Lingfei Wu, Peng Cui, Jian Pei, Liang Zhao

Graph Neural Networks

Foundations, Frontiers, and Applications

July 13, 2021

Preface

The field of graph neural networks (GNNs) has seen rapid and incredible strides over the recent years. Graph neural networks, also known as deep learning on graphs, graph representation learning, or geometric deep learning have become one of the fastest-growing research topics in machine learning, especially deep learning. This wave of research at the intersection of graph theory and deep learning has also influenced other fields of science, including recommendation systems, computer vision, natural language processing, inductive logic programming, program synthesis, software mining, automated planning, cybersecurity, and intelligent transportation.

Although graph neural networks have achieved remarkable attention, it still faces many challenges when applying them into other domains, from the theoretical understanding of methods to the scalability and interpretability in a real system, and from the soundness of the methodology to the empirical performance in an application. However, as the field rapidly grows, it has been extremely challenging to gain a global perspective of the developments of GNNs. Therefore, we feel the urgency to bridge the above gap and have a comprehensive book on this fast-growing yet challenging topic, which can benefit a broad audience including advanced undergraduate and graduate students, postdoctoral researchers, lecturers, and industrial practitioners.

This book is intended to cover a broad range of topics in graph neural networks, from the foundations to the frontiers, and from the methodologies to the applications. Our book is dedicated to introducing the fundamental concepts and algorithms of GNNs, new research frontiers of GNNs, and broad and emerging applications with GNNs.

Book Website and Resources

The website and further resources of this book can be found at: `https://graph-neural-networks.github.io/`. The website provides online preprints and lecture slides of all the chapters. It also provides pointers to useful material and resources that are publicly available and relevant to graph neural networks.

To the Instructors

The book can be used for a one-semester graduate course for graduate students. Though it is mainly written for students with a background in computer science, students with a basic understanding of probability, statistics, graph theory, linear algebra, and machine learning techniques such as deep learning will find it easily accessible. Some chapters can be skipped or assigned as homework assignments for reviewing purposes if students have knowledge of a chapter. For example, if students have taken a deep learning course, they can skip Chapter 1. The instructors can also choose to combine Chapters 1, 2, and 3 together as a background introduction course at the very beginning.

When the course focuses more on the foundation and theories of graph neural networks, the instructor can choose to focus more on Chapters 4-8 while using Chapters 19-27 to showcase the applications, motivations, and limitations. Please refer to the Editors' Notes at the end of each chapter on how Chapters 4-8 and Chapters 19-27 are correlated. When the course focuses more on the research frontiers, Chapters 9-18 can be the pivot to organize the course. For example, the instructor can make it an advanced graduate course where the students are asked to search and present the most recent research papers in each different research frontier. They can also be asked to establish their course project based on the applications described in Chapters 19-27 as well as the materials provided on our website.

To the Readers

This book was designed to cover a wide range of topics in the graph neural network field, including background, theoretical foundations, methodologies, research frontiers, and applications. Therefore, it can be treated as a comprehensive handbook for a wide variety of readers such as students, researchers, and professionals. You should have some knowledge of the concepts and terminology associated with statistics, machine learning, and graph theory. Some backgrounds of the basics have been provided and referenced in the first eight chapters. You should better also have knowledge of deep learning and some programming experience for easily accessing the most of chapters of this book. In particular, you should be able to read pseudocode and understand graph structures.

The book is well modularized and each chapter can be learned in a standalone manner based on the special interests and needs. For those readers who want to have a solid understanding of various techniques and theories of graph neural networks, you can start from Chapters 4-9. For those who further want to perform in-depth research and advance related fields, please read those chapters of interest among Chapters 9-18, which provide comprehensive knowledge in the most recent research issues, open problems, and research frontiers. For those who want to apply graph neural networks to benefit specific domains, or aim at finding interesting applications to validate specific graph neural networks techniques, please refer to Chapters 19-27.

Acknowledgements

Graph machine learning has attracted many gifted researchers to make their seminal contributions over the last few years. We are very fortune to discuss the challenges and opportunities, and often work with many of them on a rich variety of research topics in this exciting field. We are deeply indebted to these collaborators and colleagues from JD.COM, IBM Research, Tsinghua University, Simon Fraser University, Emory University, and elsewhere, who encouraged us to create such a book comprehensively covering various topics of Graph Neural Networks in order to educate the interested beginners and foster the advancement of the field for both academic researchers and industrial practitioners.

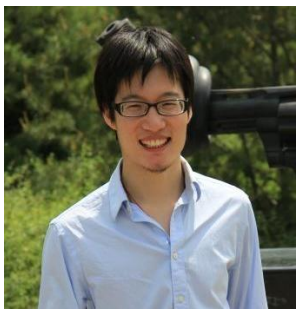
This book would not have been possible without the contributions of many people. We would like to give many thanks to the people who offered feedback on checking the consistency of the math notations of the entire book as well as reference editing of this book. They are people from Emory University: Ling Chen, Xiaojie Guo, and Shiyu Wang, as well as people from Tsinghua University: Yue He, Ziwei Zhang, and Haoxin Liu. We would like to give our special thanks to Dr. Xiaojie Guo, who generously offered her help in providing numerous valuable feedback on many chapters.

We also want to thank those who allowed us to reproduce images, figures, or data from their publications.

Finally, we would like to thank our families for their love, patience and support during this very unusual time when we are writing and editing this book.

Editor Biography

Dr. Lingfei Wu is a Principal Scientist at JD.COM Silicon Valley Research Center, leading a team of 30+ machine learning/natural language processing scientists and software engineers to build intelligent e-commerce personalization system. He earned his Ph.D. degree in computer science from the College of William and Mary in 2016. Previously, he was a research staff member at IBM Thomas J. Watson Research Center and led a 10+ research scientist team for developing novel Graph Neural Networks methods and systems, which leads to the 1 AI Challenge Project in IBM Research and multiple IBM Awards including three-time Outstanding Technical Achievement Award. He has published more than 90 top-ranked conference and journal papers, and is a co-inventor of more than 40 filed US patents. Because of the high commercial value of his patents, he has received eight invention achievement awards and has been appointed as IBM Master Inventors, class of 2020. He was the recipients of the Best Paper Award and Best Student Paper Award of several conferences such as IEEE ICC'19, AAAI workshop on DLGMA'20 and KDD workshop on DLG'19. His research has been featured in numerous media outlets, including NatureNews, YahooNews, Venturebeat, TechTalks, SyncedReview, Leiphone, QbitAI, MIT News, IBM Research News, and SIAM News. He has co-organized 10+ conferences (KDD, AAAI, IEEE BigData) and is the founding co-chair for Workshops of Deep Learning on Graphs (with AAAI'21, AAAI'20, KDD'21, KDD'20, KDD'19, and IEEE BigData'19). He has currently served as Associate Editor for IEEE Transactions on Neural Networks and Learning Systems, ACM Transactions on Knowledge Discovery from Data and International Journal of Intelligent Systems, and regularly served as a SPC/PC member of the following major AI/ML/NLP conferences including KDD, IJCAI, AAAI, NIPS, ICML, ICLR, and ACL.



Dr. Peng Cui is an Associate Professor with tenure at Department of Computer Science in Tsinghua University. He obtained his PhD degree from Tsinghua University in 2010. His research interests include data mining, machine learning and multimedia analysis, with expertise on network representation learning, causal inference and stable learning, social dynamics modeling, and user behavior modeling, etc. He is keen to promote the convergence and integration of causal inference and machine learning, addressing the fundamental issues of today's AI technology, including explainability, stability and fairness issues. He is recognized as a Distinguished Scientist of ACM, Distinguished Member of CCF and Senior Member of IEEE. He has published more than 100 papers in prestigious conferences and journals in machine learning and data mining. He is one of the most cited authors in network embedding. A number of his proposed algorithms on network embedding generate substantial impact in academia and industry. His recent research won the IEEE Multimedia Best Department Paper Award, IEEE ICDM 2015 Best Student Paper Award, IEEE ICME 2014 Best Paper Award, ACM MM12 Grand Challenge Multimodal Award, MMM13 Best Paper Award, and were selected into the Best of KDD special issues in 2014 and 2016, respectively. He was PC co-chair of CIKM2019 and MMM2020, SPC or area chair of ICML, KDD, WWW, IJCAI, AAAI, etc., and Associate Editors of IEEE TKDE (2017-), IEEE TBD (2019-), ACM TIST(2018-), and ACM TOMM (2016-) etc. He received ACM China Rising Star Award in 2015, and CCF-IEEE CS Young Scientist Award in 2018.

Dr. Jian Pei is a Professor in the School of Computing Science at Simon Fraser University. He is a well-known leading researcher in the general areas of data science, big data, data mining, and database systems. His expertise is on developing effective and efficient data analysis techniques for novel data intensive applications, and transferring his research results to products and business practice. He is recognized as a Fellow of the Royal Society of Canada (Canada's national academy), the Canadian Academy of Engineering, the Association of Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE). He is one of the most cited authors in data mining, database systems, and information retrieval. Since 2000, he has published one textbook, two monographs and over 300 research papers in refereed journals and conferences, which have been cited extensively by others. His research has generated remarkable impact substantially beyond academia. For example, his algorithms have been adopted by industry in production and popular open-source software suites. Jian Pei also demonstrated outstanding professional leadership in many academic organizations and activities. He was the editor-in-chief of the IEEE Transactions of Knowledge and Data Engineering (TKDE) in 2013-16, the chair of the Special Interest Group on Knowledge Discovery in Data (SIGKDD) of the Association for Computing Machinery (ACM) in 2017-2021, and a general co-chair or program committee co-chair of many premier conferences. He maintains a wide spectrum of industry relations with both global and local industry partners. He is an active consultant and coach for industry on enterprise data strategies, healthcare informatics, network security intelligence, computational finance, and smart retail. He received many prestigious awards, including the 2017 ACM SIGKDD Innovation Award, the 2015 ACM SIGKDD Service Award, the 2014 IEEE ICDM Research Contributions Award, the British Columbia Innovation Council 2005 Young Innovator Award, an NSERC 2008 Discovery Accelerator Supplements Award (100 awards cross the whole country), an IBM Faculty Award (2006), a KDD Best Application Paper Award (2008), an ICDE Influential Paper Award (2018), a PAKDD Best Paper Award (2014), a PAKDD Most Influential Paper Award (2009), and an IEEE Outstanding Paper Award (2007).

Dr. Liang Zhao is an assistant professor at the Department of Compute Science at Emory University. Before that, he was an assistant professor in the Department of Information Science and Technology and the Department of Computer Science at George Mason University. He obtained his PhD degree in 2016 from Computer Science Department at Virginia Tech in the United States. His research interests include data mining, artificial intelligence, and machine learning, with special interests in spatiotemporal and network data mining, deep learning on graphs, nonconvex optimization, model parallelism, event prediction, and interpretable machine learning. He received AWS Machine Learning Research Award in 2020 from Amazon Company for his research on distributed graph neural networks. He won NSF Career Award in 2020 awarded by National Science Foundation for his research on deep learning for spatial networks, and Jeffress Trust Award in 2019 for his research on deep generative models for biomolecules, awarded by Jeffress Memorial Trust Foundation and Bank of America. He won the Best Paper Award in the 19th IEEE International Conference on Data Mining (ICDM 2019) for the paper of his lab on deep graph transformation. He has also won Best Paper Award Shortlist in the 27th Web Conference (WWW 2021) for deep generative models. He was selected as "Top 20 Rising Star in Data Mining" by Microsoft Search in 2016 for his research on spatiotemporal data mining. He has also won Outstanding Doctoral Student in the Department of Computer Science at Virginia Tech in 2017. He is awarded as CIFellow Mentor 2021 by the Computing Community Consortium for his research on deep learning for spatial data. He has published numerous research papers in top-tier conferences and journals such as KDD, TKDE, ICDM, ICLR, Proceedings of the IEEE, ACM Computing Surveys, TKDD, IJCAI, AAAI, and WWW. He has been serving as organizers such as publication chair, poster chair, and session chair for many top-tier conferences such as SIGSPATIAL, KDD, ICDM, and CIKM.

List of Contributors

Miltiadis Allamanis
Microsoft Research, Cambridge, UK

Yu Chen
Facebook AI, Menlo Park, CA, USA

Yunfei Chu
Alibaba Group, Hangzhou, China

Peng Cui
Tsinghua University, Beijing, China

Tyler Derr
Vanderbilt University, Nashville, TN, USA

Keyu Duan
Texas A&M University, College Station, TX, USA

Qizhang Feng
Texas A&M University, College Station, TX, USA

Stephan Gunnemann
Technical University of Munich, München, Germany

Xiaojie Guo
JD.COM Silicon Valley Research Center, Mountain View, CA, USA

Yu Hou
Weill Cornell Medicine, New York City, New York, USA

Xia Hu
Texas A&M University, College Station, TX, USA

Junzhou Huang
University of Texas at Arlington, Arlington, TA, United States

Shouling Ji

Zhejiang University, Hangzhou, China

Wei Jin
Michigan State University, East Lansing, MI, USA

Anowarul Kabir
George Mason University, Fairfax, VA, USA

Seyed Mehran Kazemi
Borealis AI, Montreal, Canada.

Jure Leskovec
Stanford University, Stanford, CA, USA

Juncheng Li
Zhejiang University, Hangzhou, China

Jiacheng Li
Zhejiang University, Hangzhou, China

Pan Li
Purdue University, Lafayette, IN, USA

Yanhua Li
Worcester Polytechnic Institute, Worcester, MA, USA

Renjie Liao
University of Toronto, Toronto, Canada

Xiang Ling
Zhejiang University, Hangzhou, China

Bang Liu
University of Montreal, Montreal, Canada

Ninghao Liu
Texas A&M University, College Station, TX, USA

Zirui Liu
Texas A&M University, College Station, TX, USA

Hehuan Ma
University of Texas at Arlington, College Station, TX, USA

Collin McMillan
University of Notre Dame, Notre Dame, IN, USA

Christopher Morris
Polytechnique Montréal, Montréal, Canada

Zongshen Mu
Zhejiang University, Hangzhou, China

Menghai Pan

Worcester Polytechnic Institute, Worcester, MA, USA

Jian Pei
Simon Fraser University, British Columbia, Canada

Yu Rong
Tencent AI Lab, Shenzhen, China

Amarda Shehu
George Mason University, Fairfax, VA, USA

Kai Shen
Zhejiang University, Hangzhou, China

Chuan Shi
Beijing University of Posts and Telecommunications, Beijing, China

Le Song
Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

Chang Su
Weill Cornell Medicine, New York City, New York. USA

Jian Tang
Mila-Quebec AI Institute, HEC Montreal, Canada

Siliang Tang
Zhejiang University, Hangzhou, China

Fei Wang
Weill Cornell Medicine, New York City, New York, USA

Shen Wang
University of Illinois at Chicago, Chicago, IL, USA

Shiyu Wang
Emory University, Atlanta, GA, USA

Xiao Wang
Beijing University of Posts and Telecommunications, Beijing, China

Yu Wang
Vanderbilt University, Nashville, TN, USA

Chunming Wu
Zhejiang University, Hangzhou, China

Lingfei Wu
JD.COM Silicon Valley Research Center, Mountain View, CA, USA

Hongxia Yang
Alibaba Group, Hangzhou, China

Jiangchao Yao

Alibaba Group, Hangzhou, China

Philip S. Yu
University of Illinois at Chicago, Chicago, IL, USA

Muhan Zhang
Peking University, Beijing, China

Wenqiao Zhang
Zhejiang University, Hangzhou, China

Liang Zhao
Emory University, Atlanta, GA, USA

Chang Zhou
Alibaba Group, Hangzhou, China

Kaixiong Zhou
Texas A&M University, TX, USA

Xun Zhou
University of Iowa, Iowa City, IA, USA

Contents

Terminologies

This chapter describes a list of definitions of terminologies related to graph neural networks used throughout this book.

1 Basic concepts of Graphs

- **Graph**: A graph is composed of a node set and an edge set, where nodes represent entities and edges represent the relationship between entities. The nodes and edges form the topology structure of the graph. Besides the graph structure, nodes, edges, and/or the whole graph can be associated with rich information represented as node/edge/graph features (also known as attributes or contents).
- **Subgraph**: A subgraph is a graph whose set of nodes and set of edges are all subsets of the original graph.
- **Centrality**: A centrality is a measurement of the importance of nodes in the graph. The basic assumption of centrality is that a node is thought to be important if many other important nodes also connect to it. Common centrality measurements include the degree centrality, the eigenvector centrality, the betweenness centrality, and the closeness centrality.
- **Neighborhood**: The neighborhood of a node generally refers to other nodes that are close to it. For example, the k-order neighborhood of a node, also called the k-step neighborhood, denotes a set of other nodes in which the shortest path distance between these nodes and the central node is no larger than k.
- **Community Structure**: A community refers to a group of nodes that are densely connected internally and less densely connected externally.
- **Graph Sampling**: Graph sampling is a technique to pick a subset of nodes and/or edges from the original graph. Graph sampling can be applied to train machine learning models on large-scale graphs while preventing severe scalability issues.

- **Heterogeneous Graphs**: Graphs are called heterogeneous if the nodes and/or edges of the graph are from different types. A typical example of heteronomous graphs is knowledge graphs where the edges are composed of different types.
- **Hypergraphs**: Hypergraphs are generalizations of graphs in which an edge can join any number of nodes.
- **Random Graph**: Random graph generally aims to model the probability distributions over graphs that the observed graphs are generated from. The most basic and well-studied random graph model, known as the Erdos–Renyi model, assumes that the node set is fixed and each edge is identically and independently generated.
- **Dynamic Graph**: Dynamic graph refers to when at least one component of the graph data changes over time, e.g., adding or deleting nodes, adding or deleting edges, changing edges weights or changing node attributes, etc. If graphs are not dynamic, we refer to them as static graphs.

2 Machine Learning on Graphs

- **Spectral Graph Theory**: Spectral graph theory analyzes matrices associated with the graph such as its adjacency matrix or Laplacian matrix using tools of linear algebra such as studying the eigenvalues and eigenvectors of the matrix.
- **Graph Signal Processing**: Graph Signal Processing (GSP) aims to develop tools for processing signals defined on graphs. A graph signal refers to a finite collection of data samples with one sample at each node in the graph.
- **Node-level Tasks**: Node-level tasks refer to machine learning tasks associated with individual nodes in the graph. Typical examples of node-level tasks include node classification and node regression.
- **Edge-level Tasks**: Edge-level tasks refer to machine learning tasks associated with a pair of nodes in the graph. A typical example of an edge-level task in link prediction.
- **Graph-level Tasks**: Graph-level tasks refer to machine learning tasks associated with the whole graph. Typical examples of graph-level tasks include graph classification and graph property prediction.
- **Transductive and Inductive Learning**: Transductive learning refers to that the targeted instances such as nodes or edges are observed at the training time (though the labels of the targeted instances remain unknown) and inductive learning aims to learn the model which is generalizable to unobserved instances.

3 Graph Neural Networks

- **Network embedding**: The goal of network embedding is to represent each node in the graph as a low-dimensional vector so that useful information such as the

graph structures and some properties of the graph is preserved in the embedding vectors. Network embedding is also referred to as graph embedding and node representation learning.

- **Graph Neural Network**: Graph neural network refers to any neural network working on the graph data.
- **Graph Convolutional Network**: Graph convolutional network usually refers to a specific graph neural network proposed by Kipf and Welling Kipf and Welling (2017a). It is occasionally used as a synonym for graph neural network, i.e., referring to any neural network working on the graph data, in some literature.
- **Message-Passing**: Message-passing is a framework of graph neural networks in which the key step is to pass messages between different nodes based on graph structures in each neural network layer. The most widely adopted formulation, usually denoted as message-passing neural networks, is to only pass messages between nodes that are directly connected Gilmer et al (2017). The message passing functions are also called graph filters and graph convolutions in some literature.
- **Readout**: Readout refers to functions that summarize the information of individual nodes to form more high-level information such as forming a subgraph/supergraph or obtaining the representations of the entire graph. Readout is also called pooling and graph coarsening in some literature.
- **Graph Adversarial Attack**: Graph adversarial attacks aim to generate worst-case perturbations by manipulating the graph structure and/or node features so that the performance of some models are downgraded. Graph adversarial attacks can be categorized based on the attacker's goals, capabilities, and accessible knowledge.
- **Robustness certificates**: Methods providing formal guarantees that the prediction of a GNN is not affected even when perturbations are performed based on a certain perturbation model.

Notations

This Chapter provides a concise reference that describes the notations used throughout this book.

Numbers, Arrays, and Matrices

A scalar	x
A vector	$\mathbf{x}$
A matrix	X
An identity matrix	$\mathbf{I}$
The set of real numbers	$\mathbb{R}$
The set of complex numbers	$\mathbb{C}$
The set of integers	$\mathbb{Z}$
The set of real n-length vectors	$\mathbb{R}^n$
The set of real $m \times n$ matrices	$\mathbb{R}^{m \times n}$
The real interval including a and b	$[a, b]$
The real interval including a but excluding b	$[a, b)$
The element of the vector $\mathbf{x}$ with index i	$\mathbf{x}_i$
The element of matrix X's indexed by Row i and Column j	$X_{i,j}$

Graph Basics

A graph	$\mathcal{G}$
Edge set	$\mathcal{E}$
Vertex set	$\mathcal{V}$
Adjacent matrix of a graph	A
Laplacian matrix	L
Diagonal degree matrix	D
Isomorphism between graphs $\mathcal{G}$ and $\mathcal{H}$	$\mathcal{G} \cong \mathcal{H}$
$\mathcal{H}$ is a subgraph of graph $\mathcal{G}$	$\mathcal{H} \subseteq \mathcal{G}$
$\mathcal{H}$ is a proper subgraph of graph $\mathcal{G}$	$\mathcal{H} \subset \mathcal{G}$
Union of graphs $\mathcal{H}$ and $\mathcal{G}$	$\mathcal{G} \cup \mathcal{H}$

Intersection of graphs $\mathscr{H}$ and $\mathscr{G}$	$\mathscr{G} \cap \mathscr{H}$
Disjoint Union of graphs $\mathscr{H}$ and $\mathscr{G}$	$\mathscr{G} + \mathscr{H}$
Cartesian Product of graphs of graphs $\mathscr{H}$ and $\mathscr{G}$	$\mathscr{G} \times \mathscr{H}$
The join of graphs $\mathscr{H}$ and $\mathscr{G}$	$\mathscr{G} \vee \mathscr{H}$

Basic Operations

Transpose of matrix X	$X^\top$
Dot product of matrices X and Y	$X \cdot Y$ or XY
Element-wise (Hadamard) product of matrices X and Y	$X \odot Y$
Determinant of X	$\det(X)$
p-norm (also called ℓ_p norm) of $\mathbf{x}$	$\|\mathbf{x}\|_p$
Union	$\cup$
Intersection	$\cap$
Subset	$\subseteq$
Proper subset	$\subset$
Inner prodct of vector $\mathbf{x}$ and $\mathbf{y}$	$< \mathbf{x}, \mathbf{y} >$

Functions

The function f with domain $\mathbb{A}$ and range $\mathbb{B}$	$f : \mathbb{A} \to \mathbb{B}$
Derivative of y with respect to $\mathbf{x}$	$\frac{dy}{d\mathbf{x}}$
Partial derivative of y with respect to $\mathbf{x}$	$\frac{\partial y}{\partial \mathbf{x}}$
Gradient of y with respect to $\mathbf{x}$	$\nabla_{\mathbf{x}} y$
Matrix derivatives of y with respect to matrix X	$\nabla_X y$
The Hessian matrix of function f at input vector $\mathbf{x}$	$\nabla^2 f(\mathbf{x})$
Definite integral over the entire domain of $\mathbf{x}$	$\int f(\mathbf{x}) d\mathbf{x}$
Definite integral with respect to $\mathbf{x}$ over the set $\mathbb{S}$	$\int_{\mathbb{S}} f(\mathbf{x}) d\mathbf{x}$
A function of $\mathbf{x}$ parametrized by $\boldsymbol{\theta}$	$f(\mathbf{x}; \boldsymbol{\theta})$
Convolution between functions f and g	$f * g$

Probablistic Theory

A probability distribution of a	$p(\mathrm{a})$
A conditional probabilistic distribution of b given a	$p(b\|a)$
The random variables a and b are independent	$\mathrm{a} \perp \mathrm{b}$
Variables a and b are conditionally independent given c	$\mathrm{a} \perp \mathrm{b} \mid \mathrm{c}$
Random variable a has a distribution p	$\mathrm{a} \sim p$
The expectation of $f(a)$ with respect to the variable a under distribution p	$\mathbb{E}_{a \sim p}[f(a)]$
Gaussian distribution over $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance Σ	$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$