mathematics

MDPI

Article

Beyond Graph Diameter: Structural Determinants of Long-Range Information Propagation in Message-Passing Neural Networks

Marc Krueger [1], Marcus Frohme [2] and Alina Nechyporenko [2,3,*]

[1] Department of Computer Science and Media, Technical University Brandenburg (TH Brandenburg), 14770 Brandenburg, Germany; kruegmar@th-brandenburg.de
[2] Department of Molecular Biotechnology and Functional Genomics, Technical University of Applied Science Wildau (TH Wildau), 15745 Wildau, Germany; marcus.frohme@th-wildau.de
[3] Department of Computer Modeling and Intelligent Technologies, Kharkiv National University of Radio Electronics, 61023 Kharkiv, Ukraine
* Correspondence: nechyporenko@th-wildau.de

Abstract

Message-Passing Neural Networks (MPNNs) propagate information through iterative neighborhood aggregation, making long-range information exchange dependent on both network depth and graph structure. This study investigates whether graph diameter adequately characterizes effective information propagation and which topological properties provide additional explanatory information. A mathematical framework is developed to distinguish finite-depth reachability from communication redundancy, including formal results demonstrating that graphs with identical diameters can exhibit fundamentally different connectivity, articulation-point structures, and path redundancy. Controlled experiments were conducted on path, cycle, Erdős–Rényi, grid, and barbell graphs using a label-copying task and MPNNs with 2, 4, 6, and 10 message-passing layers. Graph-level structural measures were correlated with propagation accuracy, complemented by targeted node-removal experiments. The results show that graphs with comparable diameters can exhibit substantially different propagation performance. Average shortest-path length and clustering coefficient showed only weak associations with accuracy, whereas articulation points and edge connectivity exhibited stronger relationships. Removing high-betweenness nodes produced the largest accuracy reductions for several graph topologies, further demonstrating the importance of structurally critical communication pathways. Together, the theoretical and empirical results show that graph diameter alone provides an incomplete characterization of effective information propagation in MPNNs. Message-passing depth determines theoretical reachability, whereas effective long-range propagation additionally depends on graph topology, particularly communication redundancy, structural bottlenecks, connectivity, and the availability of alternative paths.

Keywords: message-passing neural networks (MPNNs); graph neural networks (GNNs); information propagation; long-range dependencies; graph topology; graph diameter; over-squashing; structural bottlenecks

MSC: 68T07; 68R10

Check for updates

Academic Editor: Andrea Scozzari

Received: 9 September 2026
Revised: 19 October 2026
Accepted: 24 October 2026
Published: 28 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Message-Passing Neural Networks (MPNNs), introduced by [1], provide a general framework for learning representations from graph-structured data through iterative ex-

change of information between neighboring nodes. The framework unified several previously proposed graph neural architectures under a common message-passing formulation consisting of message, update, and readout functions. Among the earliest and most influential graph neural architectures are Graph Convolutional Networks (GCNs) proposed by Kipf and Welling [2], which demonstrated that local neighborhood aggregation can be performed efficiently using a simplified spectral formulation.

A further important development was the introduction of Graph Attention Networks (GATs), which use attention mechanisms to assign different importance to neighboring nodes during aggregation [3].

Although originally developed for molecular property prediction, MPNNs have since become one of the dominant architectures for graph representation learning, with applications spanning social networks, recommender systems, computer vision, computational biology, and scientific machine learning. By iteratively aggregating information from neighboring nodes, MPNNs naturally exploit graph topology and enable node- and graph-level predictions across a broad range of applications.

Comprehensive surveys have documented the rapid expansion of graph neural network methodologies and applications across numerous disciplines, highlighting both their practical success and remaining theoretical challenges [4].

However, their local message-passing mechanism also imposes an inherent limitation: information originating from distant parts of a graph can influence a node only after a sufficient number of propagation steps. Consequently, tasks requiring long-range interactions remain particularly challenging for conventional MPNNs. Early studies of deep graph convolutional architectures already demonstrated that increasing the number of propagation layers does not necessarily improve their ability to use distant information. Study [5] showed that graph convolution can be interpreted as a form of Laplacian smoothing and that repeated propagation makes node representations increasingly similar. This phenomenon, commonly referred to as over-smoothing, limits the discriminative capacity of deep GCNs and illustrates that simply stacking additional layers may eventually degrade rather than improve model performance. The empirical performance of graph neural architecture has also been evaluated systematically across a wide range of benchmark datasets and tasks [6].

Subsequent theoretical work further demonstrated that graph neural networks can exponentially lose expressive power as depth increases, providing a rigorous explanation for the limitations of deep architectures [7].

The expressive capabilities of message-passing architecture have been extensively studied. Paper [8] demonstrated that the representational power of standard MPNNs is fundamentally determined by the neighborhood aggregation mechanism and is bounded by the discriminative power of the Weisfeiler–Lehman graph isomorphism test. Their analysis established a theoretical foundation for understanding both the strengths and inherent limitations of local message-passing architectures, motivating subsequent research on improving long-range information propagation.

More recent surveys identify over-squashing and information bottlenecks as among the most fundamental limitations affecting modern graph neural networks [9].

The relationship between network depth and the ability to capture distant information has been investigated theoretically. Loukas [10] demonstrated fundamental depth–width limitations of message-passing GNNs, showing that certain graph problems cannot be solved by architectures whose depth and width are insufficient relative to the graph structure. Increasing depth alone, however, does not necessarily provide an effective solution. Alon and Yahav [11] identified the phenomenon of "over-squashing", in which information from increasingly large receptive fields must be compressed into fixed-dimensional node

representations as messages propagate through the graph. This effect becomes particularly problematic when long-range dependencies pass through narrow structural bottlenecks.

Subsequent research has demonstrated that over-squashing and ineffective long-range propagation are strongly related to graph topology. Study [12] connected graph bottlenecks to discrete curvature and showed that edges associated with strongly negative curvature can indicate regions in which information flow is constrained. These findings inspired a growing body of work on graph rewiring strategies designed to alleviate topological bottlenecks and improve long-range communication [13].

Their work further demonstrated that curvature-guided graph rewiring can alleviate such bottlenecks. From an information-theoretic perspective, Banerjee et al. [14] interpreted over-squashing as information contraction during message propagation and related the phenomenon to graph expansion.

Alternative graph-theoretic approaches have further connected message propagation limitations to geometric and topological properties of graphs beyond purely distance-based descriptors [15].

These findings indicate that the ability to transmit information through a graph depends not simply on the existence of paths between nodes, but also on the structural capacity of those paths to convey information. Beyond bottlenecks, effectively integrating information from distant graph regions is also closely related to challenges arising in heterophilous graph settings, where neighboring nodes may contain substantially different information [16].

A particularly important development was provided by Di Giovanni et al. [17], who systematically investigated the roles of network width, depth, and graph topology in MPNN over-squashing. Their analysis showed that graph topology plays a central role and connects propagation difficulties to large commute times between nodes. This perspective goes beyond ordinary shortest-path distance because commute time also reflects the availability and structure of alternative routes through the graph. Similarly, Black et al. [18] characterized over-squashing using "effective resistance", a graph-theoretic quantity that captures the collective contribution of multiple paths connecting two nodes. Effective-resistance-based rewiring was consequently proposed as a mechanism for reducing communication bottlenecks. Together, these studies suggest that long-range information propagation is governed by more complex aspects of topology than pairwise distance alone.

This observation has motivated several approaches that modify the communication structure of MPNNs. For example, research work [19] dynamically expands the set of nodes allowed to communicate while retaining information about their original graph distance, thereby addressing long-range dependencies without completely discarding graph locality. More recently, Bamberger et al. [20] argued that the notion of long range itself requires a more precise characterization and introduced a formal range measure for assessing the extent to which graph tasks and GNN architectures depend on distant information. These developments underline the growing recognition that both distance and topology must be considered when analysing information propagation in graph neural networks.

The limitations of purely local message passing have additionally motivated transformer-based graph architectures that explicitly aim to capture long-range dependencies through global interactions [21]. However, the present study focuses specifically on classical message-passing neural networks. The objective is not to compare local message-passing architectures with transformer-based alternatives, but rather to investigate which graph-theoretic properties influence information propagation within the message-passing paradigm itself. Consequently, transformer-based architectures are considered outside the scope of the present work.

Despite this progress, the role of a simple and widely interpretable global distance property "graph diameter" has received comparatively limited systematic attention as an independent predictor of effective message propagation. Graph diameter represents the maximum shortest-path distance between any pair of nodes and therefore provides an intuitive description of the maximum spatial extent of a graph. One might consequently expect graphs with larger diameters to present greater difficulties for shallow MPNNs. Such an interpretation, however, ignores the fact that graphs with identical diameters may possess fundamentally different connectivity patterns, path redundancies, bottlenecks, clustering structures, and expansion properties. Whether diameter itself provides meaningful explanatory power for effective information propagation, and how its influence interacts with topology and network depth, therefore requires controlled investigation.

In this work, we systematically investigate whether graph diameter provides a meaningful predictor of effective information propagation in MPNNs. Using controlled synthetic graph families, including path, cycle, grid, Erdős–Rényi, and barbell graphs, we vary graph structure and message-passing depth while evaluating information propagation through a label-copying task. Our experiments show that graph diameter alone is a weak predictor of propagation performance: graphs with comparable diameters can exhibit substantially different propagation behavior. The observed limitations of graph diameter as a predictor of propagation performance are further supported by Proposition 2. The proposition demonstrates that graphs with identical diameters can possess substantially different edge connectivity, articulation-point structure and betweenness distributions. Consequently, diameter constrains communication distance but does not uniquely determine communication capacity. The experimental results therefore align with the theoretical observation that distance and communication structure constitute fundamentally different graph properties.

We therefore extend the analysis beyond distance-based descriptors and investigate structural properties including average shortest-path length, clustering coefficient, articulation points, edge connectivity, node degree, and betweenness centrality.

These experiments reveal a distinction between local connectivity and global communication importance. In particular, node degree alone provides limited information about propagation relevance, whereas nodes occupying critical communication positions can have a substantially greater influence on propagation performance. The node-removal experiments further demonstrate that eliminating low-degree nodes may remove globally relevant communication paths and substantially alter bottlenecked graph structures.

In contrast to prior work that develops or analyzes specific mechanisms and metrics associated with over-squashing, the present study focuses on a narrower question: whether graph diameter provides an adequate structural descriptor of effective long-range information propagation in MPNNs. Our contribution is a controlled theoretical and empirical analysis that distinguishes distance-based reachability from communication structure and examines how graphs with comparable distance characteristics can exhibit different propagation behavior due to differences in connectivity, path redundancy, and structural bottlenecks. Taken together, our theoretical and empirical results indicate that effective long-range propagation cannot be adequately characterized by graph diameter or local density alone but depends on the interaction between message-passing depth and the global communication structure of the graph. In this sense, the study complements existing topology-based analyses of over-squashing by specifically examining the limitations of graph diameter as an explanatory descriptor of long-range information propagation.

The remainder of this paper is organized as follows. Section 2 presents the mathematical framework underlying information propagation in MPNNs, including graph-theoretic concepts of reachability, connectivity, path redundancy, and structural bottlenecks. Section 3 describes the experimental methodology, including the graph families, label-copying task,

MPNN architecture, and structural measures. Section 4 presents the experimental results, while Section 5 discusses the findings and their implications for long-range information propagation. Finally, Section 6 summarizes the main conclusions and outlines directions for future research.

It is important to emphasize that the present study does not directly quantify over-squashing using established metrics such as effective resistance, commute time, discrete curvature, or Jacobian-based sensitivity measures. Instead, the study investigates whether simpler graph-theoretic descriptors, including articulation points, edge connectivity, and betweenness-centrality-related interventions, are associated with differences in information propagation performance. Consequently, the results should be interpreted as evidence regarding structural factors associated with propagation performance rather than as a direct measurement of oversquashing itself.

2. Mathematical Framework

2.1. Graph-Theoretic Preliminaries

This section introduces the mathematical notation and theoretical concept used throughout this paper. The objective is not to provide an exhaustive treatment of graph theory, linear algebra or neural network theory, but rather to establish a common framework for the analysis and interpretation of the proposed methodology and experimental results.

Definition 1. *Let V, E be sets. A graph is an ordered pair* $\boldsymbol{G = (V, E)}$*, where V is a finite set of vertices and* $\boldsymbol{E \subseteq V \times V}$ *is a set of edges connecting pairs of vertices.*

Throughout this paper, we consider simple, connected, undirected graphs unless stated otherwise. The number of vertices is denoted by $|\boldsymbol{V}|$ and the number of edges by $|\boldsymbol{E}|$.

Definition 2. *Let G = (V, E) be a graph and let* $\boldsymbol{u, v \in V}$*. The shortest-path distance between u and v is denoted by d(u, v) and provides the minimum number of edges among all paths connecting u and v. We can therefore write* $\boldsymbol{d(u, v) = min\{k \in \mathbb{N}_0 : u = v_0, v_1, \dots, v_k = v, (v_{\{i-1\}}, v_i) \in E\}}$.

While the shortest-path distance in path graphs is obvious, given that there only is one existing path connection between two distinct vertices, it becomes an important property in other graph topologies as there may be multiple alternative paths between two vertices. One of the graph metrics that shortest-path distance becomes important for is eccentricity.

Definition 3. *Let G = (V, E) be a connected graph and let* $\boldsymbol{v \in V}$*. The eccentricity of v is defined as the greatest shortest-path distance from v to any other vertex, namely* $\boldsymbol{ecc(v) = \max_{\{u \in V\}} d(v, u)}$.

The eccentricity therefore measures how far a vertex is from the most distant vertex in the graph. This motivates the definition of a graph diameter.

Definition 4. *Let G = (V, E) be a connected graph. The diameter of G is the maximum shortest-path distance between any two vertices, namely* $\boldsymbol{diam(G) = \max_{\{u,v \in V\}} d(u, v)}$.

In other words, the diameter of a graph is the maximum eccentricity among all vertices.

The shortest-path distance provides a natural notion of locality in graph-structured data. Building upon this concept, eccentricity characterizes the local reachability of individual vertices, while the graph diameter captures the largest communication distance occurring in the graph. These quantities are particularly relevant when studying long-range information propagation in Message-Passing Neural Networks.

2.2. Mathematical Formulation of Message Passing

Message-Passing Neural Networks (MPNNs) constitute a broad class of graph neural network architectures that learn node representations through iterative local information exchange between neighboring vertices. Given graph G = (V, E), each node $v \in V$ is associated with an initial feature vector $h_v^{(0)} \in \mathbb{R}^d$. The objective of message-passing is to iteratively update these node representations by aggregating information from neighboring nodes.

Formally, an MPNN layer consists of a message aggregation step followed by a node update step. At layer k + 1, each node v receives messages from its neighbors $u \in N(v)$ according to $m_v^{(k+1)} = AGG^{(k)}(\{M^{(k)}(h_v^{(k)}, h_u^{(k)}, e_{uv}) : u \in N(v)\})$, where $M^{(k)}$ denotes the message function, e_{uv} optionally represents edge attributes, and $AGG^{(k)}$ is a permutation invariant aggregation operator such as summation, averaging or maximum pooling.

The aggregated message is subsequently incorporated into the representation of node v through an update function $h_v^{(k+1)} = U^{(k)}\left(h_v^{(k)}, m_v^{(k+1)}\right)$, where $U^{(k)}$ is typically implemented as a learnable neural network layer.

By repeatedly applying these operations for K layers, information propagates through the graph and produces final node embeddings $h_v^{(K)}$. A graph-level representation may then be obtained through the readout function $h_G = R\left(\left\{h_v^{(K)} : v \in V\right\}\right)$, where R is again chosen to be permutation-invariant.

The general message-passing framework can be instantiated through different choices of message passing, aggregation and update functions. Several widely used graph neural network architectures can be expressed as special cases of the above-mentioned framework.

Graph Convolutional Networks (GCNs) employ a normalized aggregation scheme in which neighboring node representations are averaged before being transformed by a learnable weigh matrix. The layer update is given by $h_v^{(k+1)} = \sigma\left(W^{(k)} \sum_{u \in N(v) \cup \{v\}} 1 \frac{1}{\sqrt{\{\deg(v)\deg(u)}} h_u^{(k)}\right)$, where $W^{(k)}$ is a learnable weight matrix, deg(v) denotes the degree of node v and σ is a non-linear activation function.

In this formulation, $M(h_v, h_u) = h_u$ and aggregation are performed using a normalized summation operator. GCN is the framework that is used throughout this paper. Other frameworks include Graph Isomorphism Networks (GINs), Graph Attention Networks (GATs) and GraphSAGE.

2.3. Depth and Theoretical Receptive Field

The central mechanism underlying MPNNs is the iterative exchange of information between neighboring vertices. Since each message-passing layer aggregates information exclusively from one-hop neighbors, the depth of the network determines how far information can propagate through the graph.

This notion is commonly referred to as the theoretical receptive field of a node. Informally, the receptive field of a node at layer K consists of all vertices whose initial features may influence the node's representation after K rounds of message passing.

Understanding this relationship between network depth and graph distance is particularly important when studying long-range dependencies. Information originating from distant vertices must traverse multiple message-passing layers before it can affect a target node. Consequently, the depth of an MPNN imposes a fundamental limit on the range of information propagation.

The following proposition formalizes this observation.

Proposition 1. *Let G = (V, E) be a graph and consider an L-layer MPNN whose message-passing operation uses exclusively one-hop neighborhood aggregation. Then the representation $h_v^{(L)}$ of a node v can depend on the initial feature vector $h_u^{(0)}$ of a node u only if $d(u, v) \leq L$, where d(u, v) denotes the shortest-path distance between u and v.*

Proof. We proceed by induction on the number of message-passing layers.

Base case L = 0: Before any message passing is performed, $h_v^{(0)}$ contains only the initial features of node v. Therefore, $h_v^{(0)}$ can depend on $h_u^{(0)}$ only when u = v. Since d(v,v) = 0, the proposition holds for L = 0.

Induction hypothesis: Assume that after k message-passing layers, the representation $h_v^{(k)}$ can depend on $h_u^{(0)}$ only if $d(u, v) \leq k$.

Induction step: Consider layer k + 1. By the message-passing framework, $h_v^{(k+1)} = U^{(k)}\left(h_v^{(k)}, AGG\left\{M\left(h_v^{(k)}, h_w^{(k)}\right) : w \in N(v)\right\}\right)$. Thus, the representation of node v at layer k + 1 depends only on (1) its previous representation $h_v^{(k)}$ and (2) the representations $h_w^{(k)}$ of neighboring nodes $w \in N(v)$.

By the induction hypothesis, $h_w^{(k)}$ can depend only on initial node features originating from vertices u satisfying $d(u, w) \leq k$. Since w is a neighbor of v, we have $d(w, v) = 1$. Using the triangle inequality of shortest-path distances, we find $d(u, v) \leq d(u, w) + d(w, v) \leq k+1$.

Therefore, every initial feature capable of influencing $h_v^{(k+1)}$ must originate from a node within distance k + 1 of v.

Hence, $h_v^{(k+1)}$ can depend on $h_u^{(0)}$ only if $d(u, v) \leq k + 1$.

This completes the induction. □

Proposition 1 establishes that each message-passing layer expands the receptive field by exactly one graph hop. Consequently, an L-layer MPNN is unable to directly incorporate information from nodes whose shortest-path distance exceeds L. For example, a six-layer MPNN can theoretically receive information from nodes at most six hops away. Long-range dependencies therefore require increasingly deep architectures, causing information from a rapidly growing number of vertices to be compressed into fixed-dimensional hidden representations. This observation forms one of the fundamental motivations for studying oversquashing in graph neural networks.

2.4. Diameter Versus Path Redundancy and Connectivity

The diameter of a graph is frequently used as a proxy for the difficulty of long-range information propagation. Proposition 1 demonstrates that an L-layer MPNN can only propagate information between nodes separated by at most L hops. Consequently, graphs with large diameters are often regarded as challenging for message passing.

However, diameter captures only the length of the shortest communication route between two nodes and ignores the abundance of alternative routes. From the perspective of information propagation, the existence of multiple independent communication paths can be equally important. Two graphs may exhibit identical diameters while differing substantially in their robustness to edge or vertex removal and, consequently, in their ability to support information dissemination.

To formalize this distinction, we introduce the notions of path redundancy and connectivity.

Let $u, v \in V$ be two distinct vertices.

Definition 5. *Two paths between u and v are called internally vertex-disjoint if they share no common internal vertices.*

Formally, if $P_1 = (u, x_1, \ldots, x_r, v)$ and $P_2 = (u, y_1, \ldots, y_s, v)$ are two paths, then P_1 and P_2 are internally vertex-disjoint if $\{x_1, \ldots, x_r\} \cap \{y_1, \ldots, y_s\} = \emptyset$.

The maximum number of internally vertex-disjoint paths between u and v is denoted by $\kappa(u, v)$.

Definition 6. *Two paths between u and v are called edge-disjoint if they do not share any common edge.*

The maximum number of edge-disjoint paths between u and v is denoted by $\boldsymbol{\lambda}(\boldsymbol{u}, \boldsymbol{v})$. It measures the level of communication redundancy available between two vertices.

A fundamental result relating connectivity and path redundancy is Menger's theorem.

Theorem 1 (Menger [22]). *For any two distinct vertices,* $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{V}$: $\boldsymbol{\kappa}(\boldsymbol{u}, \boldsymbol{v}) = \mathbf{min}\{|\boldsymbol{S}| : \boldsymbol{S} \subseteq \boldsymbol{V} \setminus \{\boldsymbol{u}, \boldsymbol{v}\}$ *disconnects u and v}* and $\boldsymbol{\lambda}(\boldsymbol{u}, \boldsymbol{v}) = \mathbf{min}\{|\boldsymbol{F}| : \boldsymbol{F} \subseteq \boldsymbol{E}$ *disconnects u and v.*

In other words, the maximum number of independent communication paths equals the minimum number of vertices or edges whose removal separates the two vertices. Thus, high connectivity implies the existence of many alternative communication routes, whereas low connectivity indicates the presence of bottlenecks.

Diameter and connectivity characterize fundamentally different graph properties. The diameter quantifies the length of the shortest communication path $\boldsymbol{diam}(\boldsymbol{G}) = \max_{\boldsymbol{u},\boldsymbol{v} \in \boldsymbol{V}} \boldsymbol{d}(\boldsymbol{u}, \boldsymbol{v})$ whereas $\boldsymbol{\kappa}(\boldsymbol{u}, \boldsymbol{v})$ and $\boldsymbol{\lambda}(\boldsymbol{u}, \boldsymbol{v})$ quantify the number of independent communication routes available between vertices.

Consequently, two graphs may satisfy $\boldsymbol{diam}(\boldsymbol{G_1}) = \boldsymbol{diam}(\boldsymbol{G_2})$ while $\boldsymbol{\lambda}(\boldsymbol{G_1}) \neq \boldsymbol{\lambda}(\boldsymbol{G_2})$.

In such a situation, the graphs possess identical maximum shortest-path distance but differ substantially in the redundancy of communication pathways.

Proposition 1 states that depth determines whether communication between two nodes is theoretically possible. However, it does not quantify how much information can be transmitted once such communication is possible.

When a graph contains articulation points, bridges or low edge connectivity, information originating from many nodes must pass through a small number of communication channels. These bottlenecks force a large amount of information to be compressed during message aggregation and can therefore contribute to over-squashing.

This observation suggests that graph diameter alone is insufficient for characterizing information propagation. Measures of connectivity and path redundancy capture structural bottlenecks that are invisible to shortest-path distance and therefore provide a complementary theoretical explanation for the empirical importance of articulation points and edge connectivity observed in this work.

The roles of distance and connectivity must therefore be distinguished carefully. For a one-hop $\boldsymbol{L}$-layer MPNN, the inequality $\boldsymbol{d_G}(\boldsymbol{s}, \boldsymbol{t}) > \boldsymbol{L}$ implies that the initial representation of $\boldsymbol{s}$ cannot influence the representation of $\boldsymbol{t}$ after $\boldsymbol{L}$ layers. In contrast, $\boldsymbol{d_G}(\boldsymbol{s}, \boldsymbol{t}) \leq \boldsymbol{L}$ establishes only theoretical reachability. It does not guarantee that the source information is transmitted without substantial attenuation, mixing, or compression. Among node pairs lying within the same theoretical receptive field, the quantities $\boldsymbol{\kappa}(\boldsymbol{s}, \boldsymbol{t})$ and $\boldsymbol{\lambda}(\boldsymbol{s}, \boldsymbol{t})$ characterize the availability of internally vertex-disjoint and edge-disjoint communication routes, respectively. Thus, distance determines the minimum depth required for possible influence, whereas connectivity characterizes structural redundancy among the available routes. Effective information transmission additionally depends on the MPNN architecture, its parameters, and the aggregation dynamics.

The following proposition follows naturally from the discussion above.

Proposition 2. *Let* $\boldsymbol{D} \geq \mathbf{2}$ *be an integer. There exist connected graphs* $\boldsymbol{G_1}, \boldsymbol{G_2}$ *satisfying* $\boldsymbol{diam}(\boldsymbol{G_1}) = \boldsymbol{diam}(\boldsymbol{G_2}) = \boldsymbol{D}$ *while simultaneously* $\boldsymbol{\lambda}(\boldsymbol{G_1}) \neq \boldsymbol{\lambda}(\boldsymbol{G_2}), \boldsymbol{A}(\boldsymbol{G_1}) \neq \boldsymbol{A}(\boldsymbol{G_2}), \boldsymbol{B}(\boldsymbol{G_1}) \neq \boldsymbol{B}(\boldsymbol{G_2})$, *where* $\boldsymbol{\lambda}$ *denotes the edge connectivity, A denotes the number of articulation points and B denotes the betweenness-centrality distribution.*

Proof. We prove the proposition by constructing two graph families with identical diameter but different communication structures.

Graph family 1: Let $G_1 = P_{D+1}$ be the path graph on D + 1 vertices $v_0, v_1, \ldots, v_D$. The distance between v_0 and v_D equals D, so $diam(G_1) = D$. Moreover, every communication route between v_0 and v_D is unique and every internal vertex is an articulation point, meaning removal of any edge disconnects the graph. Hence $\lambda(G_1) = 1$ und $A(G_1) = D - 1$. In addition, the central vertices lie on almost every shortest path and therefore attain high betweenness centrality.

Graph family 2: Let $G_2 = C_{2D}$ be the cycle graph on **2D** vertices, where $\mathbf{D} \geq \mathbf{2}$. Consider two antipodal vertices $\mathbf{u}, \mathbf{v} \in \mathbf{V}(\mathbf{G_2})$. Since the cycle consists of two arcs of equal length connecting u and v , their shortest-path distance is $d_{G_2}(u, v) = D$. Consequently, $\mathbf{diam}(G_2) = D$. Unlike the path graph, however, the cycle graph contains two independent routes between antipodal vertices. The clockwise and counterclockwise arcs connecting u and v are internally vertex-disjoint and edge-disjoint. Therefore, $\kappa(u, v) = 2$ and $\lambda(u, v) = 2$.

Furthermore, no vertex of C_{2D} is an articulation point, since removing any single vertex leaves a connected path graph. Hence, $A(G_2) = 0$.

Similarly, removing any single edge does not disconnect the graph, implying $\lambda(G_2) = 2$.

Consequently, $\lambda(G_2) > \lambda(G_1)$ and $A(G_2) < A(G_1)$.

Although G_1 and G_2 possess identical diameters, they exhibit fundamentally different communication structures. The path graph provides only a single communication route between distant vertices, whereas the cycle graph offers redundant alternative routes. Moreover, shortest paths in the cycle graph are distributed more evenly across the graph, resulting in a substantially more uniform betweenness-centrality distribution than in the path graph.

Hence, the proposition follows. □

Proposition 3. *Let* $L \geq 2$ *. There exist connected graphs* G_1, G_2 *with source–target pairs* $(s_1 t_1)$, (s_2, t_2) *such that* $d_{G_1}(s_1, t_1) = d_{G_2}(s_2, t_2) = L$ *but* $\kappa_{G_1}(s_1, t_1) \neq \kappa_{G_2}(s_2, t_2)$ *and* $\lambda_{G_1}(s_1, t_1) \neq \lambda_{G_2}(s_2, t_2)$.

Proof. Take $G_1 = P_{L+1}$ and let s_1, t_1 be its endpoints. Then $d_{G_1}(s_1, t_1) = L$ and the unique path between them implies $\kappa_{G_1}(s_1, t_1) = \lambda_{G_1}(s_1, t_1) = 1$. Now take $G_2 = C_{2L}$ and let s_2, t_2 be antipodal vertices. Then $d_{G_2}(s_2, t_2) = L$. The two arcs of the cycle connecting s_2, t_2 are internally vertex-disjoint and edge-disjoint. Hence $\kappa_{G_2}(s_2, t_2) = \lambda_{G_2}(s_2, t_2) = 2$. Hence equal source–target distance and equal minimum required message-passing depth do not determine communication redundancy. □

2.5. Structural Bottlenecks and Information Propagation

While graph diameter characterizes the maximum shortest-path distance between vertices, it does not identify the specific structural elements through which communication must pass. In many graphs, information flow is constrained by a small subset of critical vertices whose removal significantly alters connectivity. Such structures are commonly referred to as graph bottlenecks.

From the perspective of MPNNs, bottlenecks are particularly important because information originating from large portions of the graph may be forced to propagate through a limited number of vertices or edges. Consequently, these vertices may play a disproportionately large role in determining how effectively information can be transmitted across the graph. To formalize this notion, we introduce articulation points and betweenness centrality from a mathematical point of view.

Definition 7. *Let G = (V, E) be a connected graph. A vertex* $v \in V$ *is called an articulation point if the graph obtained by removing v and all incident edges, namely G–v, contains more connected components than G.*

Equivalently, v is an articulation point if there exist vertices $u, w \in V \setminus \{v\}$ such that every path between u and w contains v.

Articulation points represent extreme communication bottlenecks. Information exchanged between different regions of the graph must pass through such vertices. The presence of many articulation points therefore indicates limited communication redundancy and increased vulnerability to information compression.

Proposition 4. *Let v be an articulation point separating two vertex sets* S_1, S_2 *. Then every path connecting a vertex in* S_1 *to a vertex in* S_2 *must contain v.*

Proof. By definition, removing v disconnects S_1 and S_2. Suppose there exists a path between a vertex in S_1 and a vertex in S_2 that does not contain v. Then this path would remain after the removal of v, contradicting the assumption that v is an articulation point. Therefore, every such path must contain v. □

While articulation points describe extreme bottlenecks, many graphs contain vertices that are not articulation points but nevertheless lie on a large proportion of communication routes. To quantify this effect, we introduce betweenness centrality.

Definition 8. *Let* σ_{st} *denote the number of shortest paths between vertices s and t and let* $\sigma_{st}(v)$ *denote the number of those shortest paths that contain vertex v as an internal vertex. The betweenness centrality of a vertex v is defined as* $BC(v) = \sum_{s,t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$.

Hence, the betweenness centrality represents the proportion of shortest communication routes between s and t that pass through v.

Vertices with high betweenness centrality participate in many shortest paths and therefore occupy strategically important positions within the graph. Such vertices influence communication throughout large portions of the network, even if their removal does not disconnect the graph entirely. Consequently, betweenness centrality also captures a broader notion of communication importance than articulation points, which represent only strict bottlenecks. However, betweenness centrality itself is derived exclusively from shortest-path structure and therefore has inherent limitations when used to characterize information propagation in MPNNs. Message passing may depend not only on shortest communication routes, but also on the availability of alternative paths and redundant communication channels. As a result, graphs may contain soft bottlenecks, namely high-betweenness vertices whose removal substantially impedes information flow without disconnecting the graph.

Articulation points and betweenness centrality capture different aspects of graph structure. Articulation points describe a local topological property: whether a single vertex is indispensable for maintaining connectivity. Betweenness centrality, in contrast, is a global quantity derived from shortest-path structure across the entire graph. A vertex may exhibit a high betweenness centrality even without being an articulation point.

Consequently, articulation points measure the existence of strict communication bottlenecks, whereas betweenness centrality measures the overall communication load carried by a vertex.

3. Materials and Methods

This section describes the experimental design, the investigated graph topologies and their structural properties, the MPNN architecture, the label-copying task, and the statistical methods used to evaluate information propagation. This study exclusively focusses on five families of synthetic graphs (path, cycle, grid, ER-random and barbell graphs). Consequently, the empirical findings should primarily be interpreted within controlled synthetic settings rather than as direct evidence for all real-world graph domains.

3.1. Graph Families

In this paper, we investigate information propagation performance of graphs with varying topology.

Path graphs are graphs that consist of one single chain of nodes. Information can only propagate along this single line. Cycle graphs are a special kind of path graph where the end points are connected to each other. Therefore, information could flow along two different paths. Erdos–Renyi Graphs are random graphs in which edges are created randomly between nodes with a fixed probability. This topology usually contains multiple alternative propagation paths. Grid graphs are designed as two-dimensional lattices with multiple edges between the nodes and therefore multiple alternative communication paths. Barbell graphs consist of two densely connected subgraphs that are connected via a narrow bridge. Information flow between these two dense subgraphs can only happen along the narrow bridge.

3.2. Label-Copying Task

For each graph instance G = (V, E), a source node $s \in V$ and a target node $t \in V$ were selected such that their shortest-path distance satisfied $d(s,t) = k$, where k denotes the propagation distance under investigation. Source–target pairs were sampled across multiple distance values to evaluate how information propagation performance changes as communication is required over increasingly larger graph distances.

The source node s was assigned to a binary label $y_s \in \{0,1\}$, while all remaining node features were initialized identically and therefore contained no information about the label. Consequently, successful prediction at the target node required the MPNN to propagate information originating from the source node through the graph structure.

For each source–target pair, the network was trained to predict the label associated with the source node using only the representation learned at the target node after L rounds of message passing. Let $\hat{y}_t$ denote the prediction obtained from the target-node representation $h_t^{(L)}$. The label-copying task is therefore defined as $\hat{y}_t = f\left(h_t^{(L)}\right)$, where f denotes the final classification layer.

Prediction accuracy was computed as Accuracy $= \frac{1}{N}\Sigma_{i=1}^{N} 1(\hat{y}_t = y_i)$, where N denotes the number of evaluated source–target pairs and 1 is the indicator function.

For each graph topology, multiple source–target pairs were sampled for every evaluated distance k. Reported accuracies were calculated across the evaluated node pairs and graph configurations. This design isolates the ability of the MPNN to transmit information through the graph and allows propagation performance to be analysed independently of application-specific prediction tasks.

3.3. MPNN Architecture and Training

To investigate the influence of graph topology on information propagation, all experiments were performed using a GCN-based MPNN. The architecture was selected because it represents one of the most widely studied instances of general MPNN framework introduced in Section 2.2 and performs message passing through iterative neighborhood aggregation.

Let A denote the adjacency matrix of a graph and let $\tilde{A} = A + I$ be the adjacency matrix with added self-loops. Furthermore, let $\tilde{D}_{ii} = \Sigma_j \tilde{A}_{ij}$ denote the corresponding degree matrix.

Each graph convolution layer updates the matrix of node pair representations $H^{(l)}$ according to $H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right)$, where $W^{(l)}$ denotes a learnable weight matrix and σ represents a non-linear activation function.

The study evaluated architecture consisting of $L \in \{2, 4, 6, 10\}$ message-passing layers in order to investigate the relationship between propagation depth and information transmission. The selected depths were chosen to cover both shallow and deep message-passing regimes while remaining computationally tractable.

Node representations were initialized using the feature vectors described in Section 3.2. Following the final message-passing layer, the representation of the target node was passed to a fully connected classification layer that predicted the source-node label in the label-copying task.

Training was performed using the cross-entropy loss function $L = -\Sigma_{i=1}^{N} y_i \, log(\hat{y}_i)$, where N_{train} denotes the number of training target nodes, $y_i \in \{0, 1\}$ is the source label associated with target node i, and $p_{i,c}$ is the predicted probability of class c. For each source node and propagation distance, the candidate target nodes were randomly divided into equally sized training and test subsets, corresponding to a train–test ratio of 0.5 : 0.5. When the number of candidate targets was odd, integer partitioning assigned $\lfloor N_k/2 \rfloor$ targets to the training set and the remaining targets to the test set. The same architecture, hidden dimensionality, optimization procedure, and training duration were used across all graph families and message-passing depths to ensure comparability between experimental conditions.

Model parameters were optimized using the Adam optimizer. The learning rate was set to $\eta = 0.001$ and the number of epochs was set to 250.

Each node was represented by a one-dimensional input feature vector, such that the input dimensionality was $d_{in} = 1$. The initial feature of the selected source node encoded its binary label, whereas all non-source nodes were assigned the same neutral input value, as described in Section 3.2. Each hidden graph-convolutional layer mapped the node representations into a 32-dimensional latent space, yielding $d_{hidden} = 32$. The final graph-convolutional layer produced two output logits for every node, corresponding to the two possible source-label classes, and therefore used an output dimensionality of $d_{out} = 2$. For a target node t, the resulting output vector was given by $z_t \in \mathbb{R}^2$, and was converted into class probabilities using the SoftMax function $p(y_t = \hat{y}_t|G, X) = \frac{exp(z_{t,c})}{\Sigma_{j=0}^{1} exp(z_{t,j})}$). The predicted label was subsequently determined by $\hat{y}_t = argmax_{c\in\{0,1\}} \, p(y_t = \hat{y}_t \mid G, X)$.

The same hidden dimensionality was maintained across all investigated graph topologies and network depths, ensuring that comparisons reflected differences in graph structure and propagation depth rather than differences in representational width.

The use of a fixed MPNN architecture across graph topologies ensured that observed differences in propagation performance could be attributed primarily to graph structure rather than architectural variation. Consequently, the experimental design isolated the interaction between message-passing depth and graph topology, which constitutes the primary focus of this study.

3.4. Experimental Design

The study comprised three complementary analyses. First, propagation performance was compared across path, cycle, Erdős–Rényi, grid, and barbell graphs. Graphs with similar or overlapping diameter values but different topological structures were evaluated using MPNNs with 2, 4, 6, and 10 message-passing layers. This analysis examined whether graph

diameter alone explains propagation performance and whether the empirical propagation boundary corresponds to the theoretical (k)-hop receptive field of a (k)-layer MPNN. Second, graph-level structural measures were related to propagation performance using Pearson correlation analysis. The analyzed measures included average shortest-path length, clustering coefficient, edge connectivity, and the number of articulation points. Correlations were calculated separately for the evaluated network depths. In particular, the two-layer and ten-layer models were analyzed as representative shallow and deep message-passing configurations. For each graph instance, mean propagation accuracy was calculated by averaging across all evaluated source–target distances k, sampled source–target node pairs, and experimental repetitions.

Third, targeted structural interventions were performed on path, Erdős–Rényi, grid, and barbell graphs. Modified graph versions were generated by removing low-degree nodes, high-degree nodes, high-betweenness nodes, or articulation points. Cycle graphs were not included in this analysis because the structural interventions investigated were designed for graph topologies containing structurally distinguishable communication bottlenecks and node-importance patterns, which are largely absent in highly symmetric cycle graphs. The intervention experiments were conducted using the 10-layer MPNN, and propagation accuracy was averaged across all evaluated graph variants and experimental repetitions.

All graphs were generated synthetically using the NetworkX library [23]. To examine propagation performance under controlled distance conditions, graph-family parameters were selected with respect to a prescribed diameter D. Path graphs were generated as $G_{\mathbf{path}}(D) = P_{D+1}$, where P_{D+1} denotes the path graph on $D + 1$ vertices. Since the two endpoints are separated by D edges, this construction satisfied $\mathbf{diam}\left(G_{\mathbf{path}}(D)\right) = D$.

Cycle graphs were generated as $G_{\mathbf{cycle}}(D) = C_{2D}$, where C_{2D} is the cycle graph on $2D$ vertices. The greatest distance in an even cycle is attained by antipodal vertices, and hence $\mathbf{diam}\left(G_{\mathbf{cycle}}(D)\right) = D$.

Grid graphs were constructed as rectangular two-dimensional lattices with dimensions $m = \left\lfloor \frac{D+2}{2} \right\rfloor, n = D + 2 - m$. Because the diameter of an $m \times n$ rectangular grid i $(m - 1) + (n - 1) = m + n - 2$, the selected dimensions yield $\mathbf{diam}\left(G_{\mathbf{grid}}(D)\right) = D$.

The two-dimensional coordinate labels produced during grid generation were subsequently replaced by consecutive integer node indices.

Barbell-like graphs were generated from two independent Erdős–Rényi components. Each component contained $n_{\mathbf{ER}} = 40$ vertices, and each possible edge was included independently with probability $p_{\mathbf{ER}} = 0.2$.

The two components were connected through an intermediate path whose construction parameter was $D - 4$. One attachment vertex was selected randomly from each Erdős–Rényi component, after which the connecting path was joined to the selected vertices. This topology was designed to combine relatively dense local subgraphs with a narrow inter-component communication structure. Because the internal diameters of randomly generated Erdős–Rényi components are stochastic, D in this construction controls the nominal bridge length rather than guaranteeing an exact realized graph diameter. The realized diameter and all other structural quantities were therefore computed directly from each generated graph before the corresponding analysis.

For the structural correlation analysis, graph instances were generated using $D \in \{4, 6, 8, 10, 12\}$.

This parameter range produced graphs with varying distance characteristics while preserving the defining structural properties of each graph family. The resulting graph instances were subsequently characterized using graph diameter, average shortest-path length, clustering coefficient, edge connectivity, and the number of articulation points. To control stochastic variation and improve reproducibility, a fixed random seed of **42** was

used throughout the experimental pipeline. The random state was initialized before graph generation, source-node selection, binary-label assignment, target-node shuffling, train–test partitioning, and model initialization. The same seed-setting procedure was applied consistently across all graph families and MPNN depths. Eligible target nodes at distance k from the selected source node were randomly shuffled and subsequently divided into training and test subsets using a **0.5** : **0.5** split. This procedure ensured that stochastic components were handled consistently across the experimental conditions investigated.

3.5. Structural Measures

3.5.1. Graph-Level Measures

Several graph-level measures were computed to characterize global distance, local density, and structural robustness.

The graph-theoretic quantities introduced in Section 2 were computed for every graph instance. Graph-level measures included graph diameter, average shortest-path length, clustering coefficient, edge connectivity, and the number of articulation points. Node-level measures included degree and betweenness centrality.

Average shortest-path length, clustering coefficient, edge connectivity, and the number of articulation points were used in the correlation analysis of propagation accuracy.

3.5.2. Node-Level Measures for Structural Interventions

Node degree and betweenness centrality were used to identify nodes for the structural-intervention experiments.

Node degree was defined as the number of edges incident to a node and represented its local connectivity. Betweenness centrality quantified the proportion of shortest paths between node pairs that passed through a given node. Nodes with high betweenness centrality were therefore considered potential global communication bottlenecks.

For the degree-based interventions, all nodes with a degree below or above the mean node degree of the graph were removed, respectively. The degree ranking was computed on the original graph and was not recomputed after node removal. If multiple nodes shared the same degree value, all nodes satisfying the corresponding removal criterion were removed simultaneously. For the betweenness-based intervention, nodes with high betweenness centrality according to the corresponding ranking were removed in the same manner.

For the articulation-point intervention, all articulation points identified in the original graph were removed. Source and target nodes were not excluded from removal and were therefore sampled only from the remaining nodes after the structural modification.

Following each intervention, graph connectivity and the number of connected components were recomputed. If source and target nodes became disconnected, the corresponding sample was analyzed separately.

This procedure ensured that the effects of structural modification and complete graph disconnection were distinguished in the subsequent analysis.

To investigate the relationship between graph structural properties and information propagation performance, Pearson correlation coefficients together with their corresponding p-values were computed. The Pearson correlation coefficient $\mathbf{r}$ measures the strength and direction of the linear relationship between two variables, where values close to 1 indicate a strong positive relationship, values close to -1 indicate a strong negative relationship, and values close to 0 indicate little or no linear association.

The p-value was used to assess the statistical significance of the observed correlations. Small p-values indicate that the observed relationship is unlikely to have occurred by chance. In this work, Pearson correlation analysis was used to evaluate how strongly graph structural

measures, such as average shortest path length, clustering coefficient, edge connectivity, and the number of articulation points, are associated with propagation accuracy.

We conducted an additional distance-controlled analysis to assess whether the accuracy differences observed in the node-removal experiments could be attributed to intervention-induced changes in source–target distance. For each intervention condition, we retained only those source–target pairs for which the shortest-path distance remained unchanged after node removal, i.e., $d_G(s,t) = d_G{}'(s,t)$. Predictions obtained before and after the intervention were compared on the same source–target pairs, resulting in a paired evaluation design. Statistical significance was assessed using the exact McNemar test, which is appropriate for paired binary classification outcomes. This procedure controls for changes in shortest-path distance when comparing propagation performance before and after structural intervention and provides a controlled follow-up analysis of the original node-removal experiments.

4. Results

The experiments used in this work were designed to determine how information propagation in message-passing neural networks is influenced by graph distance, model depth and topological properties. The results are divided into two parts. One part addresses the question of whether diameter alone is sufficient to explain the propagation performance (Sections 4.1–4.4), while the other part focuses on the topological properties, such as articulation points, connectivity patterns, etc.

4.1. Graph Diameter Alone Does Not Explain Propagation Performance

The first set of experiments investigated whether the graph diameter alone provides an explanation for the varying ability of MPNNs to propagate information across the graph. Graphs with larger diameters were expected to have more difficulties propagating information, since information must travel across longer paths before reaching distant nodes. However, the results of the experiments do not support this explanation in isolation.

Graphs with comparable diameters exhibited substantially different propagation behavior depending on their topology. This effect was most visible when comparing path-, grid-, cycle- and barbell-like structures. These graphs can be constructed with similar distance properties, but their accuracy when increasing the source–target distance was not equivalent (see Figures 1 and 2a,b).

4.2. Experimental Validation of Receptive Field Limitation

To analyze the influence of network depth on information propagation, the node label-copying task was evaluated using MPNNs with two, four, six and ten layers. Increasing the source–target distance, we measured the accuracy of the information propagation (see Figure 1).

Across all graph topologies, models with more layers propagated information more accurately along greater distances.

Accuracy clearly decreased when the propagation distance exceeded the number of network layers in accordance with the well-known k-hop limit in k-layer MPNNs.

4.3. Structural Robustness Is More Informative than Average Distance Measures

The correlation analysis further supports the conclusion that simple distance-based metrics are insufficient. For the ten-layer MPNN, mean accuracy showed only a weak positive correlation with average shortest path length, with Pearson' r equal to 0.239 and p equal to 0.310 (Figure 3a). The correlation with clustering coefficient was also weak with Pearson r equal to 0.344 and p equal to 0.137 (see Figure 3b).

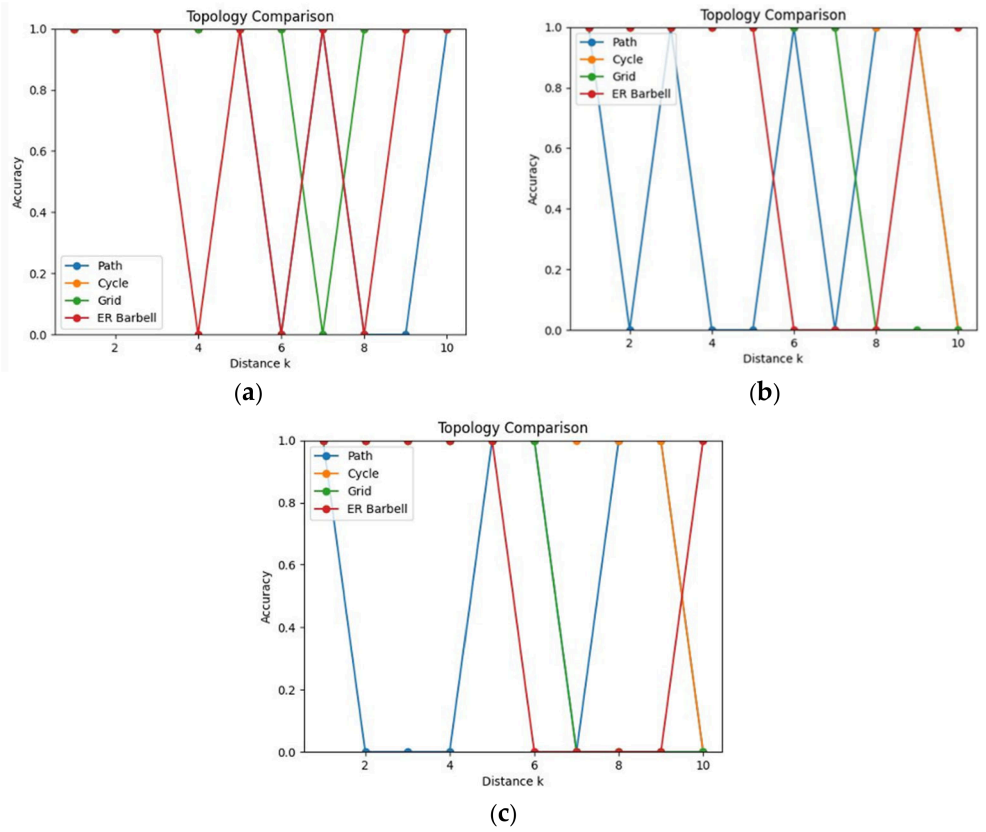


Figure 1. Propagation performance across different graph topologies and network depths ((**a**) -> 2 layers, (**b**) -> 4 layers and (**c**) -> 10 layers). Graphs with comparable distance characteristics exhibit substantially different accuracy patterns.

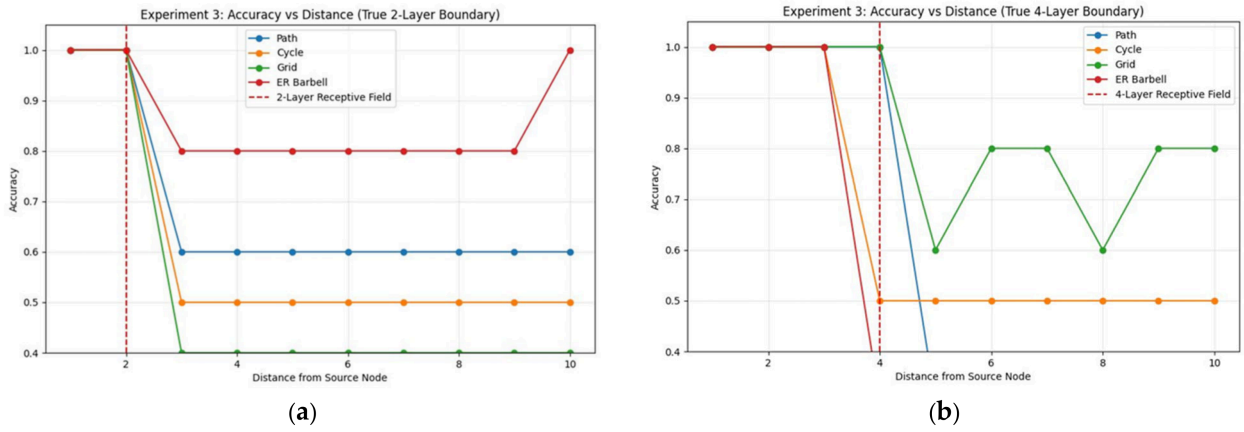
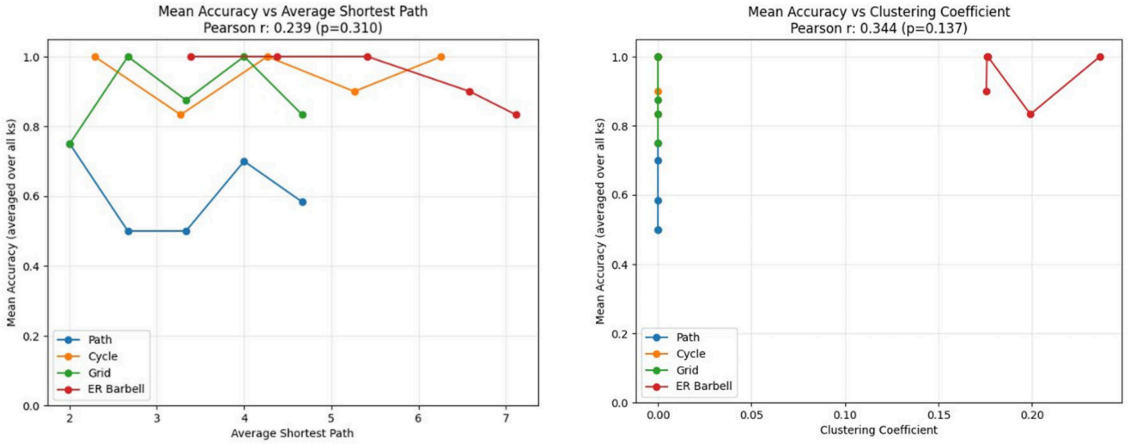


Figure 2. Propagation accuracy as a function of source–target distance for MPNNs with two (**a**) and four (**b**) layers. The dashed vertical lines indicate the theoretical receptive field boundaries of the respective models.

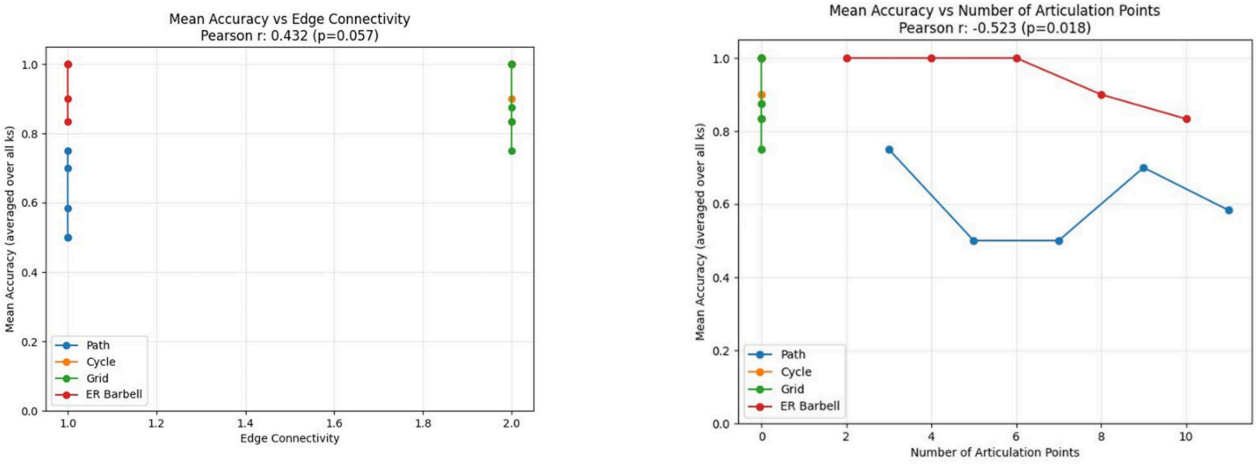
In contrast, structural robustness measures showed a stronger relationship with accuracy. For the ten-layer MPNN, the number of articulation points was negatively correlated with mean accuracy, with Pearson r equal to negative 0.523 and p equal to 0.018 (Figure 4b). Edge connectivity showed a positive relationship with mean accuracy, with Pearson r equal to 0.432 and p equal to 0.057 (Figure 4a). For the two-layer MPNN, this pattern became even more pronounced. The number of articulation points had a strong negative correlation with mean accuracy, with Pearson r equal to negative 0.775 and p equal to 0.001, while edge

connectivity had a strong positive correlation, with Pearson r equal to 0.680 and p equal to 0.001.



(**a**) (**b**)

Figure 3. Mean propagation accuracy in relation to graph structural properties. (**a**) Average shortest path length. (**b**) Clustering coefficient. Pearson correlation coefficients and associated p values are reported for both analyses.



(**a**) (**b**)

Figure 4. Correlation analysis between propagation accuracy and structural robustness metrics. (**a**) Edge connectivity. (**b**) Number of articulation points. Pearson correlation coefficients and p values are shown above each plot.

4.4. Node Removal Experiments and Structural Bottlenecks

The second set of experiments examined information propagation after systematically modifying the graph structure. For each graph topology (Path graph, Erdos–Renyi graphs, Grid graphs and barbell graphs), five settings were considered: the original graph, a graph with low-degree nodes removed, a graph with high-degree nodes removed, a graph with high-betweenness nodes removed and a graph with articulation points removed (for results, see Figure 5). The accuracy of the original graphs ranged from 0.60 to 0.71, with path graphs showing the lowest initial accuracy and grid graphs showing the highest initial accuracy.

Removing low-degree nodes reduced accuracy across all four graph topologies, although the decrease was generally modest, ranging from 0.01 to 0.03, except for the path graph. This suggests that low-degree nodes contribute to propagation performance, but their removal usually had a smaller effect than the removal of structurally central nodes.

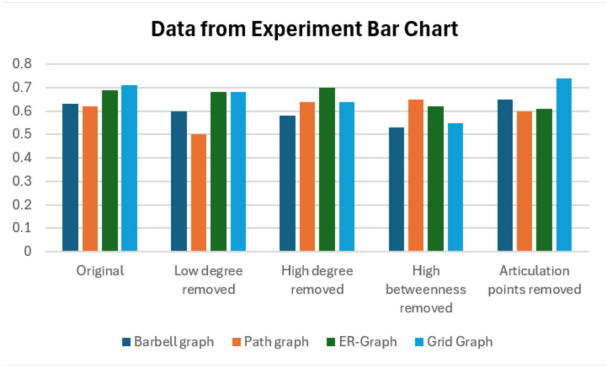


Figure 5. Accuracy (vertical axis) in different experiments across graph topologies. The figure compares the impact of different node-removal strategies on information propagation performance.

Removing high-degree nodes produced a more topology-dependent effect. The accuracy of barbell graphs and grid graphs decreased, with decreases of 0.05 and 0.07 respectively. In contrast, path graphs and Erdos–Renyi graphs showed a small increase of 0.02 and 0.01 respectively.

In the unrestricted intervention analysis, the largest accuracy differences were observed after removing high-betweenness nodes. Barbell graphs and grid graphs showed the largest decreases in accuracy, with decreases of 0.10 and 0.16 respectively. Erdos–Renyi graphs also showed a notable decrease of 0.07, while path graphs changed only slightly with a decrease of 0.01. Removal of articulation points was also associated with changes in propagation accuracy, although the effect differed across the different topologies. Erdos–Renyi graphs showed the strongest decrease in accuracy with articulation points removed, with a decrease of 0.08. Grid graphs decreased by 0.03, path graphs decreased by 0.02 and barbell graphs increased by 0.02.

To assess whether the accuracy differences observed in the structural-intervention experiments were associated with intervention-induced changes in source–target distance, we performed an additional paired analysis restricted to source–target pairs for which the shortest-path distance was preserved before and after node removal, $d_G(s,t) = d_G{}'(s,t)$. Table 1 summarizes the corresponding distance-controlled results.

Table 1. Accuracy of the trained MPNN before and after targeted node removals on source–target pairs whose shortest-path distance remained unchanged.

Graph Family	Intervention	Distance-Preserved Pairs (n)	Accuracy Before	Accuracy After	Δ Accuracy	McNemar p
Barbell	Low degree	10	0.700	0.700	0.000	1.000
Barbell	High degree	8	0.625	0.625	0.000	1.000
Barbell	High betweenness	8	0.625	0.625	0.000	1.000
Barbell	Articulation points	8	0.625	0.625	0.000	1.000
Path	Low degree	7	0.714	0.714	0.000	1.000
Path	High degree	10	0.500	0.500	0.000	1.000
Path	High betweenness	4	0.250	0.750	+0.500	0.625
Path	Articulation points	2	0.000	1.000	+1.000	0.500
Erdős–Rényi	Low degree	49	0.898	0.898	0.000	1.000
Erdős–Rényi	High degree	12	0.583	0.917	+0.333	0.125
Erdős–Rényi	High betweenness	16	0.625	0.500	−0.125	0.727
Erdős–Rényi	Articulation points	82	0.878	0.902	+0.024	0.815
Grid	Low degree	13	0.615	0.615	0.000	1.000
Grid	High degree	14	0.714	0.714	0.000	1.000
Grid	High betweenness	14	0.714	0.714	0.000	1.000
Grid	Articulation points	25	0.760	0.760	0.000	1.000

For each graph family, nodes were removed according to four intervention strategies (low degree, high degree, high betweenness, and articulation points). The table reports the number of distance-preserved evaluation pairs, classification accuracy before and after the intervention, the corresponding accuracy difference (Δ Accuracy), and the p-value of McNemar's test. After controlling for source–target distance, most intervention conditions showed little or no change in propagation accuracy. Differences were observed in several conditions; however, none of the paired comparisons reached statistical significance. The comparatively large differences observed for high-betweenness and articulation-point removal in the path graph should be interpreted cautiously because only four and two distance-preserved source–target pairs, respectively, were available for these comparisons. Overall, the distance-controlled analysis indicates that source–target distance should be considered when interpreting changes in propagation performance following structural node removal.

5. Discussion

The present study does not directly quantify oversquashing using established measures such as effective resistance, commute time, discrete curvature, or Jacobian-based sensitivity metrics. Consequently, the observed relationships between articulation points, edge connectivity, betweenness-related interventions, and propagation accuracy should not be interpreted as direct measurements of oversquashing. Instead, the results demonstrate associations between these topological properties and information propagation performance. While these observations are consistent with hypotheses proposed in the oversquashing literature, they do not establish a causal relationship between the investigated structural properties and oversquashing itself.

The results of the first set of experiments suggest that graph diameter alone is, in general, a weak predictor of information propagation performance. They indicate that information propagation is influenced by multiple topological properties rather than graph diameter alone. Given that graph diameter represents the maximum shortest-path distance between any pair of nodes, it captures only an extreme global distance property and does not reflect the overall communication structure of the graph. Path graphs represent a special case because information can propagate only along a single trajectory. Consequently, propagation performance appears to depend not only on the maximum distance between nodes but also on the availability of alternative communication paths within the graph.

Increasing the number of layers in the MPNN, the results illustrate a significant drop in accuracy after a certain distance k which corresponds to the k-hop limit. This observation is consistent with Proposition 1, which established that an L-layer MPNN can only depend on node features originating from vertices at distance at most L. The empirical propagation boundary observed in Figure 2 therefore reflects a fundamental theoretical limitation of local message-passing architectures rather than an artifact of the experimental setup.

Furthermore, even information originating from nodes within the theoretical receptive field may become increasingly compressed as it is aggregated through successive layers. Such compression is consistent with mechanisms associated with over-squashing and may contribute to reduced effective propagation.

While the k-hop limit establishes a theoretical boundary for information propagation, the experimental results demonstrate that propagation performance within this boundary remains highly dependent on graph topology. Graphs with comparable receptive field sizes exhibited substantial differences in accuracy, indicating that structural characteristics such as alternative communication paths, bottlenecks, and connectivity patterns influence how effectively information can be propagated. Consequently, propagation performance cannot be explained by propagation depth alone.

The experimental results indicate that different classes of topological measures exhibit substantially different relationships with propagation performance. While average topological measures correlate only weakly with mean accuracy, measures reflecting structural robustness show considerably stronger associations.

The average shortest path length exhibits only a weak correlation with mean accuracy. While larger average distances generally imply that information must travel further through the graph, this metric does not capture whether communication is constrained by structural bottlenecks or whether alternative propagation paths are available. Consequently, an increase in average shortest path length does not necessarily result in lower propagation performance. Likewise, reducing the average shortest path length does not guarantee improved accuracy if information flow remains limited by oversquashing or other topological constraints.

The clustering coefficient exhibits only a weak correlation with mean accuracy. While a higher clustering coefficient increases the density of local connections between neighboring nodes, it does not necessarily improve communication between distant regions of the graph. As a result, graphs with substantially different clustering coefficients may still exhibit similar propagation performance. Likewise, a lower clustering coefficient does not necessarily imply poor information propagation as long as sufficient communication paths remain available. These observations suggest that local connectivity alone is not a strong predictor of propagation performance.

On the other hand, the number of articulation points correlates strongly with mean accuracy. A larger number of articulation points indicates a greater presence of structurally critical bottlenecks through which communication between graph regions may depend. Such structures can restrict alternative propagation routes and may therefore increase susceptibility to oversquashing, providing a plausible explanation for the observed association with lower propagation accuracy. By Menger's theorem, graph regions separated by an articulation point are connected through a structurally restricted set of communication routes. In an MPNN, this can require information originating from multiple nodes to be aggregated through a limited communication structure, which may increase susceptibility to information compression and over-squashing. This provides a plausible structural interpretation of the observed association between articulation-point abundance and lower propagation accuracy.

Edge connectivity exhibited a moderate positive association with mean accuracy in the ten-layer configuration and a stronger positive association in the two-layer configuration.

In the context of Section 2.4, larger edge connectivity implies the existence of a greater number of edge-disjoint communication paths between graph regions. Information can therefore propagate through multiple alternative routes rather than being concentrated on a small set of critical edges. The positive association between edge connectivity and propagation accuracy is therefore consistent with the hypothesis that greater communication redundancy may reduce susceptibility to topology-induced information bottlenecks, including mechanisms associated with over-squashing.

The results of the second set of experiments indicate that node degree alone provides limited information about propagation relevance, whereas interventions targeting high-betweenness nodes and articulation points were associated with larger changes in propagation performance.

Removing low-degree nodes may have only a limited impact on propagation performance if sufficient alternative communication paths remain available. In such cases, information can still propagate through other regions of the graph, reducing the effect of the removed nodes on overall accuracy. The exception of path graphs is structurally meaningful

because endpoint removal changes the available source–target configurations in a topology that already has only one route for information flow.

The impact of removing high-degree nodes appears to be strongly topology-dependent. In barbell and grid graphs, the removal of high-degree nodes resulted in a noticeable decrease in accuracy. A possible explanation is that these nodes provide a large number of communication paths and therefore play an important role in maintaining efficient information propagation. Their removal may reduce connectivity, increase structural bottlenecks, and limit the availability of alternative propagation routes. Their removal may reduce connectivity, strengthen structural bottlenecks, and limit the availability of alternative propagation routes. These structural changes may increase susceptibility to information compression and other topology-induced propagation limitations, providing a plausible explanation for the observed decrease in accuracy. Path graphs and Erdos–Resnyi graphs experienced only a negligible increase in accuracy. Path graphs consist almost entirely of nodes with identical degree values, except for the endpoints. Consequently, removing high-degree nodes either removes nearly all internal nodes, disrupting the graph structure entirely, or has no effect on the graph topology at all. In contrast, Erdos–Renyi graphs typically exhibit a more homogeneous degree distribution and do not rely on a small number of high-degree nodes. As a result, removing these nodes may have only a limited effect on the overall communication structure because a large number of alternative paths often remain available for information propagation. This suggests that degree alone is not a reliable indicator of whether a node is important for propagation. A node may have many neighbors but still not be in a globally critical position in the graph.

Removing high-betweenness nodes appears to have the strongest effect on overall accuracy. Betweenness centrality measures the fraction of shortest communication paths that traverse a given vertex. Consequently, removing a node with large betweenness centrality simultaneously eliminates a substantial proportion of the shortest routes available for information propagation. The pronounced accuracy degradation observed in several graph families therefore provides empirical evidence that global communication position is more important than purely local connectivity measures such as degree.

As discussed above, the number of articulation points correlates strongly with the mean accuracy of the label-copying task. Removing articulation points affected propagation performance in a topology-dependent manner. While barbell graphs experienced an increase in accuracy, the remaining graph topologies generally exhibited decreasing accuracies. In most cases, articulation points represent critical communication bottlenecks whose removal disconnects parts of the graph and reduces the availability of alternative propagation paths. Consequently, information flow becomes more restricted, leading to lower propagation performance. The behaviour of barbell graphs suggests that the role of articulation points depends on the surrounding topology. The apparent increase in accuracy observed for barbell graphs after articulation-point removal should be interpreted cautiously. Removing articulation points can disconnect the graph and eliminate difficult source–target configurations that previously required communication across the bottleneck. The remaining evaluation set may therefore contain structurally easier propagation tasks. Consequently, higher accuracy after removal does not necessarily indicate improved global information propagation.

The additional distance-controlled analysis provides a complementary perspective on these intervention results. When source–target distance was preserved, most intervention conditions showed little or no change in propagation accuracy, and none of the paired comparisons reached statistical significance. This indicates that source–target distance and intervention-induced changes in connectivity should be considered when interpreting the accuracy differences observed in the unrestricted node-removal experiments.

Taken together, the experimental results are consistent with the mathematical framework developed in Section 2. Proposition 1 explains the observed propagation boundary in terms of finite-depth reachability, whereas Proposition 2 demonstrates that diameter alone cannot characterize communication structure. The stronger associations observed for articulation points and edge connectivity, together with the larger accuracy changes following high-betweenness node removal in several graph families, suggest that effective information propagation depends not only on how far information must travel but also on the structure and redundancy of the available communication pathways. Within the investigated experimental setting, communication redundancy and bottleneck-related measures therefore provide complementary explanatory information beyond distance-based graph measures alone.

A further limitation of the present study is that all experiments were conducted on synthetic graph families. While synthetic graphs provide controlled settings for isolating the influence of topological properties, the empirical findings should not be assumed to generalize automatically to real-world graph datasets. Future work should therefore investigate whether the observed relationships between graph topology and propagation performance persist in more complex real-world networks.

6. Conclusions

Taken together, the theoretical and experimental results show that distance-related graph properties alone provide an incomplete characterization of information propagation in Message-Passing Neural Networks. The mathematical analysis establishes that message-passing depth determines theoretical reachability, while graphs with identical diameters can differ substantially in their connectivity and communication redundancy. The experimental results complement this analysis by showing that graphs with comparable diameters can nevertheless exhibit substantially different propagation performance.

Within the investigated graph families and experimental settings, average shortest-path length and clustering coefficient showed only weak associations with propagation accuracy, whereas articulation points and edge connectivity exhibited stronger relationships. In the unrestricted node-removal experiments, removing high-betweenness nodes was associated with comparatively large reductions in propagation accuracy for several graph topologies. However, the additional distance-controlled analysis showed that most intervention conditions exhibited little or no change when source–target distance was preserved, and none of the paired comparisons reached statistical significance. These findings indicate that structural-intervention results should be interpreted in conjunction with intervention-induced changes in distance and connectivity.

The results therefore support a distinction between theoretical reachability and effective information propagation. Source–target distance determines the minimum message-passing depth required for information exchange, but satisfying this depth requirement does not guarantee effective transmission. Once nodes are theoretically reachable, propagation performance may additionally depend on the topology of the available communication pathways, including their redundancy, connectivity, and the presence of structural bottlenecks. These structural conditions are consistent with mechanisms discussed in the literature on over-squashing; however, over-squashing was not directly quantified in the present study.

Future work should investigate these relationships using more direct measures of information transmission and bottleneck severity, including curvature-based quantities, effective resistance, commute time, and Jacobian-based sensitivity measures. Such analyses could clarify how distance, connectivity, and communication redundancy jointly affect

information propagation and could support the development of topology-aware graph neural architectures.

A further prospective direction is to extend this framework to quantum and hybrid quantum–classical graph learning. In particular, controlled comparisons could investigate whether topology-induced communication limitations analogous to those observed for classical MPNNs also arise in quantum graph models, and how they interact with information encoding, circuit depth, entanglement structure, and measurement strategies.

Author Contributions: Conceptualization, A.N., M.F.; methodology, M.K. and A.N.; software, M.K.; validation, M.K. and A.N.; formal analysis, M.K. and M.F.; investigation, M.K. and A.N.; writing—original draft preparation, M.K. and A.N.; writing—review and editing, M.K., M.F. and A.N.; visualization, M.K.; supervision, M.F. and A.N.; project administration, A.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study were synthetically generated as part of the experiments described in the manuscript. No external datasets were used. The generated data can be reproduced using the graph-generation procedures and experimental settings described in the Section 3.

Acknowledgments: The authors gratefully acknowledge the Quantum Application Hub (QAH) at the Technical University of Applied Sciences Wildau for providing the scientific environment and support that contributed to this research. QAH is funded by MWFK (Ministry for Science, Research and Culture Brandenburg).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MPNNs	Message-Passing Neural Networks
GAT	Graph Attention Network
GCN	Graph Convolutional Network

References

1. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017*; PMLR: New York, NY, USA, 2017; Volume 70, pp. 1263–1272.
2. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
3. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
4. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. [CrossRef] [PubMed]
5. Li, Q.; Han, Z.; Wu, X.-M. Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018*; AAAI: Washington, DC, USA, 2018; pp. 3538–3545. [CrossRef]
6. Dwivedi, V.P.; Joshi, C.K.; Laurent, T.; Bengio, Y.; Bresson, X. Benchmarking Graph Neural Networks. *J. Mach. Learn. Res.* **2023**, *24*, 1–48. [CrossRef]
7. Oono, K.; Suzuki, T. Graph Neural Networks Exponentially Lose Expressive Power for Node Classification. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 30 April–3 May 2020.
8. Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How Powerful Are Graph Neural Networks? In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
9. Di Giovanni, F.; Rowbottom, J.; Chamberlain, B.P.; Bronstein, M.M. Understanding Graph Neural Networks: Bottlenecks, Over-Smoothing and Over-Squashing. *arXiv* **2024**, arXiv:2402.03300.

10. Loukas, A. What Graph Neural Networks Cannot Learn: Depth vs. Width. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 30 April–3 May 2020.
11. Alon, U.; Yahav, E. On the Bottleneck of Graph Neural Networks and Its Practical Implications. In Proceedings of the International Conference on Learning Representations, Virtual, 3–7 May 2021.
12. Topping, J.; Di Giovanni, F.; Chamberlain, B.P.; Dong, X.; Bronstein, M.M. Understanding Over-Squashing and Bottlenecks on Graphs via Curvature. In Proceedings of the International Conference on Learning Representations, Virtual, 25–29 April 2022.
13. Crespo, J.A.; Atwood, J.; Mahoney, M.W. Improving Graph Neural Networks with Simple Rewiring. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022*; ACM: New York, NY, USA, 2022; pp. 296–307.
14. Banerjee, P.K.; Karhadkar, K.; Wang, Y.G.; Alon, U.; Montúfar, G. Oversquashing in GNNs through the Lens of Information Contraction and Graph Expansion. In *Proceedings of the 58th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 27–30 September 2022*; IEEE Press: New York, NY, USA, 2022; pp. 1–8.
15. Bodnar, C.; Frasca, F.; Wang, Y.; Otter, N.; Montúfar, G.; Liò, P.; Bronstein, M.M. Weisfeiler and Lehman Go Cellular: CW Networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 2625–2640.
16. Zhu, J.; Yan, Y.; Zhao, L.; Heimann, M.; Akoglu, L.; Koutra, D. Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7793–7804.
17. Di Giovanni, F.; Giusti, L.; Barbero, F.; Luise, G.; Liò, P.; Bronstein, M.M. On Over-Squashing in Message Passing Neural Networks: The Impact of Width, Depth, and Topology. In *Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023*; Proceedings of Machine Learning Research; PMLR: New York, NY, USA, 2023; Volume 202, pp. 7865–7885.
18. Black, M.; Wan, Z.; Nayyeri, A.; Wang, Y. Understanding Oversquashing in GNNs through the Lens of Effective Resistance. In *Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023*; Proceedings of Machine Learning Research; PMLR: New York, NY, USA, 2023; Volume 202, pp. 2528–2547.
19. Gutteridge, B.; Dong, X.; Bronstein, M.M.; Di Giovanni, F. DRew: Dynamically Rewired Message Passing with Delay. In *Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023*; Proceedings of Machine Learning Research; PMLR: New York, NY, USA, 2023; Volume 202, pp. 12252–12267.
20. Bamberger, J.; Gutteridge, B.; Le Roux, S.; Bronstein, M.M.; Dong, X. On Measuring Long-Range Interactions in Graph Neural Networks. In *Proceedings of the 42nd International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025*; Proceedings of Machine Learning Research; PMLR: New York, NY, USA, 2025; Volume 267, pp. 2770–2789.
21. Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; Liu, T.-Y. Do Transformers Really Perform Badly for Graph Representation? *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 28877–28888.
22. Menger, K. Zur allgemeinen Kurventheorie. *Fundam. Math.* **1927**, *10*, 96–115. [CrossRef]
23. NetworkX—NetworkX Documentation (Version 3.6.1). Available online: https://networkx.org/en/ (accessed on 23 October 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.