

Chapter 56

Predictive Analytics of Social Networks: A Survey of Tasks and Techniques

Ming Yang

Kansas State University, USA

William H. Hsu

Kansas State University, USA

Surya Teja Kallumadi

Kansas State University, USA

ABSTRACT

In this chapter, the authors survey the general problem of analyzing a social network in order to make predictions about its behavior, content, or the systems and phenomena that generated it. They begin by defining five basic tasks that can be performed using social networks: (1) link prediction; (2) pathway and community formation; (3) recommendation and decision support; (4) risk analysis; and (5) planning, especially causal interventional planning. Next, they discuss frameworks for using predictive analytics, availability of annotation, text associated with (or produced within) a social network, information propagation history (e.g., upvotes and shares), trust, and reputation data. They also review challenges such as imbalanced and partial data, concept drift especially as it manifests within social media, and the need for active learning, online learning, and transfer learning. They then discuss general methodologies for predictive analytics involving network topology and dynamics, heterogeneous information network analysis, stochastic simulation, and topic modeling using the abovementioned text corpora. They continue by describing applications such as predicting “who will follow whom?” in a social network, making entity-to-entity recommendations (person-to-person, business-to-business [B2B], consumer-to-business [C2B], or business-to-consumer [B2C]), and analyzing big data (especially transactional data) for Customer Relationship Management (CRM) applications. Finally, the authors examine a few specific recommender systems and systems for interaction discovery, as part of brief case studies.

DOI: 10.4018/978-1-4666-9562-7.ch056

1. INTRODUCTION: PREDICTION IN SOCIAL NETWORKS

Social networks provide a way to anticipate, build, and make use of links, by representing relationships and propagation of phenomena between pairs of entities that can be extended to large-scale dynamical systems. In its most general form, a social network can capture individuals, communities or other organizations, and propagation of everything from information (documents, memes, rumors) to infectious pathogens. This representation facilitates the study of patterns in the formation, persistence, evolution, and decay of relationships, which in itself forms a type of dynamical system, and also supports modeling of temporal dynamics for events that propagate across a network.

In this first section, we survey goals of predictive analytics using a social network, outline the specific tasks that motivate the use of graph-based models of social networks, and discuss the general state-of-the-field in data science as applied to prediction.

1.1 Overview: Goals of Prediction

In general, time series prediction aims to generate estimates for variables of interest that are associated with future states of some domain. These variables frequently represent a continuation of the input data, modeled under some assumptions about how the future data are distributed as a function of the history of past input, plus exogenous factors such as noise. The term *forecasting* refers to this specific type of predictive task. (Gershenfeld & Weigend, 1994) Acquiring the information to support this operation is known as *modeling* and frequently involves the application of machine learning and statistical inference. A further goal of the analytical process that informs this model is *understanding* the way in which a generative

process changes over time; in some scenarios, this means estimating high level parameters or especially structural elements of the time series model.

Getoor (2003) introduces the term *link mining* to describe a specialized form of data mining: analyzing a network structure to discover novel, useful, and comprehensible relationships that are often latent, *i.e.*, not explicitly described. Prototypical link mining tasks, as typified by the three domains that Getoor surveys, include modeling collections of web pages, bibliographies, and the spread of diseases. Each member of such a collection represents one entity. In the case of web page networks, links can be *outlinks* directed from a member page to another page, *inlinks* directed from another page to a member page, or *co-citation* links indicating that some page contains outlinks to both endpoints of a link. Bibliography or *citation networks* model paper-to-paper citations, co-author sets, author-to-institution links, and paper-to-publication relationships. Epidemiological domains are often represented using *contact networks*, which represent individual organisms (especially humans or other animals) using nodes and habitual or incidental contact using links. *Spread models* extend this graphical representation by adding information about incubation and other rates and time-dependent events.

Getoor and Diehl (2005) further survey the task of link mining, taxonomizing tasks into abstract categories such as *object-based*, *link-based*, and *graph-based*. Object-based tasks, used often in information retrieval and visualization, include ranking, classification, group detection (one instance of which is community detection), and identification (including disambiguation and deduplication). Link-based tasks, which we discuss in depth in this article, include the modeling task of *link prediction* – deducing or calculating the likelihood of a future link between two candidate entities, based on their individual attributes and mutual associations. Graph-based tasks include

35 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/predictive-analytics-of-social-networks/142667

Related Content

A Robust Multivariate Outlier Detection Method for Detection of Securities Fraud

M. Fevzi Esen (2020). *International Journal of Business Analytics* (pp. 12-29).

www.irma-international.org/article/a-robust-multivariate-outlier-detection-method-for-detection-of-securities-fraud/258268

Identifying Technical Requirements for a Mobile Business Analytics Application

Tuncay Bayrak (2021). *International Journal of Business Analytics* (pp. 91-103).

www.irma-international.org/article/identifying-technical-requirements-for-a-mobile-business-analytics-application/288060

Essential Mining Approaches to Problem Solving

Stephan Kudyba and Richard Hoptroff (2001). *Data Mining and Business Intelligence: A Guide to Productivity* (pp. 58-68).

www.irma-international.org/chapter/essential-mining-approaches-problem-solving/7505

Incorporating Text OLAP in Business Intelligence

Byung-Kwon Park and Il-Yeol Song (2012). *Business Intelligence Applications and the Web: Models, Systems and Technologies* (pp. 77-101).

www.irma-international.org/chapter/incorporating-text-olap-business-intelligence/58412

Forecasting Stock Returns with Artificial Neural Networks

Suraphan Thawornwong and David Enke (2004). *Neural Networks in Business Forecasting* (pp. 47-79).

www.irma-international.org/chapter/forecasting-stock-returns-artificial-neural/27244