Graph Aggregation : Application to Social Networks

Amine Louati* Marie-Aude Aufaure**
Yves Lechevallier***

*University Paris-Dauphine - Lamsade,
Place du Maréchal de Lattre de Tassigny, F-75775 Paris 16 cedex, France
Amine.louati@dauphine.lamsade.fr
**Laboratoire MAS Ecole Centrale de Paris Grande Voie des Vignes
Chatenay-Malabry, France
Marie-Aude.Aufaure@ecp.fr,
***INRIA-Rocquencourt,Domaine de Voluceau 78150 Rocquencourt
Yves.Lechevallier@inria.fr

Abstract. In the enterprise context, people need to exploit and mainly visualize different types of interactions between heterogeneous objects. Graph model seems to be the most appropriate way to represent those interactions. However, the extracted graphs have in general a huge size which makes it difficult to analyze and visualize. An aggregation step is needed to have more understandable graphs in order to allow users discovering underlying information and hidden relationships between entities. In this work, we propose new measures to evaluate the quality of summaries based on an existing algorithm named k-SNAP that produces a summarized graph according to user-selected node attributes and relationships.

1 Introduction

Data manipulated in an enterprise context are structured data as well as unstructured content such as e-mails, documents, etc. Graphs are a natural way of representing and modeling such data in a unified manner (structured semi-structured and unstructured ones). The main advantage of such structure resides in its dynamic aspect and its capability to represent relations, even multiple ones, between objects. People need to visualize different types of interactions between heterogeneous objects (e.g. product and site, customers and products, people interaction like social networks, etc.). In order to analyze these interactions and facilitate their visualization, it is relevant to modulate such interaction by using a graph structure.

However, graphs extracted are often large, with thousands or even millions of nodes and edges. As a result, it is almost impossible to understand the information encoded in these graphs by mere visual inspection. In order, to facilitate the visualization and data interpretation, it seems interesting to perform an operation of summarization. The objective of graph summarization is to produce small and understandable summaries and can highlight communities in the network, which greatly facilitates the interpretation. Today, summarization has