Information Sciences 260 (2014) 185–199

Contents lists available at ScienceDirect

Information Sciences

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

Fast layout computation of clustered networks: Algorithmic advances and experimental analysis ☆

Walter Didimo *, Fabrizio Montecchiani

Dipartimento di Ingegneria Elettronica e dell'Informazione, Università degli Studi di Perugia, Italy

ARTICLE INFO

Article history:
Received 8 June 2012
Received in revised form 11 June 2013
Accepted 28 September 2013
Available online 9 October 2013

Keywords:
Graph drawing
Network visualization
Hierarchical clustering
Multi-scale algorithm
Force-directed algorithm
Treemap algorithm

ABSTRACT

The visual analysis of large and complex relational data sets is a growing need in many application domains, such as social sciences, biology, computer networks, and software engineering. In this respect, the capability of quickly computing two-dimensional layouts of hierarchically clustered networks plays an important role and should be a major requirement of many graph visualization systems. We present algorithmic and experimental advances on the subject, namely: (i) we propose a new drawing algorithm that combines space-filling and fast force-directed methods; it runs in $O(n \log n + m)$ time, where n and m are the number of vertices and edges of the network, respectively. This running time does not depend on the number of clusters, thus the algorithm guarantees good time performance independently of the structure of the cluster hierarchy. As a further advantage, the algorithm can be easily parallelized. (ii) We discuss the results of an experimental analysis aimed at understanding which clustering algorithms can be used in combination with our visualization technique to generate better quality drawings for small-world and scale-free networks of medium and large size. As far as we know, no previous similar experiments have been done to this aim.

1. Introduction

One of the major requirements of modern information science technologies should be the capability of handling, analyzing, and making sense of increasing volumes of data. This requirement poses severe challenges when the data are related to each other so to form a network, which is in fact the case of most data arising from a variety of application domains, including social sciences (e.g., collaboration networks, friendship networks), biology (e.g., gene interaction networks, biological pathways), computer networks (e.g., routing message exchange), and software engineering (e.g., object collaboration networks, procedure call networks). In this respect, the visual analysis of large and complex relational data sets is recognized as one of the most effective methodologies (see, e.g., [15,30]). As a consequence, the design of algorithms for the automatic layout of large networks has received considerable attention in the last decade and it is currently one of the most relevant topics in graph visualization. Often, the network to be analyzed is structured by users or by automatic procedures into hierarchical clusters (e.g., based on its semantic or topological properties), and in those cases an effective visualization algorithm should clearly convey the network structure in a reasonable computation time. Other times, the input network is not already clustered, but it exhibits an intrinsic structure of communities (i.e., dense subgraphs connected to each other by few edges),

☆ Research supported in part by the MIUR project AlgoDEEP prot. 2008TFBWL4. A preliminary extended abstract of this paper has been presented at the 16th International Conference Information Visualisation (IV'12).

* Corresponding author. Tel.: +39 0755853680; fax: +39 0755853654.
E-mail addresses: walter.didimo@unipg.it (W. Didimo), fabrizio.montecchiani@gmail.com (F. Montecchiani).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.09.048

as for many small-world and scale-free networks [39]; also in those cases, the layout should support users in community detection.

We study the problem of automatically computing 2D layouts of medium and large hierarchically clustered networks, whose cluster hierarchy is given as part of the input and cannot be changed. We aim at guaranteeing the following properties: the layout fits into a rectangular area of given aspect ratio; each cluster is completely contained in a corresponding rectangular region; inclusions between the cluster regions reflect the inclusion relationships between their corresponding clusters. Additionally, we take into account the following readability metrics, also called *aesthetic criteria* (see, e.g., [12,31]): the area of each cluster region should be proportional to the number of vertices inside it (or to the total weight of the vertices in the cluster, if the network has weighted vertices); vertices inside the deepest clusters should be uniformly distributed in the corresponding cluster regions, and large empty spaces in the drawing area should be avoided; the total number of edge crossings and the number of edges that traverse clusters should be minimized; symmetries in the graphs inside the deepest clusters should be displayed where possible. The original contribution of this paper is twofold:

(i) We propose a new drawing algorithm that computes 2D layouts of hierarchically clustered networks, according to the constraints and esthetics described above (Section 3). It combines space-filling and force-directed techniques in a new fashion, and can be easily parallelized. By exploiting and engineering fast popular layout techniques and efficient data-structures, the new algorithm runs in $O(n \log n + m)$ time, where n and m are the number of vertices and edges of the network, respectively. This theoretical bound does not depend on the number of clusters, thus the algorithm guarantees good time performance independently of the structure of the cluster hierarchy. Moreover, although the computed drawing is mainly conceived to provide an overview of the whole network (where all vertices and edges are visible at once), it can also be effectively adopted for interactive analysis based on a drill-down exploration of the clusters (see, e.g., Fig. 1(a and b)).
(ii) We present the results of an experimental study aimed at understanding which clustering algorithms can be used in combination with our technique, in order to generate better quality drawings for medium and large networks having small-world and scale-free structure (Section 4). Our benchmark consists of both artificial and real-world networks taken from different application domains. The experimental results suggest that a recent approach specifically tailored for community detection in small-world and scale-free networks is in fact suitable also for the visual representation of this kind of networks in terms of crossing reduction, within our visualization constraints. As far as we know, no previous similar experiments have been done to this aim. In the experiments, we mainly focused on measuring the number of crossings, motivated by several studies showing that edge crossings affect the drawing readability much more than other esthetic criteria (see, e.g., [42–44,50]).

Section 2 discusses previous work related to our research. Drawing examples are shown in Section 5. Conclusions and future research proposals are given in Section 6.

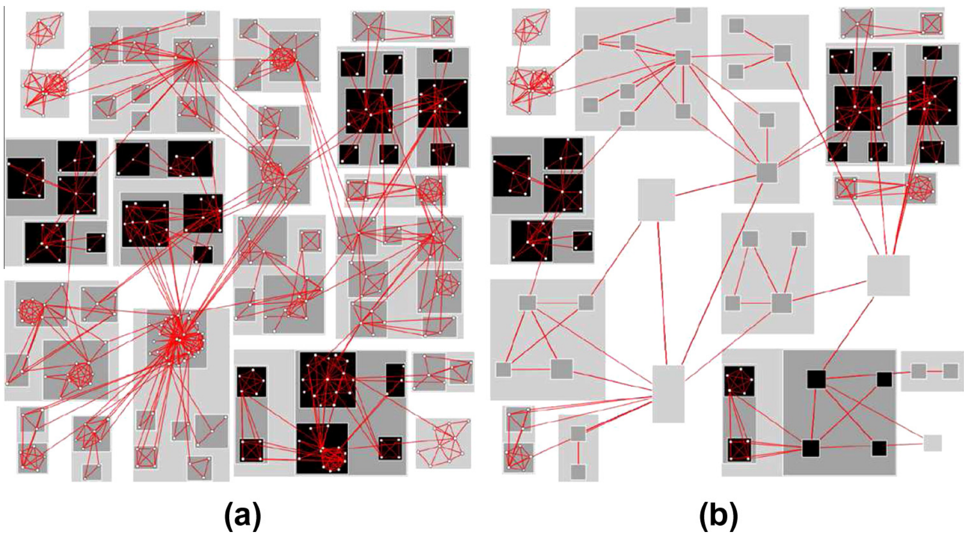

Fig. 1. (a) A clustered drawing of a social network computed with our algorithm, which gives an overview of the whole network; using a force-directed algorithm helps in symmetry display and uniform vertex distribution for the subgraphs in the deepest clusters of the hierarchy; it also reduces the size of empty spaces. (b) The same drawing used for an interactive drill-down exploration of the clusters; some clusters are contracted while others are expanded. Thanks to the geometric information already computed, cluster contractions or expansions only require quick local changes in the drawing.