Chapter VIII
Structures in Complex Bipartite Networks

Sune Lehmann
Northeastern University, USA & Harvard University, USA

ABSTRACT

A network structure of nodes and links is an informative way to study information systems. The network representation is valuable because it encodes the structure of the data. This chapter reviews recent advances in the field of network science with an emphasis on describing the structure of information networks. The author argues that bipartite networks constitute an important class of networks, and describes a method for detecting overlapping communities in bipartite networks. The author discusses the relevance of network communities to the future of organizing and understanding large datasets.

INTRODUCTION

In 1990, Tim Berners-Lee implemented the first successful communication between an HTTP client and server (CERN, 2008). Since then, the growth of the World Wide Web has been exponential. In 2008, the search engine Google increased their index to contain more than one trillion distinct pages (Alpert & Hajaj, 2008). In addition to the pages on the World Wide Web, there exists a rich ecology of servers with content that is not accessible to the search engines. This *deep web* is estimated to be orders of magnitude larger than the World Wide Web (Barker, 2004). Fifteen years ago, data was scarce and valuable, and acquisition of relevant datasets was one of the main challenges for academia and businesses alike. Today, the challenge is to filter, segment, and make sense of the prodigious amounts of available data.

Below, we take advantage of the fact that much of the data described above is *linked*. The pages of the World Wide Web are connected via hyperlinks. The Internet is the physical manifestation of the World Wide Web; this system, where routers and servers are the nodes and the physical connections between them are the links, forms a network that spans the globe. The databases that store the exa-bytes of information available to each user of the World Wide Web can be viewed as complex networks. Further, since information networks are often related (e.g. the internet acts as substrate for the world wide web), the various networks interact and influence the growth and structure of each other. The power of utilizing a network representation - of nodes and links - to describe complicated systems is that it allows us to investigate the underlying *structure* of large datasets. As it turns out, the self-organized networks described here possess structure on every level from small motifs involving only a few nodes and links through a meso-level of communities and modules that are all combined in a global organization.

In this chapter we discuss the structure of information networks. In the first part, different types of self-organized structure in complex networks are described, with a particular focus on network communities. The following part takes root in the fact that many information networks belong to a specific class of networks, called *bipartite* networks. In a bipartite network the nodes can be divided into two non-overlapping sets, where links must have one endpoint in each of the two node sets. In much of the previous work regarding complex networks, the bipartite nature of complex networks has been neglected. Here, we show that one discards important information by neglecting bipartite information. We then suggest a novel procedure for detecting communities that works directly on bipartite network data. The final part of the chapter discusses future trends in network science.

BACKGROUND

The science of networks is a relatively new field with roots in sociology, biology, mathematics, and physics. Physicists began thinking about the Laws of Networks around the same time as large databases became available via the Internet. Their way of thinking about networks was inspired by great advances in the field of non-equilibrium thermodynamics, made in the 1970's and 1980's. On the molecular level, nature tends to be uniform. Matter is made up from a huge number of particles that all behave accord-

Figure 1. Degree distributions. Panel (a) shows a Poisson distribution with mean >>1 (in this regime, the Poisson distribution tends towards the Normal distribution). In random networks, node degrees are distributed according to the Poisson/Normal distribution. Panel (b) displays a power-law distribution; many real-world networks have power-law degree distributions. When plotted on log-log axes, a power-law distribution forms a straight line.