Journal of Information and Data Management, 2026, 17:1, doi: 10.5753/jidm.2026.5727
This work is licensed under a Creative Commons Attribution 4.0 International License.

Evaluating Heterogeneous Node Embedding Compositions Using Diversity Metrics

Silvio Fernando Angonese [**Universidade Federal do Rio Grande do Sul - UFRGS** | ***sfangonese@inf.ufrgs.br***]
Renata Galante [**Universidade Federal do Rio Grande do Sul - UFRGS** | ***galante@inf.ufrgs.br***]

Institute of Informatics, Federal University of Rio Grande do Sul - UFRGS, Ave. Bento Gonçalves, 9500, Agronomia, Porto Alegre, RS, 91509-900, Brazil.

Received: 29 March 2025 • **Published:** 13 March 2026

Abstract This paper evaluates the impact of different embedding composition strategies on classification performance, analyzing local node features, neighboring node features, and metapaths. We conduct a comprehensive experimental evaluation using an authorial Person Relationships heterogeneous graph, incorporating diversity metrics to assess dataset balance and structural complexity. This approach provides deeper insights into their influence on model effectiveness and extends prior research by comparing new results against an established baseline. The experimental findings reaffirm the effectiveness of embedding compositions, with Aggregated Features + Metapaths achieving a Micro-F1 score of 94.04%, demonstrating highly accurate results, validated by diversity metrics. This outcome highlights the importance of embedding compositions in heterogeneous graph representations, reinforcing its potential to improve predictive performance in real-world graph structures.

Keywords: Heterogeneous Graph, Embeddings, Heterogeneous Embeddings, Representation Learning, Composition Embeddings, Diversity Metrics

1 Introduction

Heterogeneous graphs are essential data structures for representing both simple and complex data-driven applications. They serve as a powerful tool for representation learning and as input datasets for various downstream applications. A fundamental research question is how to enhance the representational power of heterogeneous graphs and their components, improving performance in applications that rely on these graphs.

A promising approach to enhancing this representational power is node embeddings vector representations of graph nodes that encode

A promising approach to enhancing this representational power is embedded vector representations of graph nodes. Encoding their features and relationships, enabling Deep Learning and Machine Learning algorithms to process them effectively Wang *et al*. [2023]. In heterogeneous graphs, metapaths define sequences of relationships connecting different node types, incorporating complex semantic and structural information. The combination of node embeddings with metapaths has been shown to enhance semantic representation and improve performance in graph-based applications.

Several works Hamilton *et al*. [2017]; Ying *et al*. [2018]; Wang *et al*. [2023] have introduced embedding that leverage neighboring nodes and metapaths to increase representation expressiveness and enhance performance in recommendation tasks. In our previous work, AGHE Angonese and Galante [2024a], we proposed an approach that builds heterogeneous graphs with embeddings derived from various node features, including text, images, and subgraphs. Each node has both feature and metapath-based embeddings, which serve as the foundation for the novel concept of embedding compositions.

This paper builds upon Angonese and Galante [2024b], extending those findings by creating an authorial heterogeneous graph that explicitly captures a rich network of relationships among nodes. This graph provides a comprehensive platform for evaluating the effectiveness of embedding compositions. While our previous work introduced a mechanism for aggregating semantics by combining neighbor-aggregated features and metapaths, this paper introduces diversity metrics as a key contribution to assessing the impact of dataset quality on classification performance. Diversity metrics quantify different aspects of variation within datasets, offering insights into class distribution balance, structural heterogeneity, and node attribute diversity. The experimental results demonstrate that embedding compositions consistently outperform standard embedding approaches, validating the robustness and improved representation capabilities of our method. These findings are further supported by diversity metrics, which provide a deeper understanding of the factors influencing classification performance.

The remainder of this paper is organized as follows: Section 2 presents the background techniques. Section 3 reviews related work. Section 4 introduces the AGHE approach, which is the foundation for creating the authorial heterogeneous graph and composing node embeddings. Section 5 discusses diversity metrics, a key contribution of this paper. Section 6 details the experiments, the creation of the authorial graph, and the results. Section 7 concludes the paper and outlines directions for future research.

2 Conceptualization

This section aims to present the key concepts used in this paper, providing the technical foundation applied throughout the proposed methodology.

2.1 Heterogeneous Graphs

Many real-world datasets involve complex interactions between different types of entities and relationships. Such scenarios can be naturally modeled using heterogeneous graphs—structures composed of multiple types of nodes and edges that capture complex relationships in structured data Zhang *et al*. [2019]. A key challenge lies in representing their rich, non-Euclidean structure in a form suitable for Deep Learning models, as there is no trivial way to encode such high-dimensional relationships into fixed-size feature vectors. The central problem in Machine Learning and Deep Learning on graphs is finding a way to incorporate information about graph structure into the models. From this perspective, the challenge is that there is no straightforward way to encode this high-dimensional, non-Euclidean information about graph structure into a feature vector Hamilton *et al*. [2017].

2.2 Graph Node Embeddings

A core component in graph-based learning is the representation of nodes through embeddings. The objective is to learn a function $f(x) : \mathcal{V} \rightarrow \mathcal{R}^d$ that maps each node v into a low-dimensional Euclidean space, where $d \ll |\mathcal{V}|$ Wang *et al*. [2023]. Graph embedding transforms property graphs into a vector or a set of vector spaces, as shown in **Figure 1**. Embedding should capture the graph topology, node features, node-to-node relationships, and other relevant information about graphs, subgraphs, and nodes. The similarity of embedding between nodes indicates their similarity in the network, both nodes are close to each other, connected or not by an edge, potentially used for any prediction Hamilton *et al*. [2017]; Santana and Ribeiro [2023].

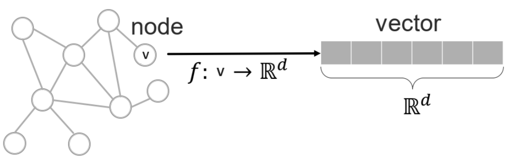

Figure 1. Embedding is saved into a dense vector space.

2.3 Metapaths

Metapaths are sequences of node types that define specific paths through heterogeneous graphs, capturing the semantics and structural correlations between different types of nodes Sun and Han [2012]. Metapaths enable the analysis of complex relationships and interactions within the graph by providing a structured way to traverse and connect different types of entities. This concept involves creating graph embeddings based on random walks that explore the heterogeneous neighborhood of a node, considering the type constraints imposed by the metapath. Skip-Gram and Node2Vec models are employed to maximize the probability of preserving both the structural and semantic properties of the graph, thus enabling the learning of desirable node representations. By leveraging metapaths, these models can capture rich contextual information and improve the accuracy and interpretability of the resulting embeddings, making them highly effective for tasks such as Link Prediction, Node Classification, and Clustering in heterogeneous graphs Dong *et al*. [2017].

Figure 2 shows an example of the application of the MetaPath2Vec architecture, which aims to generate paths that capture both the semantic and structural correlations between different types of nodes. This facilitates the transformation of heterogeneous network structures into Skip-Gram. Under the metapath scheme $CPPE$, the walker is biased towards car nodes C given its previous step on an organization node $C2$, i.e., following the semantics of this metapath. MetaPath2Vec encourages all types of nodes to appear in the context position.

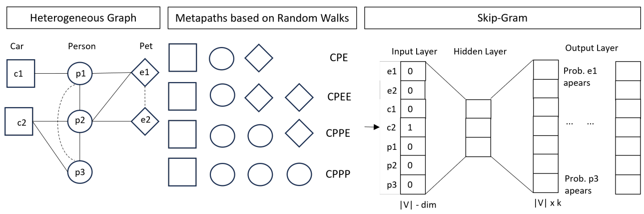

Figure 2. Example of MetaPath2Vec architecture use case.

2.4 Node Classification

A fundamental task in Recommender Systems aiming predicting the label y_v of a node, which may correspond to a type, category, or attribute, shown in **Figure 3**. The goal is to infer labels for all nodes $v \in \mathcal{V}$, given that ground-truth labels are available only for a subset of training nodes $\mathcal{V}_{\text{train}} \subseteq \mathcal{V}$. Thus, predictions $\mathcal{Z}$ for each of the nodes can be made by applying a shared function f to each of the latent vectors h, classifying nodes based on their features as $\mathcal{Z}_i = f(h_i)$ Hamilton *et al*. [2017]. Examples of Node Classification could include determining the genre of movies or type of nodes based on their features and relationships. Node Classification models can exploit node features and the concept that nodes with similar local neighborhood structures tend to have similar labels. Additionally, they leverage the heterophily concept, which suggests that nodes are more likely to connect to others with different labels and types. Another interesting approach for Node Classification is based on vector embeddings, which prove extremely useful as feature inputs. The basic idea is to use information about the neighborhood of the node in a vector embedding, which serves as the representation of nodes Hamilton *et al*. [2017].

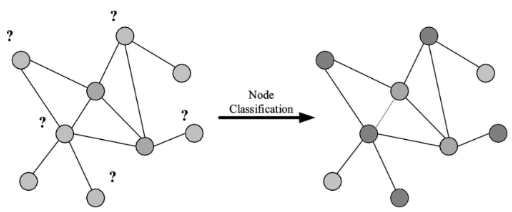

Figure 3. Graph node classification.

2.5 Diversity Metrics

A key aspect of understanding heterogeneous graphs involves assessing their balance and structural complexity. Such analysis provides valuable insights into the distributional properties of nodes and classes within the dataset. They quantify variations in class representation, node connectivity, and structural heterogeneity, which can directly impact the effectiveness of graph-based learning models. Key diversity metrics include class distribution entropy, which measures the uncertainty and balance of class distributions; class distribution Gini index, which quantifies class imbalance; degree heterogeneity, capturing variations in node connectivity patterns; and node type diversity, which evaluates the presence and variability of different node categories.

Beyond descriptive analysis, these metrics also serve as diagnostic tools for identifying biases and structural bottlenecks that may hinder model generalization. For example, high degree heterogeneity may indicate the existence of hub nodes that disproportionately influence predictions, while low node type diversity could limit the richness of semantic relationships in the graph. In the context of node classification, diversity metrics provide guidance for tailoring embedding strategies and designing more robust models that account for irregularities in graph topology or class imbalance. They also enable standardized comparisons across datasets and experimental setups, fostering reproducibility and interpretability in heterogeneous graph research. These metrics support a deeper understanding of graph properties and help refine embedding strategies, resulting in more accurate and fair models Kaminskas and Bridge [2016]; Wang *et al*. [2017]; Zhao *et al*. [2025].

3 Related Work

The heterogeneous graph can be traced back to generate data embedding from node features based on the random walks approach, citing Representation Learning on Graphs Hamilton *et al*. [2017]; Ying *et al*. [2018], improving the node expressivity. Closer to the aims of our proposal is Zhang *et al*. [2019], which defines a Heterogeneous Graph Neural Network with embedding processing. The survey Graph Neural Networks in Recommender Systems Wu *et al*. [2022] demonstrates the widespread use of GNNs in downstream applications, emphasizing their strength in graph representation learning. It specifically cites GraphSAGE Hamilton *et al*. [2017] as a key method for generating node embeddings from feature information.

The Metapath Aggregated Graph Neural Network (MAGNN) is an approach for heterogeneous graph embedding, considering the information present in heterogeneous graphs based on the Intra-Metapath aggregation extracting and combining information from metapath instances connecting nodes with their neighbors Fu *et al*. [2020].

Zhao *et al*. [2025] emphasizes that data diversity is a critical factor in ensuring the quality of recommender systems and machine learning models. They argue that evaluating dataset diversity should precede model development, as characteristics such as class concentration, structural redundancy, or low attribute variability can introduce biases in the outcomes. Metrics such as class entropy and the Gini index are recommended to quantify category distribution, while inter-item diversity and semantic variability assess the breadth of available attributes and contexts. These measures provide valuable diagnostics regarding data coverage and representativeness, and they guide the selection of more robust approaches for supervised or self-supervised learning. Thus, the careful use of diversity metrics helps mitigate overfitting, improves generalization, and supports more equitable decision-making in deep learning pipelines.

In previous work Angonese and Galante [2024b] proposed AGHE, an approach for constructing node embeddings by combining local features, features from neighboring nodes, and metapaths. Building upon this foundation, Angonese and Galante [2024a] advances embedding generation by integrating multiple data types, such as text, images, and subgraph structures embedded within nodes. Research on enhancing graph data quality through the processing of heterogeneous node features Angonese and Galante [2025] presents an algorithm for generating both individual and embedding compositions.

Our previous works collectively establish a chronological progression of research that aligns with the goals of this paper, reinforcing the importance of embedding compositions in heterogeneous graph representations. They also validate our approach of integrating local, neighbor, and metapath-based embeddings, providing a robust foundation for the contributions presented in this paper.

A key contribution of this paper is the introduction of diversity metrics as a fundamental evaluation criterion for assessing the results obtained in previous works. These metrics play a crucial role in providing a deeper understanding of embedding effectiveness and dataset quality, enabling a more comprehensive and insightful evaluation of heterogeneous graph representations. Hence, in this extended paper, both MAGNN and our previous research are used as baselines. MAGNN was chosen for its Intra-Metapath-based approach, which closely aligns our embedding compositions and enables direct comparisons.

4 AGHE - Generating Heterogeneous Embeddings

The *Approach for Generating Enhanced Heterogeneous Embeddings from Heterogeneous Graphs* (AGHE) Angonese and Galante [2024a] shown in **Figure 4**, generates heterogeneous embeddings through the processing of texts, images, and subgraphs represented in the nodes of heterogeneous graphs, such as:

1. *Graph Creation* generates the heterogeneous graph with nodes, edges, and features;
2. *Generating Text Node Embeddings* creates embeddings from node features or extracted images;
3. *Metapath and Aggregated Node Embeddings* uses random walks and MetaPath2Vec to generate feature embeddings;
4. *Graph Enhancement with Recommender Systems tasks* predicts node types, links, and clusters nodes;
5. *Rebuilding the Graph* incorporates the generated embeddings and predictions.

AGHE focuses on generating and composing embeddings but does not evaluate graph data quality. Integrating the diversity metrics proposed adds an analytical tool to assess dataset properties that impact classifier performance, ensuring embeddings are both expressive and optimized for heterogeneous graph characteristics.

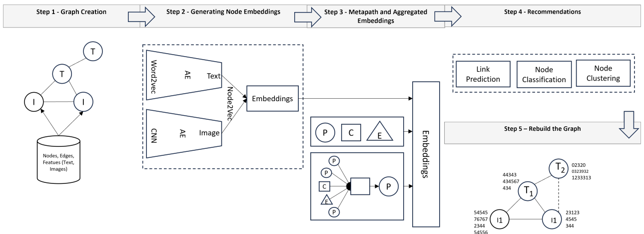

Figure 4. Steps of AGHE.

4.1 Composition of Heterogeneous Node Embeddings

The previous work Angonese and Galante [2024b] defines embedding compositions as the combination of node features and structural information from metapaths, potentially providing more expressive representations. Three compositions are considered: embeddings using only metapaths (`Metapaths`), node features with metapaths (`Features + Metapaths`), and aggregated neighboring features with metapaths (`Aggregated + Metapaths`), defined as follows:

4.1.1 Metapaths Embedding

Metapaths are a specific sequence of node types in heterogeneous graphs, capturing structural and semantic relationships. MetaPath2Vec leverages metapaths to generate embeddings by aggregating information from traversed nodes. Given a heterogeneous graph $\mathcal{HG} = (\mathcal{V}, \mathcal{E})$, a metapath $\mathcal{M}$ is a sequence:

$$\mathcal{M} = (V_1 \xrightarrow{E_1} V_2 \xrightarrow{E_2} \cdots \xrightarrow{E_{m-1}} V_m). \tag{1}$$

For a target node $v \in \mathcal{V}$, its metapath embedding $\mathbf{h}_v^M$ aggregates information from nodes reachable via $\mathcal{M}$:

$$\mathbf{h}_v^M = \text{Aggregate}\left(\{f(u) \mid u \in \text{Reachable}(v, \mathcal{M})\}\right), \tag{2}$$

where $\text{Reachable}(v, \mathcal{M})$ is the set of nodes reachable from v via $\mathcal{M}$, $f(u)$ extracts node embeddings, and Aggregate combines them into $\mathbf{h}_v^M$.

4.1.2 Features + Metapaths Embedding

This composition integrates local node features and metapath embeddings, capturing both intrinsic properties and relationships with neighbors. For a node $v \in \mathcal{V}$ with feature vector $\mathbf{x}_v$, its embedding composition $\mathbf{z}_v$ is defined as:

$$\mathbf{z}_v = \mathbf{x}_v \parallel \mathbf{h}_v^M, \tag{3}$$

where $\mathbf{h}_v^M$ Equation (2) represents the metapath embedding, and $\parallel$ denotes concatenation.

4.1.3 Aggregated + Metapaths Embedding

Composition of aggregated node features with metapath embeddings, capturing both local and neighbor semantics. For a node $v \in \mathcal{V}$, let $\mathbf{a}_v$ be the aggregated feature vector, incorporating both local and neighbor information:

$$\mathbf{a}_v = \text{Aggregate}\left(\{f(u) \mid u \in \text{Neighbors}(v) \cup \{v\}\}\right), \tag{4}$$

where $\text{Neighbors}(v)$ is the set of neighbor nodes, $f(u)$ extracts node features, and Aggregate combines them. The final embedding $\mathbf{z}_v$ is obtained by concatenating $\mathbf{a}_v$ with its metapath embedding $\mathbf{h}_v^M$:

$$\mathbf{z}_v = \mathbf{a}_v \parallel \mathbf{h}_v^M, \tag{5}$$

where $\mathbf{h}_v^M$ is defined in Equation (2) and $\parallel$ denotes concatenation.

5 Diversity Metrics

Diversity metrics are integrated into the evaluation to better understand how embedding compositions influence classification performance, as illustrated in **Figure 5**. These metrics assess how effectively node embeddings capture the structural complexity and class distribution of the data, helping to explain variations in results across datasets and embedding strategies. This approach provides a broader evaluation framework beyond traditional classification metrics and stands as a key contribution of this paper. The selected metrics offer valuable insights both when analyzing a single dataset to understand its internal distribution and when comparing multiple datasets to assess their essential characteristics.

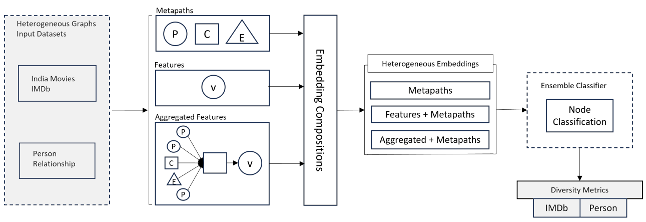

Figure 5. Embedding compositions within diversity metrics.

5.1 Class Distribution Entropy

The Class Distribution Entropy quantifies the uncertainty and diversity of class distributions in a dataset. It measures class balance, providing insight into potential imbalances that may affect classification performance. The entropy H is computed as:

$$H = -\sum_{i=1}^{C} p_i \log(p_i) \tag{6}$$

where p_i represents the proportion of nodes assigned to the i^{th} class and C is the total number of classes. The maximum possible entropy occurs when all classes are equally represented, which is:

$$H_{\max} = \log(C). \tag{7}$$

- *Low H.* Means an unbalanced distribution where certain classes dominate, potentially affecting the ability of models to generalize across different categories;
- *High H.* Approaching $\log(C)$, suggests that the class distribution is nearly uniform, indicating a more diverse and balanced set of embeddings.

The entropy score complements the quantitative analysis by summarizing the overall balance in class distributions. Rather than focusing on specific class proportions, it offers a single interpretable value that helps assess how equally the data is spread across the available classes. This aids in identifying scenarios where additional preprocessing or model adjustments may be required to ensure fair learning across categories.

5.2 Class Distribution Gini Index

The Class Distribution Gini Index measures class imbalance within a dataset. It quantifies how unequally classes are represented, aiding in the assessment of fairness and potential biases in model predictions. The Gini index G is computed as:

$$G = 1 - \sum_{i=1}^{C} p_i^2, \tag{8}$$

where p_i represents the proportion of nodes assigned to the ith class, and C is the total number of classes. The Gini index ranges from 0 to 1, where:

$$G_{\max} = 1 - \frac{1}{C}. \tag{9}$$

- *Low G.* A Gini value close to 0 indicates a balanced class distribution, where all classes are represented nearly equally, reducing bias and favoring more uniform classification outcomes;
- *High G.* A Gini value approaching $G_{\max}$ suggests a highly imbalanced dataset, where certain classes dominate significantly, potentially leading to classification challenges and biased predictions.

A higher Gini value, close to $G_{\max}$, indicates strong inequality in class distribution, where one or a few classes dominate the dataset. This can hinder the ability of a model to generalize. Conversely, a lower Gini value suggests a more balanced and fair representation of classes, improving the chances of unbiased and effective classification.

5.3 Degree Heterogeneity

The Degree Heterogeneity measures the structural variability of the graph. To assess it, we compute the heterogeneity of node degrees, defined as:

$$H_D = \frac{\sigma_D}{\mu_D + \epsilon}, \tag{10}$$

where μ_D is the mean degree, σ_D is the standard deviation of the degree distribution, and ϵ is a small constant to avoid division by zero. This metric quantifies the extent to which node connectivity varies across the graph.

- *Low $H_D \approx 0$.* The graph is homogeneous with most nodes having similar degrees;
- *Moderate H_D.* The graph has some level of degree variation, with a mix of highly connected and sparsely connected nodes;
- *High H_D.* The graph exhibits strong heterogeneity, with certain nodes acting as hubs (high-degree) while others have very few connections.

A highly heterogeneous graph can present challenges for node classification, as models may struggle to generalize across nodes with vastly different connectivity patterns. Low heterogeneity may indicate a simpler structure where learning patterns are easier.

5.4 Node Type Diversity

The Node Type Diversity measures the diversity of node types, we compute a node attribute entropy metric similar to class distribution entropy:

$$H_T = -\sum_{i=1}^{T} p_i \log(p_i), \tag{11}$$

where T represents the number of unique node types in the graph and p_i denotes the proportion of nodes belonging to the ith type.

- *High $H_T \approx \log(T)$.* The dataset has a well-balanced node type distribution;
- *Moderate $H_T < \log(T)$ but relatively high.* The dataset includes multiple node types, but some are more dominant;
- *Low $H_T \ll \log(T)$ (close to 0).* One or a few node types dominate the dataset, reducing diversity.

Higher node type diversity suggests a more complex and heterogeneous dataset, potentially requiring more expressive embedding models to capture meaningful relationships. Lower diversity implies that nodes tend to belong to a small number of dominant types, which may simplify classification but limit generalizability.

5.5 Summary of Diversity Metrics Adopted

Each of these metrics provides a unique perspective on dataset diversity:

- **Class Distribution Entropy** - measures the uncertainty and diversity in class distribution, indicating how evenly classes are represented;
- **Class Distribution Gini Index** - measures the imbalance in class distribution, indicating the presence of dominant classes;
- **Degree Heterogeneity** - captures structural complexity of graph connectivity;
- **Node Type Diversity** - assesses variability in node semantic attributes.

Together, these measures help characterize the underlying properties of the IMDb - India Movies and Person Relationships heterogeneous graph datasets, highlighting the factors that may influence node classification performance.

6 Experiments

We present the experiments conducted to investigate the effectiveness of the embedding compositions based on features and metapaths, understanding how well embedding compositions enhance node representation. The experiments incorporate diversity metrics as a fundamental tool to evaluate the results obtained for each dataset, allowing an analytical assessment of how dataset characteristics influence the outcomes.

6.1 Methodology

Figure 6 shows the main pipeline as part of the methodology used to execute the entire set of experiments. Essentially, it involves the generation of features and embeddings from nodes, and the validation of classification performance metrics based on each type of node embedding and computing diversity metrics for each dataset used in the experiments.

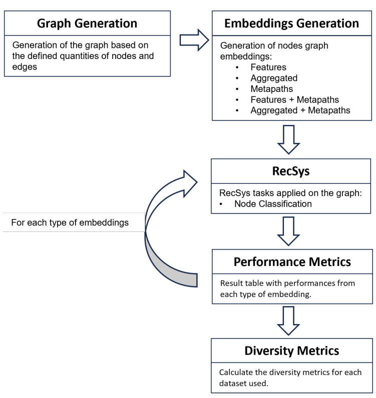

Figure 6. Steps of experiments methodology.

To evaluate classifier performance effectively, we used the Macro-F1 and Micro-F1 scores, ensuring a comprehensive assessment and compatibility with the baseline. The Macro-F1 score treats all classes equally, while the Micro-F1 score provides an overall performance measure Harbecke *et al.* [2022]. To enhance robustness, we employed an Ensemble learning approach, combining multiple classifiers to capture diverse data patterns and improve predictive accuracy. After calibration tests, the best results were obtained using Random Forest, Extreme Gradient Boosting, Support Vector Machine, Decision Tree, Bootstrap Aggregating, Adaptive Boosting, Gradient Boosting, and Logistic Regression. The evaluation used five iterations of 10-fold cross-validation with stratified validation and shuffled data.

6.2 Creation of the Person Relationships Heterogeneous Graph

This authorial Person Relationships heterogeneous graph dataset is available to the research community for further exploration and development. **Figure 7** illustrates the proposed graph, which integrates features of different data types, representing the model used in the experiment. The graph includes various node types, such as Person, Car, and Pet, each containing distinct attributes in formats such as plain text, images, or subgraphs.

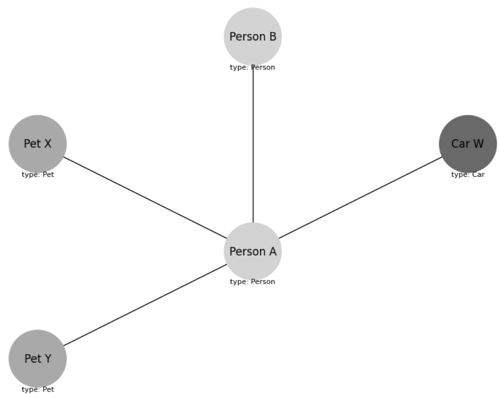

Figure 7. Graph schema data model.

The graph created for the experiments is shown in **Figure 8**, which has 2,356 nodes, 2,991 edges, and 13 defined metapaths. Type node is the field used as the multiclass of predictions, which is slightly unbalanced, where Person is at 14.8%, Car is at 43.4%, and Pet at 41.9%, representing the real-world scenario.

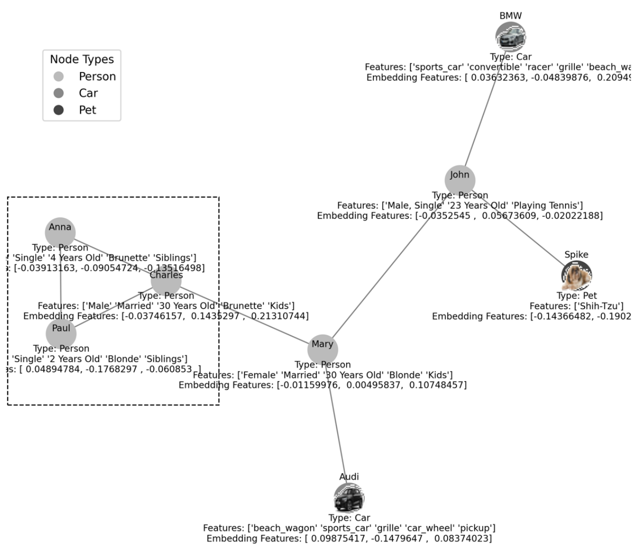

Figure 8. Person Relationships heterogeneous graph dataset.

Metapaths were defined during the graph creation process to support the embedding compositions, as formulated in Equation 1, and are defined as follows: $\mathcal{M} \leftarrow \{(Person \rightarrow Car \rightarrow Person), (Person \rightarrow Pet \rightarrow Person), (Car \rightarrow Person \rightarrow Person), (Person \rightarrow Person \rightarrow Car), (Person \rightarrow Person \rightarrow Pet), (Person \rightarrow Person \rightarrow Person), (Person \rightarrow Person \rightarrow Person \rightarrow Car), (Person \rightarrow Person \rightarrow Person \rightarrow Pet), (Person \rightarrow Pet \rightarrow Person), (Person \rightarrow Car \rightarrow Person), (Person \rightarrow Person \rightarrow Person \rightarrow Person \rightarrow Car), (Person \rightarrow Person \rightarrow Person \rightarrow Person \rightarrow Pet)\}$.

Figure 9 shows the defined metapaths and the count of nodes visited for each, indicating a relatively balanced distribution of nodes reached by them.

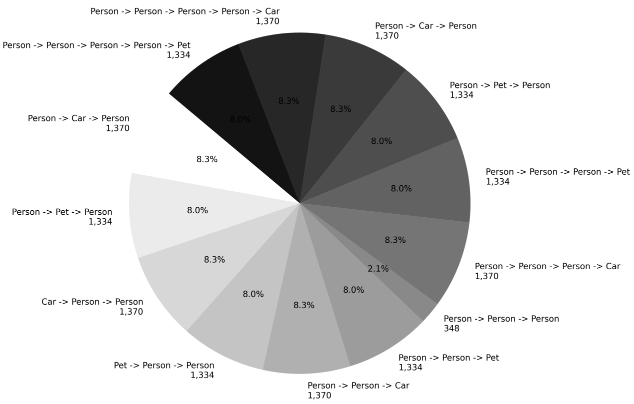

Figure 9. Distribution of graph nodes visited per metapath.

6.3 Baseline and Experiment I

To establish a reference for evaluating the effectiveness of embedding compositions, we consider the baseline model MAGNN proposed by Fu *et al.* [2020], whose results are presented in **Table 1**. This model uses Intra-Metapath embeddings and applies an external SVM classifier, achieving a Micro-F1 score of 61.53%, which serves as a benchmark for comparison.

Table 1. Baseline result from MAGNN and SVM.

IMDb Embeddings	Classifier	Mac-F1 / Mic-F1
Intra-Metapaths	MAGNN/SVM	61.44 / **61.53**

Experiment I was designed to evaluate the impact of embedding compositions proposed in this thesis. These embeddings combine structural (aggregated) and semantic (metapath) information to enrich node representations Angonese and Galante [2024b]. To ensure compatibility with the MAGNN baseline, which also employs an external SVM classifier, we adopted the same classifier in this experiment. As shown in **Table 2**, all compositions outperformed the MAGNN baseline, with the Aggregated + Metapaths composition achieving the best result, reaching 65.89% Micro-F1. This reinforces the effectiveness of integrating multiple feature perspectives into a unified embedding.

Table 2. Experiment I results using proposed embedding compositions.

IMDb Embeddings	Classifier	Mac-F1 / Mic-F1
Metapaths	SVM	60.16 / 64.19
Features + Metapaths	SVM	59.23 / 65.68
Aggregated + Metapaths	SVM	60.77 / **65.89**

6.4 Experiment II

Experiment II evaluates the impact of embedding compositions on the Node Classification task using the Person Relationships heterogeneous graph dataset. **Table 3** presents the performance metrics obtained by the Ensemble classifier, where single embeddings (Features and Aggregated) yielded the lowest classification performance, with Micro-F1 scores of 53.82% and 47.19%, respectively. This suggests that relying solely on local or aggregated node features fails to capture the complex structural and semantic relationships from the dataset. In contrast, embedding compositions that incorporate metapaths led to a significant improvement in classification performance. The Metapaths embedding alone achieved a Micro-F1 score of 93.07%, demonstrating its ability to capture meaningful structural patterns. Further improvements were observed with Features + Metapaths of 93.28%, while the best-performing approach was Aggregated + Metapaths, reaching 94.04%. These findings highlight the advantage of integrating multiple information sources, as metapaths combined with node feature aggregation lead to better classification results. These findings further validate the robustness of embedding compositions across different graph structures. The Person Relationships heterogeneous graph dataset demonstrated significantly higher classification performance, suggesting that heterogeneous node embedding compositions can better capture structural and semantic relationships.

Table 3. Experiment II results.

Person Graph Embeddings	Classifier	Mac-F1 / Mic-F1
Features	Ensemble	42.62 / 53.82
Aggregated	Ensemble	36.09 / 47.19
Metapaths	Ensemble	93.25 / 93.07
Features + Metapaths	Ensemble	93.48 / 93.28
Aggregated + Metapaths	Ensemble	94.00 / **94.04**

The results achieved through embedding compositions are encouraging, as shown in **Table 4**, demonstrating their effectiveness, particularly because models using Aggregated + Metapaths outperform the baseline used in this paper. **Table 4** presents only the best-performing Micro-F1 score metric values, highlighting a near-linear distribution of the data around the means, concentrating the standard deviations between ±1% and ±2%. Consequently, it demonstrates low variability, thus indicating that the model is reliable and robust to variations in the training data. Additionally, the experiments can be interpreted from an ablation study perspective, aiming to identify which embedding compositions have the greatest impact on model performance.

Table 4. The best Micro-F1 score achieved in Experiment II.

Datasets Embeddings	Graph Movies Mic-F1	Graph Person Mic-F1
Metapaths	64.19	93.07
Features + Metapaths	65.68	93.28
Aggregated + Metapaths	65.89	94.04

6.5 Experiment III

Experiment III presents a comparative analysis of diversity metrics across the two heterogeneous graph datasets used in Experiments I and II. Unlike the previous experiments, the goal here is not to measure the effectiveness of classifiers applied to embeddings, but rather to assess diversity metrics that characterize the datasets. This evaluation aims to identify potential improvements in the data engineering phase by analyzing class distribution balance, structural complexity, and the richness of node features. **Table 5** shows the best-performing configurations for each dataset, highlighting differences in metrics that provide insights into how graph properties may influence model outcomes. Diversity metrics help explain why performance varies across datasets, even when the same embedding compositions and classifiers are applied.

Overall, the analysis of diversity and performance metrics across both datasets shows the impact of graph structural properties on classification effectiveness. While the IMDb graph dataset presents class imbalance and structural complexity, leading to lower classification performance in terms of F1 scores, the Person Relationships graph dataset benefits from a more balanced class distribution and higher heterogeneity, resulting in superior classification outcomes. These findings highlight the importance of considering dataset-specific diversity characteristics when designing and evaluating heterogeneous graph embeddings. It is also crucial to note that different models may exhibit varying sensitivity to class distribution and graph structure.

7 Conclusion

This paper demonstrated the relevance of embedding compositions for heterogeneous graph representation, highlighting significant improvements in node classification. The combination of local features, neighboring node information, and metapaths led to more expressive representations and better predictive performance. A key contribution is the incorporation of diversity metrics to evaluate embedding techniques, offering insight into how dataset properties affect classification. The analysis showed that higher heterogeneity and well-structured class distributions lead to more reliable outcomes. Diversity metrics enable a more robust assessment of embedding compositions, complementing traditional metrics and helping researchers identify biases, refine graphs, and optimize embeddings. Another important contribution is the publicly available Person Relationships heterogeneous graph dataset, which supports the research community and future work on heterogeneous graph.

The contributions of this paper establish a foundation for future work, including evaluating embedding compositions in other graph tasks such as link prediction and clustering, integrating attention mechanisms, automating metapath selection, and applying diversity metrics as regularizers during training. These directions aim to further enhance the adaptability and effectiveness of heterogeneous embedding strategies.

Authors' Contributions

SFA and RG equally contributed to the design, execution, and writing of this manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The datasets generated and used during the current study are available at `https://github.com/SilvioAngonese/ufrgs/blob/master/Person_Relationships_Heterogeneous_Graph.zip`.

References

Angonese, S. and Galante, R. (2024a). AGHE - Approach for Generating Enhanced Heterogeneous Embeddings from Heterogeneous Graphs. In *Anais do LI Seminário Integrado de Software e Hardware*, pages 252–263, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/semish.2024.3051.

Angonese, S. and Galante, R. (2024b). Composition of heterogeneous node embeddings - unlocking the power of heterogeneous graph representation. In *Anais-XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 626–638, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbbd.2024.243436.

Angonese, S. F. and Galante, R. (2025). Enhancing Graph Data Quality by Leveraging Heterogeneous Node Features and Embeddings. In Paes, A. and Verri, F. A. N., editors, *Intelligent Systems*, pages 383–398, Cham. Springer Nature Switzerland.

Dong, Y., Chawla, N. V., and Swami, A. (2017). MetaPath2Vec: Scalable Representation Learning for Heterogeneous Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 135–144, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3097983.3098036.

Fu, X., Zhang, J., Meng, Z., and King, I. (2020). MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding. In *Proceedings of The Web Conference 2020*, WWW '20, page 2331–2341, New York, NY, USA. DOI: 10.1145/3366423.3380297.

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 1025–1035, Red Hook, NY, USA.

Harbecke, D., Chen, Y., Hennig, L., and Alt, C. (2022). Why only micro-f1? class weighting of measures for relation classification.

Table 5. Experiment III results.

Metric	IMDb Value	Interpretation	Person Value	Interpretation
Class Distribution Entropy	1.0548	The entropy value is close to the theoretical maximum ($\log(3) \approx 1.0986$), indicating a relatively high-class diversity, but some degree of imbalance is still present.	1.0119	The entropy value is relatively high, indicating that class distribution is fairly balanced, minimizing the impact of dominant classes.
Class Distribution Gini Index	0.6381	The Gini index suggests moderate class imbalance, indicating that some classes are more dominant.	0.6162	The Gini index suggests a well-balanced class distribution, with little dominance from any single class.
Degree Heterogeneity	1.5963	The degree heterogeneity value indicates a moderately complex graph structure with a mix of highly connected nodes and sparsely connected ones.	2.0499	The high degree heterogeneity value suggests a structurally complex graph with significant variation in node connectivity, possibly including important hubs.
Node Type Diversity	1.0392	The node type diversity is moderately high, meaning that different node types are present, potentially contributing to classification performance.	1.0119	The node type diversity is high, suggesting a broad range of node types, which may enhance classification performance by providing rich structural information.

Kaminskas, M. and Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems*, 7(1):2:1–2:42. DOI: 10.1145/2926720.

Santana, D. and Ribeiro, L. (2023). Approximate similarity joins over dense vector embeddings. In *Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados*, pages 51–62, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbbd.2023.232512.

Sun, Y. and Han, J. (2012). *Mining Heterogeneous Information Networks: Principles and Methodologies*. Morgan and Claypool Publishers.

Wang, H., Hu, W., Qiu, Z., and Du, B. (2017). Nodes' evolution diversity and link prediction in social networks. *IEEE Trans. on Knowl. and Data Eng.*, 29(10):2263–2274. DOI: 10.1109/TKDE.2017.2728527.

Wang, X., Bo, D., Shi, C., Fan, S., Ye, Y., and Yu, P. S. (2023). A Survey on Heterogeneous Graph Embedding: Methods, Techniques, Applications and Sources. *IEEE Transactions on Big Data*, 9(2):415–436. DOI: 10.1109/TBDATA.2022.3177455.

Wu, S., Sun, F., Zhang, W., Xie, X., and Cui, B. (2022). Graph Neural Networks in Recommender Systems: A Survey. *ACM Comput. Surv.*, 55(5). DOI: 10.1145/3535101.

Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L., and Leskovec, J. (2018). Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '18, page 974–983, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3219819.3219890.

Zhang, C., Song, D., Huang, C., Swami, A., and Chawla, N. V. (2019). Heterogeneous Graph Neural Network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '19, page 793–803, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3292500.3330961.

Zhao, Y., Wang, Y., Liu, Y., Cheng, X., Aggarwal, C. C., and Derr, T. (2025). Fairness and diversity in recommender systems: A survey. *ACM Trans. Intell. Syst. Technol.*, 16(1). DOI: 10.1145/3664928.