Chapter XII

Web Graph Clustering for Displays and Navigation of Cyberspace

Xiaodi Huang
The University of Southern Queensland, Australia

Wei Lai
Swinburne University of Technology, Australia

ABSTRACT

This chapter presents a new approach to clustering graphs, and applies it to Web graph display and navigation. The proposed approach takes advantage of the linkage patterns of graphs, and utilizes an affinity function in conjunction with the k-nearest neighbor. This chapter uses Web graph clustering as an illustrative example, and offers a potentially more applicable method to mine structural information from data sets, with the hope of informing readers of another aspect of data mining and its applications.

INTRODUCTION

A graph is suitable for World Wide Web (WWW) navigation. Nodes in a graph can be used to represent URLs and edges between nodes represent links between URLs. We can look at the entire cyberspace of the WWW as one graph — a huge and dynamic growing graph. It is, however, impossible to display this huge graph on the computer screen.

Most current research interests are moving towards using “site mapping” methods (Chen, 1997; Maarek & Shaul, 1997) in an attempt to find an effective way of constructing a structured geometrical map for a single Website (a local map). This can guide the user...
through only a very limited region of cyberspace, and does not help the user in his/her overall journey through cyberspace.

Huang et al. (1998) proposed an online exploratory visualisation approach, which provides a major departure from traditional site-mapping methods. This approach does not pre-define the geometrical structure of a specific Website (a part of cyberspace), but incrementally calculates and maintains the visualisation of a small subset of cyberspace online corresponding to the change in the user’s focus. In other words, following the user’s orientation, a sequence of Web sub-graphs is automatically displayed with the smooth animation. This feature enables the user to logically explore the entire cyberspace without requiring the whole structure of cyberspace to be known.

In real applications, graphs may be huge in terms of the number of nodes and edges. Many graph drawing algorithms have been developed (Battista, 1998), but most of them have difficulty dealing with large graphs with thousands of nodes. Clustering graphs is one efficient method to draw large graphs even though other techniques exist, such as fisheye view, hyperbolic geometry (Burchard, 1995) and distortion-oriented presentation (Leung & Apperly, 1994). A clustered graph can significantly reduce visual complexity by replacing a set of nodes in a cluster with one abstract node. Moreover, a hierarchically clustered graph can find superimposed structures over the original graph through a recursive clustering process.

The Web graph has recently been used to model the link structure of the Web. The studies of such graphs can yield valuable insights into Web algorithms for crawling, searching and discovery of Web communities. This chapter proposes a new approach to clustering the Web graph. The proposed algorithm identifies a small subset of the graph as “core” members of clusters, and then incrementally constructs the clusters using a selection criterion. Two qualitative criteria are proposed to measure the quality of graph clustering. We have implemented our algorithm and demonstrated a set of arbitrary graphs with good results.

This chapter is organized as follows. The next section introduces the background of our prototype system of Web navigation using graph layout. We then discuss some issues on graph clustering, followed by our approach described in detail. We present an empirical evaluation of a set of arbitrary graph clustering by using our approach. After the applications are briefly provided, this chapter ends with the summary of our work.

**BACKGROUND**

**Examples of Web Graph Displays**

We first have a quick look at our Web graph display system. Our system can ensure that any online Web sub-graph, as shown in Figure 1, has no overlapping node images and fits in the window. Three kinds of modes for the user’s interaction are provided. In the mode LayoutAdjust, the user can adjust a Web sub-graph layout. If the user clicks a node in the Web sub-graph, the sub-graph of this node switches between invisibility and visibility. Figure 2 shows the result after the user clicks the nodes - Phone, Fax and Teaching in the graph shown in Figure 1. If the user clicks these nodes again, their sub-graphs would become invisible (i.e., they would disappear, as shown in Figure 1). In this way the user can make a node’s sub-graph visible or invisible by direct manipulation.
Related Content

Predicting Temporal Exceptions in Concurrent Workflows
www.igi-global.com/chapter/predicting-temporal-exceptions-concurrent-workflows/61907?camid=4v1a

A Linked Neighboring Leaves N-Tree to Support Distance Range Search
Faïza Najjar and Hassenet Slimani (2009). International Journal of Information Technology and Web Engineering (pp. 31-50).
www.igi-global.com/article/linked-neighboring-leaves-tree-support/4029?camid=4v1a

A Novel Approach to Design a Scalable Air Quality Monitoring System
www.igi-global.com/article/novel-approach-design-scalable-air/72991?camid=4v1a
Benefits and Challenges for Business Process Management in the Cloud
www.igi-global.com/chapter/benefits-and-challenges-for-business-process-management-in-the-cloud/140890?camid=4v1a