Physica A 438 (2015) 48–55

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Attribute-based edge bundling for visualizing social networks

Lin Guo, Wanli Zuo *, Tao Peng, Binod Kumar Adhikari

College of Computer Science and Technology, Jilin University, Jilin 130000, China
Symbol Computation and Knowledge Engineer of Ministry of Education, Jilin University, Jilin 130000, China

HIGHLIGHTS

- Similarity analysis of edges.
- Aggregation of similar nodes.
- Link-based edge bundling.
- Conduct satisfactory experiments.

ARTICLE INFO

Article history:
Received 10 April 2015
Received in revised form 24 May 2015
Available online 3 July 2015

Keywords:
Social network
Edge bundling
Complex network
Feature extraction
Visual clutter
Graph visualization

ABSTRACT

Most nodes in complex networks have multiple attributes, which make them hard to analyze. Because general edge bundling algorithms fail to handle complex networks as a result of their intricate features, network simplification is extremely important. This paper proposes an attribute-based edge bundling algorithm that displays similar edges in nearby locations. Meanwhile, by analyzing complex networks at a community level, the overlapping clustering of nodes is well implemented, and better clustering effects can be achieved by grouping similar edges together. On the basis of datasets with different types and sizes, the experiments illustrate the simplification degree of the intricate graphs created by the algorithm proposed, which outperforms established competitors in correctness and effectiveness.

1. Introduction

Edge bundling is an image representation method that can adjust the locations and curvatures of geographically close edges and aggregate them together to alleviate clutter and reveal high-level edge patterns [1]. The existing edge bundling algorithms are limited to simplifying network structures by analyzing the locations of edges, regardless of detailed information about nodes. These methods can only make similar edges more compact in relation to each other at fixed positions. This paper largely extends the width and depth of edge bundling algorithms, which is no longer confined to geographical location analysis. The algorithm proposed modifies the shapes of edges according to attribute similarities; any two edges that are highly similar are made close; otherwise, they are not. Similar nodes can be aggregated simultaneously by dealing with edges so that both the inner and outer structures of different clusters are visualized clearly, without considering node features or comparing similarities between nodes. Under the hypothesis that the most fundamental feature of a node is its neighbors, the attribute of one edge embodies the attributes of the two nodes in the given edge. Therefore, the aggregation of similar nodes is an additional result of edge bundling.

* Corresponding author at: College of Computer Science and Technology, Jilin University, Jilin 130000, China.
E-mail address: wanli@jlu.edu.cn (W. Zuo).

http://dx.doi.org/10.1016/j.physa.2015.06.015
0378-4371/

The following is a comprehensive description of contents of this paper.

- Attribute-based edge bundling. By classifying the edges of one node according to attribute similarities, the identity of the node can be recognized. Thereafter, the curvatures and coordinates of edges can be optimized to achieve the purpose of bundling similar edges, and the nodes can be classified according to the identities.
- The node aggregation strategy. The optimization of node distribution is an additional result of edge bundling, which is feasible if the bundling is performed according to the similarities of attributes. The attribute of an edge in the complex network is simple, while the situation of a node is intricate because a node inherits all memberships of its links; therefore, link-based clustering algorithms are equal or superior to node-based processing approaches. By bundling similar edges together, similar nodes can be identified and the structures of the inner and outer of clusters can be visualized clearly.

The remainder of this paper is organized as follows: Section 2 focuses on background notions on some knowledge bases relate to our work. Section 3 proposes the concrete implementation strategy and the algorithm description in detail. Section 4 introduces the experimental results. Section 5 states the conclusions of research achievements.

2. Related work

The composite attributes of nodes in social networks result in the fact that most nodes belong to multiple communities simultaneously (different communities describe different social characteristics formed naturally through varied interests or relationships) [2]. A node belongs to multiple groups because individuals have families, friends, and colleagues, and a link exists for one dominant reason—two people that work together or have common interests are linked [3]. By classifying all the adjacent edges of a node, different roles of the node can be identified; furthermore, a node with a single role is classified into a particular community and a node occupying multiple roles are divided into corresponding clusters.

Edges are basic elements for encoding data using visualization techniques. The processing quality of edges will affect the quality of the final output. That is, the more complex the edges are, the more intricate the graph will be. The purpose of edge bundling is to group and transform similar edges into bundles in order to provide an abstract and uncluttered view of the original edge-cluttered visualization by using geometric graphs, trees, or parallel coordinate plots [4]. Edges represented by curves are close to each other. They are referred to as "compatible", and these compatible edges are then combined into a single bundle, sharing parts of their routes [1]. The output result of edge bundling is a simplified graph with reduced multiple relations and connections between nodes. The general edge bundling algorithms bundle spatially proximal edges, regardless of their directions or weights, and then reorient locations of similar edges based on distances. Edge closeness in a graph can be defined in many ways [5]: the tree-distance of edges in a hierarchy [6], the closeness of edges in a straight-line drawing of a graph [7,8], or a mix of graph-theoretic and image-space distances [9]. Edge bundling algorithms can be divided into the following categories: ink minimization algorithms [1,10], in which nodes are arranged in a circular pattern; parallel coordinate algorithms [11], which model edges in parallel coordinate plots as flexible springs; graph-based processing methods [6,8,12,13], which do not focus on putting nodes in specific locations; and geometry-based techniques [6,7,14], which build energy minimization systems by setting the midpoints of each segment as control points for reaching optimal edge-bundling results. Without any consideration of the features of nodes or edges, general edge bundling algorithms only emphasize relative geographic locations. Therefore, they deal with relatively narrow problems (nodes with no attributes or single attributes). As a result, most traditional edge bundling algorithms are incapable of handling complex networks, such as social networks.

The edge bundling algorithm proposed in this paper (denoted as EdgeB-Cluster) focuses on the attributes of nodes and edges with the purpose of solving more complex problems and expanding the range of application. This algorithm cannot only detect non-overlapping communities of edges but also overlapping clusters of nodes.

3. Edge bundling

EdgeB-Cluster iteratively classifies edges of each node, so the edges of one node are processed at a time and the graph of the social network is constructed layer-by-layer. The description of the graph is: $E = \{\bigcup p_i = \{n | n \in Adjc(ego_i)\}, ego_{i+1} \in N, i \in [0, k-1]\}$, where i represents the ith layer, n represents the adjacent node collection of node ego_i, k represents the number of iterations, and $Adjc(x)$ gets the adjacent nodes of the given node x. Each layer regards ego_i as a center node, adds all undetected adjacent nodes into graph E, and finally selects ego_{i+1} from node list N.

3.1. Analysis of edge features

Suppose that the only knowledge valuable in the complex network is the topological structure, the most fundamental characteristic of a node is the neighbors. Because the collection of the adjacent nodes can express the features of the center node, the feature of node p can be denoted as: $Node(p) = \{Adjc(p)\}$. Similarly, the information valuable of an edge is the two nodes linked, so the union set of feature lists of the two nodes represents the feature of the edge. The feature of an edge is denoted as: $Edge(R_{ij}) = \{Node_i \cup Node_{(j)}\}$, which embodies all nodes of the given link R_{ij}.