Graph Machine Learning

Take graph data to the next level by applying machine learning techniques and algorithms

Claudio Stamile

Aldo Marzullo

Enrico Deusebio

Packt

BIRMINGHAM—MUMBAI

Graph Machine Learning

Group Product Manager: Kunal Parikh
Publishing Product Manager: Devika Battike
Senior Editor: Roshan Kumar
Content Development Editor: Sean Lobo
Technical Editor: Sonam Pandey
Copy Editor: Safis Editing
Project Coordinator: Aparna Ravikumar Nair
Proofreader: Safis Editing
Indexer: Vinayak Purushotham
Production Designer: Joshua Misquitta

First published: May 2021

Production reference: 1270521

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham
B3 2PB, UK.

ISBN 978-1-80020-449-2

www.packt.com

Alla memoria di mio Zio, Franchino Avolio. Alle ruote delle bici troppo sgonfie, all'infanzia che mi ha regalato.

In memory of my uncle, Franchino Avolio. To the wheels of bikes that are too flat, to the childhood he gave me.

– Claudio Stamile

To my family, my roots.

– Aldo Marzullo

To Lili, for always reminding me with your 'learning' process how wonderful the human brain and life are.

– Enrico Deusebio

Contributors

About the authors

Claudio Stamile received an M.Sc. degree in computer science from the University of Calabria (Cosenza, Italy) in September 2013 and, in September 2017, he received his joint Ph.D. from KU Leuven (Leuven, Belgium) and Université Claude Bernard Lyon 1 (Lyon, France). During his career, he has developed a solid background in artificial intelligence, graph theory, and machine learning, with a focus on the biomedical field. He is currently a senior data scientist in CGnal, a consulting firm fully committed to helping its top-tier clients implement data-driven strategies and build AI-powered solutions to promote efficiency and support new business models.

Aldo Marzullo received an M.Sc. degree in computer science from the University of Calabria (Cosenza, Italy) in September 2016. During his studies, he developed a solid background in several areas, including algorithm design, graph theory, and machine learning. In January 2020, he received his joint Ph.D. from the University of Calabria and Université Claude Bernard Lyon 1 (Lyon, France), with a thesis entitled *Deep Learning and Graph Theory for Brain Connectivity Analysis in Multiple Sclerosis*. He is currently a postdoctoral researcher at the University of Calabria and collaborates with several international institutions.

Enrico Deusebio is currently the chief operating officer at CGnal, a consulting firm that helps its top-tier clients implement data-driven strategies and build AI-powered solutions. He has been working with data and large-scale simulations using high-performance facilities and large-scale computing centers for over 10 years, both in an academic and industrial context. He has collaborated and worked with top-tier universities, such as the University of Cambridge, the University of Turin, and the Royal Institute of Technology (KTH) in Stockholm, where he obtained a Ph.D. in 2014. He also holds B.Sc. and M.Sc. degrees in aerospace engineering from Politecnico di Torino.

About the reviewers

Kacper Kubara is a technical co-founder of Artemo and a data engineer at Annual Insight, and is currently pursuing a postgraduate degree in AI at the University of Amsterdam. Despite the focus of his research being graph representation learning, he is also interested in the tools and methods that help to bridge the gap between the AI industry and academia.

Tural Gulmammadov has been leading a group of data scientists and machine learning engineers at Oracle to tackle applied machine learning problems from various industries. He is dedicated to and motivated by the applications of graph theory and discrete mathematics in machine learning over distributed computational environments. He is a cognitive science, statistics, and psychology enthusiast, as well as a chess player, painter, seasonal horse rider, and paddler.

Table of Contents

Section 2 – Machine Learning on Graphs

3
Unsupervised Graph Learning

4
Supervised Graph Learning

5
Problems with Machine Learning on Graphs

Section 3 – Advanced Applications of Graph Machine Learning

6

Social Network Graphs

7

Text Analytics and Natural Language Processing Using Graphs

8
Graph Analysis for Credit Card Transactions

9
Building a Data-Driven Graph-Powered Application

10
Novel Trends on Graphs

Preface

Graph Machine Learning provides a new set of tools for processing network data and leveraging the power of the relationship between entities that can be used for predictive, modeling, and analytics tasks.

You will start with a brief introduction to graph theory and Graph Machine Learning, learning to understand their potential. As you proceed, you will become well versed with the main machine learning models for graph representation learning: their purpose, how they work, and how they can be implemented in a wide range of supervised and unsupervised learning applications. You'll then build a complete machine learning pipeline, including data processing, model training, and prediction, in order to exploit the full potential of graph data. Moving on, you will cover real-world scenarios, such as extracting data from social networks, text analytics, and natural language processing using graphs and financial transaction systems on graphs. Finally, you will learn how to build and scale out data-driven applications for graph analytics to store, query, and process network information, before progressing to explore the latest trends on graphs.

By the end of this machine learning book, you will have learned the essential concepts of graph theory and all the algorithms and techniques used to build successful machine learning applications.

Who this book is for

This book is for data analysts, graph developers, graph analysts, and graph professionals who want to leverage the information embedded in the connections and relations between data points, unravel hidden structures, and exploit topological information to boost their analysis and models' performance. The book will also be useful for data scientists and machine learning developers who want to build machine learning-driven graph databases. A beginner-level understanding of graph databases and graph data is required. An intermediate-level working knowledge of Python programming and machine learning is also expected to make the most out of this book.

What this book covers

Chapter 1, *Getting Started with Graphs*, introduces the basic concepts of graph theory using the NetworkX Python library.

Chapter 2, *Graph Machine Learning*, introduces the main concepts of graph machine learning and graph embedding techniques.

Chapter 3, *Unsupervised Graph Learning*, covers recent unsupervised graph embedding methods.

Chapter 4, *Supervised Graph Learning*, covers recent supervised graph embedding methods.

Chapter 5, *Problems with Machine Learning on Graphs*, introduces the most common machine learning tasks on graphs.

Chapter 6, *Social Network Analysis*, shows an application of machine learning algorithms on social network data.

Chapter 7, *Text Analytics and Natural Language Processing Using Graphs*, shows the application of machine learning algorithms to natural language processing tasks.

Chapter 8, *Graph Analysis for Credit Card Transactions*, shows the application of machine learning algorithms to credit card fraud detection.

Chapter 9, *Building a Data-Driven Graph-Powered Application*, introduces some technologies and techniques that are useful for dealing with large graphs.

Chapter 10, *Novel Trends on Graphs*, introduces some novel trends (algorithms and applications) in graph machine learning.

To get the most out of this book

A Jupyter or a Google Colab notebook is sufficient to cover all the examples. For some chapters, Neo4j and Gephi are also required.

Software/Hardware covered in the book	OS Requirements
Python	Windows, macOS X, and Linux (any)
Neo4j	Windows, macOS X, and Linux (any)
Gephi	Windows, macOS X, and Linux (any)
Google Colab or Jupyter Notebook	Windows, macOS X, and Linux (any)

If you are using the digital version of this book, we advise you to type the code yourself or access the code via the GitHub repository (link available in the next section). Doing so will help you avoid any potential errors related to the copying and pasting of code.

Download the example code files

You can download the example code files for this book from GitHub at `https://github.com/PacktPublishing/Graph-Machine-Learning`. In case there's an update to the code, it will be updated on the existing GitHub repository.

We also have other code bundles from our rich catalog of books and videos available at `https://github.com/PacktPublishing/`. Check them out!

Download the color images

We also provide a PDF file that has color images of the screenshots/diagrams used in this book. You can download it here: `https://static.packt-cdn.com/downloads/9781800204492_ColorImages.pdf`.

Conventions used

There are a number of text conventions used throughout this book.

`Code in text`: Indicates code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles. Here is an example: "Mount the downloaded `WebStorm-10*.dmg` disk image file as another disk in your system."

A block of code is set as follows:

```
html, body, #map {
 height: 100%;
 margin: 0;
 padding: 0
}
```

When we wish to draw your attention to a particular part of a code block, the relevant lines or items are set in bold:

```
Jupyter==1.0.0
networkx==2.5
matplotlib==3.2.2
node2vec==0.3.3
karateclub==1.0.19
scipy==1.6.2
```

Any command-line input or output is written as follows:

```
$ mkdir css
$ cd css
```

Bold: Indicates a new term, an important word, or words that you see on screen. For example, words in menus or dialog boxes appear in the text like this. Here is an example: "Select **System info** from the **Administration** panel."

> **Tips or important notes**
> Appear like this.

Get in touch

Feedback from our readers is always welcome.

General feedback: If you have questions about any aspect of this book, mention the book title in the subject of your message and email us at `customercare@packtpub.com`.

Errata: Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you have found a mistake in this book, we would be grateful if you would report this to us. Please visit `www.packtpub.com/support/errata`, selecting your book, clicking on the Errata Submission Form link, and entering the details.

Piracy: If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at `copyright@packt.com` with a link to the material.

If you are interested in becoming an author: If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit `authors.packtpub.com`.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions, we at Packt can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about Packt, please visit `packt.com`.