ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Foundation and Generative Models for Graphs

Davide Bacciu[1], Federico Errica[2], Stefano Moro[3], Luca Pasa[3], Davide Rigoni[3✉] and Daniele Zambon[4]

1 - University of Pisa - Largo Bruno Pontecorvo 3, 56127, Pisa, Italy
2 - NEC Laboratories Europe GmbH, Germany
3 - University of Padua - Via Trieste 63, 35121, Padua - Italy
4 - The Swiss AI Lab IDSIA & Università della Svizzera italiana, Switzerland

Abstract. The rapidly evolving field of machine learning for graph-structured data gathered significant attention due to its ability to preserve critical information inherent in complex data structures. As a result, significant efforts have been dedicated to designing advanced architectures and foundational models optimized for graph-based operations. Research in this area explores methodologies for graph representation learning and graph generation, incorporating probabilistic models such as variational autoencoders and normalizing flows. Despite increasing interest from researchers as well as their efforts in solving graph-related problems, several issues and areas remain to be addressed to improve model generalization and reliability. This tutorial reviews foundational concepts and challenges in graph representation, structure learning, and graph generation, while also summarizing the contributions accepted for publication in the special session on this topic at the 33th European Symposium on Artificial Neural Networks, Computational Intelligence, and Machine Learning (ESANN).

1 Introduction

Graphs serve as powerful and versatile abstractions that model complex systems composed of interacting entities. These interactions often signify functional or structural dependencies among the entities involved, making graphs an essential tool for understanding and analyzing complex networks in diverse domains. Two key examples of graph-structured data are molecular compounds and social networks. Molecular compounds can be viewed as networks in which atoms interact, with chemical bonds influenced by interatomic distances and energies affected by electrostatic interactions. These factors are important in shaping molecular properties and behavior. Social networks, in contrast, depict interactions among users and content like images, videos, and text. In addition to these examples, graphs are vital in combinatorial optimization problems, where they can encode symmetries and constraints.

The interest in deep learning for graph-structured data is increasing, following the recent success in a broad range of practical applications. Traditional deep learning operators, however, are often inadequate to effectively address the unique challenges inherent in structured data, thus necessitating adaptation and development of new techniques, including convolutional and pooling operators, as well as optimization methods for discrete variables. Deep Graph Networks

✉Corresponding author: davide.rigoni.1@unipd.it

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

(DGNs) [1] have emerged as a leading approach in leveraging deep learning for graph-structured data. These models are designed to learn complex representations that capture the underlying structure and relationships within graphs. DGNs are particularly effective in various applications, including: *(i)* predicting attributes of social network users based on their interactions; *(ii)* identifying potential connections or relationships within a network, such as suggesting new friends in a social media platform; and *(iii)* analyzing entire graphs to predict properties, such as the toxicity of chemical compounds, akin to evaluating molecular interactions. The evolution of DGNs can be traced back to the early use of recursive neural networks on directed acyclic graphs [2, 3], evolving to handle cyclic graphs through recurrent and convolutional paradigms [4, 5], up to the development of graph foundation models [6]. Combining deep learning with probabilistic modeling [7–12], DGNs offer powerful tools to learn distributions over graph data [13–16] and design generative models able to sample from them [17–20].

In recent years, interest in generative models has led to the development of numerous approaches for generating graph structures and estimating their properties. These approaches offer new possibilities in various domains, such as social network analysis, biological networks, and infrastructure systems [21–23]. However, the task of generating new graphs comes with new challenges associated with the discrete nature of such data structure. During the generation, ensuring that the generated graphs possess the required properties, such as degree distribution and community structures, is crucial for their validity and applicability. However, driving the learning process to optimize such properties can lead to combinatorial and non-differentiable losses, often relating to the graph isomorphism problem. Moreover, the size of the graph is, in general, prescribed with the number of edges quadratic in the number of nodes. Therefore, as the size of graphs increases, scalability becomes a pressing concern, necessitating efficient algorithms to handle large graphs. Another critical challenge is the development of robust evaluation metrics to assess the quality and realism of generated graphs, as these metrics must effectively capture the nuances and similarities between synthetic and real-world graphs. Balancing the need for diversity in generated graphs with the requirement for realism further complicates the process [17, 18, 24]. Despite these challenges, generative models have shown outstanding results, especially when applied to *de novo* drug design [19, 20].

The research field of machine learning and deep learning applied to graph data not only advances technological capabilities, but also provides valuable insights across various industries, standing at the forefront of modern data analysis, bridging complex systems, and innovative deep learning techniques to address some of the most complex challenges.

2 Deep Graph Networks and Foundation Models

Deep Graph Networks (DGNs) are architectures specifically designed for learning on graph-structured data. They are composed of multiple message passing

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

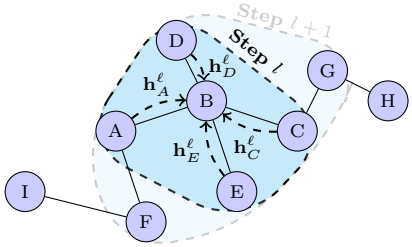

Fig. 1: Illustration of a Deep Graph Network (DGN) message-passing mechanism at step l. The highlighted subgraph (blue region) represents nodes exchanging information through propagation (dashed arrows), demonstrating how central node B aggregates features $\mathbf{h}_v^\ell$ from its neighbors $v \in \mathcal{N}_B$. At the next step, $l+1$, the information contained in node G reach the node B.

layers that iteratively update node representations by exchanging information across edges. These networks enable learning meaningful graph embeddings that can be used for node classification, link prediction, and whole-graph classification. This section introduces key background notions, describes the fundamental mechanism of graph convolutions, and discusses emerging models such as Graph Transformers [25] and foundation models.

2.1 Background Notions

A graph can be formally defined as a tuple $g = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{R})$, where $\mathcal{V}$ is the set of nodes, and $\mathcal{E}$ is the set of edges connecting them. Graphs can be either **undirected**, where edges represent bidirectional relationships, or **directed**, where edges signify asymmetric interactions. The sets $\mathcal{X}$ and $\mathcal{R}$ specify the domains of node and edge attributes, respectively. These attributes are often represented as feature vectors, though discrete edge types are also common, such as in molecular graphs where edges represent different chemical bonds.

The connectivity of a graph is often represented using an *adjacency matrix* $\mathbf{A} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$, where an entry $\mathbf{A}_{uv}$ is nonzero if an edge exists between nodes u and v. In cases where graphs are sparse (i.e., when the number of edges is significantly smaller than the number of nodes squared), adjacency lists are more efficient representations. The *neighborhood* of a node v, denoted as $\mathcal{N}_v$, consists of all nodes directly connected to v. Not rarely, temporal aspects come into play and are modeled in the form of, e.g. time series associated with the graph nodes, dynamic edges and growing graphs.

2.2 Graph Convolutions, Message Passing and Foundation Models

The fundamental operation in most DGNs is *graph convolution*, which generalizes the concept of convolution in image processing to irregular graph structures. Unlike grid-structured data such as images, graph neighborhoods are unordered

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

and variable in size. The convolution in DGNs is often formalized under the *message passing* framework, consisting of two main steps:

Message Computation: Each node v generates messages based on its current state $\mathbf{h}_v^\ell$ and the states of its neighbors $u \in \mathcal{N}_v$.

Aggregation and Update: Each node collects incoming messages and updates its representation. An illustration of this process is shown in Figure 1, where node B receives all the incoming messages from its neighbors $\mathcal{N}_B$.
This process can be written formally as:

$$\mathbf{h}_v^{\ell+1} = \phi^{\ell+1}\Big(\mathbf{h}_v^\ell,\ \Psi(\psi^{\ell+1}(\mathbf{h}_u^\ell, \mathbf{r}_{uv}) \mid u \in \mathcal{N}_v)\Big),$$

where ϕ and ψ are learnable functions (e.g., neural networks), Ψ is a permutation-invariant aggregation function such as sum or mean, and $\mathbf{r}_{uv}$, for $(u, v) \in \mathcal{E}$, are edge attributes in $\mathcal{R}$. Stacking multiple message passing layers enables nodes to incorporate information from multi-hop neighborhoods, enhancing their ability to learn expressive node embeddings. For node-level tasks, such as node classifications, predictions are finally made from the obtained node embeddings. Whereas graph-level tasks involve global pooling, where all node representations are aggregated before a prediction is made as $\mathbf{h}_g = \Psi\Big(f(\mathbf{h}_v) \mid v \in \mathcal{V}_g\Big)$, where f is another learnable function.

While message passing has several advantages – such as enabling processing graphs of different sizes, parameter sharing, keeping the computation sparse and local, and permutation equivariance – they face a few challenges. The first one is that the presented message passing has limited ability to distinguish isomorphic graphs, which relates to the concept of "expressive power". Secondly, deep stacking of message passing layers poses oversmoothing, underreaching and over-squashing issues associated with hampered information exchange among nodes.

Inspired by the success of Transformers in natural language processing, and addressing the local processing of standard DGNs, Graph Transformers have been introduced recently, which incorporate global attention mechanisms. In Graph Transformers, self-attention mechanisms are applied over nodes, allowing each node to selectively attend to important nodes across the entire graph rather than just its immediate neighbors. However, Graph Transformers also introduce higher computational complexity due to the $O(n^2)$ cost of attention computation. Hybrid models that combine local message passing with selective global attention have been proposed to mitigate this issue.

Currently, Graph Foundation Models (GFMs) [6] are emerging as a key research area, aiming to develop neural models that learn transferable graph representations applicable to diverse, unseen graphs. The usefulness of GFMs lies in their ability to perform tasks such as node classification, link prediction, knowledge graph reasoning, and algorithmic reasoning across various graphs without being limited to specific graph structures or feature dimensions. They achieve this by learning transferable representations from data leveraging large-scale pre-training on diverse graph structures. Several GFMs have been developed for diverse applications, ranging from node classification [26] and knowledge graph

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

reasoning [27] to molecular and crystal modeling [28] and 3D protein structure prediction [29]. Additionally, models such as MiniMol [30] and MolGPS [31] focus on learning molecular representations using 2D graph structures, further expanding the scope of GFMs in scientific and industrial domains.

3 Beyond Predictive DGNs: Graph Generation

Generative models are a type of algorithm designed to learn from a given training dataset and produce new data that resemble that of training. These models focus on identifying the underlying patterns and structures within the training data, allowing them to generate new samples that share characteristics with the original data. Nowadays, these models have shown great potential in diverse domains, such as the generation of novel images, videos, and structured data [21–23]. The scientific community has dedicated significant effort to advancing deep generative models for graph generation, as these models facilitate exploration and applications in areas that were previously inaccessible with decent results. Formally, let $\mathcal{G}$ denote the domain of graphs and $\mathbf{G}$ be a random variable. Generative models for graphs aim to estimate the unknown probability distribution $\mathbb{P}_{\mathcal{G}}(\mathbf{G})$ as $\widetilde{\mathbb{P}_{\mathcal{G}}(\mathbf{G})}$, which is responsible for generating the graph g in the dataset $\mathcal{D}$. Once the model has estimated the probability distribution, it can be used to generate new, unseen graphs $\widetilde{g}$ by sampling it from $\widetilde{\mathbb{P}_{\mathcal{G}}(\mathbf{G})}$, i.e. $\widetilde{g} \sim \widetilde{\mathbb{P}_{\mathcal{G}}(\mathbf{G})}$. The goal is to sample novel graph instances that were not seen during training, while preserving structures and properties similar to those in the training data.

Generative models typically consist of two key components. The first component, often based on methods like DGNs, is responsible for extracting meaningful features from a given input graph g – this is adopted especially for model training. Instead, the second component, which is adopted both in the training and generation phases, is in charge of generating a new graph $\widetilde{g}$ given some features in the input. In the learning phase, the model first extracts embeddings from the input graph g and decodes them into a structure $\widetilde{g}$ that aims to closely match the original graph. The training process is further optimized with constraints to allow the sampling of new graphs. In the generation phase, features are usually sampled from a reference probability distribution and decoded into new structures.

Numerous approaches for graph generation are documented in the literature, with many of them based on Variational Autoencoders (VAEs) [32], Generative Adversarial Networks (GANs) [24], Normalizing Flows [33], Diffusion Models [34, 35], and Score Matching [36]. Additionally, the introduction of large language models (LLMs) employing transformer architectures has broadened this field. Originally developed for natural language processing, transformer-based models have been successful in capturing complex graph patterns and have been applied also to drug generation [37, 38].

The creation of new graph-based data structures presents novel challenges. It requires addressing not only the issues associated with learning from graphs, highlighted in Section 2.2, but also the challenges involved in generating new

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

graph-structured data. One significant problem in this context is the evaluation of the generated graphs. Unlike traditional data types, graphs have complex structures and properties that need to be preserved and accurately represented in the generated outputs. Evaluating these graphs involves assessing their structural validity, diversity, and fidelity to the underlying distribution of real-world graphs. Metrics such as node and edge distribution, graph diameter, and clustering coefficients are often used, but designing comprehensive evaluation methods that capture all relevant aspects remains challenging. This makes the generation of standardized benchmarks hard especially in fields where evaluating generated graphs is particularly challenging, such as when graphs represent chemical structures. Additionally, the computational complexity involved in generating and evaluating graph data poses another layer of difficulty. As graphs grow in size and complexity, the computational resources required to generate them increase significantly, specifically scaling quadratically with the number of nodes. Algorithms must be efficiently designed to handle large-scale graphs while maintaining accuracy and performance.

A significant application of generative models for graph-structured data is the creation of chemical structures with drug-like properties. This interest arises from a pressing issue: over the years the rate of new drug discoveries [39], visible in Figure 2, has only slightly increased compared to the substantial investments made in research and development by U.S. pharmaceutical companies [40], as illustrated in Figure 3. This issue arises as the drug discovery and design process are complex and resource intensive, often extending over 10-20 years with costs exceeding $2 billion [41].

Generative models have shown great potential in tackling emerging use cases significantly advancing theoretical research in the field of deep learning. Beyond these theoretical advancements, they provide practical solutions with substantial societal benefits. One notable application is in *de novo* drug design, where their capability to efficiently explore and generate new molecular structures could transform drug discovery. This promises to speed up the development of new medications, reduce costs, and improve accessibility, ultimately enhancing public health outcomes. Nonetheless, challenges persist, particularly in maintaining the quality and reliability of the generated data.

4 Special Session's Contributions

This year's special session comprises several original contributions ranging on a set of diverse topics related to the graph representation learning field.

3-WL GNNs for Metric Learning on Graphs[1]. This study proposes a metric learning framework for graphs exploiting 3-WL GNNs. The framework consists of a Siamese GNN-based embedding block for generating graph and node-level embeddings and a metric block for distance computation. Results show that 3-WL expressivity when paired with a node-level distance strategy helps improve performance in metric learning tasks.

[1] Aldo Moscatelli, Maxime Bérar, Pierre Héroux, Florian Yger and Sébastien Adam

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

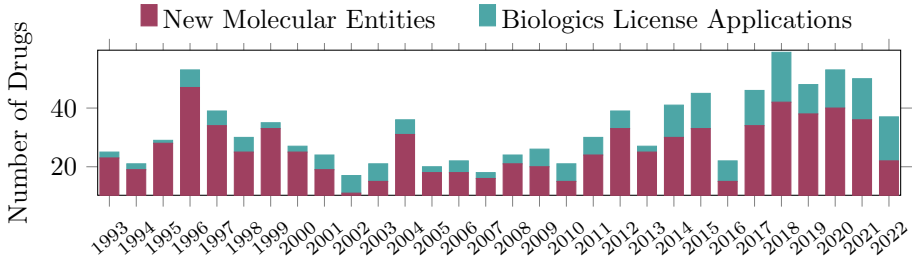

Fig. 2: Number of FDA-approved drugs per year. Products like vaccines and gene therapies are not included in this drug count.

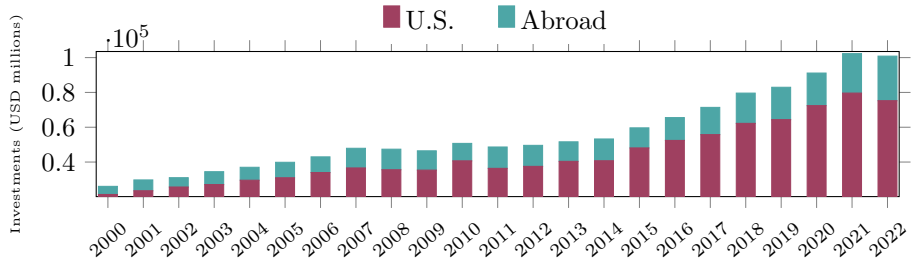

Fig. 3: Spending of the U.S. pharmaceutical industry, by PhRMA member companies, on research and development at home and abroad.

D4: Distance Diffusion for a Truly Equivariant Molecular Design[2]. This article introduces a novel Distance and Discrete Denoising Diffusion Model, called D4, which builds upon MiDi [42] to enhance the chemical validity of molecular generation. D4 uses atomic distances as a proxy for spatial information, removing the need for SE(3) equivariant architectures and thereby simplifying the overall design. The model generates molecular structures by reversing a noise process and predicts the adjacency matrix to determine bond presence. Results show that D4 matches MiDi's performance in molecule topology generation while outperforming it in spatial arrangement prediction.

Encoding Graph Topology with Randomized Ising Models[3]. The paper proposes a reservoir computing model for encoding the topological information of graph nodes, designed for potential physical implementation using neuromorphic hardware. The model leverages the spin configuration of a Lenz-Ising system to represent node embeddings. The study demonstrates that this approach effectively computes node embeddings that encode topological information, achieving node classification accuracy comparable to Graph Echo State Networks (GESN). Competitive classification results are particularly evi-

[2]Samuel Cognolato, Davide Rigoni, Marco Ballarini, Luciano Serafini, Stefano Moro and Alessandro Sperduti

[3]Domenico Tortorella, Antonio Brau and Alessio Micheli

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

dent when scaled to a large number of spin units, highlighting its potential for neuromorphic implementation.

Towards Efficient Molecular Property Optimization with Graph Energy Based Models[4]. In this article the authors propose a novel approach for optimizing chemical properties with the Unsupervised Energy-based Molecule Optimization (UEMO) model. UEMO implicitly learns chemical properties and generates molecular graphs without requiring labeled data or explicit optimization strategies. The model utilizes DGNs to process molecular graphs and Langevin dynamics for sampling molecules. The experimental results show solid performance in optimizing chemical properties, outperforming existing models, and showing improved sampling efficiency due to its lightweight architecture.

Robustness in Protein-Protein Interaction Networks: A Link Prediction Approach[5]. In this study the authors propose a method to predict the robustness of protein-protein interaction networks (PPINs) by framing it as a link prediction task. The approach bridges the gap between dynamic property inference and static network analysis, enabling the exploration of long-range functional dependencies in PPINs. The study demonstrates that a DGN can accurately predict robustness for protein pairs by leveraging PPIN topology and protein sequence embeddings.

5 Conclusions

With this tutorial and the corresponding ESANN 2025 special session, we aim to give an overview of the dynamic field of machine learning and deep learning applied to graph-structured data, focusing on foundational models like Deep Graph Networks and Graph Transformers, and their diverse applications ranging from social networks to molecular analysis. In addition, this special session aims to foster the discussion on generative models for the graph field, highlighting challenges such as scalability and evaluation complexity, especially in domains like *de novo* drug design. To conclude, this article also highlighted contributions from this year's special session including advances in metric learning, diffusion models, and protein interaction networks. Future research should enhance model expressiveness and scalability for large graphs and datasets, while focusing on establishing standardized evaluation benchmarks for generative models, which, when integrated into real-world scenarios, promise significant societal benefits, especially in healthcare and chemistry.

References

[1] Davide Bacciu, Federico Errica, Alessio Micheli, and Marco Podda. A gentle introduction to deep learning for graphs. *Neural Networks*, 129:203–221, 2020.

[2] Alessandro Sperduti and Antonina Starita. Supervised neural networks for the classification of structures. *IEEE transactions on neural networks*, 8(3):714–735, 1997.

[3] Paolo Frasconi, Marco Gori, and Alessandro Sperduti. A general framework for adaptive processing of data structures. *IEEE transactions on Neural Networks*, 9(5):768–786, 1998.

[4] Luca Miglior, Lorenzo Simone, Marco Podda and Davide Bacciu

[5] Alessandro Dipalma, Domenico Tortorella and Alessio Micheli

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

[4] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

[5] Alessio Micheli. Neural network for graphs: A contextual constructive approach. *IEEE Transactions on Neural Networks*, 20(3):498–511, 2009.

[6] Haitao Mao, Zhikai Chen, Wenzhuo Tang, Jianan Zhao, Yao Ma, Tong Zhao, Neil Shah, Mikhail Galkin, and Jiliang Tang. Position: Graph foundation models are already here. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, 2024.

[7] Claudio Gallicchio and Alessio Micheli. Graph echo state networks. In *The 2010 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2010.

[8] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2022.

[9] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[10] Luca Pasa, Nicolò Navarin, and Alessandro Sperduti. Multiresolution reservoir graph neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6):2642–2653, 2021.

[11] Davide Bacciu, Federico Errica, and Alessio Micheli. Probabilistic learning on graphs via contextual architectures. *Journal of Machine Learning Research*, 21(134):1–39, 2020.

[12] Federico Errica, Davide Bacciu, and Alessio Micheli. Graph mixture density networks. In *International Conference on Machine Learning*, pages 3025–3035. PMLR, 2021.

[13] Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. Neural relational inference for interacting systems. In *International conference on machine learning*, pages 2688–2697. PMLR, 2018.

[14] Pantelis Elinas, Edwin V Bonilla, and Louis Tiao. Variational inference for graph convolutional networks in the absence of graph data and adversarial settings. *Advances in neural information processing systems*, 33:18648–18660, 2020.

[15] Daniele Zambon, Lorenzo Livi, and Cesare Alippi. Graph iForest: Isolation of anomalous and outlier graphs. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2022.

[16] Andrea Cini, Daniele Zambon, and Cesare Alippi. Sparse Graph Learning from Spatiotemporal Time Series. *Journal of Machine Learning Research*, 24(242):1–36, 2023.

[17] Davide Rigoni, Nicolò Navarin, and Alessandro Sperduti. Conditional constrained graph variational autoencoders for molecule design. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 729–736. IEEE, 2020.

[18] Davide Rigoni, Navarin Nicolo, and Sperduti Alessandro. A systematic assessment of deep learning models for molecule generation. In *ESANN 2020-Proceedings, 28th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 547–552, 2020.

[19] Davide Rigoni, Sachithra Yaddehige, Nicoletta Bianchi, Alessandro Sperduti, Stefano Moro, and Cristian Taccioli. Tumflow: An ai model for predicting new anticancer molecules. *International Journal of Molecular Sciences*, 25(11):6186, 2024.

[20] Davide Rigoni, Nicolò Navarin, and Alessandro Sperduti. Rgcvae: Relational graph conditioned variational autoencoder for molecule design. *Machine Learning*, 114(2):47, 2025.

[21] Pedro Celard, Eva Lorenzo Iglesias, José Manuel Sorribes-Fdez, Rubén Romero, A Seara Vieira, and Lourdes Borrajo. A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. *Neural Computing and Applications*, 35(3):2291–2323, 2023.

[22] Masahiro Suzuki and Yutaka Matsuo. A survey of multimodal deep generative models. *Advanced Robotics*, 36(5-6):261–278, 2022.

[23] Xiaolin Xia, Jianxing Hu, Yanxing Wang, Liangren Zhang, and Zhenming Liu. Graph-based generative models for de novo drug design. *Drug Discovery Today: Technologies*, 32:45–53, 2019.

[24] Nicola De Cao and Thomas Kipf. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018.

[25] Erxue Min, Runfa Chen, Yatao Bian, Tingyang Xu, Kangfei Zhao, Wenbing Huang, Peilin

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Zhao, Junzhou Huang, Sophia Ananiadou, and Yu Rong. Transformer for graphs: An overview from architecture perspective. *arXiv preprint arXiv:2202.08455*, 2022.

[26] Jianan Zhao, Mikhail Galkin, Hesham Mostafa, Michael M Bronstein, Zhaocheng Zhu, and Jian Tang. Fully-inductive node classification on arbitrary graphs. In *Adaptive Foundation Models: Evolving AI for Personalized and Efficient Learning*.

[27] Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards foundation models for knowledge graph reasoning. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*, 2024.

[28] Duo Zhang, Xinzijian Liu, Xiangyu Zhang, Chengqian Zhang, Chun Cai, Hangrui Bi, Yiming Du, Xuejian Qin, Anyang Peng, Jiameng Huang, et al. Dpa-2: a large atomic model as a multi-task learner. *npj Computational Materials*, 10(1):293, 2024.

[29] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Sal Candido, et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv*, 2022:500902, 2022.

[30] Kerstin Kläser, Blazej Banaszewski, Samuel Maddrell-Mander, Callum McLean, Luis Müller, Ali Parviz, Shenyang Huang, and Andrew W. Fitzgibbon. Minimol: A parameter-efficient foundation model for molecular learning. *CoRR*, abs/2404.14986, 2024.

[31] Maciej Sypetkowski, Frederik Wenkel, Farimah Poursafaei, Nia Dickson, Karush Suri, Philip Fradkin, and Dominique Beaini. On the scalability of gnns for molecular graphs. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024.

[32] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[33] Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, and Jian Tang. Graphaf: a flow-based autoregressive model for molecular graph generation. *arXiv preprint arXiv:2001.09382*, 2020.

[34] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.

[35] Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022.

[36] Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International conference on machine learning*, pages 10362–10383. PMLR, 2022.

[37] Eyal Mazuz, Guy Shtar, Bracha Shapira, and Lior Rokach. Molecule generation using transformers and policy gradient reinforcement learning. *Scientific Reports*, 13(1):8799, 2023.

[38] Viraj Bagal, Rishal Aggarwal, PK Vinod, and U Deva Priyakumar. Molgpt: molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling*, 62(9):2064–2076, 2021.

[39] Asher Mullard. 2021 fda approvals. *Nature reviews. Drug discovery*, 2022.

[40] Statista. Spending of the u.s. pharmaceutical industry on research and development at home and abroad from 1990 to 2022 (in million u.s. dollars). Statista, August 21 2023.

[41] Chengxi Zang and Fei Wang. Moflow: An invertible flow model for generating molecular graphs. In Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash, editors, *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 617–626. ACM, 2020.

[42] Clement Vignac, Nagham Osman, Laura Toni, and Pascal Frossard. Midi: Mixed graph and 3d denoising diffusion for molecule generation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 560–576. Springer, 2023.