Graph Data Management, Modeling, and Mining

Karthik Srinivasan
University of Kansas, USA

INTRODUCTION

A graph or a network is an abstract representation of a set of objects in which some pairs of objects are connected to each other. Graphs can be a powerful medium for representation of underlying data especially if it contains multiple linked entities. Since most of the practical datasets do not have independent and identically distributed (i.e., i.i.d.) observations, graphs can be effectively used to connect different entities described in the observations. For example, if a dataset contains transactions between sellers and buyers and in a market, a graph can be created such as that sellers and buys are connected to each other based on common transactions and a graph network can be formed.

Graph analytics is the systematic computational analysis of graph data. There are numerous applications of graphs across multiple disciplines including but not restricted to world wide web, social science, cybersecurity, healthcare, ecology, finance, entertainment, political science. In order to find applications of graph analytics, one may examine if the underlying data consists of entities that may be inter-connected to each other through one or more empirical relationships. Some examples of such relationships are friendship between users of a social media application, employees reporting to other employees in an organization, products purchased with complementary, products, etc. This chapter introduces the reader to the breadth of graph analytics tools, techniques, algorithms, and software. After reading this chapter, the reader should be able to identify problems that can use a network approach as well as develop corresponding graph-based analytics solutions. While topics in graph analytics such as graph mining (Gosnell & Broecheler, 2020), graph databases (Sasaki, 2018), and graph modeling (Kolaczyk & Csardi, 2020) have been independently examined in textbooks and research papers, this chapter summarizes these components of graph analytics demonstrating basic applications using examples. It serves as an introductory text for data science enthusiasts by providing an overview of different topics as well as directions for further enquiry.

BACKGROUND

A graph or a network is made up of vertices and edges. It is mathematically represented as $G(V,E)$, where V is the set of vertices and E is the edges. In converse, vertices and edges of a graph G may be represented as $V(G)$ and $E(G)$ respectively. An edge joins two vertices in a graph. Likewise, two vertices are said to be adjacent if and only if there is an edge between them. Two vertices are said to be connected if there is a path from one to the other via any number of edges.

DOI: 10.4018/978-1-7998-9220-5.ch121

Copyright © 2023, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

A vertex or a node is a single connection point in a graph. Vertices or nodes are entities such as people, products, biological cells, organizations which could be interconnected to each other in a particular configuration and the collection of such entities, and their interconnections constitutes the graph. Nodes are usually labeled but they could be unlabeled as well. An edge, link or relationship is a line segment that connects two nodes. A node without edges is permitted. However, an edge without nodes is not. Edges may have labels as well but are usually unlabeled. Nodes as well as edges can have their own attributes or properties.

Graphs can be further categorized into the type of nodes and the type of edges. Nodes may be single mode or multi-mode. Single mode networks, also called as one-mode or unipartite or homogenous networks have nodes of the same entity type (e.g., A friend network {John, Susan, George}). On the other hand, multi-mode or heterogeneous networks have nodes of multiple types (e.g., A customer-product-store network where customers {John, Susan, George} purchase one or more fruits {apples, oranges, bananas, strawberries} from stores {Target, Walmart}). A bipartite is a two-mode network (e.g., employer-employee network) that enjoys more attention than higher-mode networks as it has a wide range of applications including recommender engines (e.g., product recommendation in e-commerce, disease prognosis in patients, etc.). Nodes could be simple scalars or vectors with numeric or categorical attributes (e.g., {Gender, Age} of people in a friend network).

Edges can be either directed (e.g., a user following another user on Twitter) or undirected (e.g., two users connected as friends on Facebook). Edges can be either unweighted or weighted, such that they not only indicate the presence of a relationship between a pair of vertices, but also indicate the strength of the relationship in terms of a meaningful measure. Lastly, edges can be explicit based on relationship information provided in the context or implicit in which case the existence of an edge needs to be inferred. For example, a network with explicit edges is a friend-friend network where the relationship is the evidence of friendship whereas the contiguous-usa network (Weisstein, 2021) is an implicit network where the edges are inferred based on the fact that two U.S. states may share a border and therefore be geographically proximate.

Table 1 shows examples of different types of graphs categorized in terms of types of edges and types of nodes.

Table 1. Different types of graph and examples

Graph type	Example(s)
Undirected unweighted single mode network	Friendship network, contiguous USA network (Weisstein, 2021)
Undirected weighted network	Zachary karate network, Disease co-occurrence network (Srinivasan et al., 2018)
Directed unweighted network	Twitter following network, Paper citation network
Directed weighted network	Peer-to-peer lending market, International trade network
Bipartite network	User-item network, Patient-disease network (Liu et al., 2020)
Multimode network / Labeled property graph	Movies graph (Sasaki, 2018)

The Zachary karate club is a popular example of a weighted undirected graph and is shown in Figure 1. It has 34 nodes and 78 edges. The nodes are members of a university karate club and the edges indicate whether they interacted outside the club. The edge weights are the number of times the corresponding pair of members interacted. The data can be download using this link: http://vlado.fmf.uni-lj.si/pub/

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/graph-data-management-modeling-and-mining/317604

Related Content

Simple Linear Iterative Clustering (SLIC) and Graph Theory-Based Image Segmentation
Chiranji Lal Chowdhary (2021). *Handbook of Research on Machine Learning Techniques for Pattern Recognition and Information Security (pp. 157-170).*
www.irma-international.org/chapter/simple-linear-iterative-clustering-slic-and-graph-theory-based-image-segmentation/279910

A Review on Time Series Motif Discovery Techniques an Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques
Ramanujam Elangovanand Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 39-56).*
www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127

Shape-Based Features for Optimized Hand Gesture Recognition
Priyanka R., Prahanya Sriram, Jayasree L. N.and Angelin Gladston (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 23-38).*
www.irma-international.org/article/shape-based-features-for-optimized-hand-gesture-recognition/266494

Intelligent System for Credit Risk Management in Financial Institutions
Philip Sarfo-Manu, Gifty Siawand Peter Appiahene (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 57-67).*
www.irma-international.org/article/intelligent-system-for-credit-risk-management-in-financial-institutions/238128

Mental Health Through Biofeedback Is Important to Analyze: An App and Analysis
Rohit Rastogi, Devendra Kumar Chaturvediand Mayank Gupta (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning (pp. 402-423).*
www.irma-international.org/chapter/mental-health-through-biofeedback-is-important-to-analyze/263330