C-MAG: Cascade Multimodal Attributed Graphs for Supply Chain Link Prediction

Yunqing Li*
AAITC, Lenovo
Morrisville, North Carolina, USA
connie555514@gmail.com

Zixiang Tang*
AAITC, Lenovo
Morrisville, North Carolina, USA
zt2292@columbia.edu

Jiaying Zhuang
AAITC, Lenovo
Morrisville, North Carolina, USA
jzhuang@lenovo.com

Zhenyu Yang
AAITC, Lenovo
Morrisville, North Carolina, USA
zyang1@lenovo.com

Farhad Ameri
School of Manufacturing Systems and
Networks, Arizona State University
Mesa, Arizona, USA
amerif@gmail.com

Jianbang Zhang
AAITC, Lenovo
Morrisville, North Carolina, USA
zhangjb2@lenovo.com

Abstract

Connecting an ever-expanding catalogue of products with suitable manufacturers and suppliers is critical for resilient, efficient global supply chains, yet traditional methods struggle to capture complex capabilities, certifications, geographic constraints, and rich multimodal data of real-world manufacturer profiles. To address these gaps, we introduce *PMGraph*, a public benchmark of bipartite and heterogeneous multimodal supply-chain graphs linking 8,888 manufacturers, over 70k products, more than 110k manufacturer–product edges, and over 29k product images. Building on this benchmark, we propose the Cascade Multimodal Attributed Graph (*C-MAG*), a two-stage architecture that first aligns and aggregates textual and visual attributes into intermediate group embeddings, then propagates them through a manufacturer–product heterograph via multiscale message passing to enhance link prediction accuracy. *C-MAG* also provides practical guidelines for modality-aware fusion, preserving predictive performance in noisy, real-world settings.

Keywords

Graph Neural Network, Link Prediction, Mutlimodal Knowledge Graph, Graph Embedding, Manufacturer Relationship

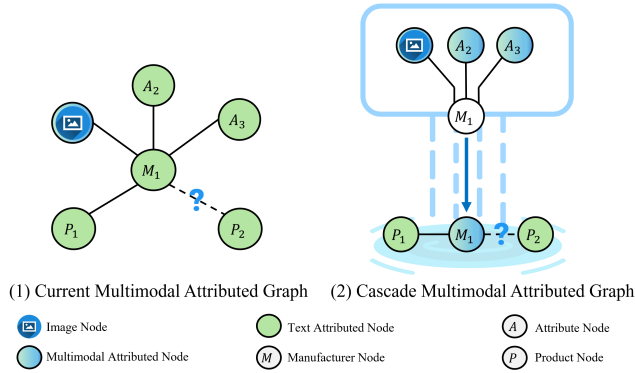*Both authors contributed equally to this research.

1 Introduction

Global supply chain disruptions[32] caused by pandemics, geopolitical tensions, and material shortages have underscored the importance of leveraging small and medium-sized enterprises (SMEs) and their potential capacities and capabilities more effectively. Beyond simply matching existing production capabilities, accurately predicting SMEs' potential to manufacture new products is essential for developing resilient and adaptable supply networks[25, 39].

Three primary challenges complicate this objective. First, the absence of scalable, structured supply chain datasets limits the ability to model and analyze manufacturer-product relationships effectively. In most cases, publicly available data on manufacturing capabilities are sparse, heterogeneous, proprietary, and locked in unstructured formats [14, 15]. As a result, large-scale computational analysis becomes difficult. Without reliable data infrastructure, it is challenging to train AI models that can accurately match production requirements with suitable SME capabilities, let alone anticipate their ability to scale or diversify into new manufacturing domains.

The second major challenge lies in the complexity and rapidly evolving nature of manufacturing technologies. New processes, equipment, and materials are continuously introduced, making it difficult to maintain a current understanding of manufacturers' capabilities [16, 40]. Often, multiple methods exist for producing the same product, each requiring different technical resources and expertise. Effectively linking product requirements to manufacturers thus demands deep knowledge of process selection, material properties, and production technologies. [44] This complexity presents a significant barrier to automating the identification of capable manufacturers, particularly among SMEs whose capabilities are rarely documented in structured, accessible formats.

Third, manufacturers in real-world supply chain knowledge graphs (SC-KG) exhibit extensive heterogeneity—spanning geographic location, certifications, production capacity, and technical capabilities—which complicates the task of constructing unified representations and predicting both present and future product relationships [2, 3, 5]. At the same time, fusing multi-modal data introduces its own set of challenges—misaligned embedding spaces, varying noise characteristics, incomplete modality coverage, and the risk of diluting key signals—demanding sophisticated alignment, weighting, and noise-robust fusion strategies to maintain predictive performance [28].

(1) Current Multimodal Attributed Graph   (2) Cascade Multimodal Attributed Graph

Image Node	Text Attributed Node	A Attribute Node
Multimodal Attributed Node	M Manufacturer Node	P Product Node

Figure 1: Overall architecture of the Cascade Multimodal Attributed Graph.

To bridge these gaps, we propose three main contributions:

- We release a suite of public-use supply chain knowledge graphs *PMGraph*[1] —including bipartite manufacturer–product graphs and heterogeneous multimodal knowledge graphs, and related variants—linking manufacturer metadata (capabilities, certifications, geographic locations) and visual assets to corresponding manufacturing products.
- We introduce Cascade Multimodal Attributed Graph (C-MAG) for manufacturer–product link prediction (Fig. 1). In the top-level graph, diverse manufacturer attribute nodes and image nodes—are algined and fused into a single group embedding. That embedding then flows into the lower heterograph, where it propagates between the manufacturer and product nodes via multiscale message passing, substantially boosting link-prediction accuracy.
- We present a graph-based link-prediction framework and comprehensive evaluation suite for quantifying the impact of visual features. Our benchmark reveals the trade-offs of multimodal integration and delivers practical, modality-aware fusion guidelines that safeguard predictive accuracy in noisy, real-world supply-chain scenarios.

For the rest of the paper, we review the related work in Sec. 2. In Sec. 3, the details of the proposed method are presented. In Sec. 4, the experiments are conducted to demonstrate the effectiveness of our method. In Sec. 5, the limitations and future work of our method are concluded.

2 Related Work

2.1 Graph Datasets

Heterogeneous Graphs. Heterogeneous graphs (or heterographs) extend standard graphs by allowing multiple node and edge types, enabling more faithful modeling of complex domains—such as supply chains—where entities and their relationships carry distinct semantics [7, 35]. Influential heterographs, like the Open Academic Graph (OAG), which merges 0.7 billion entities and 2 billion relations across papers, authors, venues, and institutions, and the

diverse, large-scale datasets of the Open Graph Benchmark (OGB) demonstrate the scale and domain heterogeneity that heterographs can capture [12, 51].

Multimodal Knowledge Graphs (MMKG). MMKGs extend heterogeneous graphs by enriching entities and relations with diverse data modalities—such as textual descriptions, images, and region-level annotations—rather than relying solely on structured triples [53]. While heterographs model multiple node and edge types to capture domain heterogeneity, MMKGs embed unstructured or semi-structured content directly into the graph, enabling joint reasoning over visual and textual signals [46]. For example, Visual Genome [22] builds scene-level graphs from 100,000 images annotated with region descriptions, Richpedia [45] links Wikidata entities to images via RDF-based visual-semantic relations, and MMEA-UMVM [6] provides 97 benchmark KGs with incomplete or ambiguous images to evaluate multimodal entity alignment.

2.2 Attributed Graphs

Attributed Graph (AG) is a directed multigraph where nodes and edges carry arbitrary key–value properties, enabling rich metadata annotation [34]. Unlike ontology-bound RDF triples in knowledge graphs, AGs allow flexible property addition without a global schema. This flexibility supports fine-grained entity–relation modeling and joint topology-feature reasoning for analytics and machine learning. [37].

Text Attributed Graphs (TAG). Text-attributed graphs enrich nodes (and optionally edges) with unstructured text. Key benchmarks include OGBN-arxiv [12], CS-TAG's eight large-scale graphs across diverse domains [49], and TEG-DB's unified text-edge graph suite with a modular evaluation pipeline [27].

Multimodal Attributed Graphs (MAG). A MAG enriches nodes and edges with heterogeneous feature modalities (e.g., structured data, text, images, audio). MAGB [50] provides the first benchmark for MAG representation learning, and Kannan et al. [17] align scientific papers with code and images to illustrate text–visual fusion challenges. MMIEA [52] introduces a cross-modal entity alignment model that addresses misalignment and missing-modality issues in attributed graphs.

2.3 Graph Based Link Prediction

Link prediction (LP) in supply chain heterographs aims to infer missing or potential links by estimating the likelihood of connections between node pairs based on observed topology and attributes. Approaches divide into two paradigms: Supervised methods train edge classifiers with cross-entropy loss on known links using Graph Neural Network (GNNs) such as GCN [18], GraphSAGE [11], GAT [43], and HGT [13]. Besides, unsupervised techiques, like GraphSAGE [11], R-GCN [36], and Node2Vec [10],leverage contrastive objectives, negative sampling, or random-walk embeddings to separate true from false edges without labels.

---

[1]*PMGraph* is publicly available at https://huggingface.co/datasets/shawntzx/PMGraph.
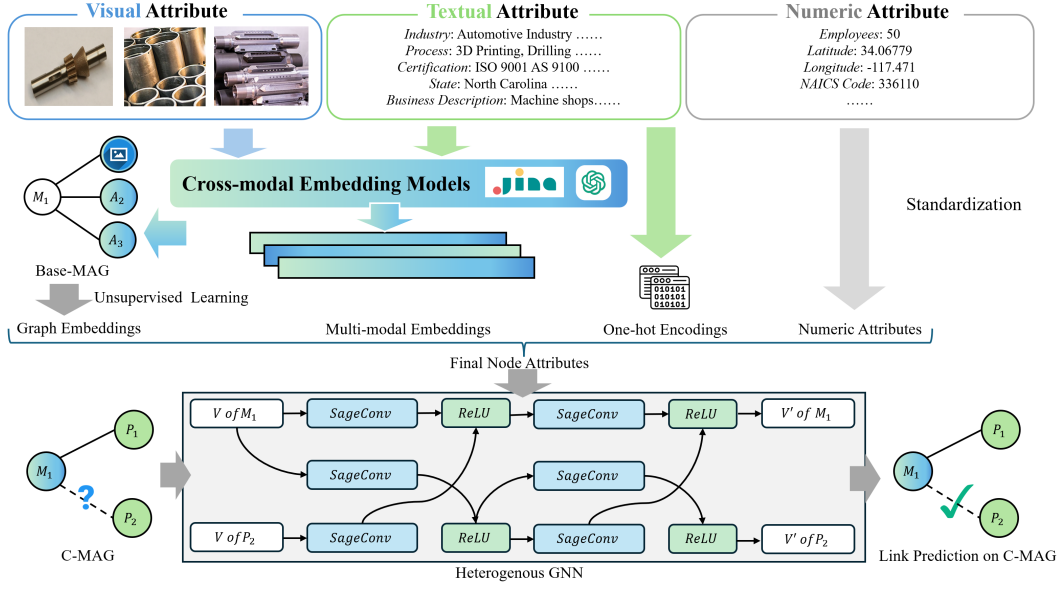
Figure 2: Overview of C-MAG pipeline

2.4 Supply-Chain Data Resources and GNN Applications

Recent supply chain graph datasets include the SC-KG by Wichmann, Brintrup, and Baker et al., which models multi-type relations (supplier–buyer, supplier–product, product–ingredient) across automotive and energy sectors for over 40,000 companies [20]; the Listed Companies Supply Chain Network of publicly traded Chinese firms capturing procurement and sales links for industry classification [48]; and SupplyGraph, a public FMCG production-planning benchmark with temporal node features from a Bangladeshi company [47]. However, most of existing supply-chain graphs are proprietary or limited to homogeneous relations within narrow domains or omit unstructured modalities (e.g., text, images)—underscoring the need for a publicly available, scalable, multimodal SC-KG.

Recent GNN-based frameworks for manufacturer and supplier prediction include a Manufacturing Service Knowledge Graph that uses neighborhood aggregation and oversampling to handle imbalanced capability classification [24]; an ontological Knowledge Graph (KG) combined with Retrieval-Augmented Generation for context-aware supplier discovery [23]; hidden-link prediction in automotive networks via GNNs augmented with Integrated Gradients for interpretability [19]; and a large-scale automotive SC-KG leveraging structured and textual embeddings for alternative supplier recommendation [42]. Despite their effectiveness, these frameworks largely rely on structural and textual embeddings, neglecting the rich visual assets and the alignment and noise challenges that arise when fusing heterogeneous modalities in supply-chain graphs.

3 Methodology

To mitigate limitations of existing frameworks—which largely rely on structural and textual embeddings, neglecting rich visual assets and challenges of alignment and noise in multimodal SC-KGs,

we propose C-MAG, a two-stage multimodal graph construction framework aimed at enhancing link prediction accuracy through hierarchical representation learning. The overall pipeline of our method is illustrated in Fig. 2, highlighting the integration of visual, textual, categorical, and numeric attributes into node representations for LP.

3.1 Stage 1: Base MAG and Auxiliary Pretraining

In the first stage, we construct a base-MAG capturing relationships among manufacturer nodes, attribute nodes (processes, certifications, materials, cities, states), and image nodes extracted directly from manufacturer websites.

Unified CLIP embeddings. Manufacturer textual descriptions, attribute labels, and product images are encoded into a unified 768-D embedding space using Jina-CLIP-v1 [21]. Embeddings are then compressed to 32-D via Truncated Singular Value Decomposition (SVD) [1], effectively reducing noise and computational complexity.

Unsupervised graph embedding. Manufacturer embeddings are pretrained on this top level MAG via unsupervised link prediction. A two-layer GraphSAGE architecture ($768 \rightarrow 64 \rightarrow 32$) aggregates neighbor embeddings, applying linear transformations and ReLU activations. Positive edges (manufacturer $\rightarrow$ attribute/image) are contrasted against sampled negative edges using binary cross-entropy loss (BCE). An R-GCN variant is also validated to ensure architecture-agnostic embedding quality.

3.2 Stage 2: Bipartite Manufacturer–Product Graph Design

In the second stage, we construct the bipartite manufacturer–product graph by initializing each node with the embeddings aggregated

Table 1: *PMGraph* statistics.

Category	Type	#
Node types	Manufacturer	8,888
	Product	72,789
	Attribute	2,918
	Image	29,178
Edge types	manufacturer→product	112,597
	manufacturer→attribute	83,105
	manufacturer→image	29,178

Table 2: Variants of heterographs used in our experiments.

Variant	Hierarchy	Text Embedding	Images	Bipartite
AG_{TFIDF}	Flat	TF–IDF	No	Yes
AG_{JINA}	Flat	Jina–CLIP	No	Yes
FAG_1	Flat	Jina–CLIP	No	No
$FMAG_2$	Flat	Jina–CLIP	Yes	No
C-MAG$_1$	2-layer	Jina–CLIP	No	Yes
C-MAG$_2$	2-layer	Jina–CLIP	Yes	Yes

from the first-stage attribute graph, thereby producing the final C-MAG representation.

Textual and categorical embeddings. Manufacturer and product textual attributes are processed with unified CLIP embeddings (768-D), compressed to 32-D with truncated SVD. Categorical metadata (business status, industry tags, certifications, NAICS codes) are one-hot encoded, numeric features (employee counts, geographic coordinates) are standardized, and the combined feature set is similarly compressed into 32-D embeddings via SVD.

Fusion and final multimodal graph assembly. Manufacturer node embeddings from Stage 1 (32-D) are concatenated with Stage 2 textual, categorical, and numeric embeddings (64-D), forming enriched 96-D vectors. SVD further compresses these embeddings back to 64-D. Product nodes retain their original 64-D embeddings derived from textual and categorical data. The resulting a bipartite MAG encodes comprehensive manufacturer–product features, integrating structural, textual, categorical, numeric, and visual information for robust link prediction.

3.3 Link Prediction

With these enriched node embeddings, we perform link prediction on manufacturer-product relationships using two heterogeneous Graph Neural Networks (heteroGNNs): HeteroSAGE and HeteroGAT. Initially, reverse edges (*product → manufacturer*) are explicitly added to complement the forward edges (*manufacturer → product*), thus creating a bidirectional heterogeneous graph structure that facilitates robust information propagation.

We implement two-layer architectures for both heteroGNN models, including intermediate nonlinear activations, dropout, and residual connections where applicable. Each model employs node-type-specific linear projection layers for embedding dimensionality alignment. To manage class imbalance, we utilize Weighted BCE [30].

4 Experiments and Analysis

4.1 Data Preparation

4.1.1 Data sources. The raw graph corpus is assembled by merging (i) the SUDOKN SC-KG [4] and (ii) the manufacturer directory and associated ontology of Li *etal.*[26]. Collectively, these sources comprise 8,888 unique manufacturer URLs (each offering multiple products) with extensive metadata, including employee counts, business descriptions, industry tags (up to 30 per manufacturer),

process and material capabilities (up to 50 each), certifications, geographic coordinates, NAICS codes, and detailed product catalogs.

Product images are scraped from the manufacturer websites referenced in SUDOKN and then passed through a multimodal LLM–based quality filter to remove off-topic or low-resolution content (see Appendix A). After filtering, 145,888 images remain. Since including all image nodes can introduce noise, we conducted an ablation study to evaluate link-prediction performance under different sampling ratios (See 4.5.2). In our main experiments, we randomly sample 20% of the filtered images to enrich *PMGraph*. The resulting graph statistics are presented in Table 1.

4.2 Graph Variants

All six heterographs share the same manufacturer–product topology and compress node features to 64-D via truncated SVD. We evaluate two baseline models, AG_{TFIDF} and AG_{JINA}, which use a flat bipartite graph with TF–IDF or Jina-CLIP embeddings for textual features; two proposed cascade models, C-MAG$_1$ and C-MAG$_2$; and two ablation variants, FAG_1 (the flat counterpart of MAG_1) and $FMAG_2$ (the flat counterpart of MAG_2). This setup allows us to compare embedding methods, assess the benefit of a cascade architecture by contrasting FAG_1 versus C-MAG$_1$ and $FMAG_2$ versus C-MAG$_2$, and evaluate the impact of visual features by comparing C-MAG$_2$ versus C-MAG$_1$ and $FMAG_2$ versus FAG_1. The six variants are summarized in Table 2.

4.3 Experimental Setup

4.3.1 Unsupervised Pretraining. Following the auxiliary pretraining in Sec. 3.1, we conduct random-seed trials of a two-layer Graph-SAGE encoder ($768 \rightarrow 64 \rightarrow 32$) for up to 1000 epochs with early stopping (patience = 20). Optimization uses Adam (lr = 10^{-3}, weight decay = 10^{-5}), with batch sizes of 32 for attribute edges and 16 for image edges. To confirm architecture-agnosticism, we repeat this procedure using a two-layer R-GCN under identical settings.

4.3.2 Supervised Link Prediction. Following most of the experimental setup of Lv et al. [29], we fine-tune two heteroGNN architectures—HeteroSAGE and HeteroGAT (4 attention heads)—on both the bipartite manufacturer–product graph and the full heterograph including attribute and image relations. Each model consists of two projection layers with 128 hidden units and applies a dropout rate of 0.5. We split edges into train/validation/test sets in an 80/10/10% ratio and use a 1:1 negative-sampling ratio. Training runs for up to 1,000 epochs with early stopping (patience = 20) monitored on the validation ROC–AUC. For hyperparameter tuning, we grid-search

Table 3: ROC-AUC and PR-AUC (%) for hetero-GNNs across graph variants (with base-MAGs pre-trained on GraphSAGE). Best and second-best scores in each column are highlighted in bold and <u>underline</u>, respectively.

Variant	HeteroSAGE		HeteroGAT	
	ROC-AUC	PR-AUC	ROC-AUC	PR-AUC
AG_{TFIDF}	54.60	56.38	70.92	70.11
AG_{JINA}	66.46	60.58	72.49	71.02
FAG_1	63.63	58.20	72.49	70.94
$FMAG_2$	57.11	51.90	71.55	69.63
$C\text{-}MAG_1$	<u>66.88</u>	<u>60.94</u>	**75.46**	<u>73.78</u>
$C\text{-}MAG_2$	**70.58**	**66.09**	<u>73.49</u>	**74.30**

the learning rate over $\{1, 5\} \times \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$ and optimize a weighted binary cross-entropy loss, selecting the configuration that maximizes validation ROC–AUC.

Flat-graph ablations. While the bipartite link-prediction setup uses only manufacturer–product edges, the flat-graph ablations (FAG_1, $FMAG_2$) trains on the full heterograph (manufacturers, products, attributes, images) but evaluates exclusively on the manufacturer–product relation. Concretely, we apply randomly link split to manufacturer-product edges only, then strip attribute and image edges from the validation and test graphs so that only bipartite links are scored.

All experiments run on NVIDIA GPUs with fixed random seeds for reproducibility.

4.3.3 Evaluation metrics. We assess link-prediction performance on held-out product–manufacturer pairs using two threshold independent metrics:

ROC-AUC. (AUC_{ROC}) The probability that a randomly chosen true edge receives a higher dot-product score than a randomly chosen non-edge:

$$AUC_{ROC} = \Pr[s_{i_+} > s_{i_-}],$$

where s_i are model scores and $y_i \in \{0, 1\}$ are binary labels for each test pair, and i_+ and i_- index a random positive and negative example, respectively.

PR-AUC. (AUC_{PR}) The area under the precision–recall curve, summarizing the trade-off between precision and recall across thresholds. Let

$$Precision(t) = \frac{TP(t)}{TP(t) + FP(t)}, \quad Recall(t) = \frac{TP(t)}{TP(t) + FN(t)},$$

then

$$AUC_{PR} = \int_0^1 Precision(Recall^{-1}(r))\, dr.$$

4.4 Link Prediction Performance

Table 3 shows that the cascade variants $C\text{-}MAG_1$ and $C\text{-}MAG_2$ occupy the top two positions for both HeteroSAGE and HeteroGAT, clearly outperforming the flat AG_{JINA} and AG_{TFIDF} baselines. This demonstrates that enriching the bipartite manufacturer–product

Table 4: ROC-AUC and PR-AUC (%) for hetero-GNNs on $C\text{-}MAG_1$ and $C\text{-}MAG_2$ (with base-MAGs pre-trained on R-GCN). Best scores in each column are highlighted in bold.

Variant	HeteroSAGE		HeteroGAT	
	ROC-AUC	PR-AUC	ROC-AUC	PR-AUC
$C\text{-}MAG_1$	**70.41**	**65.90**	**74.84**	72.59
$C\text{-}MAG_2$	67.97	62.15	74.42	**72.70**

graph with manufacturer–attribute relations via a cascade architecture can improve link-prediction, and that Jina-CLIP embeddings capture manufacturing semantics more effectively than TF–IDF.

C-MAGs also outperform their flat counterparts (FAG_1 and $FMAG_2$) due to their hierarchical design: they first aggregate attribute information and manufacturer–attribute relations into the manufacturer nodes via an unsupervised link-prediction step, and then propagate these enriched partial embeddings into the manufacturer–product prediction network. This two-stage integration captures both attribute and topological context more effectively than the flat models, which fuse all information in a single step.

We observe that adding product image nodes does not consistently improve LP: the flat image-augmented model ($FMAG_2$) underperforms its non-visual counterpart (FAG_1), and the cascade variant $C\text{-}MAG_2$ does not always outperform $C\text{-}MAG_1$. This likely stems from residual noise in the filtered image set—despite removing irrelevant or low-quality images, spurious visual features can still degrade the learned embeddings. Nevertheless, the hierarchical, two-stage design of $C\text{-}MAG_2$ helps mitigate this issue: by first refining manufacturer node representations with image information and then using those enriched embeddings for manufacturer–product prediction, it achieves greater robustness to noisy vision data than the flat $FMAG_2$ model.

4.5 Ablation Studies

4.5.1 Ablation of Auxiliary Pretraining Encoder. To assess whether our gains derive from the GraphSAGE architecture or the graph design itself, we compare unsupervised pretraining with two-layer GraphSAGE against R-GCN on the MAG_2 variant. Table 4 reports downstream ROC-AUC and PR-AUC for each hetero-GNN.

With R-GCN, the cascade models $C\text{-}MAG_1$ and $C\text{-}MAG_2$ still outperform all other variants; however, integrating multimodal features can sometimes degrade LP performance, likely due to residual noise in the filtered image data.

4.5.2 Ablation of Images Sampling Ratio. To evaluate how the proportion of image nodes affects link-prediction in the cascade model, we ablate the image sampling ratio for $C\text{-}MAG_2$ at 10%, 20%, and 50%. Figure 3 presents the ROC-AUC and PR-AUC trends obtained with the HeteroSAGE and HeteroGAT encoders.

Performance for both architectures peaks at 20% sampling, with HeteroGAT consistently outperforming HeteroSAGE. Increasing sampling beyond 20% yields only marginal improvements—likely a result of residual noise in the filtered images—but $C\text{-}MAG_2$'s hierarchical cascade maintains strong LP accuracy, outperforming all flat variants and demonstrating robustness to noisy visual data.
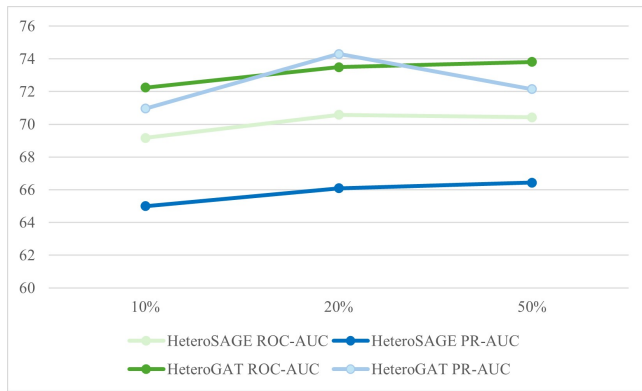
Figure 3: LP Performance with ROC-AUC(%) and PR-AUC(%) under different image sampling rates on for C-MAG$_2$.

5 Conclusion and Future Work

In this work, we introduce *PMGraph*, a large-scale, heterogeneous, multimodal supply-chain benchmark linking 8,888 manufacturers, over 70,000 products, and over 29,000 product images, and propose C-MAG architecture customized for manufacturer–product LP. By enriching the bipartite manufacturer–product graph with manufacturer–attribute relations in a two-stage cascade—first aggregating textual and visual attributes into intermediate embeddings, then propagating these through the full heterograph—C-MAG achieves state-of-the-art ROC-AUC and PR-AUC under both HeteroSAGE and HeteroGAT, substantially outperforming other graph variants. Our ablations show that C-MAG's hierarchical fusion is robust to noisy visual data, peaking at moderate image sampling ratios while maintaining strong performance even as image noise increases. These results demonstrate the importance of modality-aware, staged integration of heterogeneous information for robust SC-KG modeling. Moreover, the C-MAG framework readily generalizes beyond manufacturer–product matching to other supply-chain inference tasks—such as component prediction, domain-specific LP, and e-commerce recommendation system.

Future work will investigate adaptive image-filtering strategies using advanced vision–language models (e.g., GPT-o3 [31], Claude Sonnet [33]) and explore dynamic cascade depths to better accommodate varying data quality [9, 41]. We will also extend evaluations to a wider range of heterogeneous GNN architectures [8, 38] and conduct systematic ablations over key hyperparameters to optimize model performance in diverse, real-world supply-chain scenarios.

References

[1] Hervé Abdi. 2007. Singular value decomposition (SVD) and generalized singular value decomposition. *Encyclopedia of measurement and statistics* 907, 912 (2007), 44.

[2] Sara AlMahri, Liming Xu, and Alexandra Brintrup. 2024. Enhancing supply chain visibility with knowledge graphs and large language models. *arXiv preprint arXiv:2408.07705* (2024).

[3] Farhad Ameri, Mukund Shenoy, Ali Hasanzadeh, and Sambhav Kapoor. 2024. Open Manufacturing Capability Network Supported by Formal Ontologies. In *IFIP International Conference on Advances in Production Management Systems*. Springer, 310–324.

[4] Farhad Ameri, Mukund Shenoy, Ali Hasanzadeh, and Sambhav Kapoor. 2024. Open Manufacturing Capability Network Supported by Formal Ontologies. In *Advances in Production Management Systems. Production Management Systems for Volatile, Uncertain, Complex, and Ambiguous Environments. 43rd IFIP WG 5.7 International Conference, APMS 2024, Proceedings (IFIP Advances in Information and Communication Technology)*, Matthias Thürer, Ralph Riedel, Gregor von Cieminski, and David Romero (Eds.). Springer Nature, Cham, Switzerland, 310–324. doi:10.1007/978-3-031-71637-9_21 Conference held September 8–12, 2024.

[5] Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 2 (Feb. 2019), 423–443. doi:10.1109/TPAMI.2018.2798607

[6] Zhuo Chen, Lingbing Guo, Yin Fang, Yichi Zhang, Jiaoyan Chen, Jeff Z Pan, Yangning Li, Huajun Chen, and Wen Zhang. 2023. Rethinking uncertainly missing and ambiguous visual modality in multi-modal entity alignment. In *International Semantic Web Conference*. Springer, 121–139.

[7] Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 135–144.

[8] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. 2020. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of the web conference 2020*. 2331–2341.

[9] Xing Gao, Xiaogang Jia, Yikang Li, and Hongkai Xiong. 2023. Dynamic scenario representation learning for motion forecasting with heterogeneous graph convolutional recurrent networks. *IEEE Robotics and Automation Letters* 8, 5 (2023), 2946–2953.

[10] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.

[11] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[12] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. 2020. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems* 33 (2020), 22118–22133.

[13] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*. 2704–2710.

[14] Ahmed Ismail, Hong-Linh Truong, and Wolfgang Kastner. 2019. Manufacturing process data analysis pipelines: a requirements analysis and survey. *Journal of Big Data* 6, 1 (2019), 1–26.

[15] Eeva Järvenpää, Niko Siltala, Otto Hylli, and Minna Lanz. 2019. The development of an ontology for describing the capabilities of manufacturing resources. *Journal of Intelligent Manufacturing* 30, 2 (2019), 959–978.

[16] Simon Kamm, Nasser Jazdi, and Michael Weyrich. 2021. Knowledge discovery in heterogeneous and unstructured data of industry 4.0 systems: challenges and approaches. *Procedia CIRP* 104 (2021), 975–980.

[17] Amar Viswanathan Kannan, Dmitriy Fradkin, Ioannis Akrotirianakis, Tugba Kulahcioglu, Arquimedes Canedo, Aditi Roy, Shih-Yuan Yu, Malawade Arnav, and Mohammad Abdullah Al Faruque. 2020. Multimodal knowledge graph for deep learning papers and code. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 3417–3420.

[18] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[19] Edward Elson Kosasih and Alexandra Brintrup. 2022. A machine learning approach for predicting hidden links in supply chain with graph neural networks. *International Journal of Production Research* 60, 17 (2022), 5380–5393.

[20] Edward Elson Kosasih, Fabrizio Margaroli, Simone Gelli, Ajmal Aziz, Nick Wildgoose, and Alexandra Brintrup. 2024. Towards knowledge graph reasoning for supply chain risk management using graph neural networks. *International Journal of Production Research* 62, 15 (2024), 5596–5612.

[21] Andreas Koukounas, Georgios Mastrapas, Michael Günther, Bo Wang, Scott Martens, Isabelle Mohr, Saba Sturua, Mohammad Kalim Akram, Joan Fontanals Martínez, Saahil Ognawala, Susana Guzman, Maximilian Werk, Nan Wang, and Han Xiao. 2024. Jina CLIP: Your CLIP Model Is Also Your Text Retriever. *arXiv preprint arXiv:2405.20204* (2024). https://arxiv.org/abs/2405.20204

[22] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. 2017. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International journal of computer vision* 123 (2017), 32–73.

[23] Yunqing Li, Hyunwoong Ko, and Farhad Ameri. 2025. Integrating Graph Retrieval-Augmented Generation With Large Language Models for Supplier Discovery. *Journal of Computing and Information Science in Engineering* 25, 2 (01 2025), 021010. doi:10.1115/1.4067389 arXiv:https://asmedigitalcollection.asme.org/computingengineering/article-pdf/25/2/021010/7422522/jcise_25_2_021010.pdf

[24] Yunqing Li, Xiaorui Liu, and Binil Starly. 2024. Manufacturing service capability prediction with Graph Neural Networks. *Journal of Manufacturing Systems* 74 (2024), 291–301.

[25] Yunqing Li, Shivakumar Raman, Paul Cohen, and Binil Starly. 2021. Design of knowledge graph in manufacturing services discovery. In *International Manufacturing Science and Engineering Conference*, Vol. 85079. American Society of Mechanical Engineers, V002T07A010.

[26] Yunqing Li and Binil Starly. 2024. Building a knowledge graph to enrich ChatGPT responses in manufacturing service discovery. *Journal of Industrial Information Integration* 40 (2024), 100612. doi:10.1016/j.jii.2024.100612

[27] Zhuofeng Li, Zixing Gou, Xiangnan Zhang, Zhongyuan Liu, Sirui Li, Yuntong Hu, Chen Ling, Zheng Zhang, and Liang Zhao. 2024. Teg-db: A comprehensive dataset and benchmark of textual-edge graphs. *Advances in Neural Information Processing Systems* 37 (2024), 60980–60998.

[28] Yifan Liu, Kangning Zhang, Xiangyuan Ren, Yanhua Huang, Jiarui Jin, Yingjie Qin, Ruilong Su, Ruiwen Xu, Yong Yu, and Weinan Zhang. 2024. AlignRec: Aligning and Training in Multimodal Recommendations. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 1503–1512.

[29] Qingsong Lv, Ming Ding, Qiang Liu, Yuxiang Chen, Wenzheng Feng, Siming He, Chang Zhou, Jianguo Jiang, Yuxiao Dong, and Jie Tang. 2021. Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 1150–1160.

[30] Jishnu Mukhoti, Viveka Kulharia, Amartya Sanyal, Stuart Golodetz, Philip Torr, and Puneet Dokania. 2020. Calibrating deep neural networks using focal loss. *Advances in neural information processing systems* 33 (2020), 15288–15299.

[31] OpenAI. 2025. Introducing OpenAI o3 and o4-mini. https://openai.com/index/introducing-o3-and-o4-mini/?utm_source=chatgpt.com. Accessed: 2025-05-25.

[32] Rajat Panwar, Jonatan Pinkse, and Valentina De Marchi. 2022. The future of global supply chains in a post-COVID-19 world. *California Management Review* 64, 2 (2022), 5–23.

[33] Anthropic PBC. 2025. Claude Sonnet 4. https://www.anthropic.com/claude/sonnet?utm_source=chatgpt.com. Accessed: 2025-05-25.

[34] Joseph J Pfeiffer III, Sebastian Moreno, Timothy La Fond, Jennifer Neville, and Brian Gallagher. 2014. Attributed graph models: Modeling network structure with correlated attributes. In *Proceedings of the 23rd international conference on World wide web*. 831–842.

[35] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web – ISWC 2018*. Springer, 593–607.

[36] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15*. Springer, 593–607.

[37] Nasrullah Sheikh, Zekarias Kefato, and Alberto Montresor. 2019. gat2vec: representation learning for attributed graphs. *Computing* 101, 3 (2019), 187–209.

[38] Chuan Shi, Xiao Wang, and S Yu Philip. 2022. *Heterogeneous graph representation learning and applications*. Springer.

[39] Binil Starly, Paul Cohen, and Shivakumar Raman. 2020. Automating the search and discovery of manufacturing service providers to enable a digital supply chain network. *Smart and Sustainable Manufacturing Systems* 4, 3 (2020), 276–280.

[40] Anna Teern, Nada Elgendy, Markus Kelanti, Henna Tammia, and Tero Päivärinta. [n. d.]. From Dynamic to Evolvable Knowledge Graphs in Manufacturing: Systematic Literature Review on Learning Approaches. ([n. d.]).

[41] Yue Tian, Guanjun Liu, Jiacun Wang, and Mengchu Zhou. 2023. ASA-GNN: Adaptive sampling and aggregation-based graph neural network for transaction fraud detection. *IEEE Transactions on Computational Social Systems* 11, 3 (2023), 3536–3549.

[42] Yuchun Tu, Wenxin Li, Xiao Song, Kaiqi Gong, Lu Liu, Yunhao Qin, Songsong Liu, and Ming Liu. 2024. Using graph neural network to conduct supplier recommendation based on large-scale supply chain. *International Journal of Production Research* 62, 24 (2024), 8595–8608.

[43] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations (ICLR)*. doi:10.48550/arXiv.1710.10903

[44] Yuwei Wan, Ying Liu, Zheyuan Chen, Chong Chen, Xinyu Li, Fu Hu, and Michael Packianather. 2024. Making knowledge graphs work for smart manufacturing: Research topics, applications and prospects. *Journal of Manufacturing Systems* 76 (2024), 103–132.

[45] Meng Wang, Haofen Wang, Guilin Qi, and Qiushuo Zheng. 2020. Richpedia: a large-scale, comprehensive multi-modal knowledge graph. *Big Data Research* 22 (2020), 100159.

[46] Yanan Wang, Michihiro Yasunaga, Hongyu Ren, Shinya Wada, and Jure Leskovec. 2022. VQA-GNN: Reasoning with Multimodal Knowledge via Graph Neural Networks for Visual Question Answering. In *arXiv preprint arXiv:2205.11501*.

[47] Azmine Toushik Wasi, MD Islam, and Adipto Raihan Akib. 2024. Supplygraph: A benchmark dataset for supply chain planning using graph neural networks. *arXiv preprint arXiv:2401.15299* (2024).

[48] Desheng Wu, Quanbin Wang, and David L Olson. 2023. Industry classification based on supply chain network information using graph neural networks. *Applied Soft Computing* 132 (2023), 109849.

[49] Hao Yan, Chaozhuo Li, Ruosong Long, Chao Yan, Jianan Zhao, Wenwen Zhuang, Jun Yin, Peiyan Zhang, Weihao Han, Hao Sun, et al. 2023. A comprehensive study on text-attributed graphs: Benchmarking and rethinking. *Advances in Neural Information Processing Systems* 36 (2023), 17238–17264.

[50] Hao Yan, Chaozhuo Li, Zhigang Yu, Jun Yin, Ruochen Liu, Peiyan Zhang, Weihao Han, Mingzheng Li, Zhengxin Zeng, Hao Sun, et al. 2024. When Graph meets Multimodal: Benchmarking on Multimodal Attributed Graphs Learning. *arXiv preprint arXiv:2410.09132* (2024).

[51] Fanjin Zhang, Xiao Liu, Jie Tang, Yuxiao Dong, Peiran Yao, Jie Zhang, Xiaotao Gu, Yan Wang, Bin Shao, Rui Li, et al. 2019. OAG: Toward linking large-scale heterogeneous entity graphs. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2585–2595.

[52] Bin Zhu, Meng Wu, Yunpeng Hong, Yi Chen, Bo Xie, Fei Liu, Chenyang Bu, and Weiping Ding. 2023. Mmiea: Multi-modal interaction entity alignment model for knowledge graphs. *Information Fusion* 100 (2023), 101935.

[53] Xiangru Zhu, Zhixu Li, Xiaodan Wang, Xueyao Jiang, Penglei Sun, Xuwu Wang, Yanghua Xiao, and Nicholas Jing Yuan. 2022. Multi-Modal Knowledge Graph Construction and Application: A Survey. *arXiv preprint arXiv:2202.05786* (2022).

A Visual Data Extraction

A.1 Image Collection

We develop a scalable, high-throughput pipeline to systematically collect product images from a curated set of over 8,000 manufacturer websites. Each target URL is first normalized to ensure syntactic consistency before initiating HTTP requests to retrieve the corresponding webpage content. To robustly accommodate a wide range of webpage structures—including malformed markup or embedded binary artifacts—the retrieved HTML is processed using a fault-tolerant, two-stage parsing strategy.

Following content extraction, all intra-domain hyperlinks are identified and subjected to a lexical filtering mechanism. Specifically, links are retained if they contain keywords indicative of product-related content (e.g., "product","item" "catalog", "gallery", "prod") and discarded if they correspond to non-product sections (e.g., "about", "contact", "blog", "news", "login", "signup"). In cases where few product-specific links are detected, a supplementary set of additional internal pages is also explored to improve recall.

The homepage and selected product-relevant pages are then scanned for image elements. All discovered image URLs are resolved to their absolute forms, downloaded as binary files, and stored in structured subdirectories organized by domain. Filenames are semantically generated based on associated textual metadata, such as alt attributes, or default to the original filenames when descriptive information is unavailable—ensuring contextual relevance and traceability.

By integrating structured domain ingestion, robust HTML parsing, keyword-driven link prioritization, and hierarchical image storage, this pipeline enables efficient, scalable, and high-precision harvesting of product imagery across a broad landscape of industrial web sources.

A.2 Image Filtering

Vision–Language Filtering. Subsequently, a vision–language filtering step is introduced using the `Gemini-2.0-Flash` model. For each image in a manufacturer's directory, we issue the following templated prompt (with `product_list_str` dynamically populated from our dataset):

- You are an image verification assistant. Given an image and a list of products (`product_list_str`), respond with exactly one word: "Yes" or "No."

- Reply "Yes" only if you are absolutely certain the image shows one of the listed products.
- Reply "No" if you are uncertain, if the product is not in the list, or if image quality is too low to decide.