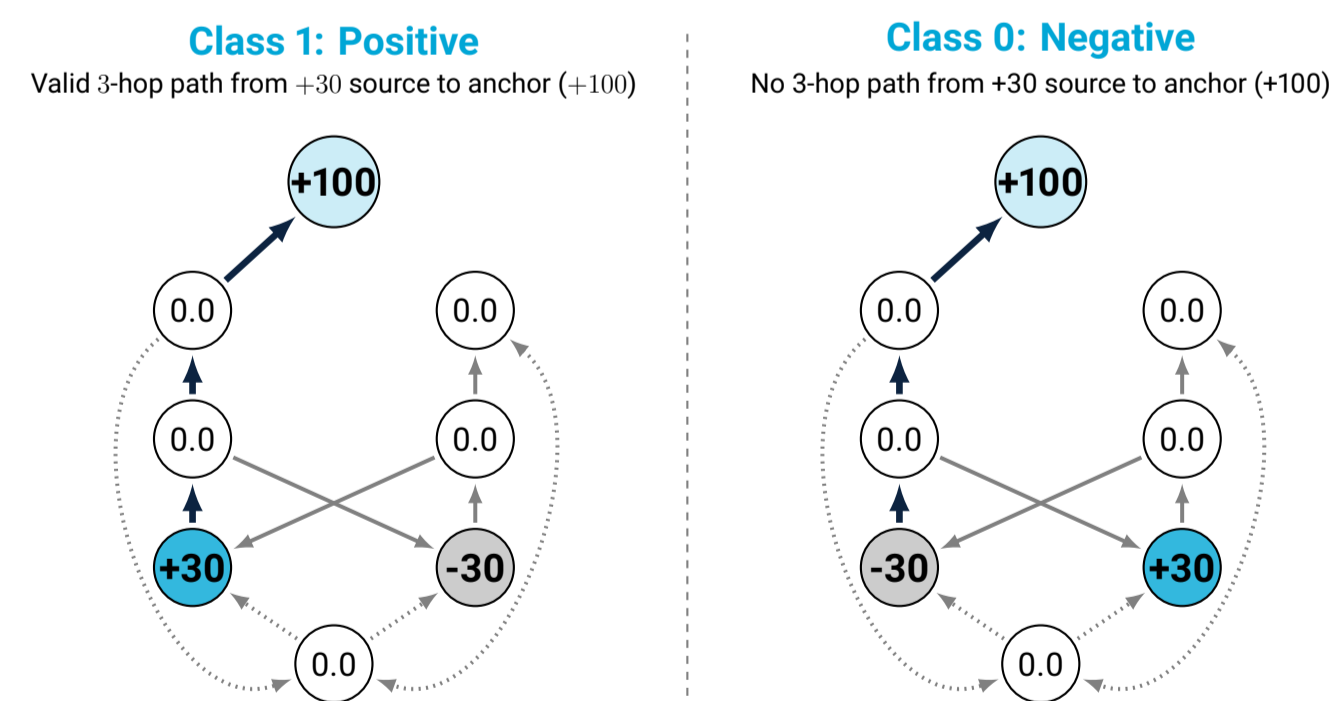


Fig. 4: The topology models a financial "smurfing" operation. Dotted lines signify background transactions generated randomly per sample.

The Anti-Money Laundering Analogy:

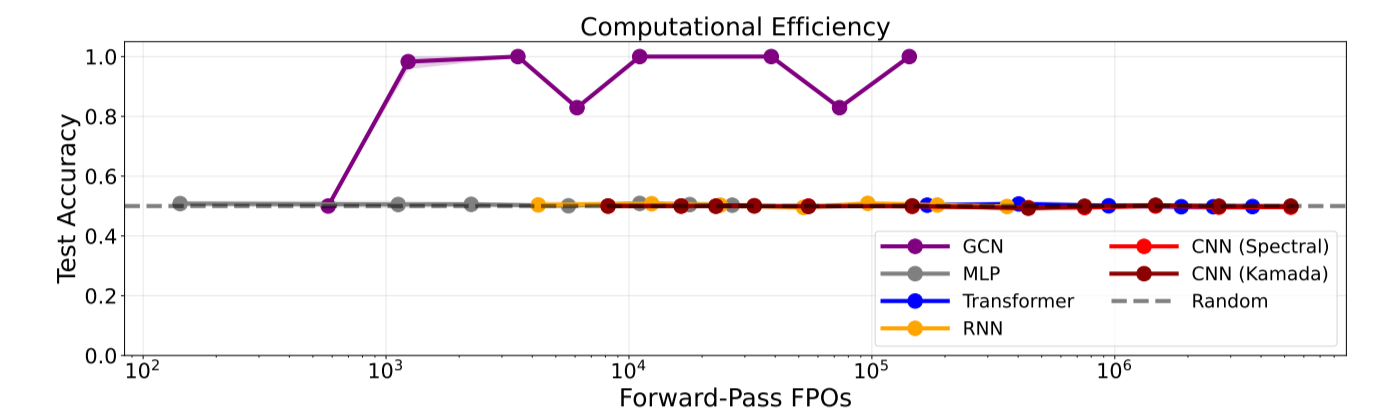
- **The Task:** Detect "smurfing" (routing illicit funds through intermediate accounts) by identifying if the illicit source (+30) possesses a 3-hop path to the anchor (+100).
- **Symmetrical Connections:** Stochastic background transactions force both the illicit and legitimate sources to share identical local financial structures.
- **The Effect:** CNNs/RNNs destroy explicit paths, MLPs suffer from permutation variance, and Transformers lack structural encodings to align anonymous edges.

Conclusion: This procedural archetype requires architectures to perform multi-hop routing to solve the task.

5. Experimental Results

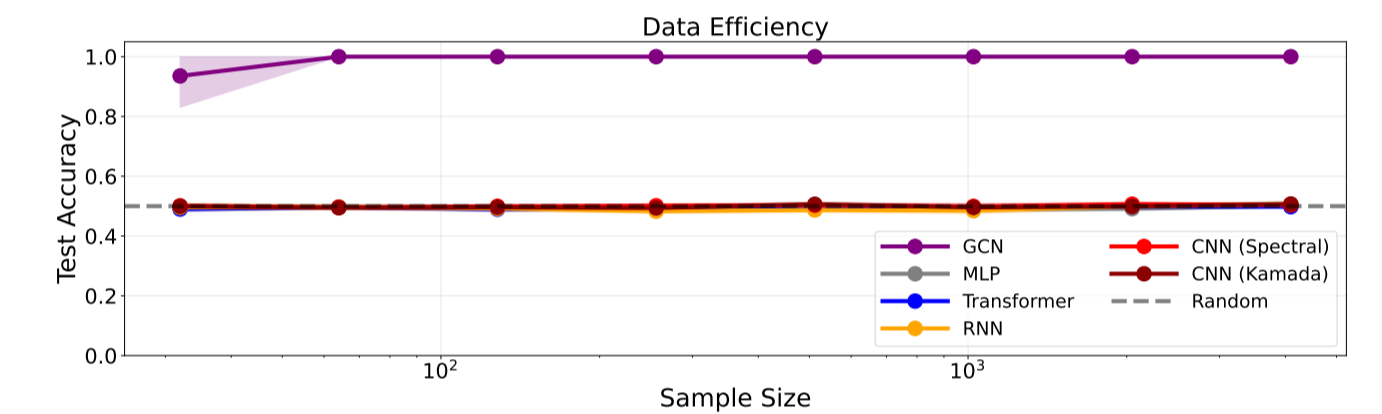
Q: Can computational scale overcome a lacking relational inductive bias?

- **A:** GCNs solve the task efficiently ($\approx 10^3$ forward-pass floating-point operations, or FPOs), with accuracy dropping only at shallow depth (< 3 layers).
- Baselines remain at chance levels despite using more than 10^6 FPOs.



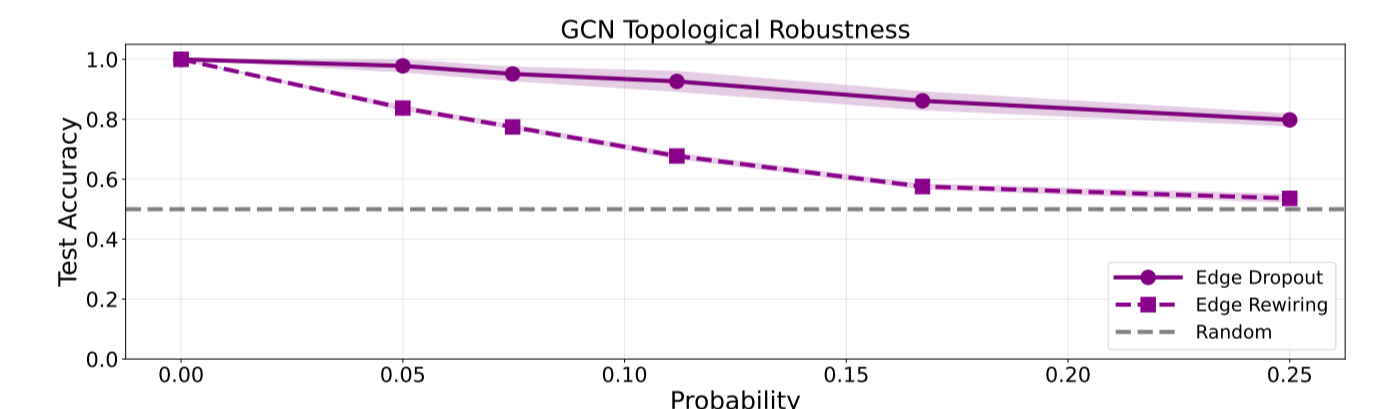
Q: How many samples are required to learn the multi-hop rule?

- **A:** GCNs converge rapidly in low-data regimes, while baselines fail to generalize even at 4 096 samples.



Q: Is the inductive bias robust to topological noise?

- **A:** GCNs exhibit resilience to moderate edge dropout but not to edge rewiring (false pathways).



Conclusion: Empirical results indicate that solving stochastic routing requires an aligned inductive bias, which is not easily substituted by raw computational capacity or data volume.

6. Conclusion & Future Work

Identified Limitations:

- **Low-Dimensionality:** The $d = 64$ constraint isolates inductive biases but abstracts away dense, real-world feature complexity.
- **Permutation Dependency:** Baseline collapse relies on random node indexing; canonical ordering could allow spatial memorization.
- **Lossy Translations:** Coercing graphs into rigid 1D/2D formats inherently severs explicit edges.

Future Work:

- **Topological Denoising:** Develop adaptive edge-dropping mechanisms to defend routing channels against adversarial rewiring.
- **Task Generalization:** Expand the evaluation framework beyond binary graph classification to tasks like (multi-class) node classification and link prediction.

Conclusion: Empirical evidence indicates that scale cannot easily substitute for relational reasoning. This framework suggests that matching an architecture's inductive bias to the structure of the data is a requirement for efficient generalization.