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 for 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 further development multilevel strategies themselves 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 to 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

* Supported by FAPESP, project 2017/05838-3

or authors [16, 14, 7], or patient and gene [13]. Furthermore, real-world networks are often bipartite, and most unipartite networks are projections of bipartite networks or exhibit 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]. The 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, as described in the next subsection. Accordingly, we present a system for the visualization of bipartite networks using multilevel strategies developed for bipartite networks. The system consists on presenting a simplified version of the network to the user, which then requests for supervertices (or collections of them) to be uncoarsened and visualized 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 concluions and further work envisioned.

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 suitable 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 (two) node partitions of a bipartite network, while levels comprise 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 is 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 (network) theory, with Vertices (nodes) and Edges (links). One may regard the network as $G = (V, E, \sigma, \omega)$, $\sigma : V \to \mathbb{R}^*$ and with $\omega : E \to \mathbb{R}^*$, where $\sigma(v)$ is the weight of the node v and $\omega(u, v)$ is the weight of the link (u, v).

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 and where $|V_i| \leq |V_{i+1}|$. 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 []. An exposition about these algorithms is found on the bibliography [] and is beyond the scope of this article. Most importantly, we are interested in coarsening suited for bipartite networks, as it has yielded simplified networks which present essential topological features of the original networks, and such result is perceived in the visualization as hinted inc̃itealan2 and further 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 Multilevel coarsening of bipartite networks

Before [22], multilevel approaches were not directly aplicable to bipartite networks. Such work introduced novel and efficient matching and coarsening algorithms, and scrutinized their validity for solving optimization problems, dimensionality reduction, and in the preservation of essential topological features of the original network. We here describe the outline of such procedures in the context of our visual analytics contribution.

The multilevel optimization is a metaheuristics used to guide, modify and potentially fix a solution obtained from a target algorithm. It is divided into three phases: coarsening, solution finding and uncoarsening, where the solution found in the coarsest network G_L is projected back to the original, uncoarsened, network G_0. Figure 1 illustrates such process.

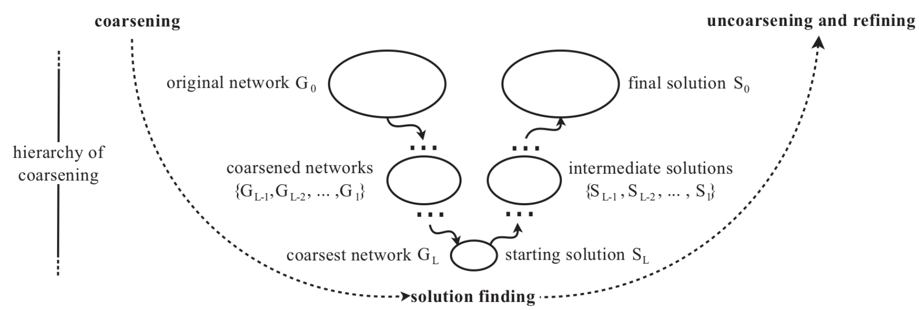

Fig. 1: Phases of the multilevel optimization. A sequence of incrementally reduced networks is obtained in the coarsening phase. An initial solution found at the coarsest network, in the solution finding phase, is projected back to the initial, uncoarsened, network, in the uncoarsening phase. Further details are given in Section 2.2 and specificities related to these phases when applied to visual analytics are in Sections 2.3 and 2.4.

The coarsening phase results in a sequence of networks G_l from the initial network G_0, on multiple levels of details. The process is carried out by two algorithms: *matching*, which decides which nodes to merge, and *contracting*, which achieves the reduced representation from the network and the matching. In general, pairs of nodes are selected to be merged into supernodes, and most often a matching M consists in a set of non-adjacent links, i.e. $\forall\, l_1, l_2 \in M, u \in l_1 \Rightarrow u \notin l_2$. The heavy-edge mathing, for example, is a matching that fits this canonical description in attempting to maximize total matching weight, but the matching may not satisfy such restrictions. Selecting clicks and other larger node sets to be merged into supernodes are possibilites under development. For bipartite network, we use the algorithms provided by [22] in which two restrictions are imposed to the matching:

– A node may only match nodes on the same layer.
– A node may only match nodes that are reachable by two successive links (i.e. the closest possible nodes on the same layer).

The matching is followed by the contraction of the network into a coarser form. Typically, the nodes matched are merged into a supernode with the weight equal to the sum of the weight of nodes it contains, and the links incident to such nodes are joined into superlinks among the supervertices, also with total weight equal to the sum of the weight of the links merged. The solution finding is usually reasoned about in terms of the computational cost, being much alleviated in the reduced network. The uncoarsening is usually performed through the complete hierarchy, from the coarsest to the original network, with successive refinement of the solutions to avoid local minima and improve solution quality. These two phases are substantially different when the multilevel strategy is applied to visualization, which motivated their separate exposition in the next subsection.

2.3 Network visualization assisted by multilevel strategies

In applying the multilevel framework to visualization, the target algorithm is the visual mapping of the network, i.e. the solution finding phase is the visualization of the network by means of the reduced representation. The uncoarsening is performed only on demand, through user requests to visualize network regions with more detail, thus avoiding unecessary complexity and, most importantly, avoiding to overload the visualization with information beyond the cognitive convenience of the user and the computational power that the machine being used is able to provide for real-time interactivity.

The interactivity is crucial for a number of reasons: the definition of the uncoarsening phase desired; for the navigation of the network, including the access to metadata and changes to the visualization achieved; for uncoarsening supernodes; and for tunning the achieved visual mapping status, such as by resizing and moving nodes. The overall procedure is delineated in Algorithm 1 and further detailed in Sections 2.4 and 3.

2.4 The navigation pathway

The abundance of information within large networks makes pertinent the application of the "visual information-seeking mantra", also known as Shneiderman's mantra: *overview first, zoom and filter, then details-on-demand*. This mantra comprises a number of visual design guidelines, such as the details-on-demand technique, and provides a very acknowledged framework for designing information visualization applications. Accordingly, our idealized exploration pathway starts with an overview, achieved by the mapping the coarsest representation of the network through standard network layout algorithms for node-link diagrams. The user may then zoom into specific regions of the network, and then request details by a number of operations, most importantly:

- reposition nodes, exposing linking patterns.
- request metadata, such as number of children, the parent in a subsequent level, or the number of neighbors.
- request visual mapping of network features, such as node size related to number of neighbors or the number of child nodes it represents.
- request the uncoarsening of specific supernodes or any arbitrary group of supernodes.

Input:
bipartite network: G
maximum number of levels: $L \in [0, n] \subset \mathbb{Z}$
reduction factor: $rf \in (0, 0.5] \subset \mathbb{R}$.
layers to be coarsened: $layers \in 1, 2$
user command given through the visual interface: C
Output:
Visual mapping of the network: V

```
i ← 1;
while i ≤ layers do
    l ← 1;
    while l ≤ L do
        M ← Matching_b(G_l, i, rf);
        G_{l+1} ← Contracting_b(G_l, M);
        increment l;
    end
    increment i;
end
V ← map_to_screen(G_L);
while 1 do
    C ← user command;
    V ← transform_visual_mapping(V, C);
end
```

Algorithm 1: An algorithmic description of the multilevel strategy adapted for the visualization of bipartite networks. The routines **Matching**$_b$ and **Contracting**$_b$ are any of the matching or contracting algorithms suited for bipartite networks, such as described in [22] which are only ones currently available to the knowledge of the authors. The **map_to_screen** routine is performed through the use of a network layout algorithm and then the rendering of the network to the screen. The visual mapping may then be transformed by the user by requesting uncoarsening of specific supernodes, or by other commands not specific to the multilevel strategy, such as requesting metadata exposition, changes to the position of nodes, the color or transparency of nodes and links, zoom and pan, or any other operation defined in Section 3.

Other operations are convenient to make the visual mapping adequate for the diverse settings possible: network size, open supervertices, levels exposed, and topological features such as commnity structures.

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)