Visual Analytics of Large Bipartite Networks assisted by Multilevel Strategies*

Renato Fabbri[0000−0002−9699−629X], Alan Valejo, and Maria Cristina Ferreira de Oliveira[0000−0002−4729−5104]

University of São Paulo, São Carlos SP, BR
renato.fabbri@gmail.com, alanvalejo@gmail.com, cristina@icmc.usp.br
http://conteudo.icmc.usp.br/pessoas/cristina/

Abstract. It is a well established fact that bipartite, or two-layer, networks are pervasive in model real-world phenomena and that they play fundamental roles in graph theory. Multilevel strategies have been developed for optimization tasks, and for the visualization of simple ("unipartite") networks, but their employment for visualizing bipartite networks were not found by the authors. In this work, we present advances in the use of multilevel strategies for the visualization of bipartite networks, allowing interactive and intuitive navigation of such structures and visual mappings of large datasets. More specifically, we developed a visual analytics web interface in which a parametrizable simplification of bipartite networks are obtained through the application of coarsening algorithms. The resulting networks are then presented to the user, providing a genuine route for the "overview first - focus on demand" process on the analysis of the underlying data, in which the analyst selects supervertices or whole network sectors for more detailed observation, i.e. performs requests for the interface to display specific structures in less simplified settings. Moreover, the application is useful for the development multilevel strategies e.g. by the specification of vertices to guide the coarsening processes and the examination of the resulting multilevel hierarchy.

Keywords: Network visualization · Multilevel strategies · Visual analytics · Big data · Complex networks · Data visualization.

1 Introduction

The visualization of large-scale networks poses challenges both in terms of computational costs and of effective presentation of the information for the user [19, 18]. These issues may be aggravated in the case of bipartite networks, due to their sparsity and topological complexities [21]. Bipartite networks are comprised of two partitions of nodes, called "layers", and links are not incident between nodes in the same partition. Such network type arises very often and naturally from the representation of relations among two kinds objects, e.g. documents and terms or authors [16, 14, 7], or patient and gene [13]. Furthermore, real-world

* Supported by FAPESP, project 2017/05838-3

networks are often bipartite, and most unipartite networks are projections of bipartite networks or may be considered as exhibiting bipartite properties [8, 9]. In order to assist the visualization and navigation of large networks, one possibility is the use of multilevel strategies, which consist on the employment of incremental coarsening of the original network to obtain a sequence of simplified representations. Multilevel strategies are most traditionally used for executing complex algorithms on large-scale networks by the application of the algorithm on a smaller version of the network [22, 15]. Their employment of multilevel strategies for the visualization of simple (i.e. "unipartite") networks have been reported, but their exploitation for visualizing bipartite networks was not found by the authors. Accordingly, we present a system for the visualization of bipartite networks using multilevel strategies developed for bipartite networks. The system consists on presented a simplified version of the network for the user, which then requests for supervertices (or collections of them) to be uncoarsened and presented in more detail.

This paper is organized as follows: in Section 1.1 the related work is examined, while in 1.2 are selected remarks about the vocabulary. The method is delineated in Section 2 and the software implementation is then described in Section 3. Results and discussion are in 4. Finally, Section 5 holds concluding and further work statements.

1.1 Related work

Multilevel strategies have been employed to visualize unipartite networks [11, 23, 12, 10, 6, 2, 20]. Also, the aggregation of clusters have been reported, and comprises an approach that resembles the coarsening procedure in creating simplified representations of the original network [1, 3, 24, 4, 5, 17]. Even so, the authors are not aware of previous reports on the use of multilevel strategies for the visualization and navigation of bipartite networks. Furthermore, only [1, 3, 24, 4, 17] reports on a system for visual analytics, i.e. the other works do not address data analysis by interactive visual interfaces, including all multilevel strategies for visualization.

1.2 Nomenclature remarks

The main vocabulary issue that arises in the context of this article is between level and layer. Layers are the sets of a bipartite network, while levels are the sequence of coarsened networks. Also, simple or homogeneous networks are opposed to heterogeneous networks, in which there are more then one type of node. Bipartite networks may be regarded as an elementary case of heterogeneous networks, but it has one additional restriction: only nodes of different types are connected. Another usual concept in this context is that of a multilayer network, in which there are layers, i.e. partitions of nodes, in addition to nodes and links. Bipartite networks, in this case, are multilayer networks with only two layers and no intra-layer links, i.e. all the links are inter-layers. Most importantly, there are many synonyms which are found in this context, e.g. supernodes are also called metanodes or supervertices. A thorough exposition of the vocabulary are beyond the scope of this article, but some attention for the issue is helpful to assist searches and newcomers.

2 Method description

2.1 Fundamental concepts

A bipartite network $G = (V, E)$ consists in a set V of vertices which is partitioned in two subsets with no links between nodes in the same set, i.e. $\exists\, V_1, V_2 :\ V_1 \cup V_2 = V,\ V_1 \cap V_2 = \emptyset$, and $E \subseteq V_1 \times V_2$. The variable names are borrowed from more traditional Graph theory, with Vertices (nodes) and Edges (links).

A multilevel strategy consists in obtaining a hierarchy of coarsened networks G_l, l integer and $l \in [0, L-1]$ where G_0 is the original network. The coarsening procedures requires two algorithms, the *matching*, that defines the nodes to be collapsed, and the *contraction*, which builds the reduced representation from the matched nodes. There are several coarsening algorithms reported in the scientific literature [], a few of them developed for bipartite networks []. A discussion of these algorithms is found on the bibliography [] and is beyond the scope of this article. Most importantly, we are interested in the MLPB algorithm, not only because it is suited for bipartite networks, but because it has yielded simplified networks which carries essential topological features of the original networks, and that result is perceived is the visualization as shown in Sections 2, 3 and 4.

Visual analytics is the scientific field dedicated to analytical reasoning assisted by interactive visual interfaces []. Therefore, the area is specially concerned with coupling interactive visual representations with sense and decision making. Of special relevance for the present work, the techniques are most often employed to amplify human capabilities in specific ways, which includes reducing the search space, enhancing the recognition of patterns and the inference of relationships, and providing a manipulable medium for the exploration of the information of interest [].

2.2 The navigation pathway

Fig. 1: .

3 Software implementation

4 Results and discussion

5 Conclusions and further work

References

1. Abello, J., Van Ham, F., Krishnan, N.: Ask-graphview: A large scale graph visualization system. IEEE transactions on visualization and computer graphics **12**(5), 669–676 (2006)
2. Archambault, D., Munzner, T., Auber, D.: Topolayout: Multilevel graph layout by topological features. IEEE transactions on visualization and computer graphics **13**(2), 305–317 (2007)
3. Archambault, D., Munzner, T., Auber, D.: Grouseflocks: Steerable exploration of graph hierarchy space. IEEE transactions on visualization and computer graphics **14**(4), 900–913 (2008)
4. Batagelj, V., Brandenburg, F.J., Didimo, W., Liotta, G., Palladino, P., Patrignani, M.: Visual analysis of large graphs using (x, y)-clustering and hybrid visualizations. IEEE transactions on visualization and computer graphics **17**(11), 1587–1598 (2010)
5. Dias, M.D., Mansour, M.R., Dias, F., Petronetto, F., Silva, C.T., Nonato, L.G.: A hierarchical network simplification via non-negative matrix factorization. In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 119–126. IEEE (2017)
6. Frishman, Y., Tal, A.: Multi-level graph layout on the gpu. IEEE Transactions on Visualization and Computer Graphics **13**(6), 1310–1319 (2007)
7. Grujić, J.: Movies recommendation networks as bipartite graphs. In: International Conference on Computational Science. pp. 576–583. Springer (2008)
8. Guillaume, J.L., Latapy, M.: Bipartite structure of all complex networks. Information processing letters **90**(5), 215–221 (2004)
9. Guillaume, J.L., Latapy, M.: Bipartite graphs as models of complex networks. Physica A: Statistical Mechanics and its Applications **371**(2), 795–813 (2006)
10. Hachul, S., Jünger, M.: Drawing large graphs with a potential-field-based multilevel algorithm. In: International Symposium on Graph Drawing. pp. 285–295. Springer (2004)
11. Harel, D., Koren, Y.: A fast multi-scale method for drawing large graphs. In: International symposium on graph drawing. pp. 183–196. Springer (2000)
12. Hu, Y.: Efficient, high-quality force-directed graph drawing. Mathematica Journal **10**(1), 37–71 (2005)
13. Hwang, T., Sicotte, H., Tian, Z., Wu, B., Kocher, J.P., Wigle, D.A., Kumar, V., Kuang, R.: Robust and efficient identification of biomarkers by classifying features on graphs. Bioinformatics **24**(18), 2023–2029 (2008)
14. Newman, M.E.: Scientific collaboration networks. i. network construction and fundamental results. Physical review E **64**(1), 016131 (2001)
15. Noack, A., Rotta, R.: Multi-level algorithms for modularity clustering. In: International Symposium on Experimental Algorithms. pp. 257–268. Springer (2009)
16. de Paulo Faleiros, T., Rossi, R.G., de Andrade Lopes, A.: Optimizing the class information divergence for transductive classification of texts using propagation in bipartite graphs. Pattern Recognition Letters **87**, 127–138 (2017)

17. Perrot, A., Auber, D.: Cornac: Tackling huge graph visualization with big data infrastructure. IEEE transactions on big data (2018)
18. Staudt, C.L., Sazonovs, A., Meyerhenke, H.: Networkit: A tool suite for large-scale complex network analysis. Network Science **4**(4), 508–530 (2016)
19. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In: Proceedings of the 24th international conference on world wide web. pp. 1067–1077. International World Wide Web Conferences Steering Committee (2015)
20. Toosi, F.G., Nikolov, N.S.: Vertex-neighboring multilevel force-directed graph drawing. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 002996–003001. IEEE (2016)
21. Valejo, A., Faleiros, T., Oliveira, M.C.F.d., Lopes, A.d.A.: A coarsening method for bipartite networks via weight-constrained label propagation (2019)
22. Valejo, A., de Oliveira, M.C.F., Geraldo Filho, P., de Andrade Lopes, A.: Multilevel approach for combinatorial optimization in bipartite network. Knowledge-Based Systems **151**, 45–61 (2018)
23. Walshaw, C.: A multilevel algorithm for force-directed graph drawing. In: International Symposium on Graph Drawing. pp. 171–182. Springer (2000)
24. Wong, P.C., Mackey, P., Cook, K.A., Rohrer, R.M., Foote, H., Whiting, M.A.: A multi-level middle-out cross-zooming approach for large graph analytics. In: 2009 IEEE Symposium on Visual Analytics Science and Technology. pp. 147–154. IEEE (2009)