MODELING TOPOLOGICAL IMPACT ON NODE ATTRIBUTE DISTRIBUTIONS IN ATTRIBUTED GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study how graph topology can condition a categorical distribution associated with node-indexed attributes. We introduce an algebraic–categorical framework based on the *point of view* of a node, represented by an under category that captures its node-centered structural perspective. Finite indexed projections of these viewpoints are aggregated into a graph-level monoidal object, separating the abstract topological representation from its concrete matrix realizations. We then introduce a distributional realization in which a supplied categorical distribution induces weighted graph matrices, yielding node-level and graph-level *point-of-view (POV) distributions* for both fixed distributions and sequences of distributions. For a connected graph and a strictly positive fixed distribution, we characterize the common limiting POV distribution by the normalized left Perron eigenvector of the induced weighted matrix. For complete graphs, we further prove recovery of the limiting supplied distribution for convergent distribution sequences, providing a sanity-check boundary when topology carries no distinguishing information. We complement the theoretical development with empirical studies that examine POV across several settings, including recovery behavior, distribution-conditioned graph propagation, topology conditioning of externally supplied node distributions, and representations of evolving policy distributions in reinforcement learning. Together, the framework provides a general mechanism for incorporating graph structure into supplied distributional information through node-centered viewpoints and path-based representations.

1 INTRODUCTION

Attributed graphs combine relational structure with node attributes, which constitute two distinct types of information. In particular, given a categorical distribution associated with node-indexed attributes, it is not immediate how the topology of the graph should modify or reweight that distribution in a principled way. In this paper, we introduce an algebraic–categorical framework for constructing *topology-conditioned distributions*. The central idea is to treat topology and attributes as distinct but interacting components: topology determines node-centered *points of view*, while the supplied categorical distribution provides the distributional information to be reweighted through those viewpoints.

Formally, let $\mathcal{G} = (V, E, X)$ be an attributed graph, where $X = (x_1, \ldots, x_n)$ is the collection of node-indexed attributes, and let P be a categorical distribution over these attribute occurrences. Our goal is to construct new distributions that represent P from the structural position of an individual node $v \in V$ and from the graph $\mathcal{G}$ as a whole. These distributions combine the supplied attribute distribution with the connectivity structure of the graph, while remaining independent of any particular downstream learning architecture.

Central to our approach is the notion that each node in a graph has its own *point of view*: a node-centered perspective on the topology determined by the ways in which that node is connected to the rest of the graph. Different nodes may therefore induce different structural perspectives, even though they belong to the same underlying graph. We formalize these node-level viewpoints and then aggregate their shared structure to obtain a graph-level point-of-view representation.

This perspective is naturally aligned with a basic categorical principle: an object can be studied through the morphisms that originate from it. In category theory, this idea is formalized by the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

under category Adámek et al. (1990), which collects the ways in which a fixed object maps to other objects. We use this construction to formalize the topological point of view of a node, interpreting the paths that originate from that node as the structural relations through which it views the graph.

To make this idea precise, we associate a category with each graph. The classical *free category* generated by a graph, whose objects are nodes and whose morphisms are paths, provides a natural starting point Mac Lane (1998). However, our goal is not only to describe node-centered viewpoints abstractly, but also to connect their aggregation to concrete matrix representations. Building on the algebraic machinery of Grothendieck Graph Neural Networks (GkGNN) Langari et al. (2026), we therefore construct a graph-associated category whose morphisms are represented by monoidal elements that encode composable paths. This allows us to define node viewpoints categorically, aggregate them independently of any particular matrix operator, and introduce matrix realizations only afterward. Our contributions are as follows:

Categorical formalization of topological points of view. We introduce a graph-associated category that determines the underlying graph up to isomorphism and use its under categories to formalize the topological point of view of each node. We then aggregate these node-level viewpoints into the monoidal object $\mathcal{G}^m$, which provides an abstract level-m representation of the graph's aggregated point of view.

Matrix and distributional realizations. We separate the abstract point-of-view object from its concrete representation by introducing a general matrix realization scheme for $\mathcal{G}^m$. As a distribution-induced realization, we associate categorical distributions over node-indexed attributes with weighted graph matrices and obtain node-level and graph-level topology-conditioned distributions. The construction accommodates both a single distribution and sequences of distributions.

Convergence and recovery guarantees. For a connected graph and a strictly positive categorical distribution, we characterize the limiting point-of-view distribution through the left Perron eigenvector of the induced weighted matrix and show that all node-level viewpoints converge to this common limit. We further show that, on complete graphs, the construction recovers the limiting distribution of any convergent sequence of categorical distributions, providing a sanity-check boundary in the absence of distinguishing topological structure.

Empirical evaluation and exploratory applications. We numerically validate the complete-graph recovery behavior, evaluate finite POV matrices as distribution-conditioned propagation representations within an existing graph-learning architecture, and further explore POV for conditioning detector-derived node distributions and constructing topology-conditioned action distributions from policy histories in reinforcement learning.

2 Related Work

Algebraic and Categorical Graph Representations. A graph classically generates a free category whose morphisms are paths (Mac Lane, 1998). More broadly, category theory has been used to describe compositional structure in machine learning, including neural architectures, optimization, attention, and bidirectional learning structures (Gavranović et al., 2024; Gavranović, 2024; Cruttwell et al., 2021; Khatri et al., 2024; Fong & Johnson, 2019). Categorical and related geometric constructions have also been applied directly to structured data, including neural sheaf diffusion (Bodnar et al., 2022), copresheaf-based neural networks (Hajij et al., 2025), hypergraph disentanglement (Lee et al., 2025), and categorical formulations of Weisfeiler–Lehman-type constructions (Choi et al., 2026). Our direct algebraic foundation is the Grothendieck Graph Neural Networks framework (GkGNN) (Langari et al., 2026), which introduces graph-derived monoids and connects their compositions with matrix representations. We build on this machinery to formalize and aggregate node-centered topological viewpoints and to endow the resulting representation with matrix and distributional realizations. The main distinction is therefore the aggregation and distributional interpretation of node-level viewpoints rather than categorical or matrix representation alone.

Matrix-Based Graph Representations and Propagation. Matrix functions provide classical representations of weighted walks. Communicability uses the matrix exponential to aggregate walks with length-dependent weights (Estrada & Hatano, 2008), while dynamic communicability uses ordered products to encode temporal constraints on walks (Grindrod et al., 2011). These constructions are related to our exponential and sequential realizations, although in our sequential setting the graph

remains fixed and the supplied distributions determine the matrices at successive factor positions. In graph learning, polynomial spectral filters (Defferrard et al., 2016), MixHop (Abu-El-Haija et al., 2019), diffusion-based propagation (Gasteiger et al., 2019b), and PPNP/APPNP (Gasteiger et al., 2019a) construct representations using different graph scales or propagation operators. HPGNN further combines personalized PageRank with higher-order topological information (Wang et al., 2025); we use its architecture experimentally to evaluate POV propagation representations while leaving the downstream learning components unchanged. POV is also related to node-dependent edge weighting and biased random walks (Lambiotte et al., 2011). In contrast, the supplied distribution in POV determines the propagation weights themselves, and accumulated weighted paths are normalized to obtain topology-conditioned distributions rather than composing separately normalized one-step transition matrices.

3 Notation and preliminaries

We consider undirected graphs denoted by $\mathcal{G} = (V, E)$ and attributed graphs denoted by $\mathcal{G} = (V, E, X)$, where $V = \{v_1, \ldots, v_n\}$ is the set of nodes equipped with a fixed ordering, E is the set of edges, and $X = (x_1, \ldots, x_n)$ is the indexed collection of node attributes, with x_i assigned to node v_i. The index i therefore identifies a node-indexed attribute occurrence rather than necessarily a distinct attribute value. In particular, $x_i = x_j$ may hold for $i \neq j$. All probability distributions considered in this paper are categorical distributions over these node-indexed attribute occurrences. We identify such a distribution with the row vector $P = (P_1, \ldots, P_n) \in \Delta^{n-1}$, where $\Delta^{n-1} = \{p \in \mathbb{R}^n_{\geq 0} : p\mathbf{1} = 1\}$ and $P_i := P(x_i)$. Here, $\mathbf{1}$ denotes the all-ones column vector of the appropriate dimension. Thus, nodes sharing the same attribute value may still receive separate probability masses because their attribute occurrences are indexed by different nodes. Subgraphs and derived multigraphs are denoted by $\mathcal{D}$ and $\mathcal{M}$, respectively, and inherit the node labels of their ambient graph without reindexing. Single-node subgraphs are allowed, and edge deletion retains the vertex set, including any vertices that become isolated. For a finite set S, $|S|$ denotes its cardinality. Matrices are denoted by uppercase bold symbols such as $\boldsymbol{A}$. The entry at row i and column j is denoted by $\boldsymbol{A}_{i,j}$, the i-th row by $\boldsymbol{A}_{i,:}$, and the j-th column by $\boldsymbol{A}_{:,j}$. We write $\boldsymbol{I}$ for the identity matrix of the appropriate dimension. For a vector z, $\|z\|_1$ denotes its ℓ_1-norm, and for a square matrix $\boldsymbol{A}$, $\rho(\boldsymbol{A})$ denotes its spectral radius. Definitions of *monoid*, *category*, *under category*, and *functor* are provided in Appendix A.

Grothendieck Graph Neural Networks (GkGNN) Framework The GkGNN framework introduces two monoidal operations, denoted by $\bullet$ and $\circ$, along with four monoids:

$$(\mathsf{Mod}(\mathcal{G}), \bullet) \subseteq (\mathsf{SMult}(\mathcal{G}), \bullet), (\mathsf{Mom}(\mathcal{G}), \circ) \subseteq (\mathsf{Mat}_{|V|}(\mathbb{R}), \circ)$$

and a monoidal homomorphism Tr that maps elements of $\mathsf{Mod}(\mathcal{G})$ to their matrix representations. We summarize the relevant concepts from Langari et al. (2026) that are used throughout the paper:

Definition 3.1. Let $\mathcal{G} = (V, E)$ be an undirected graph.

1. A **path** p from node v_{p_1} to node v_{p_m} is an ordered sequence $v_{p_1}, e_{p_1}, v_{p_2}, e_{p_2}, \ldots, v_{p_{m-1}}, e_{p_{m-1}}, v_{p_m}$, where e_{p_i} represents an edge connecting nodes v_{p_i} and $v_{p_{i+1}}$. Highlighting the edges between nodes, this definition of a path is applicable to multigraphs where multiple edges can exist between two nodes.

2. A **directed subgraph** $\mathcal{D}$ of $\mathcal{G}$ is a connected and acyclic subgraph of $\mathcal{G}$ where every edge of $\mathcal{D}$ is directed.

3. The monoid $\mathsf{SMult}(\mathcal{G})$ is the set of ordered pairs $(\mathcal{M}, S)$, where $\mathcal{M} = \bigoplus_{i=1}^{k} \mathcal{D}_i$ is a multigraph obtained by taking the union of the sets of nodes and the disjoint union of the sets of directed edges from directed subgraphs $\mathcal{D}_i$, and $S \subseteq \mathsf{Paths}(\mathcal{M})$ is a set of paths in $\mathcal{M}$. The binary operation $\bullet$ on $\mathsf{SMult}(\mathcal{G})$ is defined as $(\mathcal{M}, S) \bullet (\mathcal{N}, T) = (\mathcal{M} \bigoplus \mathcal{N}, S \star T)$, where $S \star T$ is the union of S, T (as subsets of $\mathsf{Paths}(\mathcal{M} \bigoplus \mathcal{N})$), and the collection of paths formed by composing paths in S with paths in T.

4. $\mathsf{Mod}(\mathcal{G})$ is defined as the submonoid of $\mathsf{SMult}(\mathcal{G})$ generated by the set of ordered pairs $(\mathcal{D}, \mathsf{Paths}(\mathcal{D}))$, where $\mathcal{D}$ is a directed subgraph of $\mathcal{G}$ and $\mathsf{Paths}(\mathcal{D})$ is the set of all paths in $\mathcal{D}$ adhering to the directions. Where there is no risk of confusion, we denote $(\mathcal{D}, \mathsf{Paths}(\mathcal{D}))$ simply as $\mathcal{D}$. In particular, for a directed edge e, the element $(e, \{e\})$ of $\mathsf{Mod}(\mathcal{G})$ is abbreviated as e.

5. A **cover** for a graph $\mathcal{G}$ is a collection of finitely many elements of $\mathsf{Mod}(\mathcal{G})$.

6. The monoidal operation $\circ$ on $\mathsf{Mat}_n(\mathbb{R})$, the set of all $n \times n$ matrices, is defined as follows:

$$\boldsymbol{A} \circ \boldsymbol{B} = \boldsymbol{A} + \boldsymbol{B} + \boldsymbol{A}\boldsymbol{B}.$$

4 POINTS OF VIEW OF NODES ON GRAPH TOPOLOGY

In this section, we develop a categorical framework for formalizing the *points of view* of nodes with respect to the topology of a graph $\mathcal{G} = (V, E)$. We first associate a category with $\mathcal{G}$ and use its under categories to represent node-centered topological viewpoints. We then study the relationships among these viewpoints and aggregate finite projections of them into a level-m monoidal object $\mathcal{G}^m \in \mathsf{SMult}(\mathcal{G})$. This aggregated point-of-view object is defined independently of any particular matrix representation; we subsequently introduce matrix realizations that provide concrete representations of its structure.

4.1 CATEGORICAL INTERPRETATION OF POINTS OF VIEW OF NODES

To represent the points of view of nodes categorically, we define a category associated with the graph $\mathcal{G}$. The objects of this category are the nodes of $\mathcal{G}$, and its morphisms are constructed from certain elements of $\mathsf{Mod}(\mathcal{G})$ as follows.

Let $e_1, \ldots, e_k$ be a sequence of directed edges such that the target of e_i coincides with the source of e_{i+1} for $1 \leq i < k$. We call such edges *composable*. If u is the source of e_1 and v is the target of e_k, then the monoidal element $e_1 \bullet \cdots \bullet e_k$ defines a morphism from u to v. Together with the identity morphism at each node, these morphisms define the structure $\mathsf{Cat}(\mathcal{G})$.

Theorem 4.1. *For a graph $\mathcal{G} = (V, E)$, the structure $\mathsf{Cat}(\mathcal{G})$, together with identities $0 : v \to v$, where 0 mirrors the role of the monoidal identity of $\mathsf{SMult}(\mathcal{G})$ in compositions, forms a category.*

The category $\mathsf{Cat}(\mathcal{G})$ retains the topological structure of the underlying graph, as formalized by the following theorem.

Theorem 4.2. *Two graphs $\mathcal{G}$ and $\mathcal{H}$ are isomorphic if and only if their corresponding categories $\mathsf{Cat}(\mathcal{G})$ and $\mathsf{Cat}(\mathcal{H})$ are isomorphic.*

Thus, $\mathsf{Cat}(\mathcal{G})$ determines the graph up to isomorphism. This makes the under category associated with a node a natural object for describing its topological perspective on the graph.

Points of view of nodes For a node $v \in V$, we define its *point of view* on the topology of the graph as the under category

$$v/\mathsf{Cat}(\mathcal{G}),$$

which we denote more simply by $v/\mathcal{G}$. The objects of $v/\mathcal{G}$ correspond to morphisms originating at v, and hence to composable paths through which v is connected to other nodes of the graph. Thus, $v/\mathcal{G}$ provides a node-centered topological perspective on $\mathcal{G}$. Moreover, these perspectives are disjoint at the level of objects: for distinct nodes $v_i, v_j \in V$ with $i \neq j$, the categories $v_i/\mathcal{G}$ and $v_j/\mathcal{G}$ have no objects in common.

4.2 AGGREGATION OF POINTS OF VIEW OF NODES

Although the point of view $v/\mathcal{G}$ is defined separately for each node v, these viewpoints are related through the connectivity of the graph. In particular, a directed edge $e : u \to v$ induces a map from the point of view of v to that of u by composition. The following theorem shows that this map is functorial.

Theorem 4.3. *Every directed edge $e : u \to v$ induces a functor $e \bullet - : v/\mathcal{G} \longrightarrow u/\mathcal{G}$.*

More generally, composition along a path relates the points of view of its endpoints. Indeed, for composable morphisms e and c, $(c \bullet e) \bullet - = (c \bullet -)(e \bullet -)$. Thus, graph connectivity induces structured relationships among the node-level points of view. This motivates aggregating these viewpoints into a common representation. Working directly with the under categories $v/\mathcal{G}$, however,

is difficult because they may contain infinitely many objects corresponding to paths of arbitrary length. We therefore organize finite portions of these viewpoints by aggregation level. For each $m \in \mathbb{N}$, we construct a finite indexed family of objects obtained from composable sequences of m edge or identity factors. These level-m families serve as finite projections of the node points of view and provide the bridge from the under-category description to the aggregated monoidal object $\mathcal{G}^m$.

Cover of points of view For $m \in \mathbb{N}$, let I_m denote the set of composable m-tuples $\alpha = (a_1, \ldots, a_m)$, where each a_i is either a directed edge of $\mathcal{G}$ or an identity morphism $0 : v \to v$ at some node of $v \in V$. Thus, the target of a_i coincides with the source of a_{i+1} for $1 \le i < m$. For $\alpha = (a_1, \ldots, a_m) \in \mathsf{I}_m$, define

$$X_\alpha = (\mathcal{M}_\alpha, S_\alpha) := a_1 \bullet \cdots \bullet a_m.$$

An identity factor contributes the monoidal identity and therefore contributes no directed edge to $\mathcal{M}_\alpha$. We define the level-m cover as the indexed family

$$\mathsf{Cov}(m) = \big(X_\alpha\big)_{\alpha \in \mathsf{I}_m}.$$

The index α is retained as part of the cover data, even when different composable m-tuples determine the same element of $\mathsf{Mod}(\mathcal{G})$. Hence, the m factor positions remain distinguished, while identity factors allow positions at which no directed edge is contributed. More explicitly, for $\alpha = (a_1, \ldots, a_m) \in \mathsf{I}_m$, let $J_\alpha = \{i \in \{1, \ldots, m\} : a_i \text{ is a directed edge of } \mathcal{G}\}$. Then $\mathcal{M}_\alpha$ is the directed multigraph obtained from the disjoint union of the directed-edge occurrences a_i for $i \in J_\alpha$. The set $S_\alpha \subseteq \mathsf{Paths}(\mathcal{M}_\alpha)$ consists of the nonempty paths $a_{p_1}, a_{p_2}, \ldots, a_{p_k}$, where $p_1 < \cdots < p_k, \quad p_1, \ldots, p_k \in J_\alpha$, and the selected directed-edge occurrences form a composable path in $\mathcal{M}_\alpha$. Thus, identity factors contribute neither edges nor paths, while their positions remain recorded by the index α.

For $\alpha = (a_1, \ldots, a_m), \beta = (b_1, \ldots, b_m)$ in I_m, let $a_{\alpha,i}$ and $b_{\beta,j}$ denote the directed-edge occurrences corresponding to a_i and b_j in $\mathcal{M}_\alpha$ and $\mathcal{M}_\beta$, respectively, whenever $i \in J_\alpha$ and $j \in J_\beta$. We identify occurrences that represent the same directed edge of $\mathcal{G}$ in the same factor position. Formally,

$$\text{define:} \qquad a_{\alpha,i} \sim b_{\beta,j} \quad \Longleftrightarrow \quad i = j \quad \text{and} \quad a_i = b_j \text{ as directed edges of } \mathcal{G}.$$

This is an equivalence relation on the directed-edge occurrences appearing across $\mathsf{Cov}(m)$. Identity factors contribute no edge occurrences and hence do not enter this relation.

Graph $\mathcal{G}$ as a monoidal element. Let $\mathsf{DE}(\mathcal{G})$ denote the set of all directed edges of $\mathcal{G}$. We identify $\mathcal{G}$ with the element $\Big(\bigoplus_{e \in \mathsf{DE}(\mathcal{G})} e,\ \mathsf{DE}(\mathcal{G})\Big) \in \mathsf{SMult}(\mathcal{G})$. The following theorem shows that aggregating the level-m cover through the equivalence relation $\sim$ recovers the m-fold monoidal product of this element.

Theorem 4.4. *For each $\alpha \in \mathsf{I}_m$, let $\nu_\alpha(S_\alpha)$ denote the image of S_α in $\mathsf{Paths}\left(\bigoplus_{\alpha \in \mathsf{I}_m} \mathcal{M}_\alpha\right)$ under the natural inclusion. Then*

$$\left(\bigoplus_{\alpha \in \mathsf{I}_m} \mathcal{M}_\alpha,\ \bigcup_{\alpha \in \mathsf{I}_m} \nu_\alpha(S_\alpha)\right) / \sim \ \cong\ \underbrace{\mathcal{G} \bullet \cdots \bullet \mathcal{G}}_{m \text{ times}}.$$

Definition 4.5. The *aggregated point of view of level* m is the monoidal element

$$\mathcal{G}^m := \underbrace{\mathcal{G} \bullet \cdots \bullet \mathcal{G}}_{m \text{ times}} \in \mathsf{SMult}(\mathcal{G}).$$

By Theorem 4.4, $\mathcal{G}^m$ is the quotient-level aggregation of the indexed elements of $\mathsf{Cov}(m)$. In particular, each element of the cover has an image in $\mathcal{G}^m$, so the resulting monoidal object combines the finite level-m projections of the node points of view into a single graph-level representation.

4.3 Matrix realizations of the aggregated point of view

The monoidal homomorphism Tr in the GkGNN framework maps elements of $\mathsf{Mod}(\mathcal{G})$ to matrices by counting paths between nodes. Since $\mathcal{G}^m$ belongs to the larger monoid $\mathsf{SMult}(\mathcal{G})$ and need not lie in $\mathsf{Mod}(\mathcal{G})$, Tr is not directly applicable. Nevertheless, its path-counting interpretation extends naturally to $\mathcal{G}^m$, as follows.

Theorem 4.6. *Let $\mathcal{G}^m = (\mathcal{M}, S) \in \mathsf{SMult}(\mathcal{G})$, and let $\boldsymbol{A}$ be the adjacency matrix of $\mathcal{G}$. Then the number of paths in S from v_i to v_j is $(\boldsymbol{A} \circ \cdots \circ \boldsymbol{A})_{i,j}$ ($m-$times).*

Repeated $\circ$-composition admits the following closed form.

Theorem 4.7. *For any square matrix $\boldsymbol{B}$, $\boldsymbol{B}\circ\cdots\circ\boldsymbol{B}(m-\text{times}) = (\boldsymbol{I}+\boldsymbol{B})^m - \boldsymbol{I} = \sum_{k=1}^{m} \binom{m}{k} \boldsymbol{B}^k$.*

Theorems 4.6 and 4.7 provide a canonical matrix representation of the aggregated object $\mathcal{G}^m$. More generally, we separate the abstract object $\mathcal{G}^m$ from the particular matrices used to realize it.

Definition 4.8 (Matrix realization scheme)**.** A matrix realization scheme for $\{\mathcal{G}^m\}_{m\geq 1}$ is a collection

$$\mathcal{B} = \{\boldsymbol{B}_m\}_{m\geq 1}, \qquad \boldsymbol{B}_m = (\boldsymbol{B}_{m,1}, \ldots, \boldsymbol{B}_{m,m}),$$

where each $\boldsymbol{B}_{m,r} \in \mathsf{Mat}_{|V|}(\mathbb{R})$ is derived from the graph topology, possibly together with node- or attribute-dependent weights. The realization of $\mathcal{G}^m$ under $\mathcal{B}$ is

$$\mathsf{Real}_{\mathcal{B}}(\mathcal{G}^m) := \boldsymbol{B}_{m,1} \circ \cdots \circ \boldsymbol{B}_{m,m}.$$

The realization scheme is not unique: different choices of $\mathcal{B}$ provide different matrix-level representations of the same abstract aggregated point-of-view object. In particular, taking $\boldsymbol{B}_m = (\boldsymbol{A}, \ldots, \boldsymbol{A})$ for every $m \geq 1$ recovers **the canonical structural realization**

$$\mathsf{Real}_{\mathcal{B}}(\mathcal{G}^m) = \underbrace{\boldsymbol{A} \circ \cdots \circ \boldsymbol{A}}_{m \text{ times}} = (\boldsymbol{I} + \boldsymbol{A})^m - \boldsymbol{I} = \sum_{k=1}^{m} \binom{m}{k} \boldsymbol{A}^k,$$

which, by Theorem 4.6, counts the selected paths encoded by $\mathcal{G}^m$. As a different realization, taking $\boldsymbol{B}_m = \left(\frac{\boldsymbol{A}}{m}, \ldots, \frac{\boldsymbol{A}}{m}\right)$ gives

$$\lim_{m\to\infty} \mathsf{Real}_{\mathcal{B}}(\mathcal{G}^m) = \lim_{m\to\infty} \left(\boldsymbol{I} + \frac{\boldsymbol{A}}{m}\right)^m - \boldsymbol{I} = e^{\boldsymbol{A}} - \boldsymbol{I}.$$

5 Merging points of view of nodes with distributions on attributes

In this section, we incorporate categorical distributions over the node-indexed attribute occurrences X into the aggregated point-of-view construction. Let

$$\mathcal{P} = \{P^{(k)}\}_{k\geq 1}$$

be a sequence of such distributions. At each level m, the first m distributions induce a matrix realization of $\mathcal{G}^m$, from which we obtain topology-conditioned distributions at both the node and graph levels. The single-distribution setting is recovered by taking the constant sequence $P^{(k)} = P$. We then study the asymptotic behavior of the resulting point-of-view distributions, first for a fixed distribution on connected graphs and subsequently, as a sanity-check boundary, for convergent distribution sequences on complete graphs.

5.1 Distributional realization of aggregated points of view

We specialize the matrix realization scheme of Definition 4.8 to the sequence $\mathcal{P} = \{P^{(k)}\}_{k\geq 1}$ by associating each $P^{(k)}$ with a weighted adjacency-derived matrix $\boldsymbol{W}^{(k)}$.

The choice of weighting is central to the distributional realization. We use the destination-dependent odds weight $p/(1-p)$. Its suitability is supported by the complete-graph boundary studied below: when topology provides no distinguishing structural information, the resulting point-of-view distributions recover the supplied limiting distribution.

Definition 5.1 (Distributional realization)**.** Let $\mathcal{G} = (V, E, X)$ be an attributed graph with adjacency matrix $\boldsymbol{A}$, and let $\mathcal{P} = \{P^{(k)}\}_{k\geq 1}$ be a sequence of categorical distributions over X, with $P^{(k)}_j < 1$ for every j and k. Each $P^{(k)}$ induces the weighted adjacency matrix

$$\boldsymbol{W}^{(k)} := \boldsymbol{W}(P^{(k)}), \qquad \boldsymbol{W}(P^{(k)})_{i,j} = \boldsymbol{A}_{i,j} \frac{P^{(k)}_j}{1 - P^{(k)}_j}.$$

For the level-m object $\mathcal{G}^m$, the ordered collection $\boldsymbol{B}_m = \left(\boldsymbol{W}^{(1)}, \ldots, \boldsymbol{W}^{(m)}\right)$ defines its *distributional realization*

$$\mathsf{DRA}(\mathcal{P}, m) := \boldsymbol{W}^{(1)} \circ \cdots \circ \boldsymbol{W}^{(m)}.$$

Using $\boldsymbol{I} + (\boldsymbol{A} \circ \boldsymbol{B}) = (\boldsymbol{I} + \boldsymbol{A})(\boldsymbol{I} + \boldsymbol{B})$, the distributional realization can equivalently be written as

$$\mathsf{DRA}(\mathcal{P}, m) = \prod_{k=1}^{m} \left(\boldsymbol{I} + \boldsymbol{W}^{(k)}\right) - \boldsymbol{I},$$

where the product is taken in increasing order of k.

Point-of-view distributions The distributional realization induces node-level and graph-level distributions by row normalization and aggregation.

Definition 5.2. Let $\boldsymbol{M} = \mathsf{DRA}(\mathcal{P}, m)$, and assume $\|\boldsymbol{M}_{i,:}\|_1 > 0$ for every i. The point of view of node v_i and the point of view of the graph $\mathcal{G}$, induced by $\mathcal{P}$ at level m, are defined respectively by

$$\mathsf{pov}(v_i, \mathcal{P}, m) = \frac{\boldsymbol{M}_{i,:}}{\|\boldsymbol{M}_{i,:}\|_1}, \qquad \mathsf{pov}(\mathcal{G}, \mathcal{P}, m) = \frac{1}{|V|} \sum_{i=1}^{|V|} \mathsf{pov}(v_i, \mathcal{P}, m).$$

We denote by $\mathsf{POV}(\mathcal{P}, m)$ the matrix whose i-th row is $\mathsf{pov}(v_i, \mathcal{P}, m)$.

We next consider the constant-sequence case $P^{(k)} = P$ for all k. We write $\boldsymbol{W} = \boldsymbol{W}(P)$, and use $\mathsf{DRA}(P, m)$, $\mathsf{pov}(v_i, P, m)$, $\mathsf{pov}(\mathcal{G}, P, m)$, and $\mathsf{POV}(P, m)$ for the corresponding quantities.

Theorem 5.3 (Convergence of POV on connected graphs)**.** *Let $\mathcal{G} = (V, E, X)$ be a connected undirected attributed graph with $|V| \geq 2$, and let P be a categorical distribution on X satisfying $P_i > 0$ for every i. Let $\ell > 0$ be a left Perron eigenvector of $\boldsymbol{W}$, so that $\ell \boldsymbol{W} = \rho(\boldsymbol{W})\ell$. Then, for every $v_i \in V$,*

$$\lim_{m \to \infty} \mathsf{pov}(v_i, P, m) = \lim_{m \to \infty} \mathsf{pov}(\mathcal{G}, P, m) = \frac{\ell}{\ell \mathbf{1}}.$$

We denote this common limit by $\mathsf{pov}(\mathcal{G}, P, \infty)$.

5.2 Complete-graph sanity-check boundary

We use complete graphs as a sanity-check boundary for the proposed point-of-view distributions. In a complete graph, all nodes have the same connectivity pattern, so the topology does not distinguish one node from another. We therefore adopt the following boundary assumption.

Assumption. *On a complete graph, the topology does not alter the supplied distribution over node-indexed attributes.*

Accordingly, if a sequence of categorical distributions converges to P, the corresponding point-of-view distributions should recover P as the aggregation level increases. The following theorem establishes this property.

Theorem 5.4 (Complete-graph recovery for a convergent sequence)**.** *Let $\mathcal{G} = (V, E, X)$ be a complete attributed graph with $|V| \geq 2$, and let $\mathcal{P} = \{P^{(k)}\}_{k \geq 1}$ be a sequence of categorical distributions on X satisfying $P_j^{(k)} < 1$ for all j and k. Suppose $P^{(k)} \longrightarrow P$, where P is a categorical distribution with $P_j < 1$ for all j. Then, for every $v_i \in V$,*

$$\lim_{m \to \infty} \mathsf{pov}(v_i, \mathcal{P}, m) = \lim_{m \to \infty} \mathsf{pov}(\mathcal{G}, \mathcal{P}, m) = P.$$

The constant-distribution setting is recovered by taking $P^{(k)} = P$ for all k.

Theorem 5.4 shows that, on a complete graph, the point-of-view distributions recover the limiting distribution P of every convergent sequence satisfying the assumptions above. We now relate this complete-graph boundary to a connected graph $\mathcal{G} = (V, E)$ with $|V| \geq 2$. Let $\mathcal{C}$ be the complete supergraph on the same fixed, indexed vertex set. For $(\mathcal{M}, S), (\mathcal{N}, T) \in \mathsf{SMult}(\mathcal{G})$, write $(\mathcal{M}, S) \leq (\mathcal{N}, T)$ if there exists an injective directed-multigraph morphism $\iota : \mathcal{M} \hookrightarrow \mathcal{N}$ such that $\iota_V(v_i) = v_i$ for every $v_i \in V(\mathcal{M})$ and $\iota(S) \subseteq T$. These embeddings need not be induced. Let $\mathbf{In} : \mathsf{SMult}(\mathcal{G}) \to \mathsf{SMult}(\mathcal{C})$ denote the natural inclusion. The full order-theoretic construction, under the same conventions and connected-graph assumption, is given in Appendix B.

Theorem 5.5. *Let $\mathcal{C} \supseteq \mathcal{G}$ be the complete graph obtained by adding all missing edges to $\mathcal{G}$. Then $\mathcal{G}^m$ is a greatest element, with respect to $\leq$, of $\{X \in \mathsf{SMult}(\mathcal{G}) : \mathbf{In}(X) \leq \mathcal{C}^m\}$.*

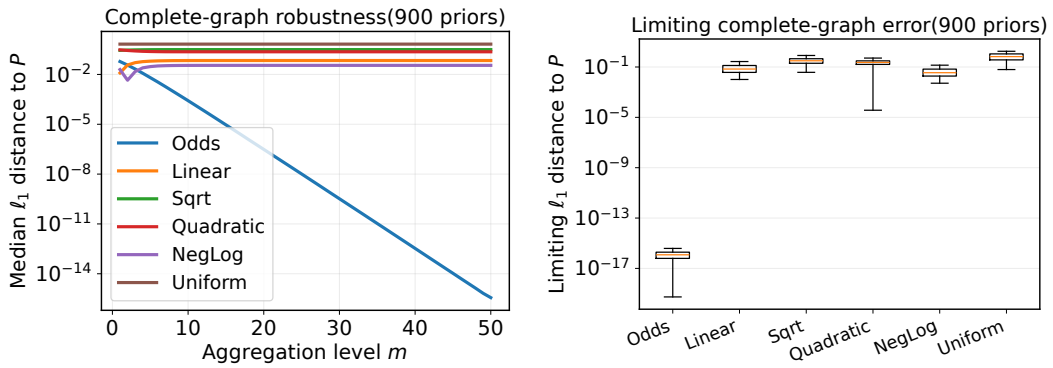


Figure 1: Complete-graph numerical validation over 900 sampled categorical distributions. **Left:** Median graph-level ℓ_1 distance from the input distribution P as the aggregation level m increases. **Right:** Distribution of the independently computed limiting ℓ_1 error. Alternative weightings can be closer to P at shallow aggregation levels, whereas Odds converges to P to numerical precision.

6 Experiments

In the main text, we evaluate two central aspects of the proposed distributional realization. We first numerically examine the complete-graph recovery boundary established in Theorem 5.4, including its finite-m behavior relative to alternative destination weightings. We then evaluate finite POV matrices as distribution-conditioned propagation representations within an existing graph-learning architecture. Additional exploratory experiments are provided in the appendix: Appendix E studies graph-level POV conditioning of externally supplied detector-derived distributions, and Appendix F explores policy-history distributions with action graphs in reinforcement learning.

6.1 Numerical Validation of Complete-Graph Recovery

We first isolate the complete-graph setting, where Theorem 5.4 guarantees recovery of the limiting input distribution. We consider its constant-sequence special case, $P^{(k)} = P$, and examine both finite aggregation levels and the limiting behavior. In addition to the proposed Odds weighting $p/(1-p)$, we consider Linear, Sqrt, Quadratic, NegLog, and Uniform destination weightings. These alternatives test whether complete-graph recovery is a generic consequence of destination-dependent weighting or a specific property of the proposed construction. We use complete graphs with $|V| \in \{5, 10, 20\}$, $P \sim \mathrm{Dirichlet}(\alpha \mathbf{1})$, $\alpha \in \{0.2, 1, 5\}$. For each of the nine $(|V|, \alpha)$ settings, we sample 100 independent categorical distributions, yielding 900 distributions in total. To control for differences in overall weight magnitude at finite aggregation levels, each alternative weighting is rescaled to match the total destination weight of Odds. For every weighting, we measure the graph-level ℓ_1 distance to P for $m = 1, \ldots, 50$ and independently compute its limiting distribution.

Figure 1 distinguishes finite-depth proximity from asymptotic recovery. Alternative weightings can be closer to P at shallow aggregation levels, whereas the Odds construction approaches P as m increases and recovers it to numerical precision in the independently computed limit. Thus, the experiment numerically illustrates the recovery behavior established by Theorem 5.4 and shows that the same asymptotic behavior does not arise automatically from the alternative destination weightings tested here. Full construction details, quantitative results, and robustness analyses are provided in Appendix C.

6.2 POV as Distribution-Conditioned Propagation in HPGNN

We next test whether finite POV matrices can replace standard propagation representations inside an existing graph-learning architecture while leaving the downstream learner unchanged. We use the official HPGNN (Wang et al., 2025) implementation and replace only the PPR and HiPPR propagation matrices supplied to its two second-order spectral-filtering branches. The downstream HPGNN architecture and training procedure are otherwise unchanged. For an attributed graph

Table 1: Node-classification accuracy (%). The HPGNN row is reproduced from Table 2 of HPGNN (Wang et al., 2025) and was not rerun by us.

Method	Cora	Citeseer	Photo	Cornell	Actor	Texas	Wisconsin
HPGNN	89.96 ± 0.13	79.89 ± 0.12	95.19 ± 0.14	88.30 ± 0.42	40.89 ± 0.10	91.48 ± 0.27	94.75 ± 0.33
HPGNN (POV views)	89.29 ± 0.56	79.95 ± 0.72	94.60 ± 0.19	89.72 ± 1.96	40.05 ± 0.50	92.02 ± 0.83	96.04 ± 1.50

$\mathcal{G} = (V, E, X)$, let $d_i = \sum_{j=1}^{|V|} \boldsymbol{A}_{i,j}$. We construct two node-indexed categorical distributions,

$$P_D(x_i) = \frac{d_i}{\sum_{j=1}^{|V|} d_j}, \qquad P_F(x_i) = \frac{\|x_i\|_1}{\sum_{j=1}^{|V|} \|x_j\|_1}, \tag{1}$$

which provide degree-conditioned and feature-conditioned views, respectively. Neither distribution uses node labels. For each distribution and aggregation level $m \in \{1, 2, 3\}$, we construct

$$\boldsymbol{D}_m := \mathsf{POV}(P_D, m), \qquad \boldsymbol{F}_m := \mathsf{POV}(P_F, m), \tag{2}$$

giving the candidate propagation views $\mathcal{Q}_{\mathrm{POV}} = \{\boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3, \boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3\}$. Since second-order HPGNN uses two propagation branches, we consider the 15 unordered pairs of distinct POV views and select one pair using validation data. No additional architectural component is introduced.

Protocol. We implement the POV variant directly within the official HPGNN repository and retain its graph loading, $60\%/20\%/20\%$ train/validation/test splits, split seeds, downstream architecture, optimizer, validation-loss checkpointing, and early stopping. Hyperparameters and the POV-view pair are selected using validation accuracy only. After selection, all choices are frozen and evaluated on all ten graph splits with ten fixed model repetitions per split. Repetitions are first averaged within each split, and we report the mean and 95% confidence interval across the ten split-level means. To remain aligned with the repository implementation, POV is constructed from the edge representation returned by its graph loader without additional symmetrization; this downstream experiment evaluates finite POV propagation matrices rather than the asymptotic undirected-graph guarantees.

Results. Table 1 shows that replacing the original HPGNN propagation views with POV yields competitive performance across all seven datasets, with higher reported point estimates than the published HPGNN results on Citeseer, Cornell, Texas, and Wisconsin. The validation-selected pairs include degree-conditioned, feature-conditioned, and mixed POV representations at different aggregation levels, indicating that no single conditioning distribution or aggregation level is preferred across all datasets. These results show that finite POV matrices can function as distribution-conditioned propagation representations within HPGNN without modifying its downstream architecture. Because HPGNN was not rerun in our environment, comparisons with its published results are descriptive rather than paired statistical tests. Full implementation, model-selection, selected-configuration, statistical, and computational details are provided in Appendix D.

7 Conclusion

We introduced an algebraic–categorical framework for modeling how graph topology can condition distributions over node-indexed attributes. The construction formalizes node-centered points of view through under categories, aggregates them into the monoidal object $\mathcal{G}^m$, and separates this abstract representation from its matrix and distributional realizations. For the odds-weighted realization, we characterize the common limiting POV distribution on finite connected graphs and establish complete-graph recovery for convergent distribution sequences under the stated assumptions. Empirically, we validate the complete-graph recovery behavior, evaluate finite POV matrices as distribution-conditioned propagation representations, and explore detector-conditioned distributions and topology-conditioned action distributions from policy histories in reinforcement learning. These studies examine selected realizations and application settings rather than the full potential of the framework; broader characterization of useful realizations, conditioning distributions, aggregation levels, and application domains remains for future work.

GENERATIVE AI USE

Generative AI tools were used during the preparation of this work to assist with checking and refining mathematical proofs, finding and reviewing related literature, providing feedback on experimental implementations, preparing and debugging experimental code, and analyzing and presenting experimental results. They were also used to improve the clarity and organization of the manuscript, review LaTeX code, and assist with drafting and rewriting portions of the text. The authors reviewed all AI-assisted content and take full responsibility for the final manuscript.

The research questions, central ideas, algebraic–categorical framework, theoretical claims and consequences, and experimental questions and designs were developed by the authors.

REPRODUCIBILITY STATEMENT

We provide detailed material to support reproducibility of both the theoretical and empirical results. All assumptions and constructions used in the theoretical development are stated in the main text, and complete proofs of the theorems are provided in Appendix H. The construction, evaluation protocol, and additional numerical checks for the complete-graph experiment are given in Appendix C. Appendix D documents the datasets, implementation, POV construction, validation-only model-selection procedure, selected configurations, statistical aggregation, and computational diagnostics for the HPGNN experiment. The detector-conditioning and reinforcement-learning experiments are specified in Appendices E and F, respectively. Finally, Appendix G provides the computational complexity analysis and matrix-free evaluation procedures for finite and limiting POV.

REFERENCES

Sami Abu-El-Haija, Bryan Perozzi, Amol Kapoor, Nazanin Alipourfard, Kristina Lerman, Hrayr Harutyunyan, Greg Ver Steeg, and Aram Galstyan. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 21–29. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/abu-el-haija19a.html.

Jiří Adámek, Horst Herrlich, and George Strecker. *Abstract and Concrete Categories – The Joy of Cats*, volume 1. John Wiley and Sons, 1990.

Cristian Bodnar, Francesco Di Giovanni, Benjamin P. Chamberlain, Pietro Liò, and Michael Bronstein. Neural sheaf diffusion: a topological perspective on heterophily and oversmoothing in gnns. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, pp. 93–104, New York, NY, USA, 2000. Association for Computing Machinery. ISBN 1581132174. doi: 10.1145/342009.335388. URL https://doi.org/10.1145/342009.335388.

Seongjin Choi, Gahee Kim, and Se-Young Yun. Weisfeiler and lehman go categorical, 2026. URL https://arxiv.org/abs/2602.06787.

G. S. H. Cruttwell, Bruno Gavranović, Neil Ghani, Paul Wilson, and Fabio Zanasi. Categorical foundations of gradient-based learning, 2021. URL https://arxiv.org/abs/2103.01931.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pp. 3844–3852, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

Ernesto Estrada and Naomichi Hatano. Communicability in complex networks. *Phys. Rev. E*, 77: 036111, Mar 2008. doi: 10.1103/PhysRevE.77.036111. URL https://link.aps.org/doi/10.1103/PhysRevE.77.036111.

Brendan Fong and Michael Johnson. Lenses and learners, 2019. URL https://arxiv.org/abs/1903.03671.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized PageRank. In *International Conference on Learning Representations*, 2019a. URL https://openreview.net/forum?id=H1gL-2A9Ym.

Johannes Gasteiger, Stefan Weiß enberger, and Stephan Günnemann. Diffusion improves graph learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019b. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/23c894276a2c5a16470e6a31f4618d73-Paper.pdf.

Bruno Gavranović, Paul Lessard, Andrew Joseph Dudzik, Tamara Von Glehn, João Guilherme Madeira Araújo, and Petar Veličković. Position: Categorical deep learning is an algebraic theory of all architectures. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 15209–15241. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/gavranovic24a.html.

Bruno Gavranović. Fundamental components of deep learning: A category-theoretic approach, 2024. URL https://arxiv.org/abs/2403.13001.

Peter Grindrod, Mark C. Parsons, Desmond J. Higham, and Ernesto Estrada. Communicability across evolving networks. *Phys. Rev. E*, 83:046120, Apr 2011. doi: 10.1103/PhysRevE.83.046120. URL https://link.aps.org/doi/10.1103/PhysRevE.83.046120.

Mustafa Hajij, Lennart Bastian, Sarah Osentoski, Hardik Kabaria, John Davenport, Dawood, Balaji Cherukuri, Joseph Kocheemoolayil, Nastaran Shahmansouri, Adrian Lew, Theodore Papamarkou, and Tolga Birdal. Copresheaf topological neural networks: A generalized deep learning framework. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 149759–149803. Curran Associates, Inc., 2025. doi: 10.52202/085713-5009. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/dc62cd4ec77f3bbec8e6245d0bd91d08-Paper-Conference.pdf.

Thomas W. Hungerford. *Algebra*. Springer New York, NY, 2011.

Nikhil Khatri, Tuomas Laakkonen, Jonathon Liu, and Vincent Wang-Maścianica. On the anatomy of attention, 2024. URL https://arxiv.org/abs/2407.02423.

R. Lambiotte, R. Sinatra, J.-C. Delvenne, T. S. Evans, M. Barahona, and V. Latora. Flow graphs: Interweaving dynamics and structure. *Phys. Rev. E*, 84:017102, Jul 2011. doi: 10.1103/PhysRevE.84.017102. URL https://link.aps.org/doi/10.1103/PhysRevE.84.017102.

Amirreza Shiralinasab Langari, Leila Yeganeh, and Kim Khoa Nguyen. Grothendieck graph neural networks framework: An algebraic platform for crafting topology-aware GNNs. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=oD3qWXgB4e.

Yoonho Lee, Junseok Lee, Sangwoo Seo, Sungwon Kim, Yeongmin Kim, and Chanyoung Park. Disentangling hyperedges through the lens of category theory. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 36673–36717. Curran Associates, Inc., 2025. doi: 10.52202/085713-1232. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/34634e7f60a26b996b1102a3ae2046e1-Paper-Conference.pdf.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation-based anomaly detection. *ACM Trans. Knowl. Discov. Data*, 6(1), March 2012. ISSN 1556-4681. doi: 10.1145/2133360.2133363. URL https://doi.org/10.1145/2133360.2133363.

Kay Liu, Yingtong Dou, Yue Zhao, Xueying Ding, Xiyang Hu, Ruitong Zhang, Kaize Ding, Canyu Chen, Hao Peng, Kai Shu, Lichao Sun, Jundong Li, George H Chen, Zhihao Jia, and Philip S Yu. Bond: Benchmarking unsupervised outlier node detection on static attributed graphs. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 27021–27035. Curran Associates, Inc., 2022.

Jacob Lurie. *Higher Topos Theory*, volume 1. Princeton University Press, 2009.

Saunders Mac Lane. *Categories for the Working Mathematician*. Springer New York, NY, 1998.

Mayu Sakurada and Takehisa Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, MLSDA'14, pp. 4–11, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450331593. doi: 10.1145/2689746.2689747. URL https://doi.org/10.1145/2689746.2689747.

Yumeng Wang, Zengyi Wo, Wenjun Wang, Xingcheng Fu, and Minglai Shao. Leveraging personalized pagerank and higher-order topological structures for heterophily mitigation in graph neural networks. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 6506–6514. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/724. URL https://doi.org/10.24963/ijcai.2025/724. Main Track.

Appendix Contents

- Appendix A: Fundamental Concepts in Category Theory
- Appendix B: Order-Theoretic Relation Between $\mathcal{G}^m$ and $\mathcal{C}^m$
- Appendix C: Additional Details for the Complete-Graph Recovery Experiment
- Appendix D: Additional Details for the HPGNN Propagation-View Experiment
- Appendix E: Additional Experiment: POV Conditioning of Detector-Derived Distributions
- Appendix F: Exploratory Application of POV to Reinforcement Learning
- Appendix G: Computational Complexity of POV
- Appendix H: Proofs of the Theorems

A Fundamental Concepts in Category Theory

We provide the fundamental categorical concepts that are used throughout the paper.

Definition A.1. Hungerford (2011) A monoid is a non-empty set M together with a binary operation $\cdot$ on M which

1) is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in \mathsf{M}$ and
2) contains identity element $e \in \mathsf{M}$ such that $a \cdot e = e \cdot a = a$

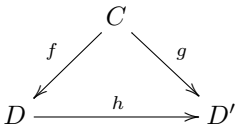
If, for all $a, b \in \mathsf{M}$, the operation satisfies $a \cdot b = b \cdot a$, then we say that M is a commutative monoid.

Definition A.2. Adámek et al. (1990) A category is a quadruple $\mathcal{A} = (\mathcal{O}, \hom, id, \cdot)$ consisting of

- a class $\mathcal{O}$, whose members are called objects,
- for each pair (A, B) of objects, a set $\hom(A, B)$, whose members are called morphisms from A to B-[the statement "$f \in \hom(A, B)$" is expressed more graphically by using arrows; e.g., by statements such as "$f : A \to B$ is a morphism" or "$A \xrightarrow{f} B$ is a morphism"]
- for each object A, a morphism $id_A : A \to A$, called the identity on A,

- a composition law associating with each morphism $A \xrightarrow{f} B$ and each morphism $B \xrightarrow{g} C$ a morphism $A \xrightarrow{g \cdot f} C$, called the composite of f and g, subject to the following conditions:
  - composition is associative; i.e., for morphisms $A \xrightarrow{f} B$, $B \xrightarrow{g} C$ and $C \xrightarrow{h} D$ the equation $h \cdot (g \cdot f) = (h \cdot g) \cdot f$ holds,
  - identities act as identities with respect to composition; i.e., for morphisms $A \xrightarrow{f} B$, we have $id_B \cdot f = f$ and $f \cdot id_A = f$,
  - the sets $\hom(A, B)$ are pairwise disjoint.

Definition A.3. Lurie (2009) For a category $\mathcal{C}$ and an object $C \in \mathcal{C}$, the under category $C/\mathcal{C}$ is the category that has morphisms with domain C, $C \to D$, as objects and the commutative triangles



as morphisms.

Definition A.4. Adámek et al. (1990) For categories $\mathcal{A}$ and $\mathcal{B}$, a functor F from $\mathcal{A}$ to $\mathcal{B}$ is a function that assigns to each $\mathcal{A}$-object A a $\mathcal{B}$-object $F(A)$, and to each $\mathcal{A}$-morphism $A \xrightarrow{f} B$ a $\mathcal{B}$-morphism $F(A) \xrightarrow{F(f)} F(B)$, in such a way that

- F preserves composition; i.e., $F(f \cdot g) = F(f) \cdot F(g)$ whenever $f \cdot g$ is defined, and
- F preserves identity morphisms; i.e., $F(id_A) = id_{F(A)}$ for each $\mathcal{A}$-object A.

B ORDER-THEORETIC RELATION BETWEEN $\mathcal{G}^m$ AND $\mathcal{C}^m$

Let $\mathcal{G}$ and $\mathcal{C}$ be as in Section 5.2, with $\mathcal{G}$ connected, $|V| \geq 2$, and $\mathcal{C}$ its complete supergraph on the same fixed, indexed vertex set. We formalize the structural relationship between $\mathsf{SMult}(\mathcal{G})$ and $\mathsf{SMult}(\mathcal{C})$ using a Galois connection.

Theorem B.1. *For an arbitrary graph $\mathcal{G}$, define a relation $\leq$ on $\mathsf{SMult}(\mathcal{G})$ by*

$$(\mathcal{M}, S) \leq (\mathcal{N}, T) \iff \exists\, \iota : \mathcal{M} \hookrightarrow \mathcal{N} \text{ such that } \iota_V(v_i) = v_i \text{ for every } v_i \in V(\mathcal{M}), \quad \iota(S) \subseteq T.$$

Then $\leq$ is a preorder on $\mathsf{SMult}(\mathcal{G})$, compatible with the monoidal operation $\bullet$.

Next, we connect the monoids $\mathsf{SMult}(\mathcal{C})$ and $\mathsf{SMult}(\mathcal{G})$ through two order-preserving morphisms. We define the inclusion $\mathbf{In} : \mathsf{SMult}(\mathcal{G}) \to \mathsf{SMult}(\mathcal{C})$, and for the reverse direction, we define the mapping R as follows:

Theorem B.2. *The mapping $\mathsf{R} : \mathsf{SMult}(\mathcal{C}) \longrightarrow \mathsf{SMult}(\mathcal{G})$, defined by $\mathsf{R}(\mathcal{M}, S) = (\hat{\mathcal{M}}, \hat{S})$, is a monoidal homomorphism, where:*

- *$\hat{\mathcal{M}}$ is obtained from $\mathcal{M}$ by removing all directed-edge occurrences corresponding to directed edges in $\mathsf{DE}(\mathcal{C}) \setminus \mathsf{DE}(\mathcal{G})$;*
- *$\hat{S}$ consists of the paths in S that traverse none of these removed directed edges.*

Both $\mathbf{In}$ and R are monotone with respect to the preorder $\leq$. Indeed, if $(\mathcal{M}, S) \leq (\mathcal{N}, T)$ via an embedding $\iota : \mathcal{M} \hookrightarrow \mathcal{N}$, then the same embedding shows

$$\mathbf{In}(\mathcal{M}, S) \leq \mathbf{In}(\mathcal{N}, T).$$

Moreover, restricting ι to the directed-edge occurrences that belong to $\mathcal{G}$ gives an embedding

$$\hat{\iota} : \hat{\mathcal{M}} \hookrightarrow \hat{\mathcal{N}}.$$

If $p \in \hat{S}$, then p traverses only directed edges of $\mathcal{G}$. Since ι preserves endpoints, $\iota(p)$ also traverses only directed edges of $\mathcal{G}$, and $\iota(p) \in T$. Hence $\hat{\iota}(\hat{S}) \subseteq \hat{T}$, so

$$\mathsf{R}(\mathcal{M}, S) \leq \mathsf{R}(\mathcal{N}, T).$$

Theorem B.3. *With respect to the preorder $\leq$, the maps*

$$\mathbf{In} : \mathsf{SMult}(\mathcal{G}) \to \mathsf{SMult}(\mathcal{C})$$

and

$$\mathsf{R} : \mathsf{SMult}(\mathcal{C}) \to \mathsf{SMult}(\mathcal{G})$$

form a Galois connection, with $\mathbf{In}$ *as the left adjoint of* R*. Explicitly, for every $(\mathcal{M}, S) \in \mathsf{SMult}(\mathcal{G})$ and $(\mathcal{N}, T) \in \mathsf{SMult}(\mathcal{C})$,*

$$\mathbf{In}(\mathcal{M}, S) \leq (\mathcal{N}, T) \quad \iff \quad (\mathcal{M}, S) \leq \mathsf{R}(\mathcal{N}, T).$$

The most significant advantage of this Galois connection is the ability to approximate an element of $\mathsf{SMult}(\mathcal{C})$ in $\mathsf{SMult}(\mathcal{G})$ using R. Specifically, for $(\mathcal{M}, S) \in \mathsf{SMult}(\mathcal{G})$ and $(\mathcal{N}, T) \in \mathsf{SMult}(\mathcal{C})$:

$$(\mathcal{M}, S) \leq \mathsf{R}(\mathcal{N}, T) \iff \mathbf{In}(\mathcal{M}, S) \leq (\mathcal{N}, T)$$

Since R is a monoidal homomorphism and $\mathsf{R}(\mathcal{C}) = \mathcal{G}$, we deduce that $\mathsf{R}(\mathcal{C}^m) = \mathcal{G}^m$. Thus, as stated in the following corollary, $\mathcal{G}^m$ approximates $\mathcal{C}^m$ in $\mathsf{SMult}(\mathcal{G})$.

Corollary B.4. *For every $X \in \mathsf{SMult}(\mathcal{G})$, we have $X \leq \mathcal{G}^m \quad \iff \quad \mathbf{In}(X) \leq \mathcal{C}^m$. Consequently, $\mathcal{G}^m$ is a greatest element, with respect to the preorder $\leq$, of*

$$\{X \in \mathsf{SMult}(\mathcal{G}) : \mathbf{In}(X) \leq \mathcal{C}^m\}.$$

C Additional Details for the Complete-Graph Recovery Experiment

This appendix provides the construction, quantitative results, and additional checks for the complete-graph experiment in Section 6.1.

Weightings and protocol. Let $\mathcal{C} = (V, E, X)$ be a complete attributed graph with $n = |V|$, and let

$$P = (P_1, \ldots, P_n)$$

be an interior categorical distribution on the node-indexed attribute occurrences X. For each sampled P, we consider the constant-sequence setting $P^{(k)} = P$ for all k.

The proposed Odds weighting is

$$f_{\mathrm{Odds}}(p) = \frac{p}{1-p},$$

and we compare it with the alternative destination-weighting functions

$$f_{\mathrm{Linear}}(p) = p, \qquad f_{\mathrm{Sqrt}}(p) = \sqrt{p}, \qquad f_{\mathrm{Quadratic}}(p) = p^2,$$
$$f_{\mathrm{NegLog}}(p) = -\log(1-p), \qquad f_{\mathrm{Uniform}}(p) = 1.$$

These alternatives include linear, sublinear, superlinear, nonlinear, and distribution-independent transformations of the input masses.

To control for differences in overall weight magnitude when comparing finite aggregation levels, each weighting is normalized to have the same total destination weight as Odds. Specifically, for a weighting function f, we define

$$c_f(P) = \frac{\sum_{r=1}^{n} \frac{P_r}{1-P_r}}{\sum_{r=1}^{n} f(P_r)}, \qquad w_{f,j}(P) = c_f(P) f(P_j).$$

Thus,

$$\sum_{j=1}^{n} w_{f,j}(P) = \sum_{j=1}^{n} \frac{P_j}{1-P_j}.$$

For Odds, $c_{\mathrm{Odds}}(P) = 1$. The corresponding complete-graph matrix is

$$\big(\boldsymbol{W}_f(P)\big)_{i,j} = \begin{cases} w_{f,j}(P), & i \neq j, \\ 0, & i = j. \end{cases}$$

In particular, $\boldsymbol{W}_{\mathrm{Odds}}(P) = \boldsymbol{W}(P)$, where $\boldsymbol{W}(P)$ is the distribution-induced matrix of Definition 5.1. At aggregation level m, the constant-sequence realization associated with f is

$$(\boldsymbol{I} + \boldsymbol{W}_f(P))^m - \boldsymbol{I}.$$

Node-level and graph-level distributions are then obtained using the row normalization and graph-level averaging in Definition 5.2.

We evaluate complete graphs with

$$n \in \{5, 10, 20\}, \qquad P \sim \mathrm{Dirichlet}(\alpha \mathbf{1}), \qquad \alpha \in \{0.2, 1, 5\}.$$

For each of the nine (n, α) settings, we sample 100 independent categorical distributions, yielding 900 distributions in total. For every weighting, we evaluate the graph-level ℓ_1 distance to P for $m = 1, \ldots, 50$.

Limiting-distribution computation. In addition to the finite-m calculation, we independently compute the limiting distribution associated with each weighting. Since every sampled distribution is interior, $\boldsymbol{W}_f(P)$ has positive off-diagonal entries for all of the tested weightings. Hence the same Perron–Frobenius argument underlying Theorem 5.3 applies.

Let $\ell_f > 0$ be a left Perron eigenvector of $\boldsymbol{W}_f(P)$, with Perron eigenvalue ρ_f, and define its normalized form by

$$q_f = \frac{\ell_f}{\ell_f \mathbf{1}}.$$

Then q_f is the limiting POV distribution associated with $\boldsymbol{W}_f(P)$. On a complete graph,

$$\rho_f q_{f,j} = w_{f,j}(P)\big(1 - q_{f,j}\big),$$

and therefore

$$q_{f,j} = \frac{w_{f,j}(P)}{\rho_f + w_{f,j}(P)}.$$

The Perron eigenvalue $\rho_f > 0$ is determined by

$$\sum_{j=1}^{n} \frac{w_{f,j}(P)}{\rho_f + w_{f,j}(P)} = 1.$$

We use

$$\|q_f - P\|_1$$

as the independently computed limiting error.

A common positive rescaling of all destination weights changes ρ_f by the same factor but leaves q_f unchanged. Thus, the normalization above controls the overall weight magnitude in the finite-m comparison without affecting the limiting distribution.

Numerical results and additional checks. Figure 1 shows that finite-depth proximity to P and asymptotic recovery are distinct. Linear is closest to P for 63.4% of the sampled distributions at $m = 1$, while NegLog is closest for 92.7% at $m = 2$. From $m = 4$ onward, however, Odds is closest for all 900 sampled distributions at every tested aggregation level through $m = 50$.

Across all 900 distributions, the maximum limiting ℓ_1 error of Odds is

$$5.64 \times 10^{-16},$$

which is at numerical precision. The median limiting errors of NegLog, Linear, Quadratic, Sqrt, and Uniform are 0.0352, 0.0688, 0.2310, 0.3122, and 0.6682, respectively. The same qualitative behavior is observed across all nine (n, α) settings.

The graph-level result is not an artifact of averaging node-level distributions. At $m = 50$, the largest worst-node ℓ_1 error for Odds over all 900 sampled distributions is

$$2.07 \times 10^{-15}.$$

Together, these results numerically verify the complete-graph recovery established for the proposed Odds construction and show that several natural alternative destination weightings do not exhibit the same limiting behavior.

D Additional Details for the HPGNN Propagation-View Experiment

D.1 Experimental implementation and datasets

Repository-aligned implementation. We implement the POV substitution directly inside the official HPGNN (Wang et al., 2025) repository rather than reimplementing HPGNN independently. We retain its graph-loading and split-generation routines, downstream architecture, optimizer, validation-loss checkpointing, early-stopping logic, and training/evaluation pipeline. In particular, we use the same ten graph-split seeds and the same $60\%/20\%/20\%$ train/validation/test protocol.

The only architectural intervention is the replacement of the PPR and HiPPR propagation matrices supplied to the two HPGNN spectral-filtering branches by a validation-selected pair of POV matrices. The shared HPGNN hyperparameters—layer k, learning rate, and weight decay—are selected within the HPGNN search setting, while the POV pair is the additional validation-selected propagation choice.

Experiments were run in Google Colab on an NVIDIA L4 GPU using PyTorch 2.11.0 and PyTorch Geometric 2.8.0.post1.

Datasets. We evaluate on the seven node-classification benchmarks used in the HPGNN comparison. Table 2 reports the graph objects loaded by the repository. To preserve alignment with HPGNN, the POV construction uses the edge representation returned by the repository loader without adding, removing, or symmetrizing edges. Accordingly, we report the number of stored PyTorch Geometric `edge_index` entries rather than interpreting this quantity as the number of unique undirected edges. The theoretical development of POV is formulated for undirected graphs. Here, we apply the finite POV matrix construction directly to the benchmark graph representations used by HPGNN; this should therefore be viewed as an empirical extension of the finite construction beyond the graph setting of the theoretical guarantees.

Table 2: Dataset statistics for the graph objects used in the HPGNN propagation-view experiment. "Edge entries" denotes the number of columns of the PyTorch Geometric `edge_index` representation.

Dataset	Nodes	Edge entries	Features	Classes
Cora	2708	10556	1433	7
Citeseer	3327	9104	3703	6
Photo	7650	238162	745	8
Cornell	183	557	1703	5
Actor	7600	30019	932	5
Texas	183	574	1703	5
Wisconsin	251	916	1703	5

D.2 POV CONSTRUCTION DETAILS

We use the two node-indexed categorical distributions P_D and P_F defined in Eq. 1. Neither distribution uses node labels. For each

$$P \in \{P_D, P_F\}$$

and aggregation level

$$m \in \{1, 2, 3\},$$

we construct the matrix $\mathsf{POV}(P, m)$ according to Definition 5.2. Using the notation of Eq. 2, the candidate propagation views are

$$\mathcal{Q}_{\mathrm{POV}} = \{\boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3, \boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3\}.$$

The six matrices yield 15 unordered pairs of distinct views, spanning degree-conditioned, feature-conditioned, and mixed representations. Thus, validation selection varies both the conditioning distribution and the POV aggregation level m.

For some benchmark graph/distribution combinations, a finite POV realization may contain an identically zero row, for which the row normalization in Definition 5.2 is undefined. For such a row only, we use the corresponding conditioning distribution P as the fallback distribution. All nonzero rows are normalized exactly as in Definition 5.2.

The resulting POV matrices replace the PPR and HiPPR matrices supplied to the two HPGNN propagation branches. All subsequent HPGNN operations are unchanged.

D.3 VALIDATION-ONLY MODEL SELECTION

All hyperparameter and POV-pair selection uses validation accuracy only; no test-set quantity enters the search. Each candidate configuration is evaluated on the first five official HPGNN graph splits using two fixed model repetitions per split. Repetitions are first averaged within each split, and configurations are ranked by mean validation accuracy across the five split-level means. The same model repetitions are reused across candidate configurations within each split.

Within each individual training run, we retain the official HPGNN validation-loss checkpointing and early-stopping procedure. Thus, validation loss determines the retained checkpoint within a run, whereas validation accuracy is used to compare hyperparameter and POV-pair configurations across runs.

The staged validation search is summarized in Table 3.

Table 3: Validation-only search procedure for the HPGNN–POV experiment.

Stage	Search
Pair/k screening	All 15 POV pairs with $k \in \{2, 4, 6, 8, 10\}$, $\mathrm{lr} = 0.1$, and $\mathrm{wd} = 10^{-3}$.
Full k refinement	Three best POV pairs with $k \in \{1, 2, \ldots, 10\}$; overlapping runs from the screening stage are reused.
Optimizer search	For each surviving pair at its validation-selected k, $\mathrm{lr} \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$ and $\mathrm{wd} \in \{0, 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-1}\}$.
Local k confirmation	For the best tuned pair, evaluate $k_0 - 1, k_0, k_0 + 1$ when these values lie in $\{1, \ldots, 10\}$, using the selected optimizer.

The resulting configuration is frozen before final test evaluation.

D.4 SELECTED CONFIGURATIONS

Table 4 reports the final validation-selected configuration for each dataset. Validation statistics are computed across the five graph-split means after averaging the two model repetitions within each split.

The selected pairs include degree-conditioned, feature-conditioned, and mixed representations and use dataset-dependent aggregation levels. Thus, neither a single conditioning distribution nor a single aggregation level is selected uniformly across the seven datasets. The selection of $m = 1$ views

Table 4: Validation-selected POV configurations. Validation accuracy is reported as mean ± standard deviation (%) across five graph-split means.

Dataset	POV pair	k	$(\mathrm{lr}, \mathrm{wd})$	Val. acc.
Cora	$(\boldsymbol{F}_1, \boldsymbol{F}_2)$	9	$(0.05, 0.005)$	90.13 ± 1.58
Citeseer	$(\boldsymbol{D}_1, \boldsymbol{F}_3)$	10	$(0.20, 0.005)$	80.33 ± 1.09
Photo	$(\boldsymbol{F}_2, \boldsymbol{F}_3)$	3	$(0.10, 0.0001)$	95.39 ± 0.46
Cornell	$(\boldsymbol{D}_1, \boldsymbol{F}_1)$	1	$(0.05, 0.001)$	94.86 ± 2.60
Actor	$(\boldsymbol{D}_2, \boldsymbol{D}_3)$	2	$(0.10, 0.001)$	40.70 ± 0.90
Texas	$(\boldsymbol{D}_3, \boldsymbol{F}_1)$	8	$(0.10, 0.001)$	95.68 ± 2.00
Wisconsin	$(\boldsymbol{D}_1, \boldsymbol{F}_1)$	1	$(0.10, 0.001)$	97.80 ± 2.68

on several datasets also indicates that the conditioning distribution can provide useful propagation representations without requiring deeper POV composition.

D.5 Final evaluation and statistical aggregation

After model selection, each dataset is evaluated on the ten official graph splits with ten fixed model repetitions per split, yielding 100 final runs per dataset. For test accuracy $a_{s,r}$, where s indexes the graph split and r the model repetition, we first average repetitions within each split,

$$\bar{a}_s = \frac{1}{10} \sum_{r=1}^{10} a_{s,r},$$

and then compute

$$\bar{a} = \frac{1}{10} \sum_{s=1}^{10} \bar{a}_s.$$

The 95% confidence interval reported for HPGNN (POV views) is

$$\bar{a} \pm t_{0.975,9} \frac{s_{\mathrm{split}}}{\sqrt{10}},$$

where s_{split} is the sample standard deviation of the ten split-level means. Thus, the graph split, rather than the individual training run, is treated as the independent statistical unit.

The HPGNN values in Table 1 are reproduced directly from Table 2 of Wang et al. (2025) and were not recomputed in our environment. Their reported uncertainty is therefore reproduced as given in the original study and should not be interpreted as having been computed by the split-level procedure above. Consequently, the two rows do not constitute a paired statistical comparison, and we interpret differences between their reported point estimates descriptively.

D.6 Computational diagnostics

Propagation-view preprocessing. Because the controlled intervention changes the propagation representation, we separately examine the cost of constructing the two propagation views, see Table 5. The HPGNN reference times are previously recorded cold measurements from the completed heterophilic-search runs, in which the timer surrounded the official graph-loading and PPR/HiPPR preprocessing pipeline and no persistent propagation-matrix cache was restored. For POV, timing begins before graph loading and ends after construction of both validation-selected POV matrices, including construction of their conditioning distributions; no POV matrix cache is used.

For each heterophilic dataset, the selected POV pair is constructed in three fresh repetitions and the median pipeline time is reported. Each HPGNN reference value is the corresponding previously recorded cold preprocessing measurement. The comparison is therefore intended as an implementation-level preprocessing diagnostic rather than a statistical runtime comparison.

Across these four datasets, the measured preprocessing-time ratio ranges from $226.3\times$ to $2782.5\times$, while the selected POV pairs contain 2.5% to 65.2% as many stored nonzeros as the corresponding PPR/HiPPR pairs. These measurements characterize the preprocessing cost of the propagation representations in the measured implementation; they do not establish a corresponding advantage in end-to-end training time.

Table 5: Propagation-view preprocessing on the four heterophilic datasets with available cold HPGNN reference measurements. HPGNN times are previously recorded cold graph-loading and PPR/HiPPR preprocessing measurements with no persistent cache restored; POV times are medians over three fresh uncached pipeline constructions. The time ratio is $t_{\mathrm{PPR/HiPPR}}/t_{\mathrm{POV}}$, and the nnz ratio is the total number of stored POV nonzeros divided by the corresponding PPR/HiPPR total. These measurements characterize propagation-view preprocessing only and do not represent end-to-end model speedups.

Dataset	POV pair	PPR/HiPPR (s)	POV (s)	Time ratio	nnz ratio
Actor	$(\boldsymbol{D}_2, \boldsymbol{D}_3)$	1718.6	0.618	2782.5×	0.420
Cornell	$(\boldsymbol{D}_1, \boldsymbol{F}_1)$	1.946	0.006	305.9×	0.031
Texas	$(\boldsymbol{D}_3, \boldsymbol{F}_1)$	2.242	0.010	226.3×	0.652
Wisconsin	$(\boldsymbol{D}_1, \boldsymbol{F}_1)$	4.508	0.007	672.1×	0.025

Final-run runtime and GPU memory. We also record wall-clock runtime and CUDA peak allocated memory for every final POV training run, see Table 6. Runtime begins immediately before the downstream HPGNN training/evaluation routine and therefore excludes the one-time POV construction. CUDA peak-memory statistics are reset immediately before each run.

Table 6: Computational resources of the final HPGNN experiments using POV views. Runtime and peak allocated GPU memory are means over 100 final training runs per dataset. POV preprocessing is excluded.

Dataset	POV pair	k	Time/run (s)	Peak GPU (GiB)
Cora	$(\boldsymbol{F}_1, \boldsymbol{F}_2)$	9	3.42	0.073
Citeseer	$(\boldsymbol{D}_1, \boldsymbol{F}_3)$	10	9.10	0.187
Photo	$(\boldsymbol{F}_2, \boldsymbol{F}_3)$	3	29.27	1.559
Cornell	$(\boldsymbol{D}_1, \boldsymbol{F}_1)$	1	1.53	0.022
Actor	$(\boldsymbol{D}_2, \boldsymbol{D}_3)$	2	13.37	0.665
Texas	$(\boldsymbol{D}_3, \boldsymbol{F}_1)$	8	3.08	0.023
Wisconsin	$(\boldsymbol{D}_1, \boldsymbol{F}_1)$	1	1.50	0.024

These measurements characterize the computational footprint of the final POV-based experiments. We do not use them to claim a direct training-time or GPU-memory advantage over HPGNN because matched HPGNN resource measurements were not collected under the same instrumentation.

E Additional Experiment: POV Conditioning of Detector-Derived Distributions

This experiment evaluates graph-level POV directly, independently of its use as a propagation representation in Section 6.2. Its purpose is not to treat POV as a generic anomaly-score enhancer, but to examine how graph topology reshapes an externally supplied categorical distribution. Graph anomaly detection provides a convenient probing task because detector scores provide node-indexed information that can be converted to a categorical distribution while preserving their ranking.

E.1 Detector-derived distributions and protocol

Ranking-preserving conditioning distribution. Let detector B assign score s_i to node v_i, and define

$$\Delta_s = s_{\max} - s_{\min}.$$

When $\Delta_s > 0$, we construct the strictly positive categorical distribution

$$P_{B,i} := P_B(x_i) = \frac{s_{\max} - s_i + \eta\Delta_s}{\sum_j (s_{\max} - s_j + \eta\Delta_s)}, \qquad \eta = 10^{-6}. \tag{3}$$

If all detector scores are numerically identical, P_B is set to the uniform distribution. The transformation preserves the detector ranking:

$$s_i > s_j \quad \Longleftrightarrow \quad -\log P_{B,i} > -\log P_{B,j}.$$

Table 7: Datasets used in the POV-conditioning experiment. Edges are unique undirected edges after symmetrization, duplicate removal, and self-loop removal. Average degree is $\bar{d} = 2|E|/|V|$. "Anomalies" denotes positive labels used for evaluation; the DGraphFin count refers to its official test subset.

Dataset	Nodes	Undirected edges	Avg. degree	Features	Anomalies
Weibo	8,405	377,271	89.8	400	347
Reddit	10,984	78,516	14.3	64	366
Disney	124	335	5.4	28	6
Books	1,418	3,695	5.2	21	28
Enron	13,533	176,987	26.2	18	5
DGraphFin	3,700,550	3,997,260	2.2	17	2,326[†]

[†]Number of anomalies in the official DGraphFin test subset.

Thus, P_B should be understood as a ranking-preserving categorical encoding of the detector output, rather than as an anomaly-probability distribution. In every trial, we verify that ROC-AUC and AP computed from $-\log P_{B,i}$ agree with those obtained from the original detector score up to numerical precision.

For finite aggregation level m, we define

$$Q_B^{(m)} := \mathsf{pov}(\mathcal{G}, P_B, m)$$

and score node v_i by

$$s_i^{\mathrm{POV},m} = -\log Q_{B,i}^{(m)}.$$

We evaluate the fixed finite level $m = 2$ together with the limiting POV distribution

$$Q_B^{(\infty)} := \mathsf{pov}(\mathcal{G}, P_B, \infty).$$

The finite level $m = 2$ is fixed for all datasets and detectors rather than selected using test performance.

Benchmark protocol. We implement the experiment on top of the BOND/PyGOD benchmark pipeline (Liu et al., 2022), retaining its detector construction, data loading, and evaluation protocol. We evaluate Isolation Forest (IF) (Liu et al., 2012), LOF (Breunig et al., 2000), and MLPAE (Sakurada & Yairi, 2014) over 20 trials on Weibo, Reddit, Disney, Books, Enron, and DGraphFin. Following the benchmark protocol, LOF is omitted on DGraphFin. We report ROC-AUC and average precision (AP).

POV is computed on an undirected, loopless, binary graph. Directed edge entries are symmetrized, duplicate edges are collapsed, and self-loops are removed without changing node ordering. This matches the undirected setting of the theoretical construction.

DGraphFin follows the benchmark-specific evaluation protocol: detectors operate on the full graph, while metrics are evaluated only on the official test subset after excluding background labels 2 and 3.

Table 7 reports the graph statistics after this preprocessing. We additionally report average degree

$$\bar{d} = \frac{2|E|}{|V|},$$

which provides structural context for interpreting the finite-versus-limiting behavior below.

Finite and limiting POV computation. For finite m, let

$$\boldsymbol{W} = \boldsymbol{A}\,\mathrm{diag}\left(\frac{P_B}{1 - P_B}\right), \qquad \boldsymbol{M} = (\boldsymbol{I} + \boldsymbol{W})^m - \boldsymbol{I},$$

where the quotient is taken elementwise. Nonzero rows of $\boldsymbol{M}$ are normalized according to Definition 5.2. If a row is identically zero, for which that normalization is undefined, we use P_B as the fallback distribution for that row.

Table 8: Graph anomaly detection before and after graph-level POV conditioning. Entries report mean ROC-AUC / AP with 95% confidence intervals over 20 trials. Confidence intervals are Student-t intervals computed across trials, and the values following $\pm$ are their half-widths. Half-widths are rounded to four decimal places; consequently, a reported value of 0.0000 indicates either identical results across trials or an interval narrower than the displayed precision. B denotes the original detector ranking; POV_2 and POV_∞ use $Q_B^{(2)}$ and $Q_B^{(\infty)}$, respectively.

Detector	Representation	Weibo	Reddit	Disney	Books	Enron	DGraphFin
IF	B	.549 ± .0105/.054 ± .0058	.457 ± .0093/.028 ± .0005	.579 ± .0201/.099 ± .0113	.430 ± .0109/.019 ± .0011	.687 ± .0174/.001 ± .0000	.608 ± .0050/.018 ± .0002
	POV_2	.372 ± .0061/.030 ± .0006	.458 ± .0014/.028 ± .0002	.756 ± .0083/.170 ± .0152	.528 ± .0060/.021 ± .0004	.611 ± .0192/.001 ± .0000	.676 ± .0011/.024 ± .0002
	POV_∞	.813 ± .0031/.233 ± .0034	.419 ± .0029/.027 ± .0001	.655 ± .0109/.113 ± .0073	.407 ± .0015/.016 ± .0000	.623 ± .0135/.004 ± .0016	.616 ± .0036/.019 ± .0002
LOF	B	.577 ± .0000/.068 ± .0000	.571 ± .0000/.042 ± .0000	.479 ± .0000/.052 ± .0000	.365 ± .0000/.015 ± .0000	.315 ± .0000/.000 ± .0000	–
	POV_2	.315 ± .0000/.027 ± .0000	.485 ± .0000/.033 ± .0000	.720 ± .0000/.139 ± .0000	.545 ± .0000/.023 ± .0000	.328 ± .0000/.000 ± .0000	–
	POV_∞	.907 ± .0000/.289 ± .0000	.418 ± .0000/.027 ± .0000	.665 ± .0000/.085 ± .0000	.427 ± .0000/.016 ± .0000	.495 ± .0000/.002 ± .0000	–
MLPAE	B	.771 ± .0228/.178 ± .0229	.506 ± .0000/.034 ± .0000	.505 ± .0376/.063 ± .0080	.416 ± .0272/.018 ± .0016	.408 ± .0352/.000 ± .0000	.385 ± .0122/.009 ± .0002
	POV_2	.316 ± .0008/.029 ± .0007	.490 ± .0000/.033 ± .0000	.729 ± .0058/.146 ± .0084	.527 ± .0119/.022 ± .0011	.324 ± .0032/.000 ± .0000	.658 ± .0003/.022 ± .0001
	POV_∞	.913 ± .0009/.300 ± .0023	.430 ± .0000/.029 ± .0000	.671 ± .0034/.085 ± .0015	.422 ± .0066/.017 ± .0005	.492 ± .0031/.002 ± .0001	.575 ± .0000/.017 ± .0000

The graph-level distribution is computed without materializing the dense $n \times n$ matrix $\boldsymbol{M}$. Let

$$r = \boldsymbol{M}\mathbf{1}, \qquad I_0 = \{i : r_i = 0\}, \qquad a_i = \begin{cases} 1/r_i, & r_i > 0, \\ 0, & r_i = 0. \end{cases}$$

Then

$$Q_B^{(m)} = \frac{1}{n}\left(\boldsymbol{M}^\top a + |I_0| P_B\right),$$

where the required products are evaluated using repeated sparse matrix–vector operations.

For connected graphs, $Q_B^{(\infty)}$ is given by Theorem 5.3. For disconnected graphs, we use the following componentwise experimental extension. Let $\mathcal{C}_1, \ldots, \mathcal{C}_K$ be the non-isolated connected components and I_0 the isolated nodes. Define

$$\boldsymbol{D}_B = \mathrm{diag}\left(\frac{P_B}{1 - P_B}\right), \qquad \boldsymbol{S}_B = \boldsymbol{D}_B^{1/2} \boldsymbol{A} \boldsymbol{D}_B^{1/2}.$$

For component $\mathcal{C}_c$, let u_c be a positive Perron eigenvector of the corresponding principal submatrix $\boldsymbol{S}_{B,c}$, and set

$$\ell_c = u_c^\top \boldsymbol{D}_{B,c}^{1/2}, \qquad q_c = \frac{\ell_c}{\ell_c \mathbf{1}}.$$

If $\widetilde{q}_c \in \mathbb{R}^n$ denotes the embedding of q_c into the full node set, we use

$$Q_B^{(\infty)} = \sum_{c=1}^{K} \frac{|\mathcal{C}_c|}{n} \widetilde{q}_c + \frac{|I_0|}{n} P_B.$$

This componentwise construction is used only as an experimental extension for disconnected graphs and is not an additional claim of Theorem 5.3.

E.2 Main results

Table 8 compares the original detector ranking with the distributions obtained from finite $m = 2$ POV and from the limiting POV.

Finite POV. The effect of finite POV varies substantially across graphs. At $m = 2$, mean ROC-AUC is higher than the original detector result for all three detectors on Disney and Books and for both evaluated detectors on DGraphFin. For example, LOF changes from 0.479 to 0.720 on Disney and from 0.365 to 0.545 on Books, while MLPAE changes from 0.385 to 0.658 on DGraphFin. In contrast, $m = 2$ lowers mean ROC-AUC for all three detectors on Weibo and for most detector configurations on Reddit and Enron. Thus, finite POV does not act as a generic anomaly-score enhancer; its downstream effect depends on the graph, the supplied conditioning distribution, and the aggregation level.

Finite versus limiting POV on Weibo. Weibo provides the clearest contrast between finite and limiting POV. It also has the largest average degree among the evaluated graphs, $\bar{d} \approx 89.8$, as reported in Table 7. On this graph, finite $m = 2$ POV lowers the mean ROC-AUC of all three detectors, whereas the limiting POV yields substantially higher values:

$$0.549 \to 0.813 \quad \text{(IF)}, \qquad 0.577 \to 0.907 \quad \text{(LOF)}, \qquad 0.771 \to 0.913 \quad \text{(MLPAE)}.$$

The relative detector ordering is also preserved:

$$\text{IF} < \text{LOF} < \text{MLPAE}$$

both before and after limiting POV for ROC-AUC, with the same ordering for AP.

The experiment does not isolate graph density as the cause of this behavior. Rather, the Weibo result demonstrates that finite aggregation and the limiting construction can produce qualitatively different transformations of the same detector-derived distribution. Average degree is therefore reported here as structural context rather than as a causal explanation.

Paired statistical evidence. The confidence intervals in Table 8 quantify uncertainty in the mean performance of each representation separately. Because the original and POV-conditioned scores are obtained from the same detector trial, we additionally analyze their paired differences. Representative ROC-AUC changes, reported as the mean paired difference with a 95% Student-t confidence interval over the 20 trial-wise differences, are

$$+0.177\,[0.164, 0.191]$$

for $m = 2$ IF on Disney,

$$+0.098\,[0.092, 0.105]$$

for $m = 2$ IF on Books, and

$$+0.273\,[0.261, 0.285]$$

for $m = 2$ MLPAE on DGraphFin. In the opposite direction, $m = 2$ MLPAE on Weibo changes by

$$-0.455\,[-0.478, -0.433].$$

For limiting POV on Weibo, IF changes by

$$+0.264\,[0.255, 0.272].$$

These paired results show that POV conditioning can produce substantial changes in either direction, depending on the graph and aggregation regime.

E.3 Finite aggregation levels versus the limiting POV

To examine whether the finite-m behavior observed at $m = 2$ is specific to that aggregation level, we evaluate

$$m \in \{1, 2, \ldots, 8\}$$

on Books and Disney for all three detectors, together with the limiting distribution $m = \infty$. These two relatively small graphs permit a broader aggregation-level sweep. The sweep is diagnostic only and is not used to select $m = 2$ in Table 8.

Figure 2 shows that the finite-m behavior on Books and Disney is not specific to $m = 2$. On Books, all three detectors remain above their original ROC-AUC throughout $m = 1, \ldots, 8$. For example, LOF remains near 0.545, compared with 0.365 before POV conditioning, whereas its limiting value is 0.427. Disney exhibits a similar separation between the tested finite levels and the limit: finite POV yields ROC-AUC values near 0.72–0.76, while the limiting values lie near 0.65–0.67. The AP curves show the same broad finite-versus-limiting distinction.

Together with the Weibo results in Section E.2, these observations show that the aggregation level m is a substantive component of the POV representation rather than merely a numerical approximation parameter. Finite aggregation and the limiting construction can produce qualitatively different downstream behavior from the same graph and conditioning distribution.

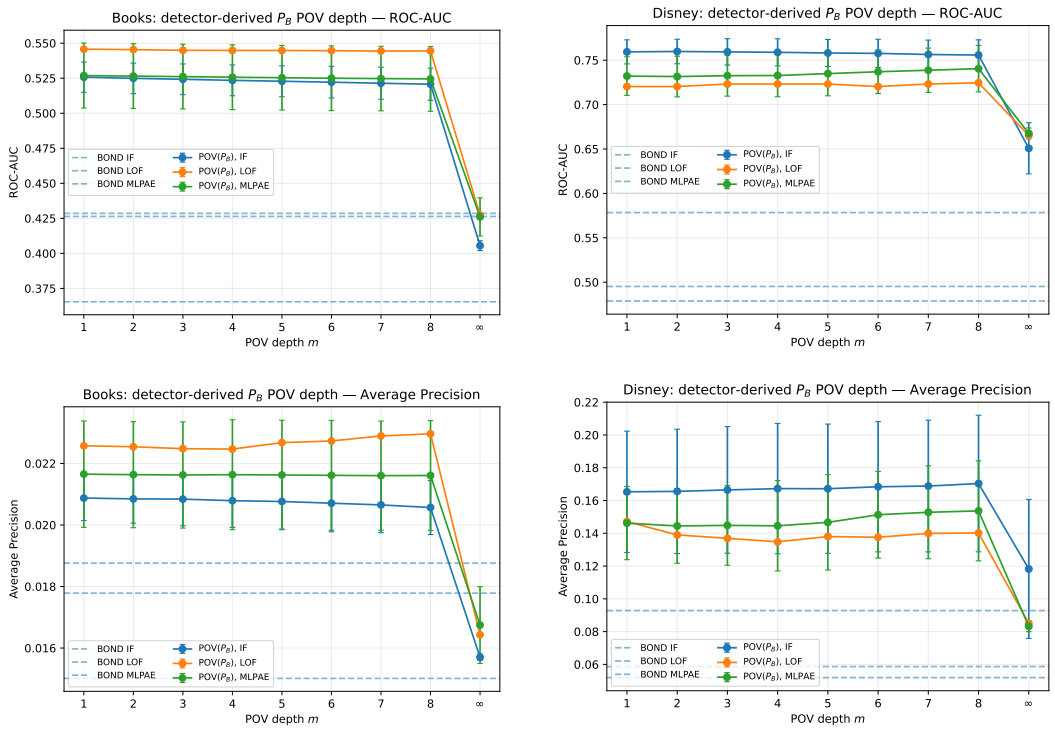


Figure 2: Effect of POV aggregation level on detector-derived distributions. **Top:** ROC-AUC on Books (left) and Disney (right). **Bottom:** average precision on Books (left) and Disney (right). Finite aggregation levels $m = 1, \ldots, 8$ are compared with the original detector and the limiting POV distribution ($m = \infty$).

Table 9: Topology-only control obtained from a uniform conditioning distribution. Entries report ROC-AUC / AP. Because this construction depends only on the fixed preprocessed graph and the uniform distribution, each dataset yields a single deterministic result and no across-trial confidence interval is applicable.

Representation	Weibo	Reddit	Disney	Books	Enron	DGraphFin
$\mathsf{pov}(\mathcal{G}, P_U, 2)$	.314/.027	.485/.033	.750/.212	.560/.024	.329/.000	.661/.022
$\mathsf{pov}(\mathcal{G}, P_U, \infty)$	.906/.290	.418/.027	.688/.095	.430/.017	.497/.002	.575/.017

E.4 Controls

Topology-only control. To examine how much structure can arise from topology independently of the detector-derived distribution, we apply POV to the uniform categorical distribution P_U. Table 9 reports the resulting ROC-AUC and AP.

The topology-only representation can itself be strongly associated with the anomaly labels on some datasets. For example, the limiting uniform-conditioned POV attains 0.906 ROC-AUC on Weibo, while the finite uniform-conditioned POV attains 0.661 on DGraphFin. Consequently, changes produced by detector-conditioned POV cannot be attributed to detector information alone. Rather, the resulting distribution reflects the interaction between the graph topology, the supplied conditioning distribution, and the aggregation regime.

E.5 Summary of the probing experiment

The probing experiment yields three main observations. First, POV can substantially reshape an externally supplied categorical distribution, with downstream effects that vary across graphs and aggregation regimes. Second, finite aggregation and the limiting construction can produce quali-

tatively different behavior from the same graph and conditioning distribution, showing that m is a substantive component of the representation rather than only an approximation parameter. Third, the topology-only and multiplicative controls show that the observed effects should not be attributed to detector information or to the full POV construction alone.

Taken together, these results support interpreting POV as a topology-conditioning mechanism whose behavior depends jointly on the graph, the supplied distribution, and the aggregation regime. They do not establish POV as a universally superior anomaly-scoring rule; rather, graph anomaly detection serves here as a controlled setting for probing how externally supplied node-indexed distributions are transformed by the POV construction.

F Exploratory Application of POV to Reinforcement Learning

The distributional realization of Definition 5.1 and the point-of-view distributions of Definition 5.2 naturally apply to settings in which learning produces a sequence of categorical distributions. Reinforcement learning provides one such setting. Let $\mathcal{A} = \{a_1, \dots, a_n\}$ be a finite action set, and let $\pi^{(k)}(\cdot \mid s)$ denote an action distribution induced by the learner at training stage k for state s. By associating action a_j with a node v_j of an action graph

$$\mathcal{G}_A = (V_A, E_A, X_A), \qquad X_A = (a_1, \dots, a_n),$$

the evolving policy induces a state-local sequence of categorical distributions

$$\mathcal{P}_s = \{P_s^{(k)}\}_{k \geq 1}, \qquad P_s^{(k)}(a_j) = \pi^{(k)}(a_j \mid s). \tag{4}$$

Definition 5.2 can therefore be applied to this policy sequence, producing node- and graph-level action distributions that depend jointly on the supplied policy history and the topology assigned to the action space.

Two cases are particularly natural. If the action space is finite and no additional relation among its actions is imposed, the complete action graph provides a structurally symmetric reference construction. In this setting, Theorem 5.4 supplies an asymptotic consistency boundary: for an increasingly aggregated convergent sequence, complete-graph POV recovers its limiting distribution. At finite aggregation levels, however, the POV distribution can differ from the most recent policy and can therefore provide a history-conditioned alternative action distribution.

If the action space possesses additional structure, one may instead encode that structure through a non-complete action graph. Definition 5.2 remains applicable, but the resulting distribution is then explicitly conditioned on the chosen action topology. We illustrate these two cases through two small reinforcement-learning experiments. KeyDoor uses a complete action graph and examines finite POV representations of recent policy history. Pendulum begins with a continuous-control problem, discretizes its action space, and considers two non-complete graphs encoding different relations among the resulting torque actions. These experiments are exploratory applications of the theoretical construction; they do not establish that POV generally improves reinforcement learning.

F.1 Discrete Action Spaces: KeyDoor

Environment and learner. We first consider a stochastic 7×7 KeyDoor grid world. After removing wall cells, the environment contains 43 traversable positions. A binary variable records whether the key has been collected, giving 86 states. The agent starts at $(6, 0)$, the key is located at $(5, 5)$, the door at $(3, 3)$, and the goal at $(0, 6)$; two terminal hazard cells are located at $(4, 1)$ and $(1, 4)$. The four actions correspond to moving up, right, down, and left. The requested movement occurs with probability 0.8, while the two adjacent directions occur with probability 0.1 each. The door cannot be crossed before the key has been collected. Reaching the goal gives reward 1.20, entering a hazard gives -0.80, collecting the key for the first time gives 0.19, and every other transition gives -0.01. Episodes terminate after at most 110 steps.

We use tabular Q-learning with discount $\gamma = 0.98$. For the n-th update of a state–action pair, the learning rate is

$$\alpha_n = \frac{0.45}{\sqrt{1 + 0.015n}}.$$

At state s, the Q-table induces the exploitation distribution

$$\pi_t(a \mid s) = \mathrm{softmax}\left(\frac{Q_t(s,\cdot)}{T}\right)_a, \qquad T = 0.25. \tag{5}$$

With probability 0.08, an action is instead sampled uniformly for explicit exploration. Historical advisors do not modify this explicit exploration.

Complete-graph historical POV. The four actions are treated as distinct choices without imposing an additional relation among them. We therefore associate them with the nodes of the complete action graph $\mathcal{G}_A = K_4$. For each state s, we retain the most recent $L = 10$ available action distributions. At a decision point, the distributions in the current window are relabeled in chronological order as

$$P_s^{(1)}, \ldots, P_s^{(m)}, \qquad m \le L.$$

Since the distributions are obtained by softmax, $0 < P_{s,j}^{(k)} < 1$, and Definition 5.1 gives

$$\boldsymbol{W}_{s,i,j}^{(k)} = \begin{cases} \dfrac{P_{s,j}^{(k)}}{1 - P_{s,j}^{(k)}}, & i \neq j, \\ 0, & i = j. \end{cases} \tag{6}$$

The chronological distributional realization is

$$\boldsymbol{M}_s = \mathsf{DRA}(\mathcal{P}_s, m) = \boldsymbol{W}_s^{(1)} \circ \cdots \circ \boldsymbol{W}_s^{(m)}. \tag{7}$$

After each training episode, the post-update action distribution is appended to the history of each state visited in that episode. Historical proposals are activated once at least three state-local distributions are available.

The complete graph connects this construction to the boundary studied in Theorem 5.4. That theorem concerns increasing aggregation of a convergent sequence and guarantees recovery of its limiting distribution. The present experiment instead uses a moving finite window, so the theorem is not invoked as a convergence guarantee for the experimental advisor. Rather, it motivates the complete graph as a symmetric setting in which recent policy history can be represented without introducing a preferred relation among the four actions.

We consider two POV proposals. The *GraphPOV* proposal is the graph-level distribution of Definition 5.2,

$$\rho_t^{\mathrm{Graph}}(\cdot \mid s) = \mathsf{pov}\big(\mathcal{G}_A, \mathcal{P}_s, m\big). \tag{8}$$

For *BestNodePOV*, we select among the valid node-level distributions using the learner's current Q-values:

$$i^\star \in \arg\max_i \left\langle \mathsf{pov}\big(v_i, \mathcal{P}_s, m\big), Q_t(s,\cdot) \right\rangle, \tag{9}$$

and set

$$\rho_t^{\mathrm{Node}}(\cdot \mid s) = \mathsf{pov}\big(v_{i^\star}, \mathcal{P}_s, m\big). \tag{10}$$

This selection uses only the learner's current $Q_t(s,\cdot)$; the optimal value function introduced below is not available to the advisor.

History baselines and intervention. To distinguish the POV construction from simpler ways of retaining policy history, we compare it with two non-graph history baselines. For each state s, *EMA* maintains an exponential moving average,

$$\rho_t^{\mathrm{EMA}}(\cdot \mid s) = (1 - \alpha_{\mathrm{EMA}})\rho_{t-1}^{\mathrm{EMA}}(\cdot \mid s) + \alpha_{\mathrm{EMA}}\pi_t(\cdot \mid s), \qquad \alpha_{\mathrm{EMA}} = 0.2, \tag{11}$$

while *WindowMean* uses the arithmetic mean of the distributions in the current state-local history window. The *Baseline* uses the same Q-learning procedure without access to policy history. Thus, EMA, WindowMean, GraphPOV, and BestNodePOV all use historical policy information but represent or aggregate it differently.

To evaluate whether a historical proposal contains useful decision information, we use a value-gated intervention rule. The rule moves the current policy in the direction of the historical proposal subject

to a bounded ℓ_1 change, and applies the intervention only when the current Q-function assigns positive expected gain to that direction. Because different history summaries can produce proposals at different natural distances from the current policy, we use a common target ℓ_1 displacement. When a proposal lies close to the current policy, its direction may therefore be extrapolated beyond the proposal, subject to the target displacement, a factor-50 cap, and the simplex boundary.

For any history-based method, let

$$d_t = \rho_t(\cdot \mid s) - \pi_t(\cdot \mid s).$$

If $d_t = 0$, no intervention is made. Otherwise, because both $\rho_t(\cdot \mid s)$ and $\pi_t(\cdot \mid s)$ are categorical distributions, $\mathbf{1}^\top d_t = 0$. The largest positive step along d_t that preserves nonnegativity is

$$\kappa_{\text{simplex}} = \min_{a:\, d_t(a)<0} \frac{\pi_t(a \mid s)}{-d_t(a)}. \tag{12}$$

We then form

$$\widetilde{\pi}_t(\cdot \mid s) = \pi_t(\cdot \mid s) + \kappa_t d_t, \qquad \kappa_t = \min\left\{ \frac{D}{\|d_t\|_1}, 50, (1 - 10^{-10})\kappa_{\text{simplex}} \right\}, \qquad D = 0.2. \tag{13}$$

The first term limits the realized intervention to $\|\widetilde{\pi}_t(\cdot \mid s) - \pi_t(\cdot \mid s)\|_1 \le D$, the second limits the amplification factor to 50, and the third keeps the adjusted distribution inside the probability simplex, with a small numerical margin.

The adjusted proposal is used only when its direction is favorable according to the current learned Q-function. When the denominator is nonzero, we compute

$$g_t = \frac{d_t^\top Q_t(s, \cdot)}{\frac{1}{2}\|d_t\|_1 \big(\max_a Q_t(s,a) - \min_a Q_t(s,a)\big)}, \tag{14}$$

and apply the history-based proposal only when $g_t > 0$. If the denominator vanishes, no history-based intervention is applied. Explicit random exploration is applied after this decision and remains unchanged.

Protocol and evaluation. We run 30 paired seeds, $1000, \ldots, 1029$, for 2500 episodes and evaluate every 50 episodes. Within each seed, all methods use common random numbers for environment stochasticity, exploration decisions, and action sampling. Evaluation is advisor-off: at each checkpoint we evaluate the greedy policy induced by the learned Q-table without using any historical advisor.

Because KeyDoor is a small finite MDP, we independently compute $V^\star$ and $Q^\star$ by value iteration. Neither quantity is used to construct proposals, select POV rows, gate interventions, choose training actions, or update Q-learning. They are used only for evaluation and post-hoc diagnostics.

Let V_t denote the exact start-state value of the current greedy policy. We normalize it relative to the uniform and optimal policies as

$$S_t = \frac{V_t - V^{\text{unif}}}{V^\star - V^{\text{unif}}}, \tag{15}$$

where V^{unif} is the exact start-state value of the uniform random policy. Thus, $S_t = 0$ corresponds to the uniform policy and $S_t = 1$ to the optimal policy. Figure 3 reports S_t throughout training, while Table 10 reports its final value and the first evaluation episode at which $S_t \ge 0.9$. We also summarize the complete learning trajectory by the trapezoidal area under S_t, normalized by the 2500-episode horizon.

Results and interpretation. Table 10, Figure 3, and Figure 4 show that retaining policy history changes the learning trajectory in this controlled environment. Relative to the history-free baseline, the paired mean AUC difference is

$$0.0213 \quad [0.0100,\ 0.0334]$$

for GraphPOV and

$$0.0270 \quad [0.0137,\ 0.0406]$$

Table 10: KeyDoor results over 30 paired seeds. AUC is computed from the advisor-off normalized greedy value in Equation 15. Intervals are 95% bootstrap confidence intervals across seeds. The final column reports the median first evaluation episode at which $S_t \geq 0.9$.

Method	AUC	95% CI	Final normalized value	Episode to 0.9
Baseline	0.8397	[0.8281, 0.8509]	0.9270 [0.9038, 0.9475]	400
EMA	0.8633	[0.8553, 0.8713]	0.9169 [0.8736, 0.9480]	350
WindowMean	0.8568	[0.8489, 0.8645]	0.9226 [0.8601, 0.9676]	350
GraphPOV	0.8610	[0.8507, 0.8708]	0.9509 [0.9378, 0.9619]	300
BestNodePOV	0.8666	[0.8576, 0.8754]	0.9507 [0.9321, 0.9657]	300

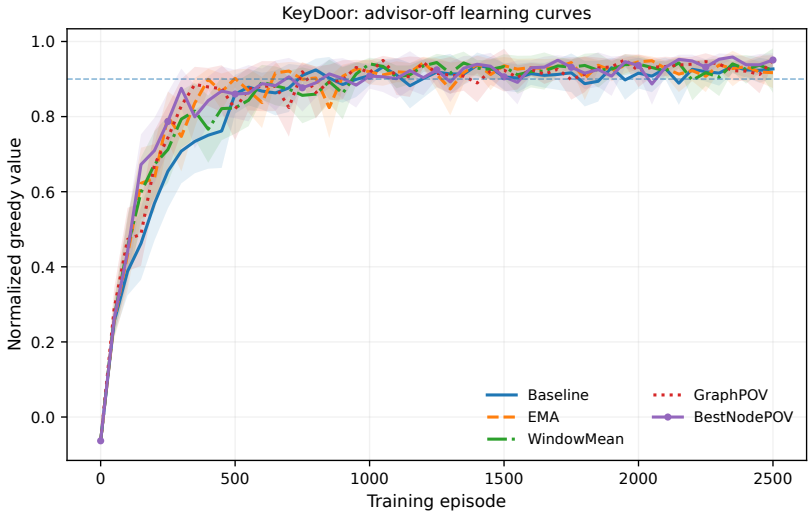


Figure 3: KeyDoor advisor-off learning curves. Curves show the mean normalized greedy value over 30 paired seeds for the history-free baseline and the four history-based methods; shaded regions show 95% confidence intervals.

for BestNodePOV, where brackets denote paired 95% bootstrap confidence intervals. The simpler history summaries also attain larger mean AUC than the history-free baseline: 0.8633 for EMA and 0.8568 for WindowMean, compared with 0.8397 for the baseline. Thus, the experiment does not identify POV as uniquely responsible for the changed learning behavior. Rather, it provides a controlled comparison of different ways of retaining and using sequences of state-local policy distributions.

The two POV variants themselves have similar learning-curve performance. The paired difference

$$\mathrm{AUC}_{\mathrm{BestNodePOV}} - \mathrm{AUC}_{\mathrm{GraphPOV}} = 0.0057$$

has 95% bootstrap confidence interval

$$[-0.0077,\ 0.0193].$$

The experiment therefore does not provide evidence for a clear performance separation between node-level selection and graph-level aggregation in this setting.

The availability of $Q^\star$ permits a separate diagnostic of whether these POV alternatives contain decision-relevant information. For BestNodePOV, 64.8% of natural proposals have positive expected direction under $Q^\star$. Among interventions that actually change the sampled action, 74.4% change it toward an action with larger $Q^\star$. When the selected node-level POV is compared directly with GraphPOV under $Q^\star$, the selected node distribution has larger oracle value in 96.9% of comparisons, with mean oracle rank 1.29 among the available POV rows. Since $Q^\star$ is never supplied to the learner or advisor, these diagnostics indicate that finite complete-graph POV can produce decision-relevant alternative distributions from the learner's own policy history in this environment.

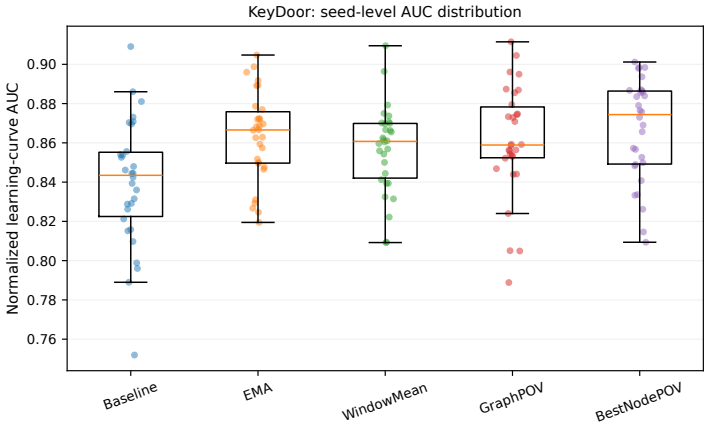


Figure 4: Seed-level distribution of normalized learning-curve AUC in KeyDoor. EMA, WindowMean, GraphPOV, and BestNodePOV use policy history, whereas the Baseline does not. The figure compares different ways of retaining and using that history rather than implying a POV-specific performance advantage.

Overall, the KeyDoor experiment illustrates one possible use of Definition 5.2 together with the complete-graph boundary of Theorem 5.4. The theorem provides a consistency reference for convergent sequences, whereas the experiment examines the distinct finite-history regime in which the resulting POV can differ from the current policy. The results show that these history-conditioned distributions can influence learning and can contain decision-relevant information in this environment, but they do not establish that POV is preferable to simpler policy-history summaries or that POV generally improves reinforcement learning.

F.2 Structured Action Spaces: Pendulum

Discretized action space and action graphs. We next consider `Pendulum-v1`. Its original control variable is continuous; for this experiment, we discretize the torque interval $[-2, 2]$ into

$$\mathcal{A}_{11} = \{-2, -1.6, -1.2, -0.8, -0.4, 0, 0.4, 0.8, 1.2, 1.6, 2\},$$

and associate these 11 executable torques with nodes $v_1, \ldots, v_{11}$. Unlike the four actions in KeyDoor, these actions inherit an ordering from the underlying continuous control space. We use this structure to define two different action graphs.

The *Path* graph connects adjacent torque values,

$$E_{\text{Path}} = \big\{\{v_j, v_{j+1}\} : j = 1, \ldots, 10\big\}, \tag{16}$$

representing local proximity along the discretized torque axis. The *ZeroHub* graph augments the Path graph by connecting the zero-torque node v_6 to every other action:

$$E_{\text{ZeroHub}} = E_{\text{Path}} \cup \big\{\{v_6, v_j\} : j \neq 6\big\}. \tag{17}$$

These graphs are not intended as optimal action-space representations. Rather, they provide two simple structural hypotheses over the same discretized action set. Because the graph adjacency determines the support of the distribution-induced matrices and hence the paths accumulated through $\circ$-composition, changing the action graph can change the resulting POV distributions.

Both graphs are non-complete, so Theorem 5.4 does not provide a recovery guarantee for either construction. The experiment instead applies Definition 5.2 directly in a setting where the chosen action topology is allowed to condition the policy-history sequence.

Neural Q learner and historical policy sequence. All methods use the same neural Q-function with architecture

$$3 \;-\; 128 \;-\; 128 \;-\; 11$$

and ReLU hidden activations. We use discount $\gamma = 0.99$, Adam with learning rate 10^{-3}, replay-buffer capacity 10^5, replay batch size 128, and a replay warmup of 2000 interactions. The target network is updated every 1000 interactions. For the bootstrapped target, the online network selects the next action and the target network evaluates it.

At observation s, a Q-network snapshot induces the categorical action distribution

$$P_s^{(k)} = \mathrm{softmax}\left(\frac{Q^{(k)}(s,\cdot)}{T}\right), \qquad T = 20. \tag{18}$$

A frozen copy of the Q-network is stored after each training episode. At the current observation, the most recent $L = 20$ snapshots are reevaluated at that same observation and ordered chronologically, giving the state-local policy-history sequence

$$\mathcal{P}_s = \{P_s^{(1)}, \ldots, P_s^{(L)}\}.$$

Historical intervention begins after a 20-episode warmup. Explicit random exploration is applied with probability 0.08.

For either action graph

$$\mathcal{G} \in \{\mathcal{G}_{\mathrm{Path}}, \mathcal{G}_{\mathrm{ZeroHub}}\},$$

with adjacency matrix $\boldsymbol{A}_{\mathcal{G}}$, Definition 5.1 gives

$$\boldsymbol{W}_{\mathcal{G},s}^{(k)} = \boldsymbol{A}_{\mathcal{G}}\,\mathrm{diag}\left(\frac{P_s^{(k)}}{1 - P_s^{(k)}}\right), \tag{19}$$

where the quotient is elementwise. The chronological realization is

$$\boldsymbol{M}_{\mathcal{G},s} = \boldsymbol{W}_{\mathcal{G},s}^{(1)} \circ \cdots \circ \boldsymbol{W}_{\mathcal{G},s}^{(L)} = \prod_{k=1}^{L}\left(\boldsymbol{I} + \boldsymbol{W}_{\mathcal{G},s}^{(k)}\right) - \boldsymbol{I}, \tag{20}$$

with the product taken in increasing chronological order.

For each topology, we consider two POV proposals. *GraphPOV* uses

$$\mathsf{pov}\big(\mathcal{G}, \mathcal{P}_s, L\big),$$

whereas *BestNodePOV* selects the valid node-level POV distribution having the largest inner product with the current learned Q-values. This gives four POV methods:

GraphPOV-Path, BestNodePOV-Path,
GraphPOV-ZeroHub, BestNodePOV-ZeroHub.

As a non-topological history baseline, *EMA* is computed from the same chronologically ordered snapshot-policy sequence with coefficient 0.2. We also include the same discrete neural Q learner without historical intervention, denoted *DQN*. Thus, for a fixed POV aggregation rule, the Path and ZeroHub variants differ only in the adjacency matrix entering Equation 19.

Capped intervention. Unlike the KeyDoor experiment, Pendulum does not extrapolate beyond the historical proposal. Let $\pi_t(\cdot \mid s)$ denote the current action distribution, $\rho_t(\cdot \mid s)$ the historical proposal, and

$$d_t = \rho_t(\cdot \mid s) - \pi_t(\cdot \mid s).$$

If $\|d_t\|_1 < 10^{-3}$, the advisor abstains. Otherwise,

$$\widetilde{\pi}_t(\cdot \mid s) = \pi_t(\cdot \mid s) + \kappa_t d_t, \qquad \kappa_t = \min\left\{1,\ \frac{0.2}{\|d_t\|_1},\ (1 - 10^{-10})\kappa_{\mathrm{simplex}}\right\}, \tag{21}$$

where $\kappa_{\mathrm{simplex}}$ is defined as in Equation 12. Hence,

$$\kappa_t \le 1, \qquad \|\widetilde{\pi}_t(\cdot \mid s) - \pi_t(\cdot \mid s)\|_1 \le \min\left\{0.2,\ \|\rho_t(\cdot \mid s) - \pi_t(\cdot \mid s)\|_1\right\}.$$

Thus, the adjusted distribution remains between the current policy and the historical proposal, with 0.2 serving as a common upper bound on the intervention magnitude rather than forcing all advisors to make perturbations of equal size.

As in KeyDoor, the proposal is accepted only when its direction has positive expected value under the current learned Q-function:

$$d_t^{\top} Q_t(s,\cdot) > 0. \tag{22}$$

Explicit random exploration is applied after this decision and overrides the historical advisor.

Table 11: Pendulum advisor-off evaluation return during the late part of training. Values are means over 30 paired seeds; larger return is better. The checkpoints are reported descriptively and are not treated as separate confirmatory tests.

Method	18k	20k	22k	24k	25k
DQN	−162.86	−168.73	−155.81	−154.59	−165.71
EMA	−162.07	−160.83	−155.19	−153.87	−161.10
GraphPOV-Path	−166.26	−158.30	−155.43	−155.84	−157.46
BestNodePOV-Path	−167.01	−156.02	−154.73	−154.98	−156.67
GraphPOV-ZeroHub	−165.25	−157.74	−153.82	−155.46	−156.73
BestNodePOV-ZeroHub	−176.04	−160.53	−154.31	−152.85	−161.96

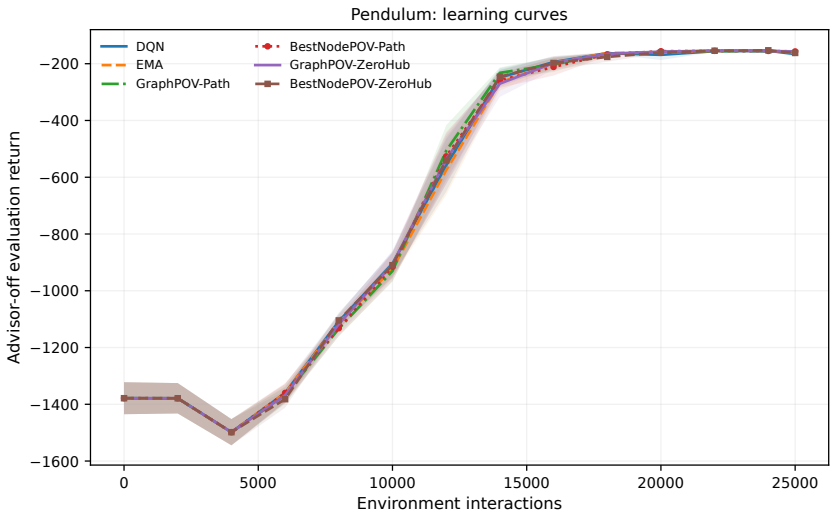


Figure 5: Pendulum advisor-off evaluation return throughout training for DQN, EMA, and the four POV variants. Curves show means over 30 paired seeds; shaded regions show 95% confidence intervals. The Path and ZeroHub variants do not exhibit a persistent return ordering.

Protocol and evaluation. We run 30 paired seeds, $10, \ldots, 39$, for 25,000 environment interactions. Methods within each seed share the corresponding reset, exploration, and action-sampling random streams. Evaluation is performed every 2000 interactions and once more at 25,000. At every checkpoint, the historical advisor is disabled and the greedy Q-policy is evaluated for 20 episodes using a fixed common set of evaluation reset seeds. The reported return therefore evaluates the controller learned under historical intervention rather than the immediate effect of using the advisor at test time.

Results and interpretation. We first examine whether the choice of action graph produces a persistent difference in downstream control performance. Table 11 and Figure 5 do not show such an ordering. From 20k onward, GraphPOV-Path and GraphPOV-ZeroHub remain close, with 25k mean returns of −157.46 and −156.73, respectively. The BestNode variants likewise do not exhibit a persistent ordering: their 25k means are −156.67 for Path and −161.96 for ZeroHub, while their relative ordering changes across the preceding late-stage checkpoints. Thus, these two action graphs do not produce a consistent return separation in this experiment.

They nevertheless produce substantially different POV advisor behavior. Figure 6 shows that GraphPOV-Path proposals are accepted by the learned-Q gate in approximately 23.6% of eligible decisions and change the sampled action in approximately 4.6%, whereas GraphPOV-ZeroHub yields corresponding rates of 63.9% and 16.6%. For BestNodePOV, the Path rates are approximately 100.0% and 52.8%, compared with 76.9% and 19.2% for ZeroHub.

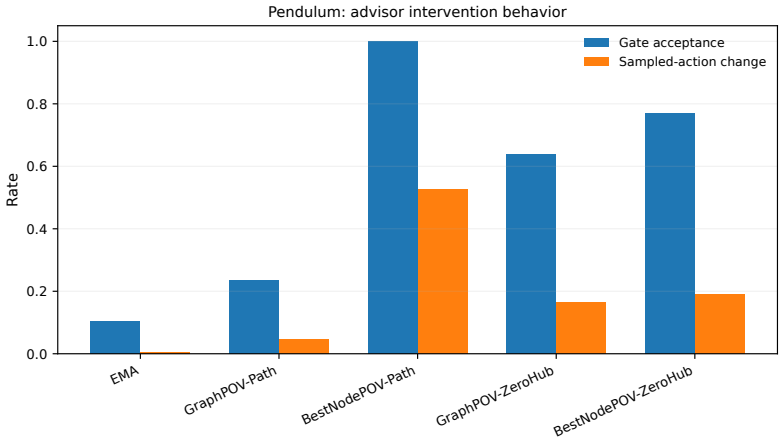


Figure 6: Pendulum advisor behavior under the common capped-intervention rule. Gate-acceptance and sampled-action-change rates are aggregated across the 30 paired seeds. Changing the action graph substantially changes the intervention behavior of both POV aggregation rules, although the two topologies do not exhibit a persistent return ordering.

Thus, changing only the action graph changes how the same form of policy-history sequence is transformed and used by the POV advisor. This provides an empirical illustration of the role of topology in Definition 5.2: the graph is not merely auxiliary metadata but participates directly in the resulting distributional representation. The experiment does not establish that either Path or ZeroHub is preferable for control; rather, it shows that different structural hypotheses over the same action set can lead to materially different topology-conditioned historical distributions.

Taken together, the KeyDoor and Pendulum experiments illustrate two complementary uses of the distributional POV construction. KeyDoor uses a complete action graph as a structurally symmetric reference for finite policy history, with Theorem 5.4 providing the associated asymptotic boundary. Pendulum instead uses non-complete action graphs to condition the same type of policy-history sequence on explicit relations among actions. The resulting advisor behavior is sensitive to the chosen topology even though the tested graph choices do not yield a persistent return ordering. These experiments therefore motivate policy histories as one possible source of distribution sequences for Definition 5.2, while leaving the design of action graphs and the role of historical POV in reinforcement learning as open questions.

G Computational Complexity of POV

Let $\mathcal{G} = (V, E, X)$ be a finite undirected loopless attributed graph, with

$$n = |V| \geq 2, \qquad e = |E|,$$

and let $\boldsymbol{A}$ be its adjacency matrix stored in sparse form. We consider an ordered sequence of categorical distributions

$$\mathcal{P} = \{P^{(1)}, \ldots, P^{(m)}\}, \qquad m \geq 1.$$

Table 12 summarizes the principal computational bounds established in this section.

We assume $0 \leq P_j^{(k)} < 1$ for every j, k. Probability distributions are represented by row vectors; auxiliary vectors used in matrix–vector products are column vectors. All time bounds below count scalar arithmetic operations. Memory bounds include the sparse graph and working storage, but exclude storage of the supplied distribution sequence unless stated otherwise. The cost of obtaining the graph and generating the conditioning distributions is not included.

For each k, define

$$d_j^{(k)} = \frac{P_j^{(k)}}{1 - P_j^{(k)}}, \qquad \boldsymbol{D}^{(k)} = \mathrm{diag}(d^{(k)}), \qquad \boldsymbol{W}^{(k)} = \boldsymbol{A}\boldsymbol{D}^{(k)},$$

Table 12: Summary of POV computational complexity. Here q is the number of shifted-power iterations, N_{mv} is the number of matrix–vector products used by the sparse symmetric eigensolver, and h is the number of retained Lanczos basis vectors. For the sparse symmetric eigensolver, the reported time bound counts matrix–vector products; Krylov-subspace orthogonalization, projected eigenproblem, and restart costs are additional.

Computation	Time	Working memory
Explicit finite DRA / full POV matrix	$O\big(mn(n+e)\big)$	$O(n^2)$
Exact matrix-free graph-level finite POV	$O\big(m(n+e)\big)$	$O(n+e)$
Limiting POV: shifted power	$O\big(q(n+e)\big)$	$O(n+e)$
Limiting POV: sparse symmetric eigensolver	$O\big(N_{\mathrm{mv}}(n+e)\big)$	$O(n+e+nh)$

and write

$$\boldsymbol{F}^{(k)} = \boldsymbol{I} + \boldsymbol{W}^{(k)}.$$

Forming $d^{(k)}$ requires $O(n)$ operations. Multiplication by the nonnegative diagonal matrix $\boldsymbol{D}^{(k)}$ cannot introduce new nonzero entries, so

$$\operatorname{supp}(\boldsymbol{W}^{(k)}) \subseteq \operatorname{supp}(\boldsymbol{A}).$$

Equality holds when every coordinate of $P^{(k)}$ is strictly positive; zero coordinates can remove nonzero entries. Since $\boldsymbol{A}$ has $2e$ nonzeros, $\boldsymbol{W}^{(k)}$ has $O(e)$ nonzeros and $\boldsymbol{F}^{(k)}$ has $O(n+e)$ nonzeros. The level-m distributional realization is

$$\boldsymbol{M} = \mathsf{DRA}(\mathcal{P}, m) = \boldsymbol{F}^{(1)}\boldsymbol{F}^{(2)}\cdots\boldsymbol{F}^{(m)} - \boldsymbol{I}, \tag{23}$$

where the product is ordered by increasing k.

Explicit finite DRA and full POV matrix. Although each factor in Equation 23 is sparse, repeated multiplication can make the intermediate products dense. For a dense $n \times n$ intermediate matrix $\boldsymbol{B}$,

$$\boldsymbol{B}\boldsymbol{F}^{(k)} = \boldsymbol{B} + (\boldsymbol{B}\boldsymbol{A})\boldsymbol{D}^{(k)}.$$

The dense–sparse multiplication $\boldsymbol{B}\boldsymbol{A}$ costs $O(ne)$, while diagonal scaling and dense addition cost $O(n^2)$. Consequently, each multiplication by a sparse factor admits the bound

$$O(ne + n^2) = O\big(n(n+e)\big).$$

Direct sequential construction of the complete level-m DRA therefore admits the worst-case time bound

$$O\big(mn(n+e)\big). \tag{24}$$

Computing row sums and normalizing all nonzero rows costs at most an additional $O(n^2)$, which is absorbed into this bound.

If the complete realization or the full matrix of node-level POV distributions is stored explicitly, the worst-case storage requirement is $O(n^2)$. The same working-memory bound suffices when intermediate products overwrite previously computed products and the sparse factors are generated as needed. Retaining all m input distributions adds $O(mn)$ storage, giving $O(n^2 + mn)$ total memory.

For a connected graph with $n \geq 2$, we have $e \geq n-1$, and hence $n + e = O(e)$. Equation 24 then simplifies to

$$O(mne) = O(m|V||E|)\,.$$

These are upper bounds for direct sequential construction, rather than optimality claims over all possible algorithms. The actual cost can be smaller when intermediate products remain sparse.

Exact matrix-free graph-level finite POV. When only the graph-level distribution is required, the $n \times n$ realization $\boldsymbol{M}$ need not be materialized. Because the factors have the form $\boldsymbol{F}^{(k)} = \boldsymbol{I} + \boldsymbol{W}^{(k)}$, with $\boldsymbol{W}^{(k)} \geq 0$, expanding their ordered product shows that $\boldsymbol{M} \geq 0$. Thus, the ℓ_1-norm of a nonzero row equals its row sum:

$$\|\boldsymbol{M}_{i,:}\|_1 = \boldsymbol{M}_{i,:}\mathbf{1}.$$

Assume first that every row is nonzero, and define the column vectors $r, a \in \mathbb{R}^n$ by

$$r = \boldsymbol{M}\mathbf{1}, \qquad a_i = \frac{1}{r_i}.$$

By Definition 5.2,

$$\begin{aligned}\mathsf{pov}(\mathcal{G}, \mathcal{P}, m) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\boldsymbol{M}_{i,:}}{r_i} \\ &= \frac{1}{n} a^\top \boldsymbol{M}. \end{aligned} \tag{25}$$

Therefore, computing the graph-level POV requires only the vectors $\boldsymbol{M}\mathbf{1}$ and $\boldsymbol{M}^\top a$, rather than all entries of $\boldsymbol{M}$.

From Equation 23,

$$\boldsymbol{M}\mathbf{1} = \boldsymbol{F}^{(1)} \cdots \boldsymbol{F}^{(m)}\mathbf{1} - \mathbf{1}, \tag{26}$$

and, after forming a,

$$\boldsymbol{M}^\top a = (\boldsymbol{F}^{(m)})^\top \cdots (\boldsymbol{F}^{(1)})^\top a - a. \tag{27}$$

The first product is evaluated by applying the factors to $\mathbf{1}$ in the order $k = m, m-1, \ldots, 1$. The second is evaluated by applying the transposed factors to a in the order $k = 1, 2, \ldots, m$. These orders matter because the factors need not commute.

For any column vector x,

$$\begin{aligned}\boldsymbol{F}^{(k)} x &= x + \boldsymbol{A}\big(d^{(k)} \odot x\big), \\ (\boldsymbol{F}^{(k)})^\top x &= x + d^{(k)} \odot (\boldsymbol{A}^\top x).\end{aligned}$$

Here, $\odot$ denotes elementwise multiplication. Although $\boldsymbol{A}^\top = \boldsymbol{A}$, the matrices $\boldsymbol{F}^{(k)}$ are generally not symmetric, so the two actions must be distinguished. Each action requires a sparse matrix–vector multiplication and $O(n)$ vector operations, giving $O(n + e)$ arithmetic work.

Each of Equations equation 26 and equation 27 requires m factor applications. Forming the odds weights as needed, computing a, and assembling the final distribution do not increase the resulting bound. Hence the exact graph-level finite POV can be evaluated in

$$O\big(m(n + e)\big) \tag{28}$$

time.

Only the sparse adjacency matrix, the current odds-weight vector, and a constant number of length-n working vectors are required. The working memory is therefore $O(n + e)$, excluding the input sequence. This bound assumes that the distributions can be revisited in reverse and forward order; it is not a single-pass streaming bound for a sequence available only once in chronological order. If all m distributions are retained in memory, total storage is $O(n + e + mn)$. For a constant sequence $P^{(k)} = P$, only one distribution and its odds-weight vector need to be stored.

For a connected graph, Equation 28 simplifies to

$$O(me) = O(m|E|)\,.$$

This computation evaluates the defining average of normalized DRA rows exactly in real arithmetic, without materializing either the full DRA matrix or the full matrix of node-level POV distributions.

Subtraction-free evaluation. The two vector products can also be evaluated by recurrences that avoid explicitly subtracting the initial vectors. For the row sums, initialize $u_{m+1} = \mathbf{0}$ and compute backward:

$$u_k = u_{k+1} + \boldsymbol{A}\big(d^{(k)} \odot (u_{k+1} + \mathbf{1})\big), \qquad k = m, m-1, \ldots, 1. \tag{29}$$

Indeed,

$$u_k = \big(\boldsymbol{F}^{(k)} \cdots \boldsymbol{F}^{(m)} - \boldsymbol{I}\big)\mathbf{1},$$

so $r = u_1 = \boldsymbol{M}\mathbf{1}$.

After forming a, initialize $v_0 = \mathbf{0}$ and compute forward:

$$v_k = v_{k-1} + d^{(k)} \odot \left[\boldsymbol{A}^\top (v_{k-1} + a)\right], \qquad k = 1, 2, \ldots, m. \tag{30}$$

The identity

$$v_k = \left[\left(\boldsymbol{F}^{(1)} \cdots \boldsymbol{F}^{(k)}\right)^\top - \boldsymbol{I}\right] a$$

gives $v_m = \boldsymbol{M}^\top a$. Thus,

$$\mathsf{pov}(\mathcal{G}, \mathcal{P}, m) = \frac{1}{n} v_m^\top.$$

Both recurrences use only the current and next vectors; their full histories need not be stored. They therefore retain the $O\big(m(n+e)\big)$ time and $O(n+e)$ working-memory bounds.

Here, "exact" means algebraically equivalent to the defining finite matrix expression, without truncation or sampling. Floating-point evaluation remains subject to roundoff, underflow, and overflow. The subtraction-free recurrences avoid the cancellation associated with explicitly subtracting the initial vectors, but they do not eliminate all numerical issues, particularly at large aggregation levels or when probabilities approach one.

Zero-row fallback. In the constant-distribution experiment of Appendix E, a zero row is replaced by the conditioning distribution P. Define

$$I_0 = \{i : r_i = 0\}, \qquad a_i = \begin{cases} 1/r_i, & r_i > 0, \\ 0, & r_i = 0. \end{cases}$$

Because $\boldsymbol{M} \geq 0$, $r_i = 0$ exactly when the corresponding row is zero. The graph-level distribution under this convention is

$$Q^{(m)} = \frac{1}{n}\left(a^\top \boldsymbol{M} + |I_0| P\right) = \frac{1}{n}\left(v_m^\top + |I_0| P\right).$$

The same recurrences apply with this definition of a. Moreover,

$$a^\top \boldsymbol{M} \mathbf{1} = n - |I_0|,$$

so the fallback expression has total mass one. Identifying zero rows and adding the fallback contribution costs $O(n)$, without materializing a separate copy of P for every zero row. The asymptotic time and memory bounds are unchanged.

Limiting POV. For a fixed strictly positive distribution, the limiting POV can be evaluated through a Perron eigenvalue problem rather than by choosing a large finite aggregation level. In the connected-graph setting of Theorem 5.3, define

$$d_i = \frac{P_i}{1 - P_i}, \qquad \boldsymbol{D} = \operatorname{diag}(d), \qquad \boldsymbol{W} = \boldsymbol{A}\boldsymbol{D}.$$

Since $P_i > 0$ for every i and $n \geq 2$, all entries of d are positive and finite. The symmetric matrix

$$\boldsymbol{S} = \boldsymbol{D}^{1/2} \boldsymbol{A} \boldsymbol{D}^{1/2} \tag{31}$$

satisfies

$$\boldsymbol{S} = \boldsymbol{D}^{1/2} \boldsymbol{W} \boldsymbol{D}^{-1/2},$$

so $\boldsymbol{S}$ and $\boldsymbol{W}$ are similar. If $z > 0$ is a column Perron eigenvector of $\boldsymbol{S}$, with

$$\boldsymbol{S} z = \rho(\boldsymbol{S}) z,$$

then the row vector

$$\ell = z^\top \boldsymbol{D}^{1/2}$$

satisfies

$$\ell \boldsymbol{W} = z^\top \boldsymbol{S} \boldsymbol{D}^{1/2} = \rho(\boldsymbol{S}) \ell.$$

It is therefore a left Perron eigenvector of $\boldsymbol{W}$. Theorem 5.3 gives

$$\mathsf{pov}(\mathcal{G}, P, \infty) = \frac{\ell}{\ell \mathbf{1}}.$$

The matrix $\boldsymbol{S}$ has the same sparsity pattern as $\boldsymbol{A}$. Its action can be evaluated without explicitly forming the diagonal matrices or $\boldsymbol{S}$:

$$\boldsymbol{S}x = \sqrt{d} \odot \left[\boldsymbol{A}\big(\sqrt{d} \odot x\big)\right].$$

Here, $\sqrt{d}$ is computed elementwise. Each application costs $O(n + e)$. A numerical approximation to the Perron eigenvector can be obtained using a sparse symmetric eigensolver, with shifted power iteration as a fallback.

Shifted power iteration. Let

$$c = \max_i (\boldsymbol{S}\mathbf{1})_i, \qquad x_0 = n^{-1/2}\mathbf{1}.$$

Connectedness and strict positivity of d imply $c > 0$. Computing c requires one application of $\boldsymbol{S}$ and an $O(n)$ maximum operation. Consider the iteration

$$x_{t+1} = \frac{x_t + c^{-1}\boldsymbol{S}x_t}{\|x_t + c^{-1}\boldsymbol{S}x_t\|_2}. \tag{32}$$

The matrix $\boldsymbol{I} + c^{-1}\boldsymbol{S}$ has the same eigenvectors as $\boldsymbol{S}$. It is nonnegative and irreducible with positive diagonal entries, and is therefore primitive. Consequently, the positive initialization ensures convergence to its normalized positive Perron eigenvector, which is also the Perron eigenvector of $\boldsymbol{S}$. The shift avoids the equal-magnitude positive and negative extreme eigenvalues that can occur for a bipartite graph.

If $q \geq 1$ iterations are performed to reach the requested numerical accuracy, the setup and iteration cost is

$$O\big((q + 1)(n + e)\big) = O\big(q(n + e)\big).$$

The final conversion to ℓ and normalization require $O(n)$ additional work, already absorbed into this bound. The working memory is $O(n + e)$. For a connected graph, the time bound simplifies to

$$O(qe).$$

The iteration count q depends on the requested accuracy, the spectral gap, and the initialization; it is not assumed to be constant.

Sparse symmetric eigensolver. A Lanczos/ARPACK-type solver can instead be applied to the standard symmetric eigenproblem for $\boldsymbol{S}$, using the matrix-free action above. The required target is the largest *algebraic* eigenvalue, rather than the largest-magnitude eigenvalue: on a bipartite graph, both $\rho(\boldsymbol{S})$ and $-\rho(\boldsymbol{S})$ can have maximal magnitude.

Each matrix-operator application costs $O(n + e)$. If N_{mv} applications are performed, their cumulative cost is

$$O\big(N_{\mathrm{mv}}(n + e)\big),$$

in addition to Krylov-subspace orthogonalization, projected eigenproblem, and restart costs. If h Lanczos basis vectors are retained, basis storage contributes $O(nh)$, giving working memory

$$O(n + e + nh).$$

Thus, the overall eigensolver cost depends on numerical convergence and the retained subspace size, rather than on the finite aggregation level m. These operator-cost bounds concern the standard matrix eigenproblem, not a shift-invert implementation requiring linear-system solves.

Disconnected-graph extension. For the constant strictly positive distribution used in the disconnected-graph extension of Appendix E, first identify the connected components in $O(n + e)$ time. Let $\mathcal{C}_1, \ldots, \mathcal{C}_K$ be the non-isolated components, let I_0 be the set of isolated nodes, and write

$$n_c = |V(\mathcal{C}_c)|, \qquad e_c = |E(\mathcal{C}_c)|.$$

For component $\mathcal{C}_c$, use the principal submatrices $\boldsymbol{A}_c$ and $\boldsymbol{D}_c$ of the original global matrices. In particular, the diagonal entries remain

$$(\boldsymbol{D}_c)_{i,i} = \frac{P_i}{1 - P_i}, \qquad i \in V(\mathcal{C}_c).$$

The input distribution is not renormalized within a component before forming the odds weights, since doing so would generally change the operator.

Define

$$\boldsymbol{S}_c = \boldsymbol{D}_c^{1/2} \boldsymbol{A}_c \boldsymbol{D}_c^{1/2}.$$

If $z_c > 0$ is its Perron eigenvector, the normalized component distribution is the row vector

$$\pi_c = \frac{z_c^\top \boldsymbol{D}_c^{1/2}}{z_c^\top \boldsymbol{D}_c^{1/2} \mathbf{1}}.$$

Let $\widetilde{\pi}_c$ denote its extension to the full node set by placing zeros outside $\mathcal{C}_c$. The experimental componentwise convention gives

$$Q^{(\infty)} = \sum_{c=1}^{K} \frac{n_c}{n} \widetilde{\pi}_c + \frac{|I_0|}{n} P.$$

This is the disconnected-graph extension specified in Appendix E, rather than an additional assertion of Theorem 5.3.

If the shifted-power solve on component $\mathcal{C}_c$ requires $q_c \geq 1$ iterations, the total arithmetic cost is

$$O\left(n + e + \sum_{c=1}^{K} q_c (n_c + e_c) \right).$$

The normalized component distributions can be written directly into their corresponding positions in the output, and the isolated-node fallback contribution can be added once. This assembly requires $O(n)$ work and does not require storing a separate length-n extension for each component.

Processing the components sequentially with shifted power iteration retains an $O(n + e)$ working-memory bound. For sequential Lanczos solves retaining h_c basis vectors on component $\mathcal{C}_c$, the corresponding bound is

$$O\left(n + e + \max_{1 \leq c \leq K} n_c h_c \right),$$

with no eigensolver storage required when $K = 0$.

H PROOFS OF THE THEOREMS

PROOF OF THEOREM 4.1

Proof. The composition of two morphisms $e_1 \bullet \cdots \bullet e_k : u \to v$ and $d_1 \bullet \cdots \bullet d_l : v \to w$ is defined by

$$(e_1 \bullet \cdots \bullet e_k) \bullet (d_1 \bullet \cdots \bullet d_l) = e_1 \bullet \cdots \bullet e_k \bullet d_1 \bullet \cdots \bullet d_l : u \to w.$$

The associativity of the composition follows directly from the associativity of the monoidal operation $\bullet$.

For each node v, we assign the morphism

$$0 : v \to v,$$

where 0 is the identity element of the monoid $\mathsf{Mod}(\mathcal{G})$. Although these identity morphisms have the same underlying monoidal element 0, they are distinguished by their source and target; in particular, for $u \neq v$, the morphisms $0 : u \to u$ and $0 : v \to v$ are distinct morphisms. Thus, $0 : v \to v$ is the identity morphism associated with the object v.

Since 0 is the identity element of $\mathsf{Mod}(\mathcal{G})$, for every morphism

$$e_1 \bullet \cdots \bullet e_k : u \to v$$

we have

$$0 \bullet (e_1 \bullet \cdots \bullet e_k) = e_1 \bullet \cdots \bullet e_k$$

and

$$(e_1 \bullet \cdots \bullet e_k) \bullet 0 = e_1 \bullet \cdots \bullet e_k,$$

where the first 0 denotes the identity morphism of u and the second denotes the identity morphism of v.

Therefore, composition is associative and every object has an identity morphism, so $\mathsf{Cat}(\mathcal{G})$ satisfies the axioms of a category. □

PROOF OF THEOREM 4.2

First, we prove the following lemma.

Lemma H.1 (Edge atomicity under $\bullet$)**.** *Let e be a directed edge, regarded as the element $(e, \{e\})$ of $\mathsf{Mod}(\mathcal{G})$. If*

$$e = a_1 \bullet \cdots \bullet a_\ell$$

for composable morphisms $a_1, \ldots, a_\ell$ in $\mathsf{Cat}(\mathcal{G})$, then exactly one factor a_{i_0} is non-identity and all other factors are identity morphisms.

Proof. Let $J = \{i \in \{1, \ldots, \ell\} : a_i \text{ is non-identity}\}$. For each $i \in J$, write $a_i = (M_i, S_i)$. Identity factors contribute the empty multigraph and therefore do not affect the multigraph component of the product. By the definition of $\bullet$, the multigraph component of $a_1 \bullet \cdots \bullet a_\ell$ is $\bigoplus_{i \in J} M_i$. Since the edge sets in this sum are combined by disjoint union, every non-identity factor contributes at least one directed-edge occurrence. On the other hand, the directed-edge element $e = (e, \{e\})$ contains exactly one directed-edge occurrence in its multigraph component. Hence, if two distinct factors a_i and a_j were non-identity, the multigraph component of their product would contain at least two directed-edge occurrences and therefore could not equal the single-edge element e. Thus, $|J| \leq 1$. Since the product equals the non-identity element e, at least one factor must be non-identity. Therefore, $|J| = 1$. Hence exactly one factor is non-identity and all remaining factors are identity morphisms. □

We now prove Theorem 4.2.

Proof. Suppose first that $f : \mathcal{G} \longrightarrow \mathcal{H}$ is a graph isomorphism. The graph isomorphism f induces a monoidal isomorphism $\mathsf{Mod}(f) : \mathsf{Mod}(\mathcal{G}) \longrightarrow \mathsf{Mod}(\mathcal{H})$ (see Langari et al. (2026)). We define $\mathsf{Cat}(f) : \mathsf{Cat}(\mathcal{G}) \longrightarrow \mathsf{Cat}(\mathcal{H})$ as follows. On objects, set $\mathsf{Cat}(f)(v) = f(v)$. For a morphism

$$e_1 \bullet \cdots \bullet e_m : u \to v,$$

define

$$\mathsf{Cat}(f)(e_1 \bullet \cdots \bullet e_m) = \mathsf{Mod}(f)(e_1 \bullet \cdots \bullet e_m) : f(u) \to f(v).$$

Indeed, since f is a graph isomorphism, it maps every directed edge $x \to y$ of $\mathcal{G}$ to a directed edge $f(x) \to f(y)$ of $\mathcal{H}$; hence composable sequences are mapped to composable sequences with the corresponding endpoints. Since $\mathsf{Mod}(f)$ is a monoidal isomorphism, it preserves the identity element and the operation $\bullet$. Therefore $\mathsf{Cat}(f)$ preserves identity morphisms and composition, and hence is a functor.

Let $g : \mathcal{H} \longrightarrow \mathcal{G}$ be the inverse graph isomorphism of f. The same construction gives a functor

$$\mathsf{Cat}(g) : \mathsf{Cat}(\mathcal{H}) \longrightarrow \mathsf{Cat}(\mathcal{G}).$$

Since $g = f^{-1}$, these two functors are inverse to one another. Therefore, $\mathsf{Cat}(\mathcal{G}) \cong \mathsf{Cat}(\mathcal{H})$.

Conversely, suppose that $L : \mathsf{Cat}(\mathcal{G}) \longrightarrow \mathsf{Cat}(\mathcal{H})$ is an isomorphism of categories, with inverse $R : \mathsf{Cat}(\mathcal{H}) \longrightarrow \mathsf{Cat}(\mathcal{G})$. The object map of L is therefore a bijection. Denote it by

$$\phi : V(\mathcal{G}) \longrightarrow V(\mathcal{H}), \qquad \phi(v) = L(v).$$

Let $e : u \to v$ be a directed edge of $\mathcal{G}$. Since e is non-identity, $L(e)$ is also non-identity. Indeed, if $L(e)$ were an identity morphism, then $e = R(L(e))$ would also be an identity morphism, which is impossible. By the definition of $\mathsf{Cat}(\mathcal{H})$, we may therefore write $L(e) = d_1 \bullet \cdots \bullet d_\ell, \quad \ell \geq 1$, where $d_1, \ldots, d_\ell$ are composable directed edges of $\mathcal{H}$.

Applying the inverse functor R and using preservation of composition gives

$$e = R(L(e)) = R(d_1) \bullet \cdots \bullet R(d_\ell).$$

By Lemma H.1, exactly one of the morphisms $R(d_i)$ is non-identity and all the others are identity morphisms. Suppose that an index $i \neq i_0$ existed. Then $R(d_i)$ would be an identity morphism. Since L preserves identity morphisms and $L \circ R$ is the identity functor, $d_i = L(R(d_i))$ would also be an identity morphism. This contradicts the fact that d_i is a directed edge. Hence no such index exists, and therefore $\ell = 1$. Consequently, $L(e) = d$ for some directed edge $d : \phi(u) \to \phi(v)$ of $\mathcal{H}$. Similarly, L maps the reverse edge $e' : v \to u$ to a directed edge $d' : \phi(v) \to \phi(u)$.

We write $E(\mathcal{G})$ and $E(\mathcal{H})$ for the undirected edge sets of $\mathcal{G}$ and $\mathcal{H}$, respectively. Thus, if $\{u, v\} \in E(\mathcal{G})$, the corresponding directed orientations $u \to v$ and $v \to u$ are mapped to directed edges between $\phi(u)$ and $\phi(v)$. Therefore,

$$\{u, v\} \in E(\mathcal{G}) \quad \Longrightarrow \quad \{\phi(u), \phi(v)\} \in E(\mathcal{H}).$$

Applying the same argument to the inverse functor $R : \mathsf{Cat}(\mathcal{H}) \longrightarrow \mathsf{Cat}(\mathcal{G})$ gives the converse implication:

$$\{\phi(u), \phi(v)\} \in E(\mathcal{H}) \quad \Longrightarrow \quad \{u, v\} \in E(\mathcal{G}).$$

Hence

$$\{u, v\} \in E(\mathcal{G}) \quad \Longleftrightarrow \quad \{\phi(u), \phi(v)\} \in E(\mathcal{H}).$$

Thus ϕ is a bijection on vertices that preserves adjacency in both directions. Therefore ϕ is a graph isomorphism, and hence $\mathcal{G} \cong \mathcal{H}$. We conclude that $\mathcal{G} \cong \mathcal{H} \quad \Longleftrightarrow \quad \mathsf{Cat}(\mathcal{G}) \cong \mathsf{Cat}(\mathcal{H})$. □

PROOF OF THEOREM 4.3

Proof. Let $e : u \to v$ be a directed edge of $\mathcal{G}$. Define $F_e := e \bullet - : v/\mathcal{G} \longrightarrow u/\mathcal{G}$. Recall that an object of the under category $v/\mathcal{G}$ is a morphism $f : v \to w$ in $\mathsf{Cat}(\mathcal{G})$. We define F_e on objects by

$$F_e(f) := e \bullet f.$$

Since $e : u \to v$ and $f : v \to w$ are composable, $e \bullet f : u \to w$ is a morphism in $\mathsf{Cat}(\mathcal{G})$, and therefore an object of $u/\mathcal{G}$. Hence F_e is well defined on objects.

Now let $h : f \to g$ be a morphism in $v/\mathcal{G}$, where $f : v \to w, \quad g : v \to w'$. By the definition of the under category, $h : w \to w'$ is a morphism in $\mathsf{Cat}(\mathcal{G})$ satisfying $f \bullet h = g$. We define

$$F_e(h) := h.$$

To verify that this is a morphism $F_e(f) \longrightarrow F_e(g)$ in $u/\mathcal{G}$, observe that, by associativity of $\bullet$,

$$F_e(f) \bullet h = (e \bullet f) \bullet h = e \bullet (f \bullet h) = e \bullet g = F_e(g).$$

Thus the required triangle in $u/\mathcal{G}$ commutes, so F_e is well defined on morphisms. It remains to verify the functoriality conditions. For an object $f : v \to w$ of $v/\mathcal{G}$, its identity morphism is the identity morphism

$$0 : w \to w$$

in $\mathsf{Cat}(\mathcal{G})$. Since F_e acts as the identity on the underlying morphisms between objects of the under categories,

$$F_e(\mathrm{id}_f) = \mathrm{id}_{F_e(f)}\,.$$

Hence F_e preserves identity morphisms. Finally, suppose

$$h : f \to g \qquad \text{and} \qquad k : g \to q$$

are composable morphisms in $v/\mathcal{G}$. Their composition is the corresponding composition $h \bullet k$ in $\mathsf{Cat}(\mathcal{G})$. Therefore,

$$F_e(h \bullet k) = h \bullet k = F_e(h) \bullet F_e(k).$$

Thus F_e preserves composition. Consequently,

$$e \bullet - : v/\mathcal{G} \longrightarrow u/\mathcal{G}$$

is a functor. □

PROOF OF THEOREM 4.4

Proof. Let

$$\widetilde{\mathcal{M}}_m = \bigoplus_{\alpha \in \mathsf{I}_m} \mathcal{M}_\alpha, \qquad \widetilde{S}_m = \bigcup_{\alpha \in \mathsf{I}_m} \nu_\alpha(S_\alpha),$$

and let

$$q : \widetilde{\mathcal{M}}_m \longrightarrow \widetilde{\mathcal{M}}_m/\sim$$

be the quotient map on directed-edge occurrences. We interpret the quotient on the left-hand side of Theorem 4.4 as

$$\left(\widetilde{\mathcal{M}}_m/\sim,\ q_*(\widetilde{S}_m)\right),$$

where q_* applies q edgewise to each selected path.

Write

$$\mathcal{G}^m = \underbrace{\mathcal{G} \bullet \cdots \bullet \mathcal{G}}_{m \text{ times}} = (\mathcal{M}, S),$$

and let $\mathcal{G}_i$ denote the i-th copy of $\mathcal{G}$ in this product. For a directed edge $e \in \mathsf{DE}(\mathcal{G})$, denote its copy in $\mathcal{G}_i$ by $e^{(i)}$.

We first characterize S. We claim that S consists precisely of the paths

$$e_1^{(i_1)}, e_2^{(i_2)}, \ldots, e_k^{(i_k)},$$

where

$$1 \le k \le m, \qquad 1 \le i_1 < \cdots < i_k \le m,$$

and $e_1, \ldots, e_k \in \mathsf{DE}(\mathcal{G})$ form a composable path in $\mathcal{G}$.

This follows by induction on m. For $m = 1$, the selected path set is exactly $\mathsf{DE}(\mathcal{G})$, so the claim holds. Suppose it holds for $m = r$. By the definition of $\bullet$, the selected paths of

$$\mathcal{G}^{r+1} = \mathcal{G}^r \bullet \mathcal{G}_{r+1}$$

consist of the selected paths already present in $\mathcal{G}^r$, the single-edge paths in $\mathcal{G}_{r+1}$, and the composable concatenations of a selected path from $\mathcal{G}^r$ with a directed edge from $\mathcal{G}_{r+1}$. These are exactly the composable edge sequences whose factor positions form a strictly increasing subset of $\{1, \ldots, r+1\}$. Thus the claim follows.

We next identify the underlying multigraphs. By definition,

$$a_{\alpha,i} \sim b_{\beta,j}$$

if and only if

$$i = j \qquad \text{and} \qquad a_i = b_j$$

as directed edges of $\mathcal{G}$. Hence an equivalence class of directed-edge occurrences is uniquely determined by a pair

$$(e, i), \qquad e \in \mathsf{DE}(\mathcal{G}), \quad 1 \le i \le m.$$

Every such pair occurs. Indeed, let $e : u \to v$ be a directed edge and fix $i \in \{1, \dots, m\}$. Define the composable m-tuple

$$\alpha = (a_1, \dots, a_m)$$

by

$$a_r = \begin{cases} 0 : u \to u, & r < i, \\ e : u \to v, & r = i, \\ 0 : v \to v, & r > i. \end{cases}$$

Then $\alpha \in \mathsf{I}_m$, and its i-th factor contributes the occurrence $a_{\alpha,i}$ of e.

Therefore we may define

$$\phi : \widetilde{\mathcal{M}}_m / \sim \longrightarrow \mathcal{M}$$

by

$$\phi([a_{\alpha,i}]) = a_i^{(i)}$$

on directed edges, while preserving the corresponding vertex labels. The definition of $\sim$ shows that ϕ is well-defined. Moreover, every directed edge of $\mathcal{M}$ is uniquely of the form $e^{(i)}$, so ϕ is bijective on directed edges. Since equivalent occurrences represent the same directed edge of $\mathcal{G}$, they have the same source and target. Thus ϕ preserves incidence and is an isomorphism of the underlying multigraphs.

It remains to compare the selected path sets. Let

$$q(a_{\alpha,i_1}), q(a_{\alpha,i_2}), \dots, q(a_{\alpha,i_k}) \in q_*(\widetilde{S}_m)$$

be the image of a selected path from some S_α. By the definition of S_α,

$$i_1 < \cdots < i_k,$$

and the directed edges

$$a_{i_1}, \dots, a_{i_k}$$

form a composable path in $\mathcal{G}$. Hence

$$a_{i_1}^{(i_1)}, a_{i_2}^{(i_2)}, \dots, a_{i_k}^{(i_k)}$$

belongs to S. Therefore the edge isomorphism ϕ sends $q_*(\widetilde{S}_m)$ into S.

The induced map on selected paths is injective. Indeed, if two paths in $q_*(\widetilde{S}_m)$ have the same image under ϕ, then their corresponding edges in every position are the same copies $e^{(i)}$. Hence the corresponding directed-edge occurrences have the same factor position and the same underlying directed edge of $\mathcal{G}$, and are therefore equivalent. Thus the two quotient paths are equal.

Finally, we prove surjectivity. Let

$$p = e_1^{(i_1)}, e_2^{(i_2)}, \dots, e_k^{(i_k)} \in S,$$

where

$$1 \le i_1 < \cdots < i_k \le m$$

and $e_1, \dots, e_k$ form a composable path in $\mathcal{G}$. Write

$$e_j : u_j \to u_{j+1}, \qquad 1 \le j \le k.$$

Then

$$t(e_j) = s(e_{j+1}) = u_{j+1} \qquad (1 \le j < k).$$

Define an m-tuple

$$\alpha = (a_1, \ldots, a_m)$$

by

$$a_r = \begin{cases} 0 : u_1 \to u_1, & r < i_1, \\ e_j, & r = i_j, \quad 1 \le j \le k, \\ 0 : u_{j+1} \to u_{j+1}, & i_j < r < i_{j+1}, \quad 1 \le j < k, \\ 0 : u_{k+1} \to u_{k+1}, & r > i_k. \end{cases}$$

Every pair of consecutive factors is composable, so $\alpha \in \mathsf{I}_m$. By the definition of S_α,

$$a_{\alpha,i_1}, a_{\alpha,i_2}, \ldots, a_{\alpha,i_k} \in S_\alpha.$$

Its image in the quotient is mapped by ϕ precisely to

$$e_1^{(i_1)}, e_2^{(i_2)}, \ldots, e_k^{(i_k)} = p.$$

Thus the induced map on selected paths is surjective.

Consequently, ϕ is an isomorphism of the underlying multigraphs and carries the quotient selected path set bijectively onto S. Therefore,

$$\left(\bigoplus_{\alpha \in \mathsf{I}_m} \mathcal{M}_\alpha, \bigcup_{\alpha \in \mathsf{I}_m} \nu_\alpha(S_\alpha) \right) / \sim \cong \underbrace{\mathcal{G} \bullet \cdots \bullet \mathcal{G}}_{m \text{ times}}.$$

□

PROOF OF THEOREM 4.6

Proof. We prove this statement by induction on m.

For $m = 1$, the claim is trivial since the number of paths from v_i to v_j in $(\bigoplus_{e \in \mathsf{DE}(\mathcal{G})} e, \mathsf{DE}(\mathcal{G}))$ corresponds exactly to the (i, j)-entry of $\boldsymbol{A}$, the adjacency matrix of $\mathcal{G}$.

Assume the statement is true for $m - 1$, i.e., the number of paths from v_i to v_j in $\mathcal{G}^{m-1} = (\mathcal{N}, T)$ equals the (i, j)-entry of $\boldsymbol{A} \circ \cdots \circ \boldsymbol{A}$ (the composition of $\boldsymbol{A}$ with itself ($(m-1)$-times). We will show the statement holds for m.

For m, let:

$$(\mathcal{M}, S) = (\mathcal{N}, T) \bullet \Big(\bigoplus_{e \in \mathsf{DE}(\mathcal{G})} e,\ \mathsf{DE}(\mathcal{G}) \Big) = \Big(\bigoplus_{e \in \mathsf{DE}(\mathcal{G})} e \oplus \mathcal{N},\ T \star \mathsf{DE}(\mathcal{G}) \Big).$$

A path in $S = T \star \mathsf{DE}(\mathcal{G})$ can be categorized as follows: 1. It is a path in T (i.e., a path entirely within $\mathcal{G}^{m-1}$), or 2. It is a path in $\mathsf{DE}(\mathcal{G})$, or 3. It is a composition of a subpath in T from v_i to some intermediate node v_l, followed by a subpath in $\mathsf{DE}(\mathcal{G})$ from v_l to v_j.

Using this decomposition, the total number of paths from v_i to v_j in S is given by: (Number of paths in T from v_i to v_j) + (Number of paths in $\mathsf{DE}(\mathcal{G})$ from v_i to v_j) + (Composition of subpaths from v_i to v_j).

By the inductive hypothesis, the number of paths in T equals the (i, j)-entry of $\boldsymbol{A} \circ \cdots \circ \boldsymbol{A}$ ($(m-1)$ times). Additionally, the number of paths in $\mathsf{DE}(\mathcal{G})$ is given by $\boldsymbol{A}$, and the composition of subpaths corresponds to the multiplication of $(m-1)$ copies of $\boldsymbol{A}$ with $\boldsymbol{A}$, i.e., $(\boldsymbol{A} \circ \cdots \circ \boldsymbol{A})\boldsymbol{A}$.

Therefore, the total number of paths from v_i to v_j in S equals the (i, j)-entry of:

$$\boldsymbol{A} + (\boldsymbol{A} \circ \cdots \circ \boldsymbol{A})_{(m-1)\text{-times}} + (\boldsymbol{A} \circ \cdots \circ \boldsymbol{A})_{(m-1)\text{-times}} \boldsymbol{A} = (\boldsymbol{A} \circ \cdots \circ \boldsymbol{A})_{(m)\text{-times}}.$$

Thus, the statement is true for m. By induction, the proof is complete. □

PROOF OF THEOREM 4.7

Proof. In Langari et al. (2026), it is proved that for $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_m \in \mathsf{Mat}_n(\mathbb{R})$ with $m \in \mathbb{N}$:

$$\begin{aligned}\boldsymbol{A}_1 \circ \boldsymbol{A}_2 \circ \cdots \circ \boldsymbol{A}_m = &\sum_{i=1}^{m} \boldsymbol{A}_i \\ &+ \sum_{\sigma \in O(m,2)} \boldsymbol{A}_{\sigma_1} \boldsymbol{A}_{\sigma_2} + \cdots \\ &+ \sum_{\sigma \in O(m,j)} \boldsymbol{A}_{\sigma_1} \cdots \boldsymbol{A}_{\sigma_j} + \cdots \\ &+ \boldsymbol{A}_1 \boldsymbol{A}_2 \cdots \boldsymbol{A}_m\end{aligned}$$

where $O(m,i)$ is the set of all strictly monotonically increasing sequences of i numbers of $\{1, \cdots, m\}$. By setting $\boldsymbol{A}_i = \boldsymbol{B}$ for all $1 \leq i \leq m$, we get

$$\boldsymbol{B} \circ \cdots \circ \boldsymbol{B}(\ m\textit{-times}\ ) = \sum_{i=1}^{m} \binom{m}{i} \boldsymbol{B}^i = (\boldsymbol{I} + \boldsymbol{B})^m - \boldsymbol{I}$$

□

PROOF OF THEOREM 5.3

Proof. Let

$$\boldsymbol{M} = \mathsf{DRA}(P, m) = \underbrace{\boldsymbol{W} \circ \cdots \circ \boldsymbol{W}}_{m \text{ times}} = (\boldsymbol{I} + \boldsymbol{W})^m - \boldsymbol{I}.$$

Since P is categorical, $|V| \geq 2$, and $P_j > 0$ for every j, we have $0 < P_j < 1$. Hence

$$\frac{P_j}{1 - P_j} > 0$$

for every j, and therefore

$$\boldsymbol{W}_{i,j} > 0 \quad \Longleftrightarrow \quad \boldsymbol{A}_{i,j} = 1.$$

Thus, $\boldsymbol{W}$ has the same support as the adjacency matrix $\boldsymbol{A}$. Because $\mathcal{G}$ is connected and undirected, $\boldsymbol{A}$ is irreducible, and therefore $\boldsymbol{W}$ is an irreducible nonnegative matrix.

Moreover, since $|V| \geq 2$ and $\mathcal{G}$ is connected, every node has at least one neighbor. Hence every row of $\boldsymbol{W}$ contains a positive entry. Using

$$\boldsymbol{M} = (\boldsymbol{I} + \boldsymbol{W})^m - \boldsymbol{I} = \sum_{k=1}^{m} \binom{m}{k} \boldsymbol{W}^k,$$

we see that $\boldsymbol{M}$ is nonnegative and contains the term $m\boldsymbol{W}$. Consequently,

$$\|\boldsymbol{M}_{i,:}\|_1 > 0$$

for every i and every $m \geq 1$, so the node-level point-of-view distributions are well defined.

Now set

$$\boldsymbol{K} = \boldsymbol{I} + \boldsymbol{W}.$$

Since $\boldsymbol{W}$ is irreducible, so is $\boldsymbol{K}$, and every diagonal entry of $\boldsymbol{K}$ is positive. Therefore $\boldsymbol{K}$ is primitive.

Let $r > 0$ and $\ell > 0$ be right and left Perron eigenvectors of $\boldsymbol{W}$, respectively:

$$\boldsymbol{W} r = \rho(\boldsymbol{W}) r, \qquad \ell \boldsymbol{W} = \rho(\boldsymbol{W}) \ell.$$

Then

$$\boldsymbol{K} r = (\boldsymbol{I} + \boldsymbol{W}) r = (1 + \rho(\boldsymbol{W})) r,$$

and similarly

$$\ell \boldsymbol{K} = (1 + \rho(\boldsymbol{W})) \ell.$$

Thus, r and ℓ are respectively a positive right eigenvector and a positive left eigenvector of $\boldsymbol{K}$ associated with the eigenvalue $1+\rho(\boldsymbol{W})$. Since $\boldsymbol{K}$ is irreducible and nonnegative, Perron–Frobenius implies

$$\rho(\boldsymbol{K}) = 1 + \rho(\boldsymbol{W}).$$

Hence r and ℓ are respectively right and left Perron eigenvectors of $\boldsymbol{K}$.

Moreover, every row of $\boldsymbol{W}$ contains a positive entry and $r > 0$. Hence

$$\boldsymbol{W}r > 0.$$

Since

$$\boldsymbol{W}r = \rho(\boldsymbol{W})r$$

and $r > 0$, it follows that

$$\rho(\boldsymbol{W}) > 0.$$

Consequently,

$$\rho(\boldsymbol{K}) = 1 + \rho(\boldsymbol{W}) > 1.$$

Because $\boldsymbol{K}$ is primitive, the Perron–Frobenius asymptotic formula gives

$$\frac{\boldsymbol{K}^m}{\rho(\boldsymbol{K})^m} \longrightarrow \frac{r\ell}{\ell r}.$$

Since

$$\boldsymbol{M} = \boldsymbol{K}^m - \boldsymbol{I}$$

and $\rho(\boldsymbol{K}) > 1$, we have

$$\frac{\boldsymbol{I}}{\rho(\boldsymbol{K})^m} \longrightarrow 0,$$

and therefore

$$\frac{\boldsymbol{M}}{\rho(\boldsymbol{K})^m} \longrightarrow \frac{r\ell}{\ell r}.$$

Fix $v_i \in V$. Taking the i-th row yields

$$\frac{\boldsymbol{M}_{i,:}}{\rho(\boldsymbol{K})^m} \longrightarrow \frac{r_i}{\ell r}\,\ell.$$

Since $r_i > 0$, $\ell > 0$, and hence $\ell r > 0$, continuity of the ℓ_1-norm gives

$$\frac{\|\boldsymbol{M}_{i,:}\|_1}{\rho(\boldsymbol{K})^m} \longrightarrow \frac{r_i}{\ell r}\,\ell\mathbf{1} > 0.$$

Dividing the two limits, we obtain

$$\frac{\boldsymbol{M}_{i,:}}{\|\boldsymbol{M}_{i,:}\|_1} \longrightarrow \frac{\ell}{\ell\mathbf{1}}.$$

By Definition 5.2,

$$\mathsf{pov}(v_i, P, m) = \frac{\boldsymbol{M}_{i,:}}{\|\boldsymbol{M}_{i,:}\|_1},$$

and hence

$$\lim_{m\to\infty} \mathsf{pov}(v_i, P, m) = \frac{\ell}{\ell\mathbf{1}}.$$

Finally,

$$\mathsf{pov}(\mathcal{G}, P, m) = \frac{1}{|V|}\sum_{i=1}^{|V|} \mathsf{pov}(v_i, P, m).$$

Since the sum is finite and every node-level point of view converges to the same distribution,

$$\lim_{m\to\infty} \mathsf{pov}(\mathcal{G}, P, m) = \frac{\ell}{\ell\mathbf{1}}.$$

□

PROOF OF THEOREM 5.4

We first establish the complete-graph recovery result for the constant-sequence case.

Theorem H.2 (Complete-graph recovery for a constant sequence)**.** *Let $\mathcal{G} = (V, E, X)$ be a complete attributed graph, and let $P = (P_1, \ldots, P_{|V|})$ be a categorical distribution over X satisfying $P_j < 1$ for every j. Then, for every node $v_i \in V$,*

$$\lim_{m\to\infty} \mathsf{pov}(v_i, P, m) = \lim_{m\to\infty} \mathsf{pov}(\mathcal{G}, P, m) = P.$$

We begin with the constant-sequence setting, which provides the basic complete-graph recovery property needed for the general sequential result.

Proof. Let

$$S = \{j \in \{1, \ldots, n\} : P_j > 0\}, \qquad Z = \{1, \ldots, n\} \setminus S.$$

Since P is categorical and $P_j < 1$ for every j, the support S contains at least two indices.

For each j, write

$$w_j = \frac{P_j}{1 - P_j}.$$

Since $\mathcal{G}$ is complete, the distribution-induced matrix satisfies

$$\boldsymbol{W}_{i,j} = \begin{cases} w_j, & i \neq j, \\ 0, & i = j. \end{cases}$$

Set

$$\boldsymbol{K} = \boldsymbol{I} + \boldsymbol{W}.$$

Then

$$\mathsf{DRA}(P, m) = \boldsymbol{K}^m - \boldsymbol{I}.$$

We first determine the dominant spectral structure of $\boldsymbol{K}$. After reordering the vertices so that the indices in S come first, $\boldsymbol{K}$ has the block form

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}_S & 0 \\ \boldsymbol{B} & \boldsymbol{I} \end{pmatrix},$$

where $\boldsymbol{K}_S = \boldsymbol{I} + \boldsymbol{W}_S$ is the principal submatrix indexed by S.

Let P_S denote the restriction of P to S. Since $\sum_{i\in S} P_i = 1$, for every $j \in S$,

$$\begin{aligned} (P_S \boldsymbol{W}_S)_j &= \sum_{\substack{i\in S \\ i\neq j}} P_i \frac{P_j}{1 - P_j} \\ &= \frac{P_j}{1 - P_j} \sum_{\substack{i\in S \\ i\neq j}} P_i \\ &= \frac{P_j}{1 - P_j}(1 - P_j) \\ &= P_j. \end{aligned}$$

Hence

$$P_S \boldsymbol{W}_S = P_S,$$

and therefore

$$P_S \boldsymbol{K}_S = 2P_S.$$

Every coordinate of P_S is strictly positive. Moreover, $\boldsymbol{K}_S$ is a positive matrix: its diagonal entries equal 1, and for $i \neq j$,

$$(\boldsymbol{K}_S)_{i,j} = \frac{P_j}{1 - P_j} > 0.$$

Thus, by the Perron–Frobenius theorem,

$$\rho(\boldsymbol{K}_S) = 2,$$

the eigenvalue 2 is algebraically simple, and every other eigenvalue λ of $\boldsymbol{K}_S$ satisfies

$$|\lambda| < 2.$$

Because $\boldsymbol{K}$ is block triangular,

$$\det(\lambda \boldsymbol{I} - \boldsymbol{K}) = \det(\lambda \boldsymbol{I} - \boldsymbol{K}_S)(\lambda - 1)^{|Z|}.$$

Therefore the eigenvalues of $\boldsymbol{K}$ consist of those of $\boldsymbol{K}_S$, together with 1 with algebraic multiplicity $|Z|$. Consequently, 2 is the unique eigenvalue of $\boldsymbol{K}$ having maximal modulus, and it is algebraically simple.

Furthermore,

$$P = (P_S, 0)$$

is a left eigenvector of $\boldsymbol{K}$, since

$$P\boldsymbol{K} = (P_S \boldsymbol{K}_S, 0) = 2P.$$

Let $r_S > 0$ be a right Perron eigenvector of $\boldsymbol{K}_S$, so that

$$\boldsymbol{K}_S r_S = 2r_S.$$

If $Z \neq \varnothing$, define

$$r_Z = \boldsymbol{B} r_S.$$

Every entry of $\boldsymbol{B}$ is strictly positive, because for $i \in Z$ and $j \in S$,

$$\boldsymbol{B}_{i,j} = \frac{P_j}{1 - P_j} > 0.$$

Hence $r_Z > 0$. Define

$$r = \begin{cases} r_S, & Z = \varnothing, \\ \begin{pmatrix} r_S \\ r_Z \end{pmatrix}, & Z \neq \varnothing. \end{cases}$$

Then $r > 0$, and when $Z \neq \varnothing$,

$$\boldsymbol{K} r = \begin{pmatrix} \boldsymbol{K}_S r_S \\ \boldsymbol{B} r_S + r_Z \end{pmatrix} = \begin{pmatrix} 2r_S \\ 2r_Z \end{pmatrix} = 2r;$$

the case $Z = \varnothing$ is immediate. Thus r is a right eigenvector of $\boldsymbol{K}$ associated with the eigenvalue 2.

We now determine the asymptotic behavior of $\boldsymbol{K}^m$. Since 2 is algebraically simple and every other eigenvalue λ of $\boldsymbol{K}$ satisfies $|\lambda| < 2$, the Jordan decomposition of $\boldsymbol{K}$ shows that every Jordan block associated with such a λ contributes

$$O\big(m^{d-1}|\lambda|^m\big)$$

to $\boldsymbol{K}^m$, where d is the size of the block. After division by 2^m, every such contribution tends to zero. Hence

$$\frac{\boldsymbol{K}^m}{2^m}$$

converges to the spectral projection associated with the simple eigenvalue 2. Since r and P are respectively right and left eigenvectors of $\boldsymbol{K}$ for the eigenvalue 2, and

$$Pr > 0$$

because P is a nonzero nonnegative vector and $r > 0$, this spectral projection is

$$\frac{rP}{Pr}.$$

Therefore

$$\frac{\boldsymbol{K}^m}{2^m} \longrightarrow \frac{rP}{Pr}.$$

Fix $i \in \{1, \dots, n\}$. Taking the i-th row yields

$$\frac{(\boldsymbol{K}^m)_{i,:}}{2^m} \longrightarrow \frac{r_i}{Pr} P.$$

Since $P\mathbf{1} = 1$,

$$\frac{(\boldsymbol{K}^m)_{i,:}\mathbf{1}}{2^m} \longrightarrow \frac{r_i}{Pr} > 0.$$

Now let

$$\boldsymbol{M} = \mathsf{DRA}(P, m) = \boldsymbol{K}^m - \boldsymbol{I}.$$

Since $\boldsymbol{W} \geq 0$,

$$\boldsymbol{M} = (\boldsymbol{I} + \boldsymbol{W})^m - \boldsymbol{I} = \sum_{k=1}^{m} \binom{m}{k} \boldsymbol{W}^k \geq m\boldsymbol{W}.$$

Every row of $\boldsymbol{W}$ contains a positive entry. Indeed, if $i \in S$, there exists $j \in S \setminus \{i\}$ because $|S| \geq 2$, and if $i \in Z$, any $j \in S$ satisfies $\boldsymbol{W}_{i,j} > 0$. Hence

$$\|\boldsymbol{M}_{i,:}\|_1 > 0$$

for every i and every $m \geq 1$, so the node-level POV distributions are well defined.

Since

$$\boldsymbol{M}_{i,:} = (\boldsymbol{K}^m)_{i,:} - e_i^\top,$$

we have

$$\frac{\boldsymbol{M}_{i,:}}{2^m} = \frac{(\boldsymbol{K}^m)_{i,:}}{2^m} - \frac{e_i^\top}{2^m} \longrightarrow \frac{r_i}{Pr} P.$$

Moreover, because $\boldsymbol{M} \geq 0$,

$$\|\boldsymbol{M}_{i,:}\|_1 = \boldsymbol{M}_{i,:}\mathbf{1} = (\boldsymbol{K}^m)_{i,:}\mathbf{1} - 1,$$

and therefore

$$\frac{\|\boldsymbol{M}_{i,:}\|_1}{2^m} \longrightarrow \frac{r_i}{Pr} > 0.$$

Dividing the two limits gives

$$\frac{\boldsymbol{M}_{i,:}}{\|\boldsymbol{M}_{i,:}\|_1} \longrightarrow P.$$

By Definition 5.2,

$$\lim_{m \to \infty} \mathsf{pov}(v_i, P, m) = P \qquad \text{for every } v_i \in V.$$

Finally,

$$\mathsf{pov}(\mathcal{G}, P, m) = \frac{1}{|V|} \sum_{i=1}^{|V|} \mathsf{pov}(v_i, P, m).$$

Since V is finite and every node-level point-of-view distribution converges to P,

$$\lim_{m \to \infty} \mathsf{pov}(\mathcal{G}, P, m) = P.$$

□

The preceding theorem gives convergence when the same distribution P is used at every aggregation level. To extend this result to a convergent sequence of distributions, we first strengthen the static convergence to uniform convergence over all initial categorical distributions.

The following lemma provides this uniform-attraction property for the normalized action of the limiting matrix.

Lemma H.3 (Uniform attraction on complete graphs)**.** *Under the assumptions of Theorem H.2, let* $n = |V|$*, let*

$$\boldsymbol{W} = \boldsymbol{W}(P), \qquad \boldsymbol{K} = \boldsymbol{I} + \boldsymbol{W},$$

and define

$$T(x) = \frac{x\boldsymbol{K}}{x\boldsymbol{K}\mathbf{1}}, \qquad x \in \Delta^{n-1},$$

where

$$\Delta^{n-1} = \{x \in \mathbb{R}^n_{\geq 0} : x\mathbf{1} = 1\}.$$

Then

$$\sup_{x \in \Delta^{n-1}} \|T^{[m]}(x) - P\|_1 \longrightarrow 0 \qquad \text{as } m \to \infty, \tag{33}$$

where $T^{[m]}$ *denotes the* m*-fold iterate of* T*.*

Proof. Since $\boldsymbol{W} \geq 0$, we have

$$\boldsymbol{K} = \boldsymbol{I} + \boldsymbol{W} \geq \boldsymbol{I}.$$

Thus, for every $x \in \Delta^{n-1}$,

$$x\boldsymbol{K} \geq 0$$

and

$$x\boldsymbol{K}\mathbf{1} \geq x\mathbf{1} = 1.$$

Hence T is well defined and maps Δ^{n-1} into itself. Moreover, since its denominator is bounded below by 1, T is continuous on Δ^{n-1}.

We first show that iterating T is equivalent to multiplying by $\boldsymbol{K}^m$ and normalizing once at the end. Specifically,

$$T^{[m]}(x) = \frac{x\boldsymbol{K}^m}{x\boldsymbol{K}^m\mathbf{1}} \qquad \text{for every } x \in \Delta^{n-1}. \tag{34}$$

Indeed, the claim is immediate for $m = 1$. If it holds for some $m \geq 1$, then

$$\begin{aligned} T^{[m+1]}(x) &= T\left(\frac{x\boldsymbol{K}^m}{x\boldsymbol{K}^m\mathbf{1}}\right) \\ &= \frac{\frac{x\boldsymbol{K}^m}{x\boldsymbol{K}^m\mathbf{1}}\boldsymbol{K}}{\frac{x\boldsymbol{K}^m}{x\boldsymbol{K}^m\mathbf{1}}\boldsymbol{K}\mathbf{1}} \\ &= \frac{x\boldsymbol{K}^{m+1}}{x\boldsymbol{K}^{m+1}\mathbf{1}}. \end{aligned}$$

Thus equation 34 follows by induction.

For each $i \in \{1, \ldots, n\}$, let $e_i \in \mathbb{R}^n$ denote the i-th standard basis vector, and define

$$q_i^{(m)} = \frac{e_i^\top \boldsymbol{K}^m}{e_i^\top \boldsymbol{K}^m\mathbf{1}}.$$

We first show that

$$q_i^{(m)} \longrightarrow P \qquad \text{for every } i.$$

Since P is categorical and $P_j < 1$ for every j, its support contains at least two indices. Hence, for every i,

$$\sum_{j \neq i} \frac{P_j}{1 - P_j} > 0.$$

Since n is finite, define

$$c = \min_{1 \leq i \leq n} \sum_{j \neq i} \frac{P_j}{1 - P_j} > 0.$$

Therefore

$$\boldsymbol{K}\mathbf{1} \geq (1 + c)\mathbf{1}.$$

Because $\boldsymbol{K} \geq 0$, induction gives

$$\boldsymbol{K}^m\mathbf{1} \geq (1 + c)^m\mathbf{1}.$$

Thus, for every i,

$$s_{i,m} := e_i^\top \boldsymbol{K}^m \mathbf{1} \longrightarrow \infty.$$

In the static setting,

$$\mathsf{DRA}(P, m) = \boldsymbol{K}^m - \boldsymbol{I}.$$

Moreover, since

$$\boldsymbol{K} = \boldsymbol{I} + \boldsymbol{W} \qquad \text{with} \qquad \boldsymbol{W} \geq 0,$$

we have

$$\boldsymbol{K}^m - \boldsymbol{I} = (\boldsymbol{I} + \boldsymbol{W})^m - \boldsymbol{I} = \sum_{\ell=1}^{m} \binom{m}{\ell} \boldsymbol{W}^\ell \geq 0.$$

By Theorem H.2,

$$\frac{e_i^\top (\boldsymbol{K}^m - \boldsymbol{I})}{e_i^\top (\boldsymbol{K}^m - \boldsymbol{I})\mathbf{1}} \longrightarrow P.$$

By definition of $s_{i,m}$,

$$e_i^\top (\boldsymbol{K}^m - \boldsymbol{I})\mathbf{1} = s_{i,m} - 1.$$

Moreover, since $\boldsymbol{K}^m - \boldsymbol{I} \geq 0$,

$$\left\| e_i^\top (\boldsymbol{K}^m - \boldsymbol{I}) \right\|_1 = e_i^\top (\boldsymbol{K}^m - \boldsymbol{I})\mathbf{1} = s_{i,m} - 1.$$

Consequently,

$$\begin{aligned} &\left\| \frac{e_i^\top \boldsymbol{K}^m}{s_{i,m}} - \frac{e_i^\top \boldsymbol{K}^m - e_i^\top}{s_{i,m} - 1} \right\|_1 \\ &\qquad \leq \left| \frac{1}{s_{i,m}} - \frac{1}{s_{i,m} - 1} \right| \|e_i^\top \boldsymbol{K}^m\|_1 + \frac{\|e_i^\top\|_1}{s_{i,m} - 1} \\ &\qquad = \frac{2}{s_{i,m} - 1} \longrightarrow 0. \end{aligned}$$

By Theorem H.2,

$$\frac{e_i^\top (\boldsymbol{K}^m - \boldsymbol{I})}{s_{i,m} - 1} \longrightarrow P.$$

Therefore, by the triangle inequality,

$$\begin{aligned} \|q_i^{(m)} - P\|_1 \leq &\left\| \frac{e_i^\top \boldsymbol{K}^m}{s_{i,m}} - \frac{e_i^\top (\boldsymbol{K}^m - \boldsymbol{I})}{s_{i,m} - 1} \right\|_1 \\ &+ \left\| \frac{e_i^\top (\boldsymbol{K}^m - \boldsymbol{I})}{s_{i,m} - 1} - P \right\|_1 \longrightarrow 0. \end{aligned}$$

Hence

$$q_i^{(m)} \longrightarrow P \qquad \text{for every } i.$$

Now let $x = (x_1, \ldots, x_n) \in \Delta^{n-1}$. By equation 34,

$$T^{[m]}(x) = \frac{\sum_{i=1}^n x_i e_i^\top \boldsymbol{K}^m}{\sum_{i=1}^n x_i e_i^\top \boldsymbol{K}^m \mathbf{1}}.$$

Hence

$$T^{[m]}(x) = \sum_{i=1}^{n} \alpha_i^{(m)}(x)\, q_i^{(m)},$$

where

$$\alpha_i^{(m)}(x) = \frac{x_i\, e_i^\top \boldsymbol{K}^m \mathbf{1}}{\sum_{j=1}^n x_j\, e_j^\top \boldsymbol{K}^m \mathbf{1}}.$$

These coefficients satisfy

$$\alpha_i^{(m)}(x) \geq 0, \qquad \sum_{i=1}^{n} \alpha_i^{(m)}(x) = 1.$$

Therefore

$$\|T^{[m]}(x) - P\|_1 \le \sum_{i=1}^{n} \alpha_i^{(m)}(x)\|q_i^{(m)} - P\|_1 \le \max_{1\le i\le n} \|q_i^{(m)} - P\|_1.$$

The right-hand side is independent of x, and it tends to zero because n is finite and $q_i^{(m)} \to P$ for every i. Thus

$$\sup_{x\in\Delta^{n-1}} \|T^{[m]}(x) - P\|_1 \longrightarrow 0.$$

□

We can now return to the varying sequence $\mathcal{P} = \{P^{(k)}\}_{k\ge1}$. Since $P^{(k)} \to P$, the corresponding normalized maps become uniformly close to the limiting map associated with P. Combining this asymptotic closeness with the uniform-attraction property above yields the general complete-graph recovery result.

Proof. For each $k \ge 1$, let

$$\boldsymbol{W}^{(k)} = \boldsymbol{W}(P^{(k)}), \qquad \boldsymbol{K}^{(k)} = \boldsymbol{I} + \boldsymbol{W}^{(k)},$$

and let

$$\boldsymbol{W} = \boldsymbol{W}(P), \qquad \boldsymbol{K} = \boldsymbol{I} + \boldsymbol{W}.$$

For the complete graph,

$$\boldsymbol{W}^{(k)}_{i,j} = \begin{cases} \dfrac{P_j^{(k)}}{1-P_j^{(k)}}, & i \ne j, \\ 0, & i = j, \end{cases} \qquad \boldsymbol{W}_{i,j} = \begin{cases} \dfrac{P_j}{1-P_j}, & i \ne j, \\ 0, & i = j. \end{cases}$$

Since $P_j^{(k)} \to P_j < 1$ and $t \mapsto t/(1-t)$ is continuous on $[0, 1)$,

$$\boldsymbol{K}^{(k)} \longrightarrow \boldsymbol{K}. \tag{35}$$

Since the matrices are finite-dimensional, this entrywise convergence also holds in any matrix norm.

Let

$$\Delta^{n-1} = \{x \in \mathbb{R}^n_{\ge0} : x\mathbf{1} = 1\},$$

and let $e_i \in \mathbb{R}^n$ denote the i-th standard basis vector. For $x \in \Delta^{n-1}$, define

$$T_k(x) = \frac{x\boldsymbol{K}^{(k)}}{x\boldsymbol{K}^{(k)}\mathbf{1}}, \qquad T(x) = \frac{x\boldsymbol{K}}{x\boldsymbol{K}\mathbf{1}}.$$

Since $\boldsymbol{K}^{(k)} \ge \boldsymbol{I}$ and $\boldsymbol{K} \ge \boldsymbol{I}$,

$$x\boldsymbol{K}^{(k)}\mathbf{1} \ge 1, \qquad x\boldsymbol{K}\mathbf{1} \ge 1.$$

Hence all these maps are well defined and map Δ^{n-1} into itself.

We first show that

$$\sup_{x\in\Delta^{n-1}} \|T_k(x) - T(x)\|_1 \longrightarrow 0. \tag{36}$$

Let

$$a_k = x\boldsymbol{K}^{(k)}, \qquad a = x\boldsymbol{K},$$

and

$$d_k = a_k\mathbf{1}, \qquad d = a\mathbf{1}.$$

Then $d_k, d \ge 1$, and

$$\|T_k(x) - T(x)\|_1 \le \frac{\|a_k - a\|_1}{d_k} + \|a\|_1 \left|\frac{1}{d_k} - \frac{1}{d}\right|.$$

Since $a \ge 0$,

$$\|a\|_1 = d,$$

and

$$|d - d_k| \le \|a - a_k\|_1.$$

Therefore

$$\|T_k(x) - T(x)\|_1 \le 2\|a_k - a\|_1.$$

Moreover,

$$a_k - a = x(\boldsymbol{K}^{(k)} - \boldsymbol{K}).$$

If $\|\cdot\|_\infty$ denotes the induced matrix ℓ_∞-norm, then

$$\|a_k - a\|_1 \le \|\boldsymbol{K}^{(k)} - \boldsymbol{K}\|_\infty.$$

Hence

$$\sup_{x\in\Delta^{n-1}} \|T_k(x) - T(x)\|_1 \le 2\|\boldsymbol{K}^{(k)} - \boldsymbol{K}\|_\infty \longrightarrow 0,$$

which proves equation 36.

We next define the iterates of the limiting map. Let

$$T^{[0]}(x) = x,$$

and, for $L \ge 1$,

$$T^{[L]}(x) = T\big(T^{[L-1]}(x)\big).$$

Thus $T^{[L]}$ is the L-fold iterate of T.

Similarly, for $\nu \ge 0$, define the consecutive block maps $\mathcal{T}_{\nu,L} : \Delta^{n-1} \to \Delta^{n-1}$ recursively by

$$\mathcal{T}_{\nu,0}(x) = x,$$

and

$$\mathcal{T}_{\nu,L}(x) = T_{\nu+L}\big(\mathcal{T}_{\nu,L-1}(x)\big), \qquad L \ge 1. \tag{37}$$

We claim that, for every fixed $L \ge 1$,

$$\sup_{x\in\Delta^{n-1}} \|\mathcal{T}_{\nu,L}(x) - T^{[L]}(x)\|_1 \longrightarrow 0 \qquad \text{as } \nu \to \infty. \tag{38}$$

We prove this by induction on L.

For $L = 1$,

$$\mathcal{T}_{\nu,1}(x) = T_{\nu+1}(x), \qquad T^{[1]}(x) = T(x),$$

so equation 38 follows directly from equation 36.

Suppose now that equation 38 holds for some fixed $L \ge 1$. From equation 37,

$$\mathcal{T}_{\nu,L+1}(x) = T_{\nu+L+1}\big(\mathcal{T}_{\nu,L}(x)\big),$$

while

$$T^{[L+1]}(x) = T\big(T^{[L]}(x)\big).$$

Hence

$$\begin{aligned}
&\|\mathcal{T}_{\nu,L+1}(x) - T^{[L+1]}(x)\|_1 \\
&\le \big\|T_{\nu+L+1}\big(\mathcal{T}_{\nu,L}(x)\big) - T\big(\mathcal{T}_{\nu,L}(x)\big)\big\|_1 \\
&\quad + \Big\|T\big(\mathcal{T}_{\nu,L}(x)\big) - T\big(T^{[L]}(x)\big)\Big\|_1 .
\end{aligned}$$

Since $\mathcal{T}_{\nu,L}(x) \in \Delta^{n-1}$, the first term tends uniformly to zero by equation 36. The map T is continuous on the compact simplex Δ^{n-1}, hence uniformly continuous. The induction hypothesis therefore implies that the second term also tends uniformly to zero. This proves equation 38 for $L + 1$, and hence for every fixed $L \ge 1$.

We now prove convergence of the sequential normalized product. Let $x_0 \in \Delta^{n-1}$, and define recursively

$$x_m = T_m(x_{m-1}), \qquad m \ge 1.$$

By induction, the intermediate normalization factors cancel, and

$$x_m = \frac{x_0 \boldsymbol{K}^{(1)} \boldsymbol{K}^{(2)} \cdots \boldsymbol{K}^{(m)}}{x_0 \boldsymbol{K}^{(1)} \boldsymbol{K}^{(2)} \cdots \boldsymbol{K}^{(m)} \mathbf{1}}. \tag{39}$$

Let $\varepsilon > 0$. By Lemma H.3, choose a fixed $L \geq 1$ such that

$$\sup_{x \in \Delta^{n-1}} \|T^{[L]}(x) - P\|_1 < \frac{\varepsilon}{2}.$$

For this fixed L, by equation 38, there exists N such that, for every $\nu \geq N$,

$$\sup_{x \in \Delta^{n-1}} \|\mathcal{T}_{\nu,L}(x) - T^{[L]}(x)\|_1 < \frac{\varepsilon}{2}.$$

Now let $m \geq N + L$ and set

$$\nu = m - L.$$

By the recursive definitions of x_m and $\mathcal{T}_{\nu,L}$,

$$x_m = \mathcal{T}_{\nu,L}(x_\nu).$$

Since $x_\nu \in \Delta^{n-1}$,

$$\begin{aligned} \|x_m - P\|_1 &= \|\mathcal{T}_{\nu,L}(x_\nu) - P\|_1 \\ &\leq \|\mathcal{T}_{\nu,L}(x_\nu) - T^{[L]}(x_\nu)\|_1 + \|T^{[L]}(x_\nu) - P\|_1 \\ &< \varepsilon. \end{aligned}$$

Therefore

$$x_m \longrightarrow P \qquad \text{for every } x_0 \in \Delta^{n-1}. \tag{40}$$

It remains to pass from the normalized product to the distributional realization. Define

$$\boldsymbol{R}_0 = \boldsymbol{I}, \qquad \boldsymbol{R}_m = \boldsymbol{K}^{(1)} \boldsymbol{K}^{(2)} \cdots \boldsymbol{K}^{(m)} \quad (m \geq 1).$$

Taking $x_0 = e_i^\top$ in equation 39 and using equation 40, we obtain

$$\frac{(\boldsymbol{R}_m)_{i,:}}{\|(\boldsymbol{R}_m)_{i,:}\|_1} \longrightarrow P. \tag{41}$$

We now show that every row sum of $\boldsymbol{R}_m$ diverges. Since P is categorical and $P_j < 1$ for every j, its support contains at least two indices. Thus

$$c = \min_{1 \leq i \leq n} \sum_{j \neq i} \frac{P_j}{1 - P_j} > 0.$$

Therefore

$$\boldsymbol{K}\mathbf{1} \geq (1 + c)\mathbf{1}.$$

By equation 35, there exists $N_0 \geq 1$ such that, for every $k \geq N_0$,

$$\boldsymbol{K}^{(k)}\mathbf{1} \geq \left(1 + \frac{c}{2}\right)\mathbf{1}.$$

Let

$$r_{i,m} = \|(\boldsymbol{R}_m)_{i,:}\|_1 = e_i^\top \boldsymbol{R}_m \mathbf{1}.$$

Since $\boldsymbol{W}^{(k)} \geq 0$, we have

$$\boldsymbol{K}^{(k)} = \boldsymbol{I} + \boldsymbol{W}^{(k)} \geq \boldsymbol{I}$$

for every k. Consequently,

$$\boldsymbol{R}_m \geq \boldsymbol{I}$$

for every $m \geq 0$, and hence

$$r_{i,m} = e_i^\top \boldsymbol{R}_m \mathbf{1} \geq 1.$$

In particular,

$$r_{i,N_0-1} \geq 1 > 0.$$

For $m \geq N_0$,

$$\begin{aligned} r_{i,m} &= e_i^\top \boldsymbol{R}_{m-1} \boldsymbol{K}^{(m)} \mathbf{1} \\ &\geq \left(1 + \frac{c}{2}\right) e_i^\top \boldsymbol{R}_{m-1} \mathbf{1} \\ &= \left(1 + \frac{c}{2}\right) r_{i,m-1}. \end{aligned}$$

Hence

$$r_{i,m} \geq \left(1 + \frac{c}{2}\right)^{m-N_0+1} r_{i,N_0-1},$$

and therefore

$$r_{i,m} \longrightarrow \infty. \tag{42}$$

By the ordered-product identity,

$$\begin{aligned} \boldsymbol{R}_m &= \prod_{k=1}^{m} (\boldsymbol{I} + \boldsymbol{W}^{(k)}) \\ &= \boldsymbol{I} + \sum_{\ell=1}^{m} \sum_{1 \leq k_1 < \cdots < k_\ell \leq m} \boldsymbol{W}^{(k_1)} \cdots \boldsymbol{W}^{(k_\ell)}. \end{aligned}$$

Every summand is nonnegative, so

$$\boldsymbol{R}_m - \boldsymbol{I} \geq 0.$$

For each $m \geq 1$, write

$$\boldsymbol{M} := \mathsf{DRA}(\mathcal{P}, m) = \boldsymbol{R}_m - \boldsymbol{I}.$$

The node-level point-of-view distribution is well defined for every $m \geq 1$. Indeed, $P^{(1)}$ is categorical and no coordinate equals 1, so its support contains at least two indices. Hence every row of $\boldsymbol{W}^{(1)}$ contains a positive entry. Since $\boldsymbol{W}^{(1)}$ is a nonnegative summand of $\boldsymbol{R}_m - \boldsymbol{I} = \boldsymbol{M}$, every row of $\boldsymbol{M}$ has positive 1-norm. In particular,

$$\|\boldsymbol{M}_{i,:}\|_1 > 0.$$

Since

$$\boldsymbol{M}_{i,:} = (\boldsymbol{R}_m)_{i,:} - e_i^\top,$$

and $\boldsymbol{M} \geq 0$, we have

$$\|\boldsymbol{M}_{i,:}\|_1 = r_{i,m} - 1.$$

Consequently,

$$\begin{aligned} &\left\| \frac{\boldsymbol{M}_{i,:}}{\|\boldsymbol{M}_{i,:}\|_1} - \frac{(\boldsymbol{R}_m)_{i,:}}{\|(\boldsymbol{R}_m)_{i,:}\|_1} \right\|_1 \\ &\leq \left| \frac{1}{r_{i,m} - 1} - \frac{1}{r_{i,m}} \right| \|(\boldsymbol{R}_m)_{i,:}\|_1 + \frac{\|e_i^\top\|_1}{r_{i,m} - 1} \\ &= \frac{2}{r_{i,m} - 1} \longrightarrow 0 \end{aligned}$$

by equation 42.

Combining this with equation 41, we obtain

$$\mathsf{pov}(v_i, \mathcal{P}, m) = \frac{\boldsymbol{M}_{i,:}}{\|\boldsymbol{M}_{i,:}\|_1} \longrightarrow P.$$

Finally,

$$\mathsf{pov}(\mathcal{G}, \mathcal{P}, m) = \frac{1}{|V|} \sum_{i=1}^{|V|} \mathsf{pov}(v_i, \mathcal{P}, m).$$

Since V is finite and every node-level point-of-view distribution converges to P,

$$\mathsf{pov}(\mathcal{G}, \mathcal{P}, m) \longrightarrow P.$$

□

PROOF OF THEOREM B.1

Proof. Reflexivity follows from the identity embedding $\mathrm{id}_{\mathcal{M}} : \mathcal{M} \hookrightarrow \mathcal{M}$. For transitivity, if

$$(\mathcal{M}, S) \leq (\mathcal{N}, T) \quad \text{and} \quad (\mathcal{N}, T) \leq (\mathcal{Q}, R),$$

with corresponding embeddings ι and κ, then $\kappa \circ \iota : \mathcal{M} \hookrightarrow \mathcal{Q}$ is an embedding and

$$(\kappa \circ \iota)(S) \subseteq \kappa(T) \subseteq R.$$

Hence $\leq$ is a preorder.

To verify compatibility with $\bullet$, suppose

$$(\mathcal{M}, S) \leq (\mathcal{N}, T)$$

via an embedding $\iota : \mathcal{M} \hookrightarrow \mathcal{N}$, and let $(\mathcal{Q}, R) \in \mathsf{SMult}(\mathcal{G})$. Then

$$\iota \oplus \mathrm{id}_{\mathcal{Q}} : \mathcal{M} \oplus \mathcal{Q} \hookrightarrow \mathcal{N} \oplus \mathcal{Q}$$

is an embedding, and the definition of $\star$ gives

$$(\iota \oplus \mathrm{id}_{\mathcal{Q}})(S \star R) \subseteq T \star R.$$

Therefore

$$(\mathcal{M}, S) \bullet (\mathcal{Q}, R) \leq (\mathcal{N}, T) \bullet (\mathcal{Q}, R).$$

The analogous argument with $\mathrm{id}_{\mathcal{Q}} \oplus \iota$ gives compatibility on the other side. □

PROOF OF THEOREM B.2

Proof. First, $\mathsf{R}(\mathcal{M}, S)$ is an element of $\mathsf{SMult}(\mathcal{G})$. Indeed, write the multigraph component of $\mathcal{M}$ as a disjoint union of connected acyclic directed subgraphs of $\mathcal{C}$. After removing the directed edges not belonging to $\mathcal{G}$, each remaining nonempty connected component is a connected acyclic directed subgraph of $\mathcal{G}$. Their disjoint union is $\hat{\mathcal{M}}$, and every path in $\hat{S}$ is a path in $\hat{\mathcal{M}}$.

Let $(\mathcal{M}, S), (\mathcal{N}, T) \in \mathsf{SMult}(\mathcal{C})$. Removing the forbidden directed edges commutes with disjoint union, so

$$\widehat{\mathcal{M} \oplus \mathcal{N}} = \hat{\mathcal{M}} \oplus \hat{\mathcal{N}}.$$

Moreover,

$$\widehat{S \star T} = \hat{S} \star \hat{T}.$$

Indeed, a path in $S \star T$ survives the restriction precisely when all of its directed edges belong to $\mathcal{G}$; for a composed path pq, this is equivalent to both p and q surviving.

Therefore,

$$\mathsf{R}\big((\mathcal{M}, S) \bullet (\mathcal{N}, T)\big) = \mathsf{R}(\mathcal{M}, S) \bullet \mathsf{R}(\mathcal{N}, T).$$

Finally, the monoidal identity contains no directed edges or paths, so it is fixed by R. Hence R is a monoidal homomorphism. □

PROOF OF THEOREM B.3

Proof. Let

$$(\mathcal{M}, S) \in \mathsf{SMult}(\mathcal{G}), \qquad (\mathcal{N}, T) \in \mathsf{SMult}(\mathcal{C}).$$

Suppose first that

$$\mathbf{In}(\mathcal{M}, S) \leq (\mathcal{N}, T).$$

Since R is monotone,

$$\mathsf{R}(\mathbf{In}(\mathcal{M}, S)) \leq \mathsf{R}(\mathcal{N}, T).$$

Because restriction removes no edge or path from an element already in $\mathsf{SMult}(\mathcal{G})$,

$$\mathsf{R}(\mathbf{In}(\mathcal{M}, S)) = (\mathcal{M}, S).$$

Hence

$$(\mathcal{M}, S) \leq \mathsf{R}(\mathcal{N}, T).$$

Conversely, suppose

$$(\mathcal{M}, S) \leq \mathsf{R}(\mathcal{N}, T) = (\hat{\mathcal{N}}, \hat{T}).$$

Then there exists an embedding

$$\iota : \mathcal{M} \hookrightarrow \hat{\mathcal{N}}$$

such that

$$\iota(S) \subseteq \hat{T}.$$

Since

$$\hat{\mathcal{N}} \hookrightarrow \mathcal{N} \qquad \text{and} \qquad \hat{T} \subseteq T,$$

the same embedding, followed by the inclusion, gives

$$\mathbf{In}(\mathcal{M}, S) \leq (\mathcal{N}, T).$$

Therefore

$$\mathbf{In}(\mathcal{M}, S) \leq (\mathcal{N}, T) \quad \Longleftrightarrow \quad (\mathcal{M}, S) \leq \mathsf{R}(\mathcal{N}, T),$$

so $\mathbf{In} \dashv \mathsf{R}$. □

PROOF OF COROLLARY B.4

Proof. By definition of the restriction map, $\mathsf{R}(\mathcal{C}) = \mathcal{G}$. Since R is a monoidal homomorphism, $\mathsf{R}(\mathcal{C}^m) = \mathsf{R}(\mathcal{C})^m = \mathcal{G}^m$. Now let $X \in \mathsf{SMult}(\mathcal{G})$. Applying Theorem B.3 with the element $\mathcal{C}^m \in \mathsf{SMult}(\mathcal{C})$ gives

$$\mathbf{In}(X) \leq \mathcal{C}^m \quad \Longleftrightarrow \quad X \leq \mathsf{R}(\mathcal{C}^m) = \mathcal{G}^m.$$

This proves the stated equivalence. Taking $X = \mathcal{G}^m$, reflexivity gives $\mathcal{G}^m \leq \mathcal{G}^m$, and hence $\mathbf{In}(\mathcal{G}^m) \leq \mathcal{C}^m$. Thus $\mathcal{G}^m$ belongs to the displayed set. Conversely, every element X of that set satisfies $X \leq \mathcal{G}^m$. Therefore $\mathcal{G}^m$ is a greatest element of that set with respect to the preorder. □

The second assertion of Corollary B.4 is precisely Theorem 5.5.