Departement of Information Systems
Hochschule Heilbronn

Master Thesis

Towards Quantum Graph Neural Networks for Molecular Property Prediction

submitted by

Tobias Roth

March 5, 2024

First Supervisor	Prof. Dr.-Ing. Carsten Lanquillon
Second Supervisor	Prof. Dr.-Ing. Jochen Günther
Advisor	Philipp Wagner, M. Sc.

Eidesstattliche Erklärung

Hiermit erkläre ich eidesstattlich,

1. dass die vorliegende Arbeit von mir selbstständig und ohne unerlaubte Hilfe angefertigt wurde,
2. dass ich alle Stellen (einschließlich Abbildungen und Tabellen), die wörtlich oder annähernd wörtlich oder dem Gedanken nach aus Veröffentlichungen und unveröffentlichten Unterlagen und Gesprächen entnommen worden sind, als solche an den entsprechenden Stellen innerhalb der Arbeit durch Zitate bzw. Verweise kenntlich und nachvollziehbar gemacht habe, einschließlich der Kenntlichmachung KI- generierter Texte oder Abbildungen,
3. die Arbeit noch bei keiner anderen Prüfung vorgelegt wurde.

Zum Einsatz von KI-basierten Instrumenten (wie z.B. ChatGPT) erkläre ich, dass für die vorliegende Arbeit

- ☒ kein Einsatz erfolgte
- ☐ ein Einsatz in folgender Form erfolgte:

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird. Insbesondere ist mir bewusst, dass eine Note auch nach der Veröffentlichung angepasst werden kann, wenn die Täuschung erst zu einem späteren Zeitpunkt bekannt geworden ist, z.B. durch fortgeschrittene Prüfprogramme. Die nachträgliche Anpassung kann auch noch nach Abschluss des Studiums erfolgen. In diesem Fall wird das unrichtige Zeugnis (inkl. Urkunde und Diploma Supplement) eingezogen.

Ort, Datum	*Name, Vorname*	*Unterschrift*

Abstract

Contents

List of Abbreviations

CRISP-DM	Cross Industry Standard Process for Data Mining
GAE	Graph Auto Encoder
GAN	Graph Attention Network
GCN	Graph Convolutional Neural Network
GNN	Graph Neural Network
GIN	Graph Isomorphism Networks
GraphSAGE	Graph Sample and Aggregated
GraphSNN	Graph Symmetric Neural Network
MPNN	Message Passing Neural Network
PES	Potential Energy Surface
RecGNN	Recurrent Graph Neural Network
STGNN	Spatial-temporal Graph Neural Networks

List of Figures

List of Tables

1 Global Introduction

1.1 Problem Definition

In view of the steadily advancing climate change, efforts to reduce environmental pollution are increasingly coming into focus both in the public eye and within the scientific community [1]. This includes the search for scalable and cost-effective renewable energy storage solutions, which is essential to meet the world's growing energy demand while mitigating climate change [2]. The conversion of electricity to hydrogen, as well as the reverse combustion process, can play an important role here [1]. Therefore, new materials are constantly being investigated to enable catalytic processes in the field of hydrogen production to run efficiently [3]. Machine learning methods are already being used to simulate and calculate catalytic properties. In particular, graph neural networks (GNN) are proving to be especially promising here [4]. Since the prediction of potential areas and other relevant properties proceeds at the molecular and atomic level, the use of quantum computers is also being investigated. In the literature, first approaches to realize GNNs on quantum computers already exist [5–7].

1.2 Aim of this work

This master's thesis investigates the extent to which quantum GNNs are suitable for predicting molecular properties. To achieve this, the first step is to acquire an understanding of GNNs and the foundations of quantum computing, which will be discussed in this work. Subsequently, a classical GNN will be implemented to predict molecular properties using available data. Later in this thesis, a quantum GNN will be designed and developed that uses the same input data as the classical GNN. The quantum GNN will be tested and evaluated on a quantum computer. Having a classical and a quantum GNN, a comparison of performance between these models will take place.

1.3 Global Research Questions

In order to solve the presented problem, the following research question and sub-questions are created:

How can classical and quantum graph neural networks be used for the prediction of molecular properties?

Q1: How do GNNs and quantum GNNs work?
Q2: How must molecular data be processed to be used with a quantum computer?
Q3: Are quantum GNNs suitable for predicting molecular properties in the context of electrocatalysis?
Q4: How does the performance compare to classical GNNs?

1.4 Global Research Methodology

For the scientific design of the research process, this work will follow the established procedure according to Peffers et al. [8], as illustrated in Figure 1.1. Based on this, the research process is structured into different steps, which are explained below.

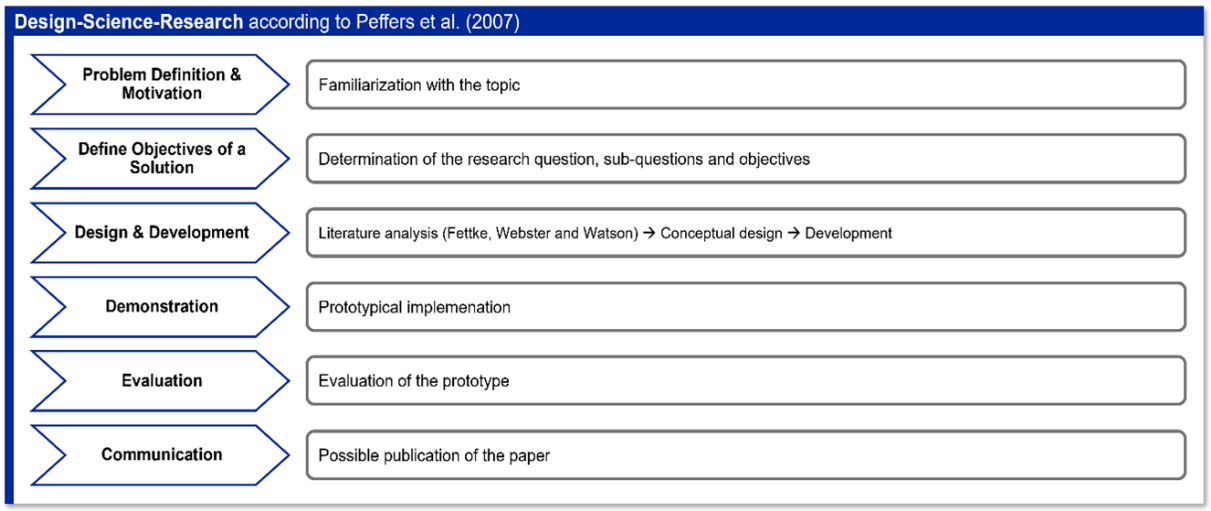


Figure 1.1: Overview of the general research methodology [8]

Previously, the problem upon which this work is based on and the objectives to be achieved within this work were defined in cooperation with the Frauenhofer IPA.

In the next steps, a comprehensive and structured literature review will be conducted with the aim of identifying the necessary foundations for GNNs, quantum computing, and the relevant molecular properties. The literature review will follow the methods of Fettke [9], Webster and Watson [10], as well as vom Brocke et al. [11]. Relevant literature databases and other elements will be defined at the time of conducting the literature review. The results of the literature review form the first part of this work and will be documented in the form of a state-of-the-art paper. Afterwards a classical GNN will be designed and developed to predict molecular properties using the existing data set. This serves to establish a basic understanding of GNNs and the dataset as well as possible

results of the prediction. Furthermore, the classical GNN will be utilized to compare its results and performance with the quantum GNN, including various performance metrics such as accuracy. After the successful implementation of the classical GNN, a quantum GNN will be designed and developed, which is the core focus of this thesis. The development of a quantum GNN involves using different tools and frameworks compared to a classical GNN, including the preprocessing of the available dataset. This prototype will undergo testing, evaluation, and potential optimization using a quantum computer. Specific requirements for the GNN and potential evaluation criteria will be defined as part of the research process. This can be achieved through the conducted literature review or other methods, such as expert interviews. With both classical and quantum GNNs in place, an extensive comparison between these models will be conducted. The results of this procedure constitute the second part of this thesis. The entire process and its findings will be documented.

2 Architectures of Graph Neural Networks

Abstract

As deep learning algorithms increasingly adress graph-structured data, Graph Neural Networks (GNNs) play a more and more important role in various fields, including computational chemistry and the prediction of molecular properties. However, currently in the literature different architectures and approaches of GNNs exist. Therefore, this paper analyzes the current state-of-art by exploring architectures and approaches of GNNs in the literature. By creating an overview, it will show that convolutional GNNs are commonly used and particulary reliable when working with molecular properties. With the help of structured process CRISP-DM, later a convolutional GNN will be developed that predicts the potential energy surface of molecules with the QM9 dataset and a simpler water dataset. The evaluation shows high accuracy, therefore the GNN was succesfully developed.

Keywords: *graph neural networks, architectures, literature review, potential energy surface prediction*

2.1 Introduction

While deep learning algorithms effectively capture hidden patterns of Euclidean data, there is an increasing number of application areas where the data is represented in the form of graphs or structures similar to graphs [12]. Due to the expressive capabilities of graphs, researches on analyzing these kind of structures with machine learning have been receiving more and more attention. This is evident in the areas of social science, such as social networks, natural science like physical systems, knowledge graphs and other research areas [13]. Deep learning based methods that operate in a graph domains are called Graph Neural Networks (GNNs) [14].
In computational chemistry, molecules are modeled as graphs enabling various experiments [12]. This includes the prediction of potential energy surfaces for molecules made

of hydrogen compounds or other materials [15]. In the current literature, different approaches or architectures for GNNs exist [12]. This raises the question of which approaches are available and which of these approaches is well suited for the prediction of molecular properties such as potential energy surfaces.

In order to solve the presented problem, the following research question and sub-questions are created:

"How can a Graph Neural Network for the prediction of molecular properties be developed?"

Q1: *What approaches for Graph Neural Networks currently exist in the Literature?*
Q2: *What architecture suits the prediction of molecular properties best?*

The goal of this paper is to develop an understanding of graph neural networks in general and the different types of architectures that exist in the literature. Besides, in the scope of this thesis, an understanding of the prediction of potential energy is created. Based on this procedure, the literature review will show what architecture of graph neural networks suits the prediction of potential energy best. In summary, to answer the developed research question and sub-questions, the following artifacts will be created as part of the research:

- literature review of different graph neural network architectures
- prototypical implementation of a graph neural network for the potential energy prediction

The structure of this paper is as followed: first, the scientific foundations and theoretical background are discussed. Then the research methodology is presented in detail. Based on this, the findings and their results are presented and then discussed. Finally, a conclusion and discussion of the results is provided, as well as an overview of possible research topics based on this work.

2.2 Theoretical Background

2.2.1 Graph Neural Networks

Graph Neural Networks (GNNs) are a type of neural networks that are used to work with graph-structured data. They provide a framework for tasks that focus on learning dependencies and interactions between data points. GNNs use the structure of graphs to capture and process information in way that is not possible using traditional neural networks. In the context of GNNs a graph consists of nodes that represent entities and

edges that describe relationships between those entities. The objective of GNNs is to learn important node representations that contain both local and global graph information. [12] GNNs are effective in dealing with non-Euclidean and irregular data structures. This makes GNNs very suitable for domains where relationships between data points are as important as, or even more important then, the individual data points itself [13].

The architecture of a GNN consists of layers that update node representations by aggregating information from neighboring nodes iteratively. This process enables nodes to relearn their understanding of the local context while considering the wider structure of the graph. The aggregation step involves weighted summation or concatenation of features from neighboring nodes, allowing each node to summarize information from its surroundings. [14, 16] A key aspect of a GNN is the neighborhood aggregation function, which is also known as the message-passing mechanism [17]. This function defines how information between nodes in each layer is exchanged. The power of GNNs lies in their ability to progressively enhance node representations through multiple layers, capturing detailed dependencies within the graph.

2.2.2 Potential Energy Surface Prediction

The potential energy surface (PES), in the context of molecules and atoms, describes the energy of a molecule in regard to certain parameters, e.g. to the position of its atoms. Predicting the potential energy surface of atoms is a task in computational chemistry and materials science. It gives information about the stability and properties of molecules and materials [15]. This is helpful when to decide whether materials can be used to enable efficient catalytic processes in the field of hydrogen production [3].

2.3 Research Methodology

First, the given problem of the prediction of molecular properties with GNNs was investigated. For this purpose, an initial investigation of the literature took place and the topic was explored. This is followed by the definitions of the research questions as well as the sub-questions and the delineation of the research topic.

Afterwards, a comprehensive and structured literature review is conducted with the aim of identifying different approaches of GNNs and finding the most suitable for molecular property prediction. The findings of this literature review are used for the creation of a common understanding of GNNs in the given context. In the next step, a GNN is prototypically implemented with the available dataset. This serves the goal of being able to directly test and evaluate the developed GNN. For better structuring and traceability of

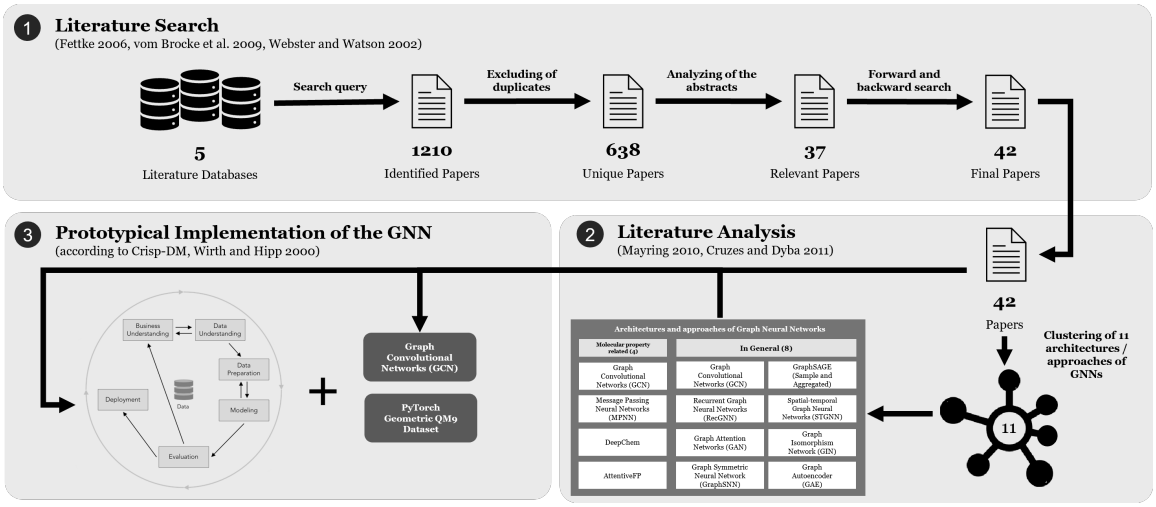


Figure 2.1: Overview of the conducted research process

the procedure, the exact research process is shown in Figure 2.1, which is also described in detail in the following section.

In the **first step**, the literature search based on Fettke [9], vom Brocke et al. [11] and Webster and Watson [10] was performed. The following literature databases were used: IEEE Xplore Digital Library, ScienceDirect, SpringerLink, Emeralt Insight and Google Scholar. Electronic searches of titles using the search terms [("Graph Neural Networks") AND (("architectures") OR ("approaches"))] and [("Graph Neural Networks") AND ("*molecular*")]. Moreover, no specific period of time was selected. According to the search query performed, these searches resulted in a total of 1210 publications. After analyzing abstract and keywords, 37 relevant publications remained for the given research focus. In addition to the database search, Webster and Watson [10] recommend performing a forward and backward search. These were performed in a final step and increased the number of final publications to 42.

In the **second step**, a total number of 42 papers were read in full and coded. For the clustering of the architectures and approaches of GNNs the procedure of Fettke [9] was performed. After this process of clustering, a total of 11 different architectures and approaches for GNNs could be identified. According to the research topic, this results are assigned into two different higher-level categories, "molecular property related" and "In General".

The **third step** covers the prototypical implementation of the GNN for molecular property prediction. Based on the findings of the literature review, a GNN is implemented and evaluated that predicts potential energy surfaces with the QM9 Dataset from Pytorch Geometric [18]. The implementation of the GNN follows the established procedure

of CRISP-DM [19] and is explained later in this paper.

2.4 Findings

In this section, the results and findings are presented and discussed in detail. Especially the literature review and the prototypical implementation of the GNN for the prediction of potential energy will be considered.

2.4.1 Literature Review

After the conduction of the literature review, two categories of GNN architectures and approaches could be identified, namely "molecular property related" and "In General". The first category lists all GNN architectures that are applied in computational chemistry, especially in working with molecular properties. The second category "In General" describes general GNN architectures that occur in literature and also for other application areas. For both categories, different architectures and approaches were found, which are shown in Figure 2.2. In the following, the findings will be explained in more detail.

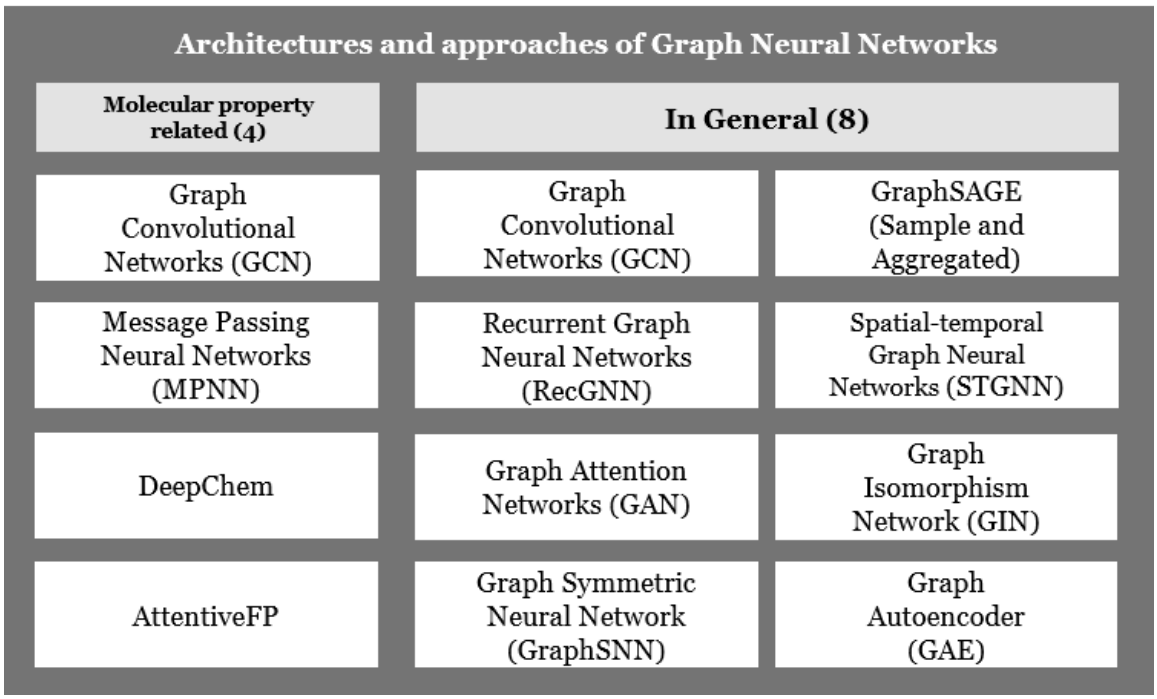


Figure 2.2: Overview over die identified architectures and approaches for GNNs

For the processing of data related to molecular property or computational chemistry, in the literature for different types of GNN approaches could be identified: Graph Convolutional Networks (GCN), Message Passing Neural Networks (MPNN), DeepChem and AttentiveFP.

Proposed by Gilmer et al. in 2020 [20], Message Passing Neural Networks (MPNNs) are a class of graph neural networks designed for processing graph-structured data, targeting the application areas of chemistry and molecular property prediction. The architecture of MPNNs consists of message passing, updating node and edge representations iteratively through message passing functions, enabling the model to capture relationships reliable within the graph. The key components include message passing, aggregation, and readout. During message passing, information from neighboring nodes and edges is aggregated and transformed through learnable functions. Aggregation mechanisms, often involving summation or attention, allow nodes to incorporate information from their local environments. A readout function then processes the final node representations to make predictions about the entire graph. This design enables MPNNs to effectively handle variable-sized graphs and adapt to diverse molecular structures.

Unlike the other findings, DeepChem is no GNN itself, but a Python library for machine learning and deep learning on molecular and quantum datasets. The library consists of a extensive toolkit for working with molecular property prediction and related tasks [21]. For example, DeepChem offers different deep learning models, including graph convolutional networks, graph attention networks, and recurrent neural networks. DeepChem is used for research in the field of computational chemistry and drug discovery [22].

AttentiveFP is a deep learning model for graph data, which integrates attention mechanisms into GNNs. In the area of computational chemistry, AttentiveFP captures and values the importance of specific atoms in a molecule. This model gives attention weights to highlight relevant atoms during information aggregation. In this way, the model is able to focus on specific regions within a molecular graph. The architecture of AttentiveFP involves message-passing steps where atoms exchange information based on their local environments. The attention mechanism assigns an importance scores to neighboring atoms, enabling the model to prioritize interactions between them. The focus on local context is particularly beneficial when handling various molecular structures. The attention-based approach enhances the capacity to recognize subtle chemical patterns, making it useful for applications in computational chemistry. According to its architecture, AttentiveFP is a specific Graph Attention Network (GAN). [23, 24]

The literature review showed that Convolutional Graph Neural Networks (GCNs) are both used for working in the field of molecular property prediction and in other application areas. Introduced by Kipf and Welling in 2016 [25] GCNs are one of the earliest and commonly used architectures of GNNs [12].
GCNs enable the modeling of molecular structures by learning node representations through a certain amount of graph convolutional layers. These layers aggregate infor-

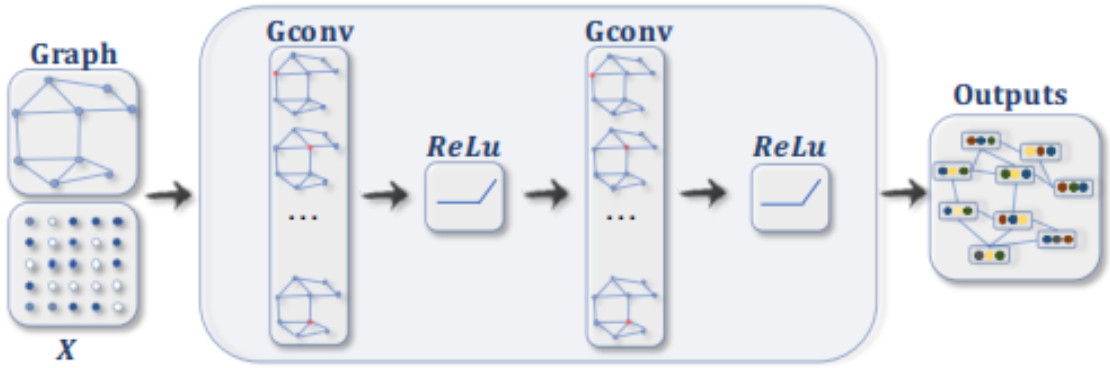


Figure 2.3: Examplary GCN architecture for node classification [12]

mation from neighboring atoms, allowing the model to learn relationships within graph data. The main point of the functionality of a GCN is the ability to weight these aggregations, highlighting more important atoms and interactions. [26] Figure 2.3 shows an examplary architecture for a GCN: graph data is used as input, the GCN consists of two convolutional layers where ReLu is used as activation function and the output is a predicted property of the input data.
The architecture of GCNs involves iterative information propagation through the graph. Each layer computes a weighted sum of neighboring node features, producing node representations. The weight parameters are learned during the training of the model, allowing the GCN to adapt to the characteristics of the respective graph dataset. GCNs are especially effective when working with molecular structures, because the hierarchical structure of a GCN enables the learning of global information, which is required for having a holistic understanding of molecular structures and perform tasks effectively. [12, 26]

For the general use of GNNs and other application fields, a total of eight different approaches of GNNs could be identified: the previously explained GCNs, GraphSAGE (Sample and Aggregated), Recurrent Graph Neural Networks (RecGNN), Spatial-temporal Graph Neural Networks (STGNN), Graph Attention Networks (GAN), Graph Isomorphism Networks (GIN), Graph Symmetric Neural Network (GraphSNN), and Graph Autoencoder (GAE). While AttentiveFP is a specific GAN, the structure and functionality of GANs is similar to that of AttentiveFP described above [23]. For reasons of focus, only the RecGNNs and GAEs are discussed below.

Recurrent Graph Neural Networks (RecGNNs) extend the possibilities of traditional GNNs by using both graph-based and recurrent components. Thereby, temporal information and sequential patterns within graphs can be captured. At each time step, the model updates node representations using graph convolutional layers and captures the structure of the evolving graph. Simultaneously, recurrent units capture a memory of past states,

which enables the model to save information for a certain time step. The recurrent nature of RecGNNs allows them to learn and represent evolving graph structures. This is especially helpful for a dataset where the relationships between the data show temporal dynamics, e.g. in time-sensitive domains. [27, 28]

Designed for learning representations of graph-structured data, Graph Autoencoders are neural network models that are a tool especially helpful for the reconstruction of nodes and graphs. GAEs encode graph-structured input into low-dimensional vectors. By doing so, the essential structural features are captured while the graph typology is preserved. This approach is to be classified as unsupervised learning and supports GAEs in applications like link prediction or anomaly detection. The architecture of GAEs consist of an encoder and a decoder. The encoder maps nodes or graphs to lower-dimensional representations, while the decoder reconstructs the original input from the previously created embeddings. Similar to other variants of GNNs like the RecGNNs, GAEs leverage graph convolutional layers to aggregate information from neighboring nodes during encoding, facilitating the learning of meaningful representations. [29, 30]

2.4.2 Development of the GNN

After gaining insight into different architectures and approaches of GNNs in the literature, in the following the development of the GNN for the prediction of potential energy surfaces will be explained. In the course of this paper, a GCN is developed and used with the QM9 Dataset [18]. In addition, the developed model was trained a second time using a data set containing only water molecules. This will be used to compare the performance of different data sets on the models. The development of the GCN was based on the established CRISP-DM process [19], which is explained step by step below.

In the initial Phase, the Business Understanding, the goal is to understand the research objectives and requirements. In view of the advancing climate change, the reduction environmental pollution plays an important role for humankind [1]. This includes the search for scalable and cost-effective renewable energy storage solutions, e.g. like the conversion of electricity to hydrogen as well as the reverse combustion process [2]. For this purpose, new materials are being investigated, to enable efficient catalytic processes in the field of hydrogen production [3]. Machine Learning Methods, like GNNs, are used to simulate and calculate catalytic properties like the prediction of potential energy surfaces [4]. Therefore, the goal of the developed GNN in this paper is to predict the potential energy surfaces of different molecules.

The focus of the second step, namely Data Understanding, is to gain insight into the given

dataset. The QM9 Dataset consists of approximately 130,000 molecules with 19 regression targets [18]. Within all the different molecules, there are five potential atom types like Carbon and Oxygen and four different types of chemical bonds between them (single bond, double bond, etc.) [31]. For the usage of the GNN, the molecules are transformed, whereas the atoms form the nodes of the graph and the edges are described through the chemical bonds between the atoms. The relevant regression target for the prediction of potential energy surface is the seventh target of the QM9 Dataset, the Internal energy at 0K. The water dataset contains only water molecules with oxygen and hydrogen atoms. In summary, there are approximately 36,000 water molecules in this dataset with different atomic positions or coordinates. Therefore, compared to the QM9 dataset, the water dataset is not as complex in terms of the number of different atoms and chemical bonds and also in the general size of the dataset. The water data set is also used to predict the potential energy surface.

In the third step, the available data will be prepared to be able to train the developed GCN. The data preparation includes at first the extraction of all relevant information from the dataset, which is in this case the atoms, the cartesian coordinates, and the Internal energy at 0K. For better results of with the model prediction, the target variables are normalized with the help of their mean and standard deviation. Another important step in the data preparation is the conversion of the coordinates into inverse pairwise distance, so that the GCN understands the spatial relationships like translation and rotation invariance between the atoms in the molecular structure. The same procedure goes for the water dataset.

After the data preparation, the GNN architecture is implemented during the modeling phase of the CRISP-DM process. The GNN consists of two convolutional layers, with the before prepared nodes, edges and features as well as the initialized weights as input parameters. The activation function is LeakyRelu. The neural network is implemented with the PyTorch Python package. Before the training, the dataset is randomly split into training and validation dataset, for the evaluation of the model in the next step.

The fifth phase describes the evaluation of the developed model, where the performance of the model is assessed with the validation dataset. With the help of the Adam optimizer algorithm, also implemented in PyTorch, the model is optimized during training. With this procedure, the ability of the model to generalize and make accurate prediction on data other than the training data is enhanced. Since the prediction of the potential energy surfaces is a regression task, the performance of the model can be measured by using different regression performance metrics. The following performance metrics are used: Model Training Time (lower is better), Mean Squared Error (lower is better), Root Mean

Squared Error (lower is better) and the R^2-Score (higher is better). In the following, the performance metrics are evaluated by comparing the performance of a baseline (non-graph-based) neural network architecture with the implemented GCN architecture, where the different approaches are trained individually for each dataset. Table 2.1 lists the computed metrics for the training with the QM9 dataset.

Parameter	**Baseline Model**	**Graph-based Model**
Training time	6 min 17.1s	9 min 2.4s
Mean Squared Error (MSE)	365484.59	11513.64
Root Mean Squared Error (RMSE)	604.553	107.30
R^2-Score (R^2)	0.6904	0.9902

Table 2.1: Comparison of performance metrics of the developed model [QM-9 dataset]

As shown in the table, training the baseline model with similar hyperparameters to the GCN takes about 3 minutes less than training the GCN. However, the measurement of MSE and RMSE shows that the results of the GCN are much more accurate than the results of the baseline model, e.g. the RMSE of the GCN is about 6 times smaller than the RMSE of the baseline architecture. Comparing the overall accuracy using the R^2 score shows that the baseline model has an accuracy of 69% while the GCN has an accuracy of 99%, indicating a very accurate fit between the predictions and the actual data. The difference in accuracy between the two models can be seen by comparing the figures 3.1a and 3.1b.

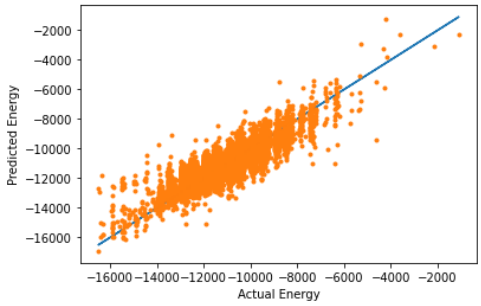
The following table 2.2 compares the performance metrics of the baseline and GCN models trained on the water dataset. It can be seen that there is not much difference in the training time of the different models, while interestingly the GCN trains faster than the baseline model with the same hyperparameters. However, the baseline neural network provides more accurate results, with the MSE and RMSE values being smaller than those of the GCN. Also, the accuracy of the baseline model is 99.1%, while the accuracy of the GCN is only 97.6%.

Parameter	**Baseline Model**	**Graph-based Model**
Training time	1 min 38.3s	1 min 26.9s
Mean Squared Error (MSE)	0.0000612	0.0014
Root Mean Squared Error (RMSE)	0.0078	0.0373
R^2-Score (R^2)	0.9991	0.9756

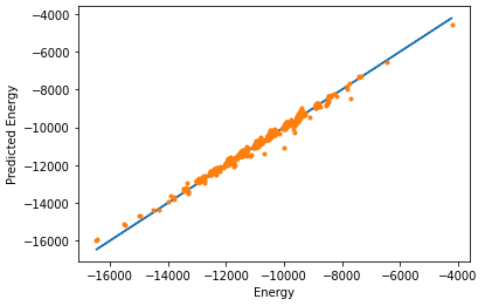
Table 2.2: Comparison of performance metrics of the developed model [water dataset]

While the training of the models with the QM9 dataset shows very different results, the

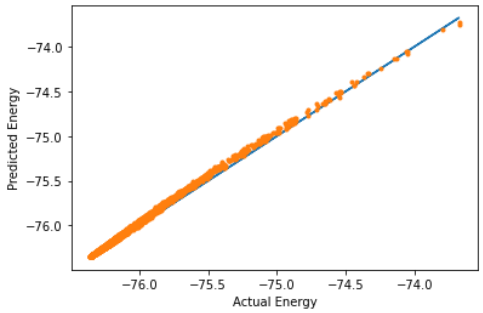
figures of 2.4c and 2.4d show that there is only a small difference in terms of accuracy when training a baseline and a GCN model with the simpler water dataset.



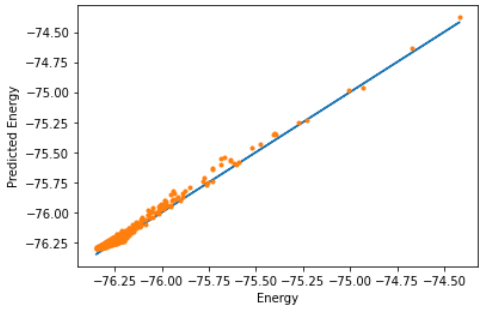
(a) QM9 dataset classical neural network approach



(b) QM9 dataset graph-based neural network approach



(c) Water dataset classical neural network approach



(d) Water dataset graph-based neural network approach

Figure 2.4: Comparison of actual and predicted energy from the different datasets

To summarize the evaluation, it can be said that the use of a GCN architecture is particularly useful when the dataset is complex and contains different atoms and chemical compounds, as is the case with the QM9 dataset. In contrast, the use of a basic neural network architecture is recommended for rather simple datasets, such as the water dataset used previously.

During the last phase of the CRISP-DM process, the developed and evaluated GCN can be deployed and make use of the knowledge gained. In the course of this paper, the developed GCN is a prototype and will be further used for the comparison to other machine learning models in the context of potential energy surface prediction.

2.5 Conclusion

Overall, the conducted research process has proven to be suitable for the literature review, followed by the development of the GCN for molecular property prediction. In this chapter, the results are summarized and critically reflected upon, and future research potential is identified.

2.5.1 Summary

Within the scope of the literature review, in total 11 different architectures or approaches for the implementation of GNNs could be identified and further categorized. Furthermore, an overview over the differences between the identified approaches and architectures was given. Utilizing the information gained in the scope of the literature review, the convolutional approach of a GNN was used for the development of a GNN that predicts the potential energy surface of a molecule. By using the in the discipline of computational chemistry frequently used QM9 dataset for the development of the model, the results can be adapted to other datasets that contain different information, e.g. in form of other molecules.

2.5.2 Critical Reflection

As mentioned before, the research methodology was suitable for analyzing the findable literature and for the prototypical implementation of the GNN. However, this work also has its limitations. Further approaches or architectures of GNNs could have been investigated to be used for the prediction of potential energy surfaces. By doing so, a comparison between the results, e.g. with measured performance metrics, could have taken place. In addition, the developed GNN could have been tested with other datasets than the QM9 dataset, to be able to evaluate its general validity. These aspects might have identified important requirements that are not included in the current iteration of the prototypical GNN.

2.5.3 Outlook

The insights in gained in this paper provide a starting point for future research. The QM9 dataset, that was used in this paper, provides quantum chemical properties at DFT level. Besides, the prediction of potential energy surfaces and other relevant properties proceeds at the molecular and atomic level. The advancing development of quantum computers raises the question, if machine learning models like GNNs that compute quantum properties like potential surface areas can be realized on quantum computers. While there already exist first approaches of quantum graph neural networks in the literature [5–7], the results of the literature review and the developed GNN in this paper serve as a starting point for a further investigation in the field of quantum graph neural networks.

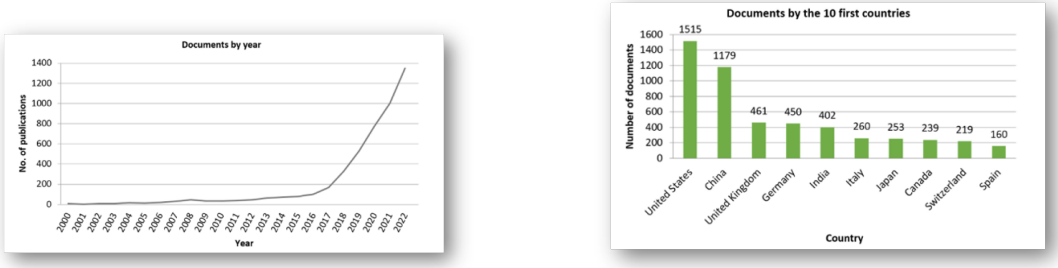
3 Quantum Graph Neural Networks for Molecular Property Prediction

Abstract

Hydrogen research, particularly in the fields of energy storage and catalysis, hinges on a deep understanding of molecular interactions and properties. The surface area of molecules is a crucial parameter influencing hydrogen adsorption and catalytic efficiency. Traditional computational methods for predicting molecular surface areas, while effective, face challenges in terms of computational complexity and accuracy when dealing with large and complex molecular structures. The accurate prediction of molecular surface area is paramount in optimizing hydrogen storage and catalysis processes. Conventional computational models, primarily based on classical algorithms, struggle with the vast complexity and quantum nature of molecular systems. This gap necessitates a novel approach that can handle the quantum mechanical properties inherent in molecules more naturally and efficiently.

Keywords: *quantum machine learning, quantum graph neural networks, hydrogen research, potential energy surface prediction*

3.1 Introduction



(a) Number of documents released per year in the field of QML

(b) Number of documents released per country in the field of QML

Figure 3.1: Results of QML-Literature-Overview from Tychola et al. 2023 [32] (excerpt)

While there is growing research on the application of quantum computing in various do-

mains, including chemistry and materials science, the specific application of QGCNNs for predicting molecular surface area in hydrogen research remains underexplored. This presents an opportunity to bridge the gap between quantum computing and molecular science, particularly in the context of hydrogen storage and catalysis. Recent advancements in quantum computing offer promising avenues for addressing complex problems in molecular sciences. Quantum Graph Convolutional Neural Networks (QGCNNs) have emerged as a potential tool for processing graph-structured data, which is naturally suited for molecular representations (Bronstein et al., 2017). The application of quantum computing in neural networks has shown significant improvements in handling high-dimensional data and complex interactions (Biamonte et al., 2017). Moreover, the representation of molecules as graphs aligns well with the capabilities of GNNs in capturing the intricate relationships and features within such structures (Duvenaud et al., 2015). The accurate prediction of molecular surface area is paramount in optimizing hydrogen storage and catalysis processes. Conventional computational models, primarily based on classical algorithms, struggle with the vast complexity and quantum nature of molecular systems. This gap necessitates a novel approach that can handle the quantum mechanical properties inherent in molecules more naturally and efficiently.
In order to solve the presented problem, the following research question and sub-questions are created:

"How can a Quantum Graph Neural Network for the prediction of molecular properties be developed?"

Q1: *How do Quantum Graph Neural Networks work?*
Q2: *How must molecular data be processed to be used with a quantum computer?*
Q2: *How does the performance compare to classical Graph Neural Networks?*

The goal of this paper is to develop an understanding of quantum graph neural networks and how they work. Besides, in the scope of this paper, an understanding of the prediction of potential energy is created. Based on this procedure, the analysis of different foundation papers will show how a quantum graph neural network can be developed. In summary, to answer the developed research question and sub-questions, the following artifacts will be created as part of the research:

- foundation paper analysis of different papers in the context of quantum graph neural networks
- requirement analysis for prediction of a quantum graph neural network for the prediction of molecular properties based on the analysis of the foundation papers
- attempt to develop a quantum graph neural network based on the conducted requirement analysis

The structure of this paper is as followed: first, the scientific foundations and theoretical background are discussed. Then the research methodology is presented in detail. Based on this, the findings and their results are presented and then discussed. Finally, a conclusion and discussion of the results is provided, as well as an overview of possible research topics based on this work.

3.2 Theoretical Background

3.2.1 Quantum Computing

In order to understand the fundamentals of quantum computing, it is first necessary to explain the basic building blocks of quantum computing. These are also known as quantum bits, quantum gates and quantum circuits, and are explained in the following. In quantum computing, quantum bits, or qubits, are the basic units of information. Unlike classical bits, which can only take the states 0 or 1, qubits operate according to the principles of quantum mechanics, which allows them to take both states simultaneously. This overlapping is called superposition and enables an exponential increase in information density. [33] Quantum gates are the building blocks for quantum circuits and perform operations on qubits. They are the analogue of logic gates in classical computer science, but with the difference that quantum gates perform continuous transformations, also known as unitary transformations, which change the superposition states of the qubits. An example of single quantum gates are Pauli rotation gates. These are applied by extracting the exponentials of the Pauli operator.

Gate	**Matrix Representation**
R_x gate	$\begin{pmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$
R_y gate	$\begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$
R_z gate	$\begin{pmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{pmatrix}$

Table 3.1: Pauli-rotation gates according to Zheng et al. 2021 [34]

There are also two-bit gates, which perform operations on two qubits and thus logically link them together. One example of this is the CNOT gate, which is used frequently in the field of quantum computing. [34] Quantum circuits are a combination of the previously introduced quantum gates and qubits. This is a specific sequence of quantum gates that are applied to a specific number of qubits in order to perform complex calculations. A quantum circuit is always designed for a specific application. This enables a variety of specific applications such as the simulation of molecular interactions. [33]

3.2.2 Quantum Graph Neural Networks

3.2.3 Potential Energy Prediction

The potential energy surface (PES), in the context of molecules and atoms, describes the energy of a molecule in regard to certain parameters, e.g. to the position of its atoms. Predicting the potential energy surface of atoms is a task in computational chemistry and materials science. It gives information about the stability and properties of molecules and materials [15]. This is helpful when to decide whether materials can be used to enable efficient catalytic processes in the field of hydrogen production [3].

3.3 Research Methodology

First, the given problem of constructing QGNNs for the prediction of molecular properties was investigated. For this purpose, an initial investigation of the literature took place and the topic was explored. This is followed by the definitions of the research questions as well as the sub-questions and the delineation of the research topic.

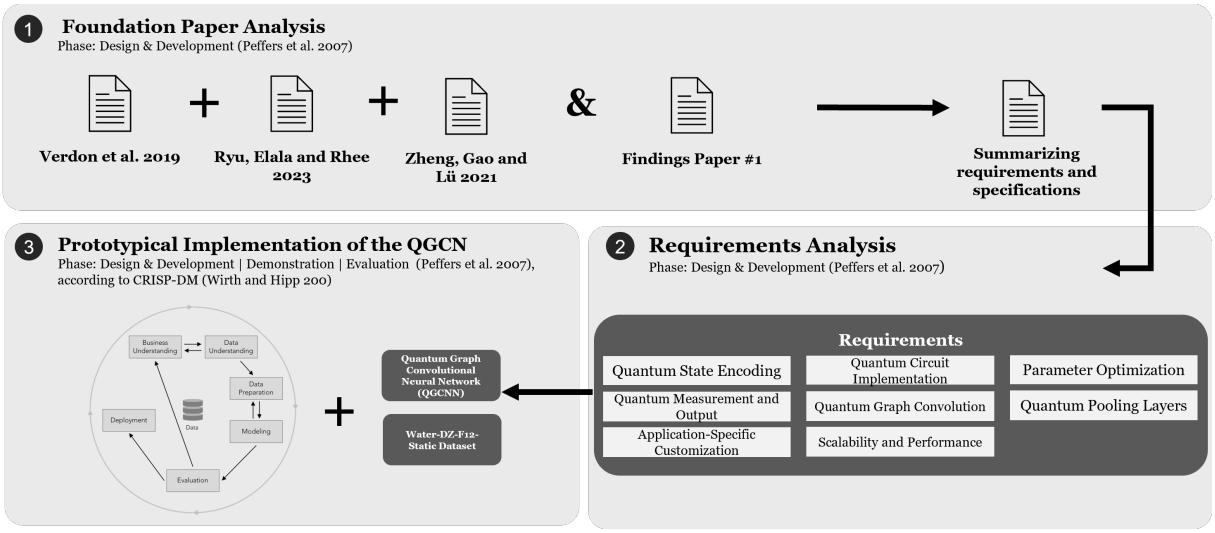


Figure 3.2: Overview of the conducted research process

Three foundation papers were then analyzed to investigate different approaches to the development of QGNNs and to identify a possible approach to the development of a QGNN for the prediction of molecular properties. Together with the results of the first paper of this thesis and the results of the foundation paper analysis, requirements for the QGNN to be developed are defined. Using these requirements, this paper will attempt to develop a prototype QGNN. This will enable the prediction of molecular properties using QGNNs to be tested and a comparison to classical GNNs to be made. For a better structure and traceability of the procedure, the exact research process is shown in Figure 3.2 on the

previous page, which is described in detail in the following section.

The **first step** was to analyze the foundation papers. The papers by Verdon et al. [5], Zheng, Gao and Lü [34] and the paper by Ryu, Elala and Rhee [35] were used. Since the literature on QGNNs is limited to a very small number of papers, a structured literature analysis is not useful in this case. Accordingly, various literature databases were searched for literature on QGNNs and the papers most relevant to the topic of this thesis were selected for the foundation paper analysis. The results of the analysis of these three papers were combined with the results of the first paper of this thesis and summarized for the next step of the conducted research process.

In the **second step**, a requirements analysis was carried out based on the results of the foundation paper analysis. A total of 11 different requirements were identified for the QGNN to be developed. Basically, these requirements can be divided into two categories: "data preparation" and "architecture of the QGNN". The results of this requirements analysis are discussed in more detail in Section 3.5.2.

The **third step** covers the prototypical implementation of the QGNN for the prediction of molecular properties. Based on the previously identified requirements, an attempt was made to implement a QGNN capable of predicting the potential energy surface using the water data set mentioned in the first paper of this thesis. Similar to the first paper, the implementation followed the structured procedure of CRISP-DM [19] and is explained in Section 3.5.3 of this thesis. Although the QGNN could not be implemented completely successfully, the development still provides interesting insights and further work can be based on the findings of this paper.

3.4 Findings

In this section, the results and findings are presented and discussed in detail. Especially the literature review and the prototypical implementation of the GNN for the prediction of potential energy will be considered.

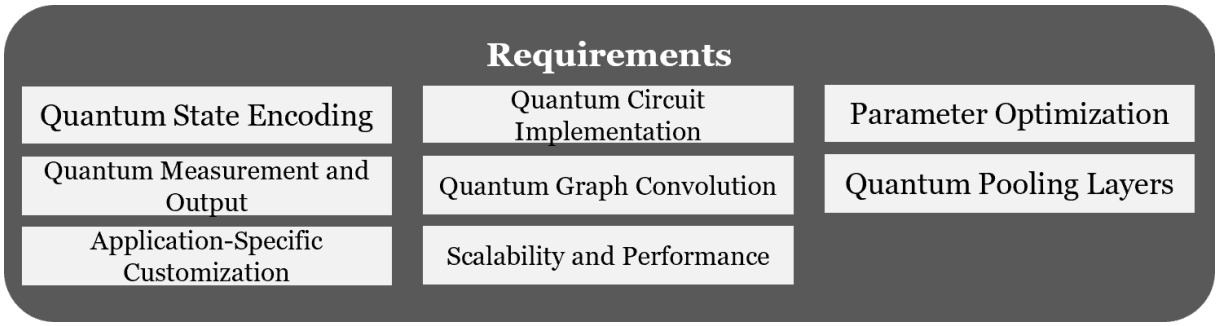


Figure 3.3: Identified requirements from the requirement analysis

3.5 Findings

3.5.1 Foundation Paper

3.5.2 Requirement Analysis

3.5.3 Development of the QGNN

3.6 Conclusion

4 Global Conclusion

4.1 Summary

4.2 Critical Reflection

4.3 Outlook

References

[1] M. Amin *et al.*, "Hydrogen production through renewable and non-renewable energy processes and their impact on climate change," *International Journal of Hydrogen Energy*, vol. 47, no. 77, pp. 33 112–33 134, Sep. 2022, ISSN: 0360-3199. DOI: 10.1016/j.ijhydene.2022.07.172.

[2] Ş. Kılkış *et al.*, "Research frontiers in sustainable development of energy, water and environment systems in a time of climate crisis," *Energy Conversion and Management*, vol. 199, p. 111 938, Nov. 2019, ISSN: 0196-8904. DOI: 10.1016/j.enconman.2019.111938.

[3] Z. Chen *et al.*, "Waste-Derived Catalysts for Water Electrolysis: Circular Economy-Driven Sustainable Green Hydrogen Energy," en, *Nano-Micro Letters*, vol. 15, no. 1, p. 4, Dec. 2023, ISSN: 2311-6706, 2150-5551. DOI: 10.1007/s40820-022-00974-7.

[4] R. Tran *et al.*, "The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysts," *ACS Catalysis*, vol. 13, no. 5, pp. 3066–3084, Mar. 2023, arXiv:2206.08917 [cond-mat, physics:physics], ISSN: 2155-5435, 2155-5435. DOI: 10.1021/acscatal.2c05426.

[5] G. Verdon, T. McCourt, E. Luzhnica, V. Singh, S. Leichenauer, and J. Hidary, *Quantum Graph Neural Networks*, arXiv:1909.12264 [quant-ph], Sep. 2019.

[6] K. Beer, M. Khosla, J. Köhler, and T. J. Osborne, *Quantum machine learning of graph-structured data*, arXiv:2103.10837 [quant-ph], Mar. 2021.

[7] X. Ai, Z. Zhang, L. Sun, J. Yan, and E. Hancock, *Decompositional Quantum Graph Neural Network*, arXiv:2201.05158 [quant-ph], Mar. 2023.

[8] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, Dec. 2007, Publisher: Routledge _eprint: https://doi.org/10.2753/MIS0742-1222240302, ISSN: 0742-1222. DOI: 10.2753/MIS0742-1222240302.

[9] P. Fettke, "State-of-the-Art des State-of-the-Art: Eine Untersuchung der Forschungsmethode „Review" innerhalb der Wirtschaftsinformatik," de, *WIRTSCHAFTSINFORMATIK*, vol. 48, no. 4, p. 257, Aug. 2006, ISSN: 0937-6429, 1861-8936. DOI: 10.1007/s11576-006-0057-3.

[10] J. Webster and R. T. Watson, "Analyzing the Past to Prepare for the Future: Writing a Literature Review," en, *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002.

[11] J. vom Brocke, A. Simons, B. Niehaves, and K. Reimer, "RECONSTRUCTING THE GIANT: ON THE IMPORTANCE OF RIGOUR IN DOCUMENTING THE LITERATURE SEARCH PROCESS," en, 2009.

[12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," en, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, Jan. 2021, ISSN: 2162-237X, 2162-2388. DOI: 10.1109/TNNLS.2020.2978386.

[13] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," en, *AI Open*, vol. 1, pp. 57–81, 2020, ISSN: 26666510. DOI: 10.1016/j.aiopen.2021.01.001.

[14] P. Veličković, "Everything is Connected: Graph Neural Networks," *Current Opinion in Structural Biology*, vol. 79, p. 102 538, Apr. 2023, arXiv:2301.08210 [cs, stat], ISSN: 0959440X. DOI: 10.1016/j.sbi.2023.102538.

[15] J.-C. Liu *et al.*, "Computational Prediction of Graphdiyne-Supported Three-Atom Single-Cluster Catalysts," en, *CCS Chemistry*, vol. 5, no. 1, pp. 152–163, Jan. 2023, ISSN: 2096-5745. DOI: 10.31635/ccschem.022.202201796.

[16] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, *How Powerful are Graph Neural Networks?* arXiv:1810.00826 [cs, stat], Feb. 2019.

[17] J. Gasteiger, J. Groß, and S. Günnemann, *Directional Message Passing for Molecular Graphs*, arXiv:2003.03123 [physics, stat], Apr. 2022.

[18] PyTorch Geometric Documentation - Torch geometric datasets, *Qm9 dataset description*, 2023.

[19] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Manchester, vol. 1, 2000, pp. 29–39.

[20] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural Message Passing for Quantum Chemistry," en, in *Proceedings of the 34th International Conference on Machine Learning*, ISSN: 2640-3498, PMLR, Jul. 2017, pp. 1263–1272.

[21] *DeepChem Documentation*, 2023.

[22] H. Altae-Tran, B. Ramsundar, A. S. Pappu, and V. Pande, "Low Data Drug Discovery with One-Shot Learning," en, *ACS Central Science*, vol. 3, no. 4, pp. 283–293, Apr. 2017, ISSN: 2374-7943, 2374-7951. DOI: 10.1021/acscentsci.6b00367.

[23] Z. Xiong *et al.*, "Pushing the boundaries of molecular representation for drug discovery with graph attention mechanism," *Journal of Medicinal Chemistry*, vol. 63, Aug. 2019. DOI: 10.1021/acs.jmedchem.9b00959.

[24] D. Jiang *et al.*, "Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models," en, *Journal of Cheminformatics*, vol. 13, no. 1, p. 12, Feb. 2021, ISSN: 1758-2946. DOI: 10.1186/s13321-020-00479-8.

[25] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[26] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: A comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp. 1–23, 2019.

[27] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.

[28] E. Hajiramezanali, A. Hasanzadeh, K. Narayanan, N. Duffield, M. Zhou, and X. Qian, "Variational graph recurrent neural networks," *Advances in neural information processing systems*, vol. 32, 2019.

[29] X. Li, H. Zhang, and R. Zhang, "Adaptive graph auto-encoder for general data clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 9725–9732, 2021.

[30] X.-M. Zhang, L. Liang, L. Liu, and M.-J. Tang, "Graph neural networks and their current applications in bioinformatics," *Frontiers in genetics*, vol. 12, p. 690 049, 2021.

[31] S. Menon, "Understanding molecular data using graph convolutional networks,"

[32] K. A. Tychola, T. Kalampokas, and G. A. Papakostas, "Quantum machine learning—an overview," en, *Electronics*, vol. 12, no. 11, p. 2379, May 2023, ISSN: 2079-9292. DOI: 10.3390/electronics12112379.

[33] D. Claudino, "The basics of quantum computing for chemists," *International Journal of Quantum Chemistry*, vol. 122, no. 23, e26990, 2022.

[34] J. Zheng, Q. Gao, and Y. Lü, "Quantum graph convolutional neural networks," in *2021 40th Chinese Control Conference (CCC)*, IEEE, 2021, pp. 6335–6340.

[35] J.-Y. Ryu, E. Elala, and J.-K. K. Rhee, "Quantum graph neural network models for materials search," *Materials*, vol. 16, no. 12, p. 4300, 2023.