SupplyGraph: A Benchmark Dataset for Supply Chain Planning using Graph Neural Networks

Azmine Toushik Wasi, MD Shafikul Islam, Adipto Raihan Akib

Shahjalal University of Science and Technology, Bangladesh
azminetoushik.wasi@gmail.com, shafikul37@student.sust.edu, aqibadipto27@gmail.com

Abstract

Graph Neural Networks (GNNs) have gained traction across different domains such as transportation, bio-informatics, language processing, and computer vision. However, there is a noticeable absence of research on applying GNNs to supply chain networks. Supply chain networks are inherently graph-like in structure, making them prime candidates for applying GNN methodologies. This opens up a world of possibilities for optimizing, predicting, and solving even the most complex supply chain problems. A major setback in this approach lies in the absence of real-world benchmark datasets to facilitate the research and resolution of supply chain problems using GNNs. To address the issue, we present a real-world benchmark dataset for temporal tasks, obtained from one of the leading FMCG companies in Bangladesh, focusing on supply chain planning for production purposes. The dataset includes temporal data as node features to enable sales predictions, production planning, and the identification of factory issues. By utilizing this dataset, researchers can employ GNNs to address numerous supply chain problems, thereby advancing the field of supply chain analytics and planning. Source: https://github.com/CIOL-SUST/SupplyGraph

Introduction

Supply chain is a dynamic network of organizations that participate in the various processes and activities that produce value in the form of products and services for consumers via upstream and downstream linkages (Mentzer et al. 2001), entailing a continuous flow of information, goods, and money among its various stages (Higuchi and Troutt 2004). As supply chain consists of various entities interconnected in a complex network, it involves intricate inter-dependencies and complex decision-making processes. Besides, the modern supply chain produces an enormous amount of data; relationships and dependencies between all entities are not straightforward and require sophisticated models to capture (Sadeghiamirshahidi et al. 2014; Chaovalitwongse and Furman 2013).

Production planning is a crucial component of supply chain management as predicting the future demand for products or services, which helps organizations optimize their inventory levels, production schedules, and resource allocation (Aamer,

Eka Yani, and Alan Priyatna 2020; Zougagh, Charkaoui, and Echchatbi 2020). Company revenue greatly relies on accurately predicting demand and planning accordingly which has led to the exploration of various deep learning and machine learning models (Alves, Silva, and Mateus 2021) (Pirhooshyaran and Snyder 2020) (Pacella and Papadia 2021) to tackle this issue. However, GNNs can model network-like structures, such as trade flows in global trade or social networks, allowing for a better understanding of complex supply chain dynamics (Kosasih and Brintrup 2022). An illustration explaining Supply chain as a graph is presented in Figure 1. Although few studies have been conducted on applying GNNs to supply chain contexts only to predict hidden link prediction to mitigate risk (Kosasih and Brintrup 2022) and uncover hidden dependency links in supply chains (Aziz et al. 2021). The absence of publicly available datasets in this domain obstructs progress and benchmarking. By furnishing a dedicated dataset, we bridge the gap between supply chain dynamics and the potential of GNNs, enabling novel models that can facilitate robust supply chain strategies for businesses.

Related Works

Supply Chain and Machine Learning

In the domain of production planning within supply chain management, extensive research has been dedicated to harness machine learning (Feizabadi 2022; Zhu et al. 2021; Aamer, Eka Yani, and Alan Priyatna 2020). Numerous studies aim to improve demand prediction and optimize production processes through these methodologies (Nitsche et al. 2021; Younis, Sundarakani, and Alsharairi 2022; Filali, Filali et al. 2021). For instance, deep learning techniques including Artificial Neural Networks (ANNs) (Lunardi and Lima 2021; Mrad and Boujelbene 2019; Vairagade et al. 2019), Convolutional Neural Networks (CNNs) (Husna, Amin, and Shah 2021; Tang and Ge 2022; Li 2021), and Long Short-Term Memory networks (LSTMs) (Bousqaoui, Slimani, and Achchab 2021; Lingelbach et al. 2021). Researchers have increasingly leveraged these methods to enhance the efficiency and effectiveness of production planning and supply chain management.
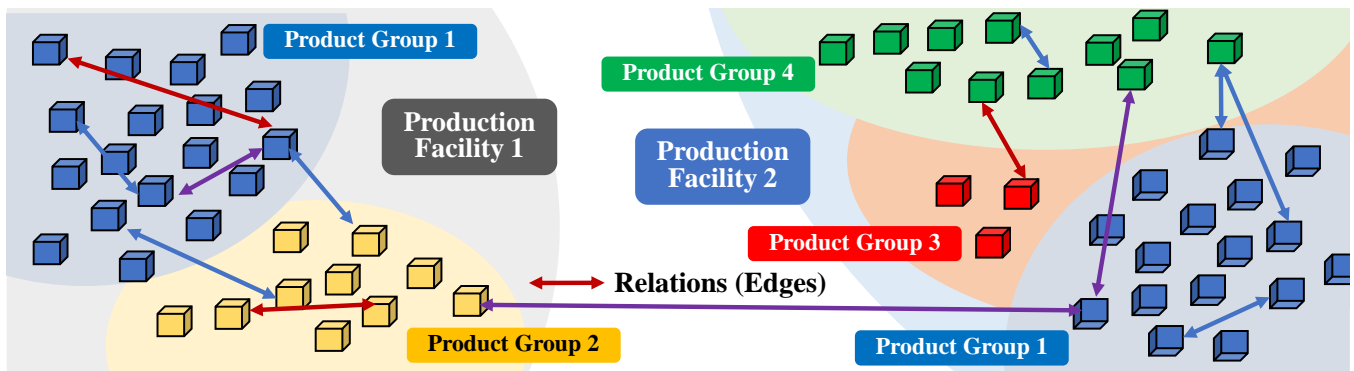
Figure 1: **Supply Chain Problem Formulation in Homogeneous Graph.** Boxes represent various product types, with color indicating different groups. They are closely located based on product groups and production facilities. Different relational connections denote shared raw material requirements, interdependence between products, and other impacts.

Table 1: Dataset Information

Node		Edge Classes		Edge Count	
Total Products (Nodes)	41	Total Edge Types	62	Total Unique Edges	684
Number of Groups	5	Class (Group)	5	Count (Group)	188
Number of Sub-groups	19	Class (Sub-group)	19	Count (Sub-group)	52
Number of Plants	25	Class (Plant)	25	Count (Plant)	1647
Number of Storage	13	Class (Storage)	13	Count (Storage)	3046

Supply Chain and Graph Neural Networks

Supply Chain machine learning with Graph Neural Networks holds significant promise by enabling the modeling of complex supply chain structures, optimizing logistics, and enhancing decision-making through data-driven insights. Recent advancements in this field leverage GNNs to improve demand forecasting, and enhance supply chain resilience, paving the way for more efficient and adaptive supply chain operations. Also, the utilization of graph representation learning methodologies has led to breakthroughs in revealing hidden dependency links within supply chain networks, further elevating the performance of link prediction tasks (Aziz et al. 2021; Kosasih and Brintrup 2022) inspired by previous GNN link prediction works (Zhang and Chen 2018, 2017; Cai et al. 2020; Cai and Ji 2020). This research highlights machine learning's effectiveness in refining demand forecasting and production planning, with GNNs showing potential in addressing supply chain challenges. However, growth in GNN-based supply chains research is very slow. It shows the significance of specialized datasets and cutting-edge methodologies for fully utilizing GNNs' advantages in addressing the complex demand prediction and production planning issues that arise in the supply chain domain.

Data Collection

We obtained data from the central database system of one of the biggest and most prominent FMCG (Fast Moving Consumer Goods) companies in Bangladesh. We reorganize it for temporal graph utilization, extracting nodes and features. We are unable to publish the actual product names, product codes, and the name of the company, as doing so could potentially compromise the competitive standing of the concerned company. Table 1 summarizes the dataset information.

Quality Control. Each individual node, edge, and node feature undergoes a thorough manual examination and validation process. This involves scrutinizing the data for any irregularities, such as anomalies or missing information. Additionally, even zero values are meticulously reviewed to ensure their accuracy and legitimacy.

Temporal & Location Coverage. The occurrences in this database cover the period from January 1, 2023, to August 9, 2023, and include four features: production, sales orders, delivery to distributors, and factory issues in two metrics: unit and weight produced in metric tons. The company operates across the entirety of Bangladesh.

Feature Description

Within SupplyGraph, **nodes** pertains to distinct products, while **edges** represent various connections linking these products: same product group or sub-group, same plant or storage location. Figure 2 presents some examples of homogeneous graphs.

In the temporal data, node features include production, sales orders, delivery to distributors, and factory issues.

Production, which quantifies product output considering sales orders, customer demand, vehicle fill rate, and delivery urgency. This quantity is typically measured in units or Metric Tons. *Sales Order* signifies distributor-requested quantities, pending approval from the accounts department. It reflects overall product demand. *Delivery to Distributors* denotes dispatched products aligning with orders, impacting company
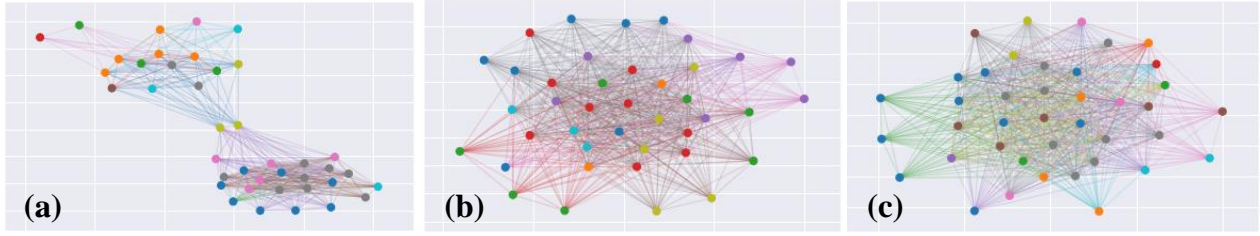
Figure 2: Homogeneous Graphs from the dataset. (a) Nodes are sub-group products, and plants are edges. (b) Nodes plant-products and storage locations are edges. (c) Nodes are sub-group products and storage locations are edges.

revenue significantly. *Factory Issue* covers total products shipped from manufacturing facilities, with some going to distributors and the rest to storage warehouses. We have all the temporal data in two modes: number of units produced (example: 500 units) and total weight example: 10 metric tons) produced.
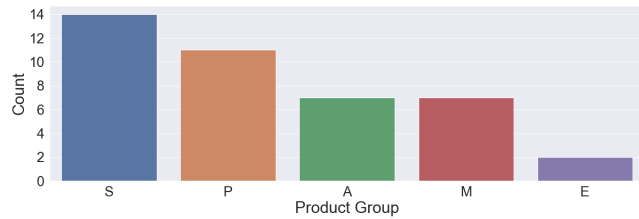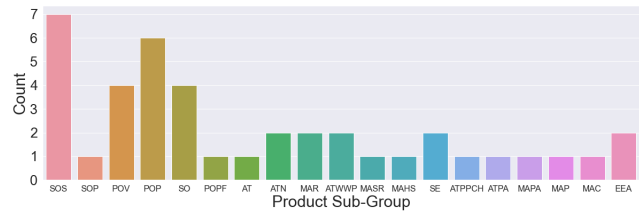
Data Statistics

Figure 3: Product group count.

Figure 4: Product subgroup count.

Product Group and Sub-groups   Figure 3 shows that product group "S" has the highest number of products reflecting its highest variety and product group "E" has the lowest number of products. Notably, a majority of the product groups exhibit a range of around 8 to 10 products each. Figure 4 reveals the diversity within product sub-groups. As previously noted, the "s" product group stands out for its remarkable variety, and this characteristic is consistently maintained within its corresponding sub-group. Among our product subgroups, "SOS" emerges as the one encompassing the largest array of product categories. Interestingly, despite the elevated diversity exhibited by the product group, the corresponding sub-group maintains a relatively stable count.

Analyzing Temporal Trends.   Figure 5, A particularly attractive pattern emerges: the production curve exhibits distinct and sharp fluctuations. Notably, the production unit of the company occasionally closes, which can be attributed to factors such as scheduled vacations, adjustments in the product's MRP rate, demand variations, and optimizations in transportation policies. In contrast, the other aspects of the graph, encompassing sales orders, delivery to distributors, and factory issuance, remain notably stable and consistent. Their minimal deviations stand in contrast to the variabilities characterizing the production curve, providing an insight into the dynamics of the supply chain activities.
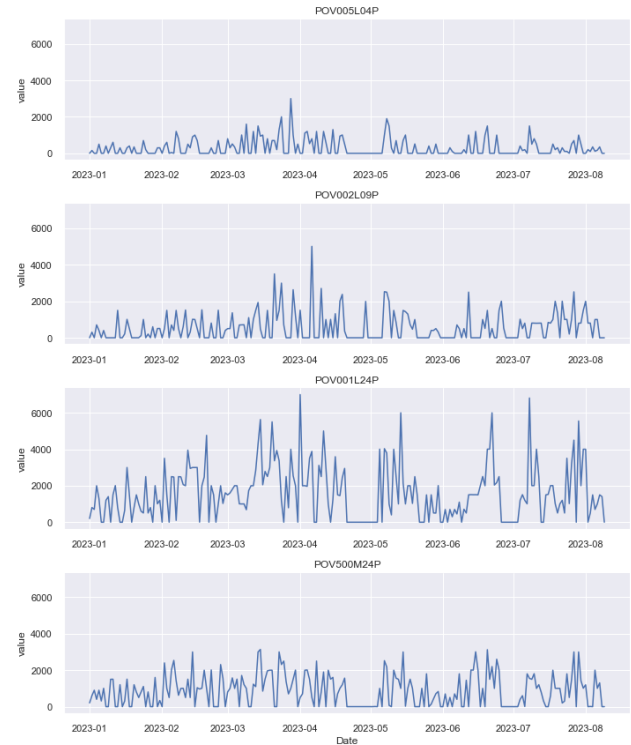

Figure 5: All product's production (MT) of POV sub-group.

In, Figure 6, we can notice a very interesting pattern. The product is manufactured in large batches, with orders being accumulated over several days before production begins. The processes of delivering to distributors, handling factory is-

sues, and fulfilling sales orders appear to be running smoothly and consistently. The company has opted to operate the production plant intermittently and maintain product inventory. When you observe straight lines in the production graph, it indicates that the plant is not actively engaged in production during those periods.
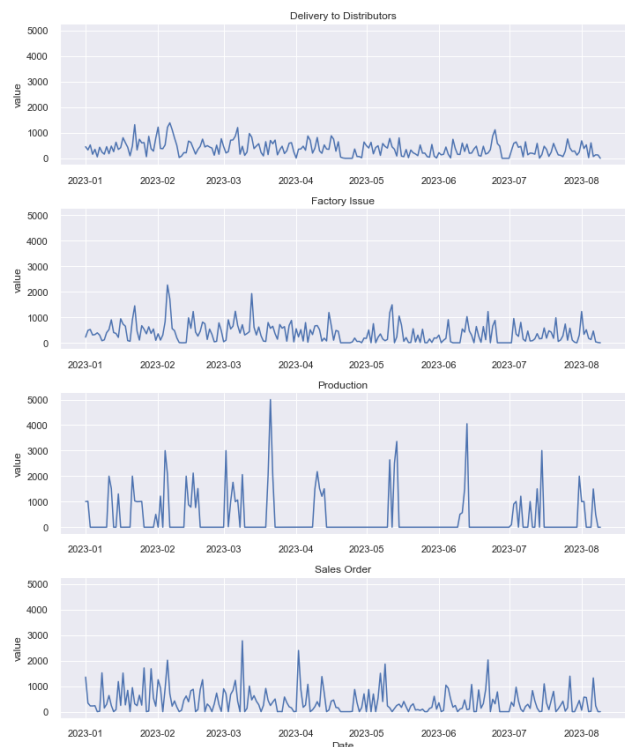

Figure 6: All 04 temporal values of product SOS008L02P.

Analyzing Temporal Correlations. The correlation plot portrayed in Figure 7 serves to demonstrate the ways in which diverse subgroups of product "A" are interconnected in a multitude of facets. It is worth noting that the relationship between delivery to distributors and sales orders is particularly robust, whereas the production unit exhibits a distinct and prominent correlation pattern. Furthermore, a noteworthy observation pertains to the correlation between the production quantities of two subgroups, specifically "ATWW-POO2K12P" and "ATWWPOO1K24P". Additionally, it is worth highlighting that factory issues and sales orders similarly display correlation trends, albeit with some discernible variations.

Upon examining the correlation plot of the "S" subgroup in Figure 8, a distinct pattern becomes apparent: there is a significant and consistent correlation between the "SOSOO1L12P" and "SOS500M24P" subgroups across all variables. Notably, an intriguing observation arises in the context of sales orders, where a majority of the products within the "S" subgroup demonstrate a relatively higher correlation compared to delivery to distributors, factory issuances, and production.

The comprehensive analysis of correlations and temporal activities strongly indicates a substantial opportunity for the
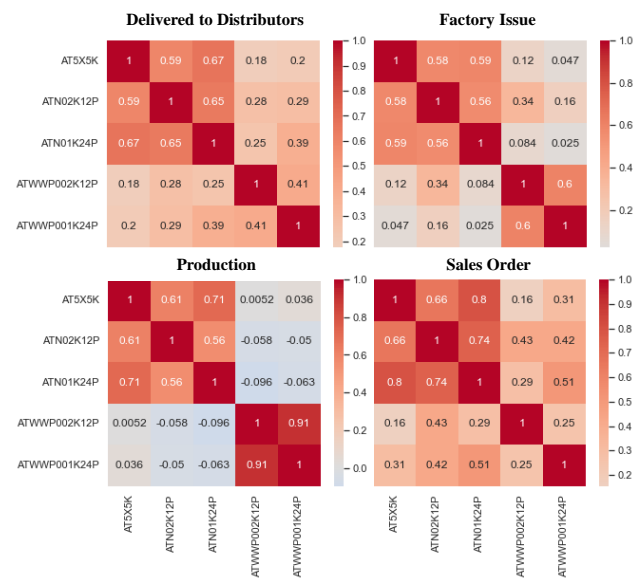

Figure 7: Temporal correlation of some products from Product Group A.

application of graph-based machine learning methodologies in the realm of supply chains. Leveraging the inherent structures and relationships within supply chain networks, graph-based machine learning holds great promise in enhancing decision-making processes, optimizing resource allocation, and fostering adaptability in the dynamic landscape of supply chain operations.

Dataset Potential and Practical Applications

Our dataset's structural design models products as nodes, while the interrelationships, such as shared product groups, subgroups, production facilities, and storage locations, are represented as edge connections. This dataset equips us to apply GNNs to complex supply chain predicaments, such as:

Supply chain planning involves the utilization of historical data pertaining to production and demand as node attributes, with the integration of time-related features to capture seasonality and trends. To enhance demand forecasting accuracy, related entities' influence can be harnessed through neighborhood nodes. The hierarchical structure of nodes, including product groups, subgroups, plants, and storage locations, can be used for hierarchy-aware forecasting. In data-scarce scenarios, improved forecasting can be achieved by applying transfer learning across nodes of the same classification.

Product Classification involves classifying products by product groups, sub-groups, facilities; and also by economic profitability and production similarities for economic decision-making.

Product Relation Classification involves classifying or predicting product relations in the supply chain graph. Product Relation detection involves detecting a missing edge or relation in the supply chain graph by a binary prediction objective.

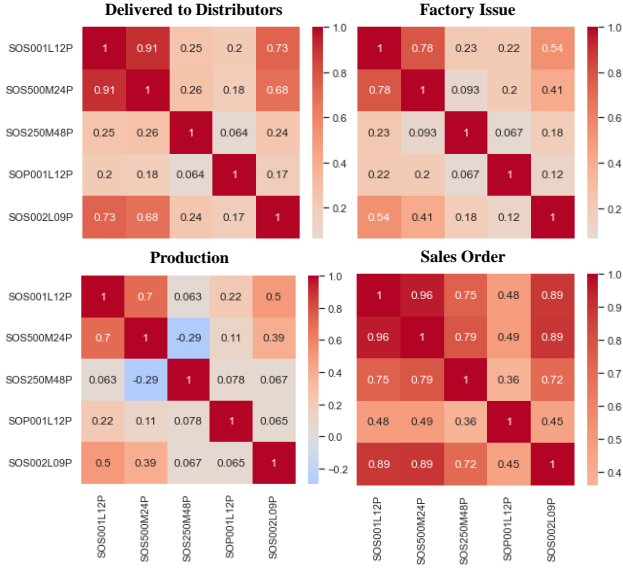**Supply chain optimization** involves modeling goods flow

Figure 8: Temporal correlation of some products from Product Group A.

between nodes, considering demand, lead times, and reorder points. Optimal routes and quantities can be recommended using edges as transportation links. Nodes representing plants can also aid in suggesting production adjustments based on demand projections, constraints, and capacity.

Anomaly detection is achieved by contrasting predicted demand with actual data to identify disruptions, stockouts, or unusual demand spikes. Deviations in demand from aggregated neighborhood demand can signal inconsistencies within neighborhoods, which can be detected as anomalies.

Event classification involves the training of GNNs to classify events based on edges and nodes, such as production capacity changes, new product launches, and disruptions.

Fluctuation Detection entails training GNNs to identify and classify fluctuations in the network based on impacts of global price hikes, supply chain issues, and disruptions. Fluctuations in supply chains refer to disruptions that can occur due to various factors and can significantly impact the operations of businesses. These disruptions can amplify negative shocks, affecting not just the firm experiencing the failure, but also its suppliers and customers, and even firms in other parts of the production network (Acemoglu and Tahbaz-Salehi 2020). Using temporal data, the model can recognize patterns in these fluctuations over time and determine if any demand and supply fluctuation can occur in near future for better planning.

Combinatorial (Cappart et al. 2022) and **Constrained** (Kotary et al. 2021) **Optimization** involves utilizing GNNs to optimize complex decision-making within the supply chain, considering various factors and constraints. GNNs can effectively model the intricate relationships between nodes and edges, enabling efficient allocation of resources, route planning, and inventory management, ultimately leading to improved performance.

Heterogeneous Graphs. All these tasks can be executed in both heterogeneous and homogeneous forms. In the context of the heterogeneous approach, the incorporation of nodes representing distinct entities, such as products, storage locations, and production facilities, interconnected by edges that signify their intricate relationships, becomes a pivotal strategy. This approach facilitates the construction of a heterogeneous graph model, which effectively encapsulates the intricate web of interactions inherent in supply chain networks. Embracing these heterogeneous models affords us the capacity to encompass the varied characteristics of different supply chain components and their complex interdependencies. This approach proves instrumental in addressing the multifaceted challenges that characterize supply chain management. The ability to comprehensively represent and analyze these heterogeneous relationships equips decision-makers with a deeper understanding of the dynamics within the supply chain, empowering them to devise optimized strategies and adapt to the dynamic nature of supply chain operations. In doing so, overall supply chain performance and resilience can be significantly enhanced.

Hypergraphs. Efficient task execution can also be achieved through hypergraph models. Nodes, representing entities like products, storage locations, and production facilities, intricately interconnect through hyperedges, capturing the intricate relationships within supply chain networks. This strategic use of hyperedges facilitates the construction of a tailored hypergraph model, adept at encapsulating the complex web of interactions inherent in supply chain networks. The adoption of hypergraph models can provide a dynamic framework to address the diverse characteristics and intricate interdependencies of supply chain components, serving as a pivotal strategy for managing the multifaceted challenges inherent in supply chain management and fostering a deeper understanding of its dynamics, which can contribute to the enhancement of overall supply chain performance and resilience.

Conclusion

Supply chain networks present a potent framework for optimization and problem-solving for GNNs. However, the lack of real-world benchmark datasets impedes progress in this field. To address this issue, our work aims to provide a real-world benchmark graph-based dataset from a prominent FMCG firm in Bangladesh, with a focus on production-oriented supply chain planning. This dataset includes interconnected product nodes, as well as temporal features for sales prediction, production planning, and issue identification. Experimental results establish the importance of GNN-based methods in supply chains. Researchers can utilize this dataset to leverage graph neural networks in resolving multifaceted supply chain issues, thus catalyzing advancements in supply chain analytics and planning.

Acknowledgement

References

Aamer, A.; Eka Yani, L.; and Alan Priyatna, I. 2020. Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management: An International Journal*, 14(1): 1–13.

Acemoglu, D.; and Tahbaz-Salehi, A. 2020. Firms, Failures, and Fluctuations: The Macroeconomics of Supply Chain Disruptions. Working Paper 27565, National Bureau of Economic Research.

Alves, J. C.; Silva, D. M. d.; and Mateus, G. R. 2021. Applying and comparing policy gradient methods to multi-echelon supply chains with uncertain demands and lead times. In *International Conference on Artificial Intelligence and Soft Computing*, 229–239. Springer.

Aziz, A.; Kosasih, E. E.; Griffiths, R.-R.; and Brintrup, A. 2021. Data considerations in graph representation learning for supply chain networks. *arXiv preprint arXiv:2107.10609*.

Bousqaoui, H.; Slimani, I.; and Achchab, S. 2021. Comparative analysis of short-term demand predicting models using ARIMA and deep learning. *International Journal of Electrical and Computer Engineering*, 11(4): 3319.

Cai, L.; and Ji, S. 2020. A Multi-Scale Approach for Graph Link Prediction. In *AAAI Conference on Artificial Intelligence*.

Cai, L.; Li, J.; Wang, J.; and Ji, S. 2020. Line Graph Neural Networks for Link Prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44: 5103–5113.

Cappart, Q.; Chételat, D.; Khalil, E.; Lodi, A.; Morris, C.; and Veličković, P. 2022. Combinatorial optimization and reasoning with graph neural networks. arXiv:2102.09544.

Chaovalitwongse, W. A.; and Furman, K. C. 2013. Computational optimization and logistics challenges in industrial applications. *Annals of Operations Research*, 203: 1–2.

Feizabadi, J. 2022. Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, 25(2): 119–142.

Filali, A. E.; Filali, S. E.; et al. 2021. Exploring applications of Machine Learning for supply chain management. In *2021 Third International Conference on Transportation and Smart Technologies (TST)*, 46–52. IEEE.

Higuchi, T.; and Troutt, M. D. 2004. Dynamic simulation of the supply chain for a short life cycle product—Lessons from the Tamagotchi case. *Computers & Operations Research*, 31(7): 1097–1114.

Husna, A.; Amin, S. H.; and Shah, B. 2021. Demand forecasting in supply chain management using different deep learning methods. In *Demand forecasting and order planning in supply chains and humanitarian logistics*, 140–170. IGI Global.

Kosasih, E. E.; and Brintrup, A. 2022. A machine learning approach for predicting hidden links in supply chain with graph neural networks. *International Journal of Production Research*, 60(17): 5380–5393.

Kotary, J.; Fioretto, F.; Hentenryck, P. V.; and Wilder, B. 2021. End-to-End Constrained Optimization Learning: A Survey. arXiv:2103.16378.

Li, T. 2021. Algorithm optimization of large-scale supply chain design based on FPGA and neural network. *Microprocessors and Microsystems*, 81: 103790.

Lingelbach, K.; Lingelbach, Y.; Otte, S.; Bui, M.; Künzell, T.; and Peissner, M. 2021. Demand forecasting using ensemble learning for effective scheduling of logistic orders. In *Advances in Artificial Intelligence, Software and Systems Engineering: Proceedings of the AHFE 2021 Virtual Conferences on Human Factors in Software and Systems Engineering, Artificial Intelligence and Social Computing, and Energy, July 25-29, 2021, USA*, 313–321. Springer.

Lunardi, A. R.; and Lima, F. R. 2021. Comparison of artificial neural networks learning methods to evaluate supply chain performance. *Gestão & Produção*, 28.

Mentzer, J. T.; DeWitt, W.; Keebler, J. S.; Min, S.; Nix, N. W.; Smith, C. D.; and Zacharia, Z. G. 2001. Defining supply chain management. *Journal of Business logistics*, 22(2): 1–25.

Mrad, M.; and Boujelbene, Y. 2019. An Improved Neural Approaches for Forecasting Demand in Supply Chain Management. *International Journal of Computer Applications*, 975: 8887.

Nitsche, A.-M.; Schumann, C.-A.; Franczyk, B.; and Reuther, K. 2021. Mapping supply chain collaboration research: a machine learning-based literature review. *International Journal of Logistics Research and Applications*, 1–29.

Pacella, M.; and Papadia, G. 2021. Evaluation of deep learning with long short-term memory networks for time series forecasting in supply chain management. *Procedia CIRP*, 99: 604–609.

Pirhooshyaran, M.; and Snyder, L. V. 2020. Simultaneous decision making for stochastic multi-echelon inventory optimization with deep neural networks as decision makers. *arXiv preprint arXiv:2006.05608*.

Sadeghiamirshahidi, N.; Afshar, J.; Firouzi, A. R.; and Hassan, S. 2014. Improving the efficiency of manufacturing supply chain using system dynamic simulation. *Jurnal Teknologi*, 69(2): 10–11113.

Tang, Z.; and Ge, Y. 2022. CNN model optimization and intelligent balance model for material demand forecast. *International Journal of System Assurance Engineering and Management*, 13(Suppl 3): 978–986.

Vairagade, N.; Logofatu, D.; Leon, F.; and Muharemi, F. 2019. Demand forecasting using random forest and artificial neural network for supply chain management. In *Computational Collective Intelligence: 11th International Conference, ICCCI 2019, Hendaye, France, September 4–6, 2019, Proceedings, Part I 11*, 328–339. Springer.

Younis, H.; Sundarakani, B.; and Alsharairi, M. 2022. Applications of artificial intelligence and machine learning within supply chains: systematic review and future research directions. *Journal of Modelling in Management*, 17(3): 916–940.

Zhang, M.; and Chen, Y. 2017. Weisfeiler-Lehman Neural Machine for Link Prediction. *Proceedings of the 23rd ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining.

Zhang, M.; and Chen, Y. 2018. Link prediction based on graph neural networks. *Advances in neural information processing systems*, 31.

Zhu, X.; Ninh, A.; Zhao, H.; and Liu, Z. 2021. Demand forecasting with supply-chain information and machine learning: Evidence in the pharmaceutical industry. *Production and Operations Management*, 30(9): 3231–3252.

Zougagh, N.; Charkaoui, A.; and Echchatbi, A. 2020. Prediction models of demand in supply chain. *Procedia computer science*, 177: 462–467.