



Multilevel Coarsening for Interactive Visualization of Large Bipartite Networks

Alan Demétrius Baria Valejo^{1*}, Renato Fabbri², Alneu de Andrade Lopes², Liang Zhao³ and Maria Cristina Ferreira de Oliveira²

¹ Department of Computing, Federal University of São Carlos, São Carlos, Brazil, ² Department of Computer Science, Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, Brazil, ³ Department of Computing and Mathematics, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto (FFCLRP), University of São Paulo, Ribeirão Preto, Brazil

OPEN ACCESS

Edited by:

Henrique Ferraz de Arruda,
ISI Foundation, Italy

Reviewed by:

Dakota Scott Murray,
Indiana University, United States
Guoqiang Liang,
Beijing University of Technology, China

*Correspondence:

Alan Demétrius Baria Valejo
alanvalejo@ufscar.br

Specialty section:

This article was submitted to
Research Assessment,
a section of the journal
Frontiers in Research Metrics and
Analytics

Received: 14 January 2022

Accepted: 16 May 2022

Published: 16 June 2022

Citation:

Valejo ADB, Fabbri R, de Andrade Lopes A, Zhao L and Oliveira MCF (2022) Multilevel Coarsening for Interactive Visualization of Large Bipartite Networks. *Front. Res. Metr. Anal.* 7:855165. doi: 10.3389/frma.2022.855165

Bipartite networks are pervasive in modeling real-world phenomena and play a fundamental role in graph theory. Interactive exploratory visualization of such networks is an important problem, and particularly challenging when handling large networks. In this paper we present results from an investigation on using a general multilevel method for this purpose. Multilevel methods on networks have been introduced as a general approach to increase scalability of community detection and other complex optimization algorithms. They employ graph coarsening algorithms to create a hierarchy of increasingly coarser (reduced) approximations of an original network. Multilevel coarsening has been applied, e.g., to the problem of drawing simple (“unipartite”) networks. We build on previous work that extended multilevel coarsening to bipartite graphs to propose a visualization interface that uses multilevel coarsening to compute a multi-resolution hierarchical representation of an input bipartite network. From this hierarchy, interactive node-link drawings are displayed following a genuine route of the “overview first, zoom and filter, details on demand” visual information seeking mantra. Analysts may depart from the coarsest representation and select nodes or sub-graphs to be expanded and shown at greater detail. Besides intuitive navigation of large-scale networks, this solution affords great flexibility, as users are free to select different coarsening strategies in different scenarios. We illustrate its potential with case studies involving real networks on distinct domains. The experimental analysis shows our strategy is effective to reveal topological structures, such as communities and holes, that may remain hidden in a conventional node-link layout. It is also useful to highlight connectivity patterns across the bipartite layers, as illustrated in an example that emphasizes the correlation between diseases and genes in genetic disorders, and in a study of a scientific collaboration network of authors and papers.

Keywords: complex network visualization, bipartite network visualization, exploratory network visualization, multilevel network visualization, multilevel network coarsening

1. INTRODUCTION

Bipartite networks are a special type of network in which the set of nodes is split into two independent partitions called “layers” (they are hence called “two-layer” networks) and the links connect nodes from different layers. They emerge in important data mining problems that require modeling relations between two types of entities, e.g., documents and terms, papers, and authors (Newman, 2001; Grujić, 2008; Faleiros et al., 2017), or patients and genes (Hwang et al., 2008). Some authors even argue that many real-world unipartite networks are actually projections of bipartite networks (Guillaume and Latapy, 2004, 2006).

Exploratory visualization of network topology plays a relevant role in many data mining and optimization problems modeled as networks. Effectively conveying the relevant topological information of large-scale networks is challenging in interactive visualization (Tang et al., 2015; Staudt et al., 2016), and even more so in handling bipartite networks, in view of their peculiar organization and inherent topological complexity. A few recent contributions discuss strategies for interactive visualization of such networks (Chan et al., 2018; Garcia-Algarra et al., 2018; Pezzotti et al., 2018; Steinbock et al., 2018; Sun et al., 2019; Zhao et al., 2019; Waldner et al., 2020), as further discussed in Section 3.

Analysis assisted by interactive visualization is arduous because human perceptual abilities are limited to rather small network sizes even when executing simple tasks such as locating a node or finding the links between a pair of nodes (Ghoniem et al., 2005). As network sizes increase, even simple data exploration tasks become increasingly difficult. Furthermore, besides the cognitive burden on users attempting to grasp both local and global information, a heavy rendering load impairs smooth navigation on node-link views (Ghoniem et al., 2005). Limitations of node-link representations encouraged many research efforts on devising alternative visualization techniques, with relative success (Nobre et al., 2020; Di Giacomo et al., 2021). Yet, node-link views remain a classic approach for interactive visualization of networks, even if to be used in combination with other techniques (Di Giacomo et al., 2021).

In order to reduce clutter and obtain representations more amenable to user interaction, general-purpose interactive visualization of large unipartite networks often relies on node aggregation into clusters or communities (Abello et al., 2006; Archambault et al., 2008; Wong et al., 2009; Batagelj et al., 2010; Dias et al., 2017b; Perrot and Auber, 2018) (see Section 3). Nevertheless, a single level partition of the set of nodes may prove ineffective, as important topological structures often manifest themselves at multiple observation levels. Indeed, alternative solutions have been introduced that rely on hierarchical algorithms for clustering or community detection to obtain a hierarchical aggregation to support user navigation and exploratory tasks. Thus, it is possible to interact with higher-level representations in order to make early assessments before proceeding to detailed investigation. This strategy is likely to be more effective than continuously rendering the full network in situations where the user focus is directed to a specific region. A hierarchical network model offers a natural support to the

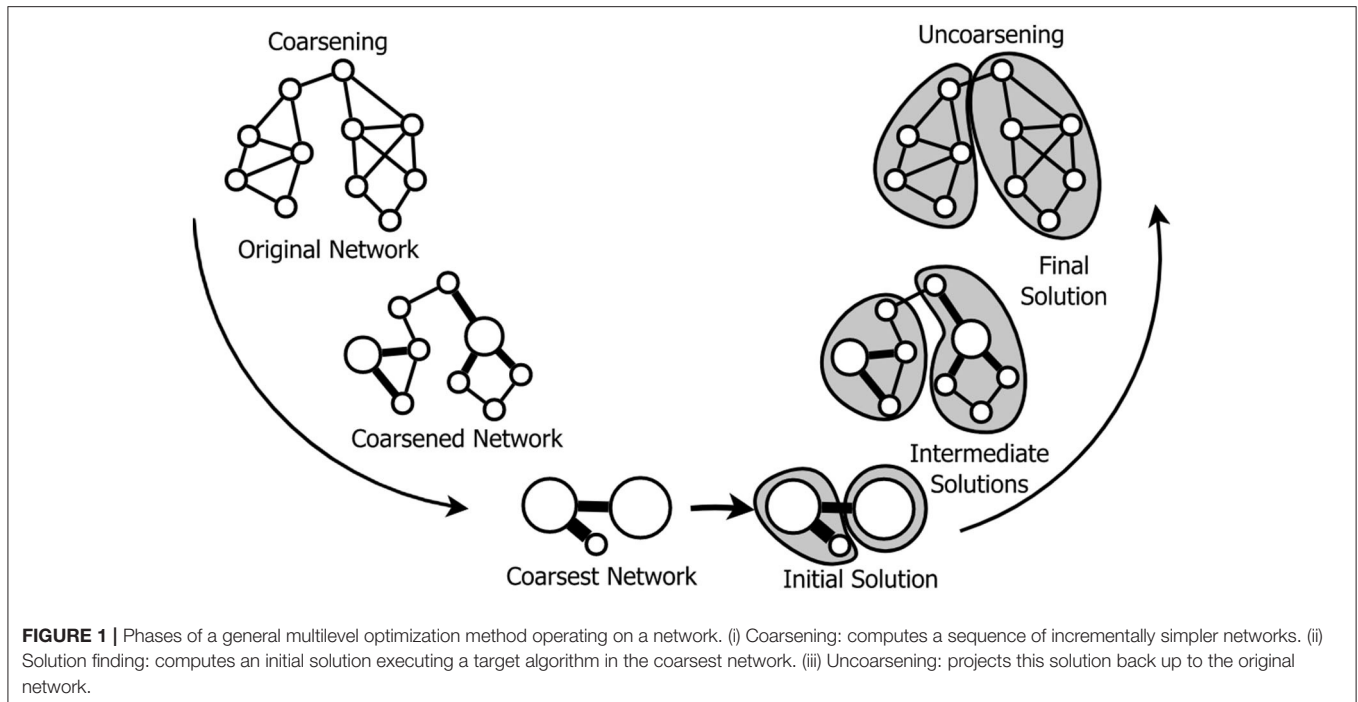
well-known “visual information seeking mantra” of information visualization (Shneiderman, 1996), where users depart from an overview and then navigate and interact in the visual space to focus on target regions and demand further details. This mantra embeds a number of visual design guidelines and is widely acknowledged as a general interaction framework in designing information visualization applications.

Alternatively, a hierarchical aggregation of a complex network may be obtained with the multilevel method, a meta-heuristic employed to modify and potentially fix a solution obtained with a target optimization algorithm (Karypis and Kumar, 1998; Wong et al., 2009). A multilevel algorithm performs an incremental coarsening of an original network to yield a hierarchy of simplified networks from the original (Brandt, 1988; Karypis and Kumar, 1998). It yields a hierarchical representation, where each level depicts a network with fewer nodes and links than its previous one. Multilevel methods have been originally introduced to enable executing expensive algorithms on large-scale networks.¹ The rationale is to compute an initial solution executing a target algorithm on a simpler network that still preserves the relevant properties of the original. This initial solution is then incrementally refined over the inverse hierarchy of coarsened networks to yield an approximate solution to the original problem (Noack and Rotta, 2009; Valejo et al., 2018). Applying the strategy successfully required two essential capabilities, namely finding a coarsest network that preserves the essential features of the original, and generalizing the initial solution obtained to the full original model. Thus, identifying a coarsening strategy that yields a suitable trade-off between solution accuracy, generalization capability, and speed is critical. This is commonly approached with empirical investigation on each problem and data set, sometimes departing from existing coarsening algorithms, sometimes introducing novel ones. Yet, multilevel methods have been successfully employed in a range of important problems defined in unipartite networks (Brandt, 1988). It has also been considered to accelerate the computation of node-link layouts for purposes of network drawing and visualization (Harel and Koren, 2000a; Walshaw, 2000; Hachul and Jünger, 2004; Hu, 2005; Archambault et al., 2007; Frishman and Tal, 2007; Bartel et al., 2010; Arleo et al., 2016; Toosi and Nikolov, 2016; Hinge et al., 2017).

In earlier work, authors of this paper introduced a general multilevel framework for bipartite networks, i.e., it preserves the bipartite restrictions and ensures efficient coarsening in this context (Valejo et al., 2017a, 2018, 2020a). The solution was motivated by the growing occurrence of bipartite networks as models for solving important real-world problems, e.g., the method has been applied in problems of link prediction (Ferreira et al., 2019), trajectory mining (Minatel et al., 2018, 2019), and community detection (Valejo et al., 2014a,b, 2020a).

In this contribution, we address the problem of employing the multilevel method for interactive visualization of large bipartite networks. As multilevel network coarsening can preserve the topological structures relevant to problem solving, we propose

¹By “large-scale” we mean networks with up to a few hundreds of thousands nodes, in contrast to massive-scale networks with millions to billions nodes.



using coarsening algorithms to obtain multi-resolution networks, which can be rendered as node-link diagrams at distinct levels of detail for purposes of interactive visualization. To the best of our knowledge, multilevel coarsening has not been considered so far for interactive visualization of real-world bipartite networks (Valejo et al., 2020b). We follow this premise and introduce a proof-of-concept solution that supports user-driven navigation in a hierarchical multilevel representation. The distinctive points of our contribution are the following:

1. We define a visualization strategy on top of an existing general multilevel framework for bipartite networks. Our strategy allows plugging-in any network coarsening algorithm, yielding maximum flexibility in obtaining multi-resolution representations of a given network. Thus, given an input network, alternative multi-resolution models can be created based on user convenience and her choice of coarsening algorithm.
2. We use our strategy as the underlying model of a system that supports interactive visualization and navigation on the multi-resolution network models, starting with a node-link view of the coarsest network. User interaction triggers gradual, on-demand expansion or contraction of selected sub-sets nodes. Thus, at any moment a node link view may be rendered with different regions shown at distinct resolution levels. The system also offers functionalities for zooming in/out and requesting complementary information on nodes or links.
3. The system is introduced as a proof-of-concept to demonstrate the feasibility of the strategy to support effective exploratory visualization of complex bipartite networks. The operation of the strategy and its potential are illustrated with examples in which real-world bipartite networks with varied topological properties are explored on-demand.

The remainder of this paper is organized as follows: in Section 2 we present a brief overview of our previously introduced multilevel framework for bipartite networks, which provides the underlying framework for the proposed visualization strategy. In Section 3, we briefly review related work on interactive network visualization, with an emphasis on solutions aimed at bipartite networks. In Section 4, we describe the proposed visualization strategy built on top of the general multilevel method, and describe how it has been implemented into a proof-of-concept system. In Section 5, we report illustrative case studies using the strategy in different scenarios. Finally, in Section 6, we summarize our findings and discuss future work.

2. MULTILEVEL METHOD ON BIPARTITE NETWORKS

The overall strategy of a multilevel method in the context of solving problems in networks is illustrated in **Figure 1**. The method operates in three phases, identified as coarsening, solution finding and uncoarsening. A description of the phases requires establishing some definitions and notation, as follows.

A unipartite network is represented by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \sigma, \omega)$, wherein \mathcal{V} and \mathcal{E} define the set of nodes and links, respectively, and a link $(v, u) = \{(u, v) = (v, u) \mid u, v \in \mathcal{V}\}$. $n = |\mathcal{V}|$ and $m = |\mathcal{E}|$ denote the number of nodes and links, respectively. A link (u, v) or a node u can be associated with a weight, denoted, respectively, as $\omega(u, v): \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}^*$ or $\sigma(u): \mathcal{V} \rightarrow \mathbb{R}^*$. The network is bipartite if \mathcal{V} is split into two sub sets \mathcal{V}^1 and \mathcal{V}^2 , such that $\mathcal{V}^1 \cap \mathcal{V}^2 = \emptyset$ and $\mathcal{E} \subseteq \mathcal{V}^1 \times \mathcal{V}^2$.

The h -hop neighborhood of a node u , denoted by $\Gamma_h(u)$, is given by the nodes distant from u by h or less links. Thus, $\Gamma_1(u)$ is the set of nodes adjacent to u ; $\Gamma_2(u)$ is the set of nodes 2-hops

away from u , and so forth. The degree of u , denoted by $\kappa(u)$, is the number of its incident links, i.e., $|\Gamma_1(u)|$. A similarity measure (henceforth called similarity *index*) is a function that quantifies common properties of a pair of nodes (u, v) , yielding values (scores) in the range $[0, 1] \subset \mathbb{R}$, from lowest (0) to highest (1) similarity (Valejo et al., 2014a). A typical structural index is the number of common neighbors, defined as $CN(u, v) = |\Lambda(u, v)|$, wherein $\Lambda(u, v) = \{\Gamma_1(u) \cap \Gamma_1(v)\}$.

Back to the multilevel method, in the coarsening phase a given network \mathcal{G}_0 is iteratively coarsened into a hierarchy of increasingly simpler representations, yielding a multilevel hierarchy $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_H\}$ with H levels. The hierarchy describes \mathcal{G}_0 at multiple levels-of-detail, wherein \mathcal{G}_H is the *coarsest network*. Specifically, a coarsening algorithm constructs a coarser network by collapsing a set of matched nodes $\{u_1, \dots, u_n\}$ into a single *super-node* sV_i . The nodes in set $\{u_1, \dots, u_n\} \in \mathcal{V}_h$ collapsed into a super-node $sV_i \in \mathcal{V}_{h+1}$ are called *predecessors* of sV_i , denoted $\Theta(sV_i)$, whilst the super-node (sV_i) is a *successor* of its originating nodes, denoted $\Delta(u_i)$. A successor network \mathcal{G}_{h+1} will inherit the non-matched nodes from its predecessor network \mathcal{G}_h . In order for \mathcal{G}_{h+1} to be a good proxy to its predecessor network, the weight $\sigma(sV_i)$ of a super-node $sV_i \in \mathcal{V}_{h+1}$ is computed as the sum of weights of its predecessor nodes. Furthermore, any links incident to nodes in $\Theta(sV_i)$ are collapsed into the so-called *super-links* incident to sV_i .

In the solution finding phase, a target task, such as community detection, dimension reduction, link prediction or node-link layout computation, is evaluated in \mathcal{G}_H yielding an initial solution \mathcal{S}_H . In the uncoarsening phase \mathcal{S}_H is gradually refined through the intermediate networks $\mathcal{G}_{H-1}, \mathcal{G}_{H-2}, \dots, \mathcal{G}_1$ up to the original network \mathcal{G}_0 .

Notably, effective application of the strategy to any problem is highly dependent on the quality of the hierarchical multilevel representation built. Coarsening requires a policy to collapse nodes and links into super-nodes and super-links; uncoarsening requires a policy to expand collapsed nodes or links. Clearly, the steps of computing an initial solution and refining it are task dependent. As opposite, policies for merging nodes and links may be defined in terms of the network topology only, regardless of the task (Valejo et al., 2020b). Here we briefly introduce a few coarsening strategies for bipartite networks defined in earlier work, namely algorithms *OPM*, *RGMb*, *Gmb*, and *MLPb*.

OPM (one-mode projection-based matching algorithm) is one of the earliest coarsening algorithms applied to bipartite networks (Valejo et al., 2017a,b). It decomposes the bipartite network \mathcal{G} into two unipartite networks, one relative to each layer, i.e., \mathcal{G}^1 and \mathcal{G}^2 , which allows employing solutions devised for unipartite networks, such as the popular Heavy-Edge Matching (*HEM*) algorithm (Karypis and Kumar, 1998). In *HEM*, a random node u is matched with an adjacent node v for which the edge (u, v) has maximum weight overall all edges adjacent to u .

Valejo et al. (2018) designed two coarsening algorithms specifically for bipartite networks. Algorithms *RGMb* (Random greedy matching) and *Gmb* (Greedy matching) extend earlier link-based matching strategies which operate by collapsing node pairs into a single super node. Both inspect the 2-hop neighborhood of a node u to identify matching candidates,

picking the node with the greatest similarity index with u . *RGMb* picks candidate nodes for matching randomly, whereas *Gmb* picks nodes from a priority queue holding the most similar pairs.

A third algorithm called *MLPb* (Multilevel label propagation for bipartite networks) (Valejo et al., 2020a), has been introduced based on the well-known *Label Propagation Algorithm (LPA)* (Raghavan et al., 2007). Each node is initially assigned a unique label, and at each iteration node labels are updated with the node's most frequent label in its two-hop neighborhood. Groups of nodes with the same label are collapsed into a single super-node. Coarsening with *MLPb* typically requires less iteration steps than *RGMb* or *Gmb*, as unlike these the algorithm is not restricted to collapsing just a pair of nodes per iteration. Preliminary studies provided empirical evidence that multilevel representations obtained with *MLPb* preserve the essential topological features of a bipartite network (Valejo et al., 2020a).

Of course, the choice of coarsening algorithm affects the multilevel representation obtained, which is key to the interactive visualization process. There is no better algorithm *a priori*, as the suitability of a representation must be considered in the context of the user problem, tasks and goals. This is a problem that deserves further investigation. Interestingly, the impact of distinct coarsening policies on the effectiveness of the multilevel strategy applied in different data mining tasks is also a relevant research topic where interactive visualization itself can play an important role (Valejo et al., 2021).

3. INTERACTIVE VISUALIZATION OF LARGE NETWORK

Whilst node-link representations are intuitive to convey network topology, they have very limited scalability (Ghoniem et al., 2004; Nobre et al., 2020; Di Giacomo et al., 2021). Rendering large networks can be slow and the result can be cumbersome. For unipartite networks it is common to use aggregations of nodes and/or links to mitigate clutter and reduce rendering load. Link bundling is possibly the best established solution for implicit link aggregation (Holten, 2006; van der Zwan et al., 2016; Lhuillier et al., 2017), whereas node aggregation strategies are usually explicit and domain-driven, with a diversity of strategies reported in the literature. In general terms, most solutions use clustering or community detection algorithms to group nodes into meta-nodes, super-nodes, clusters or communities that may be handled as individual entities (Von Landesberger et al., 2011), yielding a hierarchical representation that can be navigated and explored for visualization and analysis purposes. Hierarchical algorithms have been employed to create simplified representations of large networks (Dias et al., 2017a), and multilevel methods have been applied, for instance, to reduce the computational cost of computing node-link layouts (Harel and Koren, 2000b; Gajer and Kobourov, 2001; Hachul and Jünger, 2004; Wong et al., 2009). An experimental evaluation has been reported on using multilevel algorithms in association with energy-based layout algorithms (Bartel et al., 2011).

Still on the realm of unipartite networks, Abello et al. (2006) uses clustering algorithms to define a hierarchy over the network, so that users can navigate in a top-down strategy by

interactively expanding individual clusters. Archambault et al. (2008) proposes using domain-specific attributes associated with the nodes and edges to create different possible hierarchies on the same unipartite network, instead of a single, fixed hierarchy as obtained with clustering. Wong et al. (2009) introduces a coarsening algorithm to create a sequence of reduced networks that retain the most important structural features at each level. A coarser graph is guaranteed to have no less than half the number of nodes than the previous (less coarsened) graph, i.e., ensuring a 50% reduction rate at each level. Similarly, Dias et al. (2017b) uses a non-negative matrix factorization (NMF) to rewrite the edge weights and then employ a coarsening algorithm to create a sequence of simpler networks. Batagelj et al. (2010) introduced a two-level clustering algorithm that guarantees the intra and the inter-cluster edges satisfy a set of desired topological properties. This strategy makes it possible to exploit and combine different visualization algorithms. Perrot and Auber (2018) presents a technique to visualize huge graphs using a client running in a Web browser. For this purpose, the authors introduce novel scalable algorithms based on the well-known k -means clustering to produce multiple levels of abstraction of the network.

Another category of related work comprises solutions introduced for exploratory visualization of data modeled as large scale bipartite graphs, which are becoming more prevalent recently. Several systems employ the biclustering algorithm (Heinrich et al., 2011) to generate node aggregations in this context (Dörk et al., 2012; Xu et al., 2016; Steinbock et al., 2018; Zhao et al., 2018; Sun et al., 2019). Using a co-clustering algorithm, the interactive visualization system by Xu et al. (2016) assists user recommendation on cluster models to be input into a machine learning algorithm (Hoi et al., 2010) that will learn acceptable clusterings according to user preferences. Chan et al. (2019) also use co-clustering to facilitate interactive data exploration. Waldner et al. (2020) deal with interactive exploration of time-dependent large bipartite graphs. They use two clustering algorithms to build a hierarchical aggregation: a biclustering algorithm to group nodes so as to maximize the graph modularity, and a time series clustering algorithm to group nodes based on their temporal correlation. The visual encoding adopts the usual approach of presenting two vertical lists of nodes, laid out in parallel. Other approaches are also employed to assist visualization of bipartite graphs, as in Garcia-Algarra et al. (2018), who use a k -core decomposition to identify and aggregate groups of nodes that share connectivity properties in order to simplify the network structure. The rationale of aggregating groups of nodes that share connectivity properties is also at the core of multilevel coarsening, employed in this paper.

Yet, other approaches are possible. The system WAOW-Vis (Pezzotti et al., 2018), for instance, adopts a hierarchical dimensionality reduction technique to create hierarchical representations of bipartite graphs. Users may interact to expand nodes in a particular area. The BiCFlows system (Steinbock et al., 2018) relies a novel visualization to support exploration of large weighted bipartite graphs, using biclustering for a hierarchical aggregation. Selecting a clustering yields a new execution of the biclustering algorithm to support a more detailed visualization. The ViBr tool (Chan et al., 2019) uses the Minimum Description

Length (MDL) principle to create a clustered network, which can be explored selecting specific clusters and re-executing the MDL approach. The system employs adjacency lists for visualization, rather than node-link views. Zhao et al. (2019) also prefer to adopt an alternative representation to the node-link view, in this case a matrix-based visualization to assist in interactive analysis of missing links in bipartite networks.

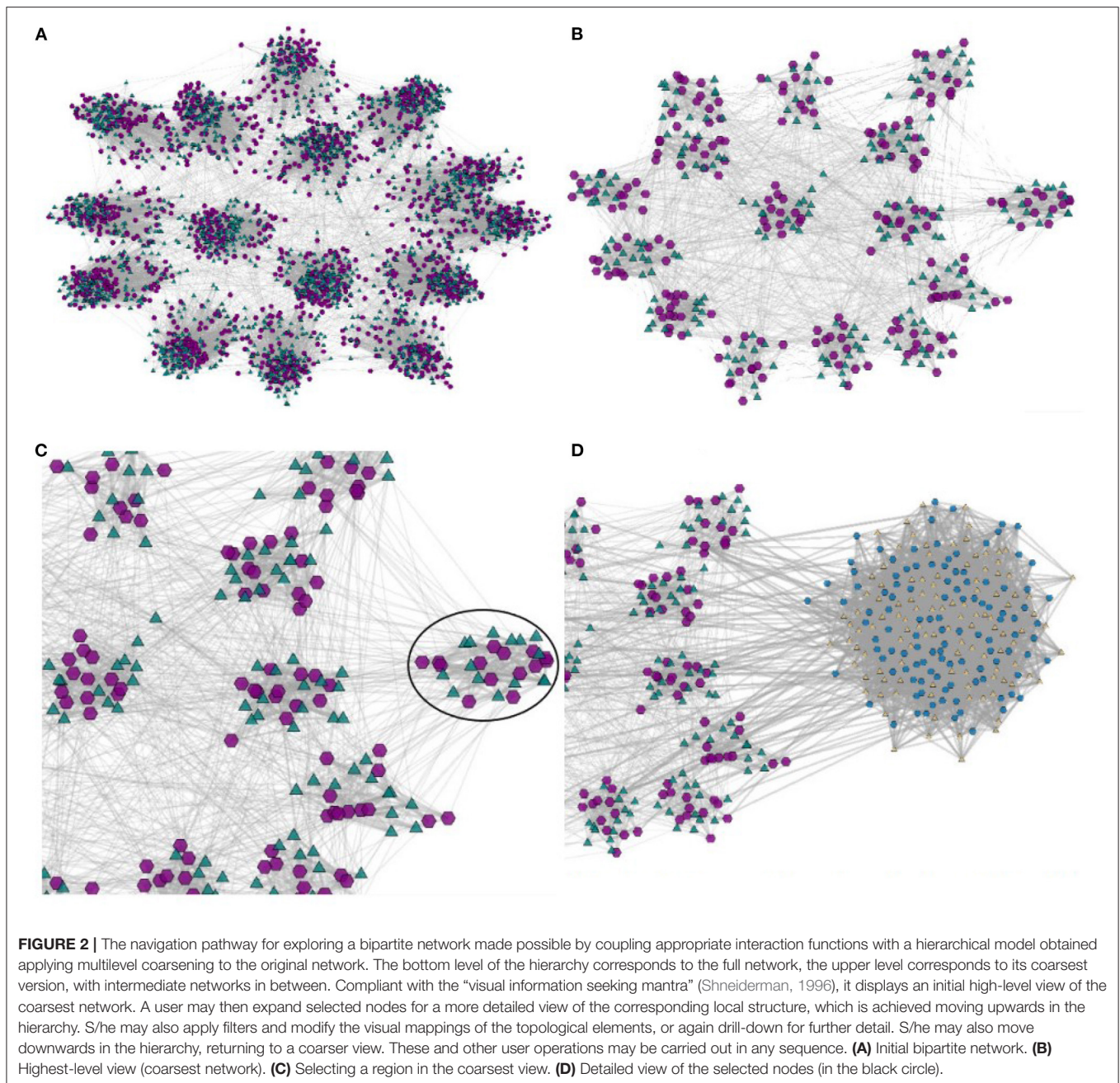
Regarding multilevel methods in association with visualization of bipartite networks, very few specific strategies have been designed. Cintra et al. (2019) introduces a visualization metaphor in association with the multilevel method, however the focus is not on exhibiting network topology, but in assisting developers of multilevel methods to compare the outcomes of different coarsening algorithms. As opposite, Valejo et al. (2020a) illustrate how the *MLPb* coarsening algorithm can support node-link visualizations of large bipartite networks, however the authors only consider static visualizations, rather than interactive exploratory visualization scenarios.

In this work, we contribute a novel approach for interacting with node-link views of large bipartite networks. Navigation of the network at varying levels-of-detail is afforded by a multi resolution hierarchical representation of the network obtained with multilevel coarsening algorithms. We are not aware of previous efforts of using multilevel network coarsening in this context.

4. MULTILEVEL VISUALIZATION OF BIPARTITE NETWORKS

In this section, we describe how a multi-resolution network representation $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{\mathcal{H}}\}$ obtained with a multilevel coarsening algorithm can provide an underlying hierarchical representation for interactive exploration of an original large-scale bipartite network \mathcal{G}_0 . We specialize the general framework depicted in **Figure 1** to address the target task of constructing a multiresolution visual mapping of \mathcal{G}_0 . In this context, the initial solution $\mathcal{S}_{\mathcal{H}}$ corresponds to any suitable node-link layout of its coarsest representation $\mathcal{G}_{\mathcal{H}}$. However, rather than pre computing a layout for each intermediate network \mathcal{G}_h , we introduce a strategy that performs uncoarsening locally over sub-graphs defined by user selections. In other words, whenever the user wishes a more detailed view of a certain region of the graph than the one currently displayed, s/he identifies and selects the corresponding super-nodes and the uncoarsening algorithm is triggered on demand, on the selected super-nodes. The opposite is also possible, i.e., changing from a detailed current view of a sub-set of nodes to a coarser view by collapsing a selected sub-set of nodes into super-nodes.

As further detailed in Section 4.2, expanding selected super-nodes for a more detailed view requires identifying their predecessor nodes at the upper hierarchical level, which must be incorporated into the rendering layout. Likewise, collapsing user selected nodes requires identifying their successor super-nodes, then updating the rendering accordingly. This solution prevents unnecessary complexity and, most importantly, avoids overloading the visualization with information beyond the



cognitive convenience of the user and the computational power available to ensure real-time interactivity. On the other hand, the node-link layout of a network displayed at any moment may be depicted with different regions shown at distinct resolution levels, following the user interactions.

The approach is defined as a general framework that is independent of specific choices of either coarsening algorithms or node-layout algorithms; an implementation may thus offer multiple choices of such algorithms to meet distinct domain and/or user specific demands. Additionally, it seamlessly supports the “visual information-seeking mantra” (Shneiderman, 1996) of *overview first, zoom and filter, then details-on-demand*.

Exploration starts with the user observing a node-link layout of the highest level network (its coarsest representation \mathcal{G}_H), rendered with any efficient algorithm. S/he may interact to inspect topology, nodes or associated content, and eventually select an arbitrary sub-set of nodes to obtain an expanded view. In other words, the user is able to zoom into the region defined by the selected nodes, which are rendered at the next level of the hierarchical multilevel network $\{\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_{H-1}\}$, as illustrated in **Figure 2**. In the example, the initial node-link layout exhibits \mathcal{G}_0 , after a user interaction the layout is updated so that the selection is rendered at the level of detail given by \mathcal{G}_1 . Similar interactions can be applied to arbitrary regions

Algorithm 1: BiNetVis algorithm for multilevel visualization of a bipartite network.

Input: input network \mathcal{G}_0 , layers $\in \{1, 2\}$
Output: Visual mapping of the current network

```

1 for  $h \leftarrow 0$  to  $(\mathcal{H} - 1)$  by 1 do
2   | foreach  $l \leq$  layers do
3     |  $\mathcal{G}_{h+1} \leftarrow$  Coarsen( $\mathcal{G}_h, l$ )
4 screen  $\leftarrow$  render coarsest network  $\mathcal{G}_H$ 
5 while  $C$  is not “exit” do
6   |  $h \leftarrow$  current user focus level
7   |  $\mathcal{V}' \leftarrow$  user selection of nodes
8   |  $C \leftarrow$  user command
9   | if  $C$  is “to expand” and  $h \neq 0$  then
10    |  $\mathcal{G}_h \leftarrow$  Expand( $\mathcal{G}_h, \mathcal{G}_{h+1}, \mathcal{V}'$ )
11   | else if  $C$  is “to collapse” and  $h \neq \mathcal{H}$  then
12    |  $\mathcal{G}_h \leftarrow$  Collapse( $\mathcal{G}_h, \mathcal{G}_{h-1}, \mathcal{V}'$ )
13   | else
14    |  $\mathcal{G}_h \leftarrow$  update visual mapping according to  $C$ 
15   | screen  $\leftarrow$  render network  $\mathcal{G}_h$ 

```

of the network rendered at any level of detail captured by the multilevel representation, therefore allowing a user to switch between coarser to finer levels of detail, or vice-versa.

We named our approach BiNetVis (from Multilevel Bipartite Network Visualization), and it is described in **Algorithm 1**. The algorithm instantiates the general multilevel framework depicted in **Figure 1** to the specific problem of multilevel visualization of bipartite networks. Initially, an input network \mathcal{G}_0 is coarsened with an arbitrary user selected algorithm, indicated by the routine *Coarsen* (lines 1-3). Coarsening is performed independently on each network layer $\{1, 2\}$, or the user may choose to coarsen just one of the layers. The number of levels of the resulting network hierarchy, which defines the extent of the desired simplification, is typically a parameter of the coarsening algorithm. Next, any choice of node-link layout algorithm can be employed to render \mathcal{G}_H to the screen, line 4. For purposes of interactive exploratory visualization, network \mathcal{G}_H , which depicts the coarsest representation of \mathcal{G}_0 , provides the initial focus for the user to interact. S/he can move the focus up or down the multilevel hierarchy while interacting, and the visual mapping is adjusted accordingly to the current focus. Possible user operations in the interaction loop (lines 5–15) would be requests to expand selected (super-)nodes or to collapse selected (super-)nodes (routines *Expand* and *Collapse* in lines 10 and 12, respectively). Although not explicit in the algorithm, additional operations can be incorporated into an implementation, e.g., to request metadata relative to nodes or links, modify graphical properties such as position, color or transparency of nodes and/or links; apply zooming or panning to the rendering, etc.

Algorithms 2, 3 detail the routines *Expand* and *Collapse*, which take as inputs two networks \mathcal{G}_h and \mathcal{G}_{h+1} (or \mathcal{G}_h and \mathcal{G}_{h-1}) where h is the level of the current viewing focus, and a user selection of nodes $\mathcal{V}' \in \mathcal{V}_h$. Recall that set $\Theta(u)$ refers to the predecessor nodes of u at level $h - 1$, and $\Delta(u)$ refers to the successor super-node of u at level $h + 1$. Expanding a selection

Algorithm 2: Expand selected nodes one level up.

Input: $\mathcal{G}_h, \mathcal{V}' \subset \mathcal{V}_h$
Output: Current visual mapping of the network

```

1 foreach  $u \in \mathcal{V}'$  do
2   |  $\mathcal{G}_h \leftarrow$  remove  $u$  and its incident links
3   | foreach  $p \in \Theta(u)$  do
4     |  $\mathcal{G}_h \leftarrow$  add node  $p$ 
5     | foreach  $n \in \Gamma_1(p)$  do
6       |  $sV \leftarrow \Delta(n)$ 
7       | if  $(sV, u) \in \mathcal{E}_h$  then
8         |  $\mathcal{E}_h = \omega_h(sV, u) + \omega_{h-1}(n, u)$ 
9       | else
10      |  $\mathcal{E}_h =$  insert new link  $(sV, u)$  with weight
          |  $\omega_{h-1}(n, u)$ 

```

Algorithm 3: Collapse selected nodes one level down.

Input: $\mathcal{G}_h, \mathcal{V}' \subset \mathcal{V}_h$
Output: Current visual mapping of the network

```

1 foreach  $v \in \mathcal{V}'$  with  $v \in \mathcal{V}_h$  do
2   |  $\mathcal{G}_h \leftarrow$  remove  $v$  and its incident links
3   |  $sV \leftarrow \Delta(v) \in \mathcal{V}_{h+1}$ 
4   | if  $sV \notin \mathcal{V}_h$  then
5     |  $\mathcal{G}_h \leftarrow$  recover  $sV$  and its incident links

```

\mathcal{V}' requires incorporating into the current network layout the predecessor nodes, at level $h + 1$, of all nodes $v \in \mathcal{V}'_h$ and their adjacent links. Similarly, collapsing \mathcal{V}' implies in replacing its nodes with their corresponding successor nodes and adjacent links at level $h - 1$. In both cases, the node-layout rendering will be updated with a hybrid network model which simultaneously includes nodes and links from networks at multiple levels of the hierarchical multilevel representation.

4.1. BiNetVis Interface

We developed a proof-of-concept implementation of the BiNetVis approach²; the system interface is illustrated in **Figure 3**. An analyst initially selects a network model from a drop-down menu (1), which also admits uploading a new network. BiNetVis currently admits networks described in the NCOL³ readable file format, often associated with biological data graphs, though other formats could be incorporated.

The next step once a network has been loaded is to select a coarsening algorithm and set its parameters in the corresponding menu (area 2 in **Figure 3**). The user can choose from OPM, Gmb, RGmb, and MLPb coarsening algorithms (introduced in Section 2), and set their parameters—reasonable default settings are provided to newcomers. The options indicated in area 3 can be checked to hide/show the “layout options” menu and to hide/show edges in the layout rendering, in addition to showing the predecessor nodes of a selected super-node. Once

²source code and installation instructions at <https://github.com/ttm/ml-net-gui> (runs on Linux platforms).

³<http://lgl.sourceforge.net/#FileFormat>

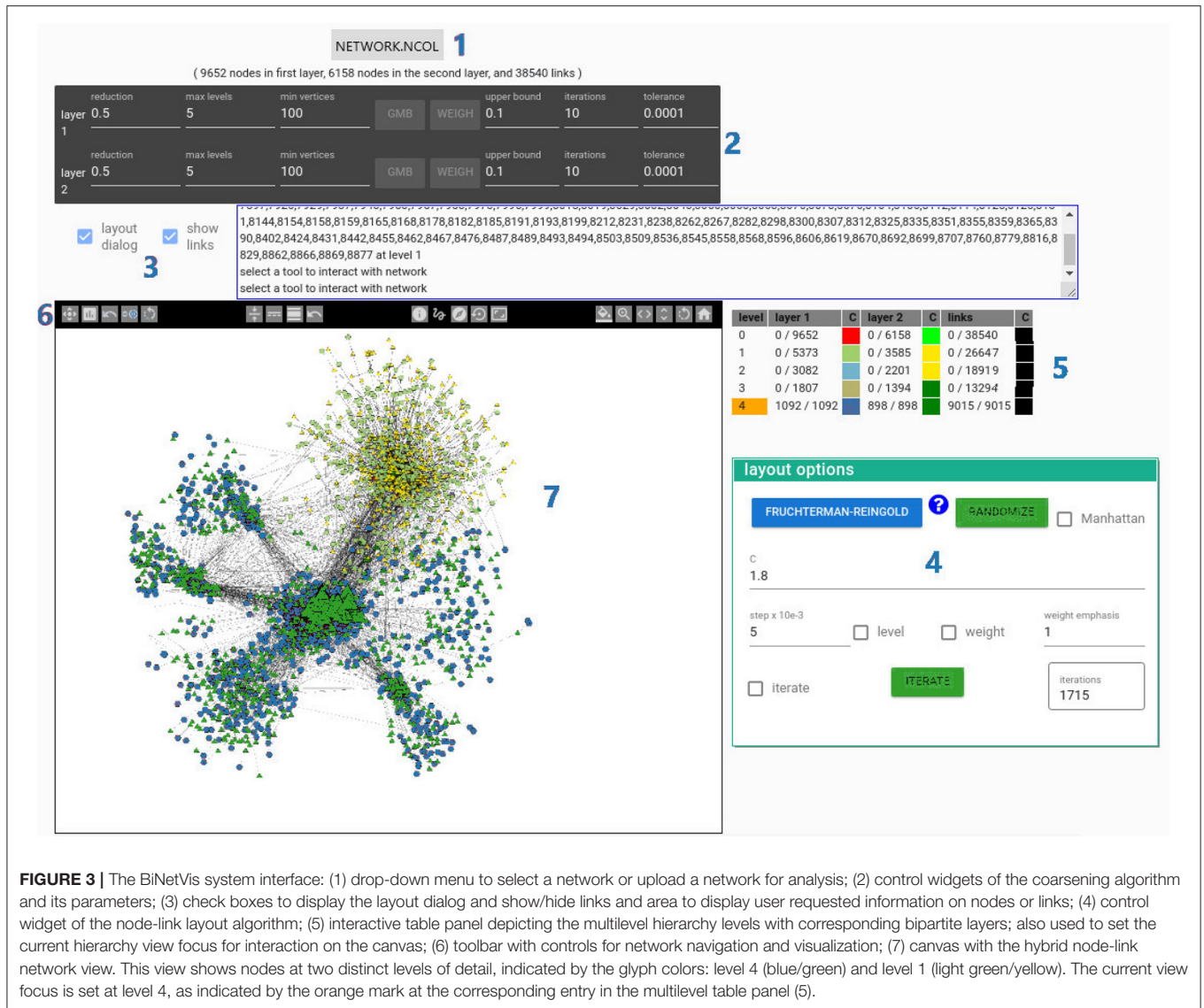


FIGURE 3 | The BiNetVis system interface: (1) drop-down menu to select a network or upload a network for analysis; (2) control widgets of the coarsening algorithm and its parameters; (3) check boxes to display the layout dialog and show/hide links and area to display user requested information on nodes or links; (4) control widget of the node-link layout algorithm; (5) interactive table panel depicting the multilevel hierarchy levels with corresponding bipartite layers; also used to set the current hierarchy view focus for interaction on the canvas; (6) toolbar with controls for network navigation and visualization; (7) canvas with the hybrid node-link network view. This view shows nodes at two distinct levels of detail, indicated by the glyph colors: level 4 (blue/green) and level 1 (light green/yellow). The current view focus is set at level 4, as indicated by the orange mark at the corresponding entry in the multilevel table panel (5).

the “render network” button is hit (area 4) the coarsening algorithm is activated to build the hierarchical representation, and a node-link layout of the coarsest network is computed with the selected layout algorithm and rendered to the canvas (area 7). The system currently incorporates two node layout algorithms, namely Fruchterman and Reingold (1991) and Kamada and Kawai (1989). The most relevant interface components to support user interaction in BiNetVis are the multilevel table panel (area 5) and the toolbar (area 6).

The table panel (5) provides the necessary controls to navigate in the multilevel hierarchy. The color pairs used in rendering the nodes change to identify the different levels, as indicated in the table panel. Because different regions of the network may be rendered at distinct levels of detail, a user must explicitly inform the view focus level h to which level the current operations apply. The default focus level is the coarsest one, and it may be changed by the user clicking in any line of the table panel that corresponds

to the desired level focus. The current focus is signaled by the orange marker in the corresponding line, e.g., in **Figure 3** the marker appears at the line corresponding to level $h = 4$, which is the coarsest representation of the input network. For each layer of G_i at level $h = i$ the table panel also informs how many nodes and links are currently visible, vs. their total numbers. The user also has options to modify node colors or shapes, as well as to toggle on/off drawing of links by means of left/right clicks on the corresponding colored square. Any current user actions apply only to the nodes/links at the currently defined view focus level.

Interaction is governed by the controls in the toolbar (6), which are organized in four groups according to their target element (node, link, multilevel hierarchy, canvas). The leftmost group of controls affect the mapping of node properties to graphical attributes of its corresponding graphical marker. Moving right, the next group of controls affect the mapping of link properties to graphical attributes of the corresponding

lines. The next group is of controls to navigate in the multilevel hierarchy, e.g., obtain information relative to a selected (clicked) node, its predecessors and successor, move nodes in the canvas by clicking and dragging; define a rectangular region in the canvas and drag to move all nodes within, expand super-nodes to show their predecessor nodes and links at a finer level of detail, or collapse super-nodes that have been previously expanded. Finally, the rightmost set of controls affect the canvas, with tools to change the background color and perform zooming, panning and rotation of the network.

4.2. BiNetVis Implementation

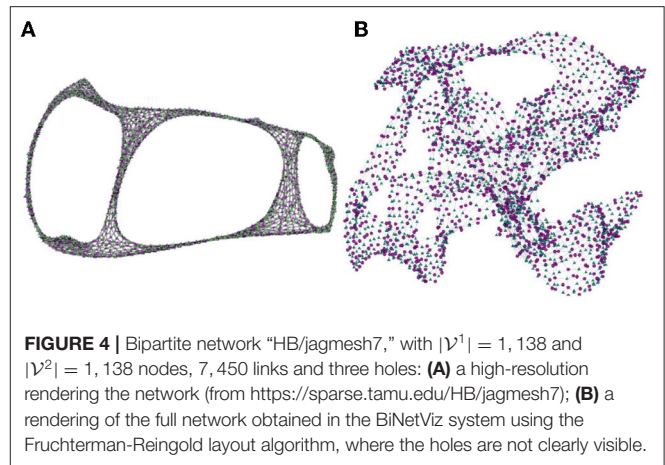
Components of the BiNetVis software are mostly written in a combination of JavaScript and Python: it uses Vue.js (set up by Nuxt.js) in the front-end client, the back-end is a Flask Python server, used to perform specialized or heavy calculations. A secondary server, a FeatherJS, is used to facilitate contact with the database and real-time multi-user interaction. The data is stored in a MongoDB database and ordinarily in the file system, while the multi-user interaction is currently deactivated to avoid unnecessary complexity. Multiple bipartite network coarsening algorithms available from a previous implementation (Valejo et al., 2018) are accessed by the Flask server. Pixi.js is employed for fast WebGL 2D rendering on the canvas. The choice of features and technologies supports visualization of networks with tens of thousands of nodes without any perceptible lag in the interactive navigation and transformations, with links rendered, even when running the system on ordinary machines, e.g., with 8GB RAM DDR3, a first generation i7 processor and a 1GB GPU.

In order to comply with the goal of handling large bipartite networks, the glyphs on the canvas use only triangles as geometric primitives (for nodes) and straight line segments (for links). It is possible to choose the colors and shapes associated with the bipartite layers at the different hierarchy levels, and also to alleviate the computational burden by setting the option of not rendering the links.

5. USE CASE SCENARIOS

We present four use case scenarios to illustrate the flexibility afforded by the proposed visualization solution.

1. A first example illustrates how network coarsening can reveal fundamental topological properties of a complex network, which may remain hidden in a full rendering of the full network.
2. A second study investigates how user-driven uncoarsening of specific regions of the n-reactome biochemical network can reveal hidden local structures with no need to iterate over the entire network.
3. A third study explores a gene-disease network to illustrate the possibility of switching between different perspectives by coarsening just one layer at a time, indicating that users can easily obtain relevant contextual information by on-demand navigation and exploration of the network.
4. A fourth study explores a scientific collaboration network (authors and papers) obtained from papers published related



to COVID-19 and the “CoronaVac” vaccine. This study illustrates how users can explore groups in different contexts after obtaining a global visual perspective of how they are connected.

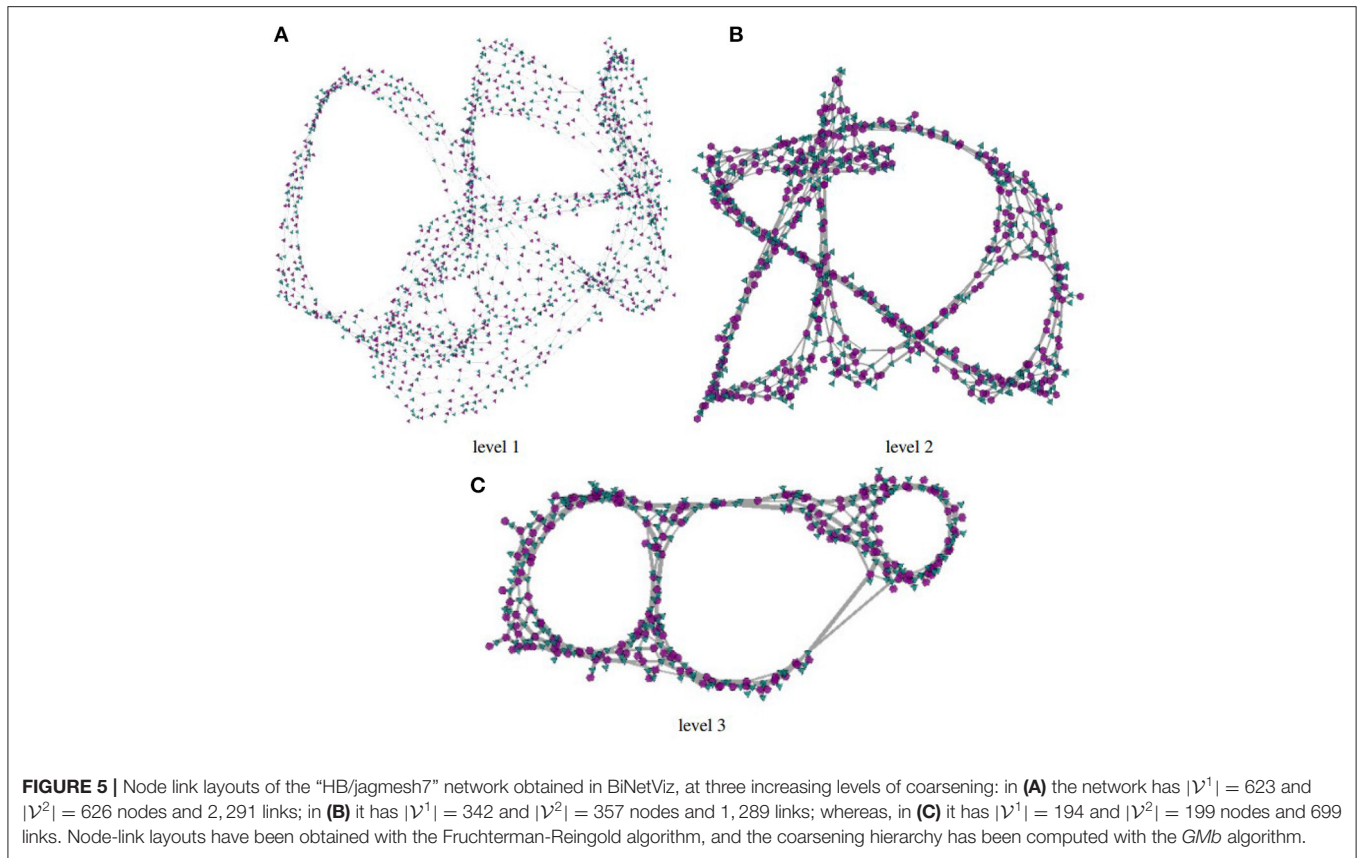
5.1. Revealing Hidden Patterns

The overlapping of graphical elements in node-link visualizations of large networks contributes to blurring important topological patterns. Consider, as an example, the bipartite network “HB/jagmesh7,” from the SuitSparse Matrix Collection.⁴ This network, with 1,138 nodes and 3,156 links, has a peculiar topology with three holes—a hole is a loop formed by nodes with cyclic links only and no links crossing the loop. Holes may indicate a relevant large-scale event occurring at a particular region of the network.

Figure 4A, replicated from the SuitSparse Matrix site, illustrates a static high-resolution drawing of this network. In this large-sized high-resolution image the node layout algorithm had sufficient space to place the graphical elements so that the three holes are clearly visible. However, such rendering would hardly afford real-time user interaction. More commonly, users interact with node-link layouts displayed in limited-size screens. **Figure 4B** illustrates a rendering of the full “HB/jagmesh7” network obtained in BiNetVis with the Fruchterman-Reingold layout algorithm (Fruchterman and Reingold, 1991). In this view the hole topology of the network is not as clear. Even though it is possible to rotate a layout, or to create and render alternative layouts, they will not necessarily convey the existing holes, of which an analyst may not be aware. Moreover, a node-link screen rendering of a network this size also suffers from severe overlapping of graphical elements, which impairs legibility, interpretation, and interactivity.

Figure 5 shows coarsened versions of the original network obtained with the *Gmb* algorithm using the parameters as following: three levels of coarsening and a reduction factor of 50% at both levels. Again, node-link layouts have been computed with the Fruchterman-Reingold algorithm. The coarsest network model shown in **Figure 5C** has just $|\mathcal{V}^1| = 194$ and $|\mathcal{V}^2| =$

⁴<https://sparse.tamu.edu/>



199. Yet, the holes in the network topology remain evident in the three reduced models, and the layout in **Figure 5C** has a similar shape to the one shown in **Figure 4A**. In both the largest hole is placed in the central area, whereas the smallest hole appears to the right. Whilst detailed topological information is not accessible in the coarsened models, the renderings still preserve the essential global topological structure of the full network. Once the global structure is conveyed, a user may further investigate local topological structures by interactively selecting and expanding nodes.

5.2. On-Demand User Interaction

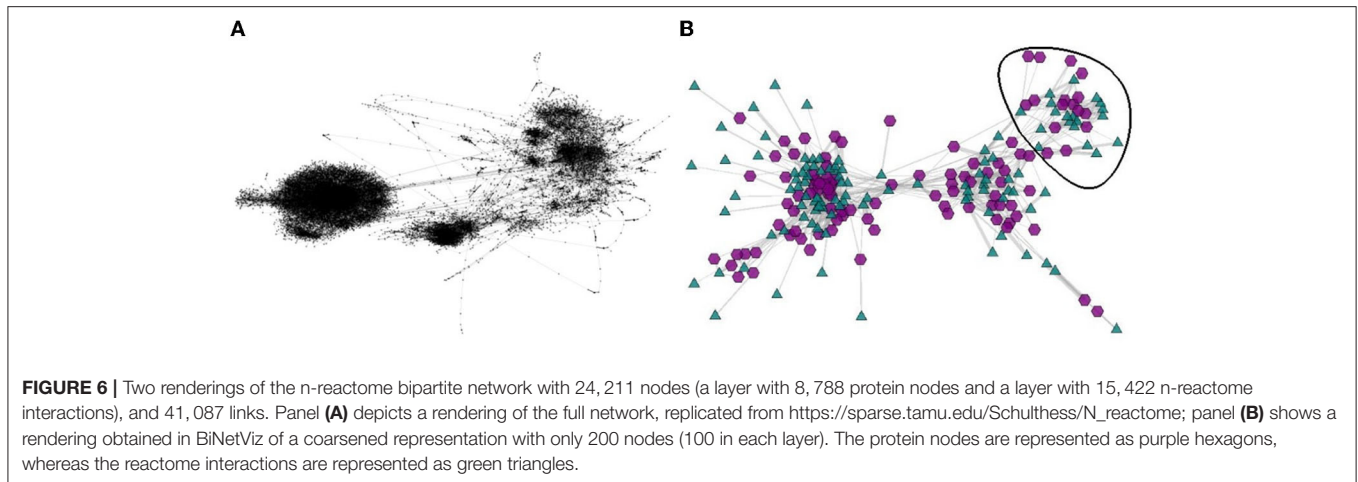
The interaction patterns between genes and proteins establish the basis of molecular biology and disease pathogenesis, i.e., the study of how diseases arise and evolve (Pawson and Linding, 2008; AlQuraishi et al., 2014). Protein groups are defined by the topological properties of the networks they entail, which allows for the isolation of functional and disease pathways (Sharan et al., 2007; Barabási et al., 2011). We illustrate a brief case study of BiNetViz on the n-reactome biochemical network, also available at the SuiteSparseMarix (https://sparse.tamu.edu/Schulthess/N_reactome).

Figure 6A depicts a high quality rendering of the original two-layer network, formed by 8,788 proteins in one layer, 15,422 reactome interactions in the second layer, and 41,087 links between proteins and interactions. In this high-quality rendering

one observes three more salient communities and several smaller ones. **Figure 6B** illustrates a coarsened representation obtained in BiNetViz, with 100 nodes in each layer (protein nodes represented as purple hexagons, and reactome interactions represented as green triangles). The coarser networks have been obtained executing three steps of algorithm *MLPb*, setting a 50% reduction factor at each coarsening level with an upper bound of 0.1 to allow for unbalanced super-nodes. Notice how the coarsest network, even after severe reduction, still mirrors the predominant characteristics of the original one, emphasizing the presence of three major communities and their connections.

For purposes of illustrating the possibilities afforded by BiNetViz we chose to inspect the connectivity node patterns of the smallest community, identified by the nodes within black circle depicted in **Figure 6B**. **Figures 7A–C** depict this particular community rendered, respectively, at coarsening levels 2, 1 and as in the original level, whereas the remaining nodes are still rendered at coarsened level 3. One observes how the complex topology is gradually unfolded in the sequence of views, each one revealing more detail.

A zoomed view of the community rendered at level 0 is depicted in **Figure 7D**, where one observes a group of densely connected nodes in the central area. As these nodes clearly correspond to numerous protein reactome interactions, they are likely to play a central role in the network, and an analyst might wish to further investigate the topological structures in that area.



Or, she might become interested in examining the less connected nodes, in an attempt to understand their behavior or specific roles.

This example illustrates how BiNetVis could be useful for revealing localized patterns and topological relations involving groups of nodes from one or both layers of a large bipartite network. The framework allows focusing on particular areas for detail on the connections, preserving interactivity without losing the global context provided by the remainder of the network. For this particular network, it could support, e.g., studying groups of proteins that could be targeted for therapeutic or other purposes, or groups of genes associated with metabolic deficiency.

5.3. Exploring a Gene-Disease Network

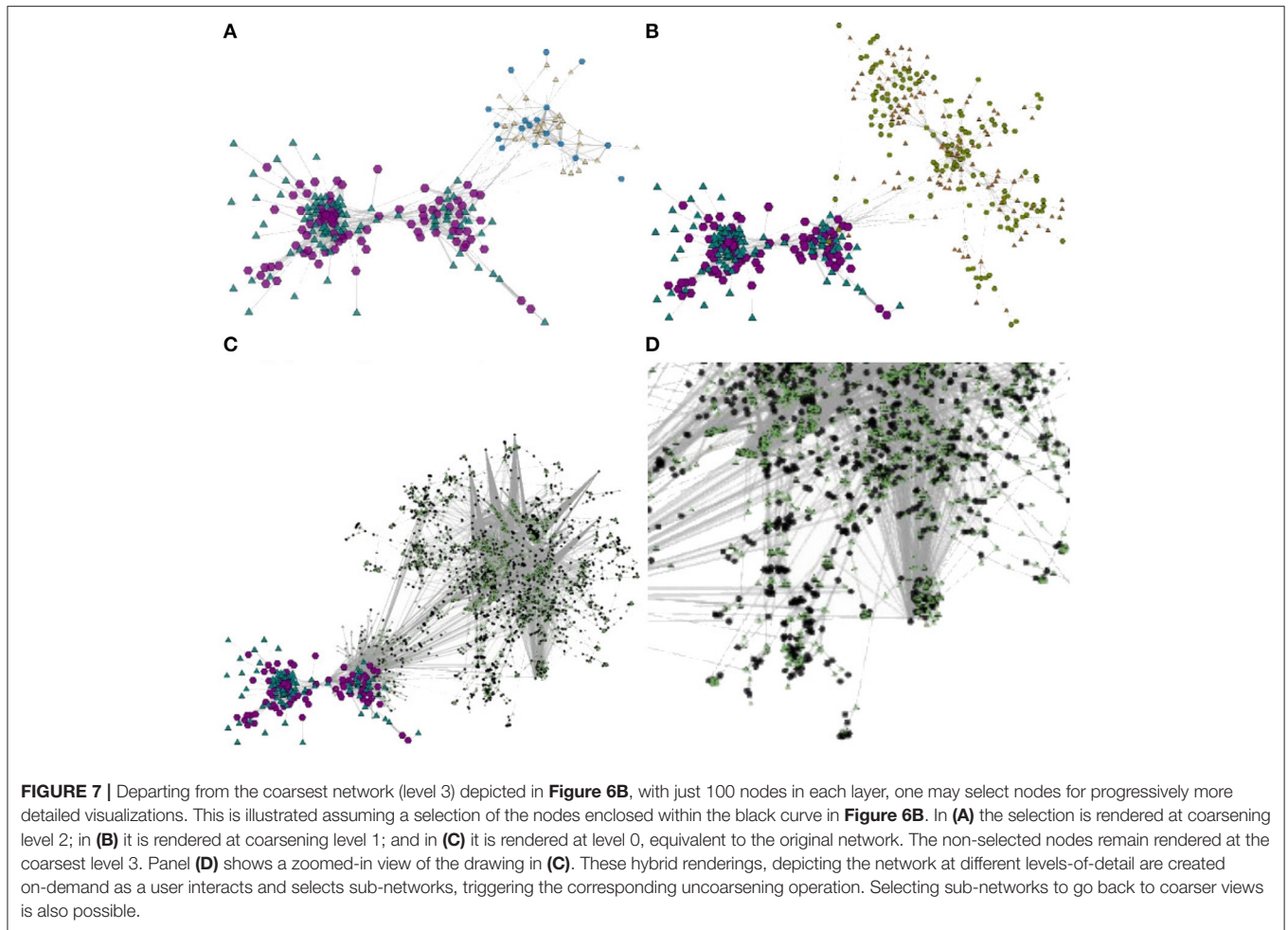
Goh et al. (2007) discuss how a network of human diseases and genes linked by known disease-gene associations offers a platform to investigating “whether human genetic disorders and the corresponding disease genes might be related to each other at a higher level of cellular and organism organization.” Authors use an available repository with data on diseases, genes and their associations to construct a bipartite network in which one layer has nodes corresponding to known genetic disorders, and the other layer has nodes corresponding to all known disease genes in the human genome. At the time, repository listed 1,284 diseases and 1,777 genes. A disease and a gene are connected with a link if mutations in the gene are implicated in the manifestation of the disease.

The authors manually classified each disease into one of 22 classes, based on the physiological system affected. They analyzed one-mode projections, i.e., derived unipartite models of the bipartite network, creating two projections: a human disease network, obtained connecting any two disease nodes that share at least one gene in common, i.e., gene mutations are associated to both diseases; a gene disease network connecting any two gene nodes connected to a common disease, i.e., known to be associated with the same disorder.

BiNetViz offers an alternative approach for direct investigation of the bipartite network topology. **Figure 8A** illustrates a giant component of the original network formed by $|\mathcal{V}^1| = 516$ nodes representing diseases (depicted as purple hexagons) and $|\mathcal{V}^2| = 903$ nodes representing genes (depicted as green triangles), connected by 1,550 links. **Figures 8B–E** illustrate this giant component at increasing levels of coarsening (levels 1, 2, and 3, respectively). The network was coarsened with the *Gm* algorithm, set with three coarsening levels, each level reduced by a factor of 50% relative to the previous one. **Figure 8E** illustrates the expansion of the super-nodes outlined with a black circle in **Figure 8D** to level 1 (expanded nodes depicted in blue and yellow representing genes and diseases, respectively). This feature enables a user to navigate through multiple levels-of-detail, e.g., getting an overview of the whole network and, at the same time, locally expanding (or contracting) selected regions for a more detailed observation. Therefore, multiple levels-of-detail may be rendered in a single node-link view.

Another interesting possibility afforded by BiNetViz is to conduct exploratory investigations by coarsening just one of the network layers at a time, possibly alternating the analysis focus between the gene and the disease layers. Based on this premise, the node-link view can display the layers at distinct granularity levels, whilst still preserving the full connectivity information between both layers. For instance, coarsening the disease layer, while keeping the gene layer at its original resolution, allows identifying groups of diseases associated with similar genes, which will become super nodes in the coarsened representation. This strategy allows identifying the genes related with similar diseases. Furthermore, a specific super-node representing a group of diseases can be uncoarsened to reveal their connectivity to a particular gene. Next we illustrate such a hypothetical scenario.

Figure 9 illustrates a rendering of the giant component of the network, in which only the gene layer (green triangles) has been coarsened. Coarsening the gene layer allows exploring the network with a focus on the diseases and, if necessary, it is possible to expand the gene-layer



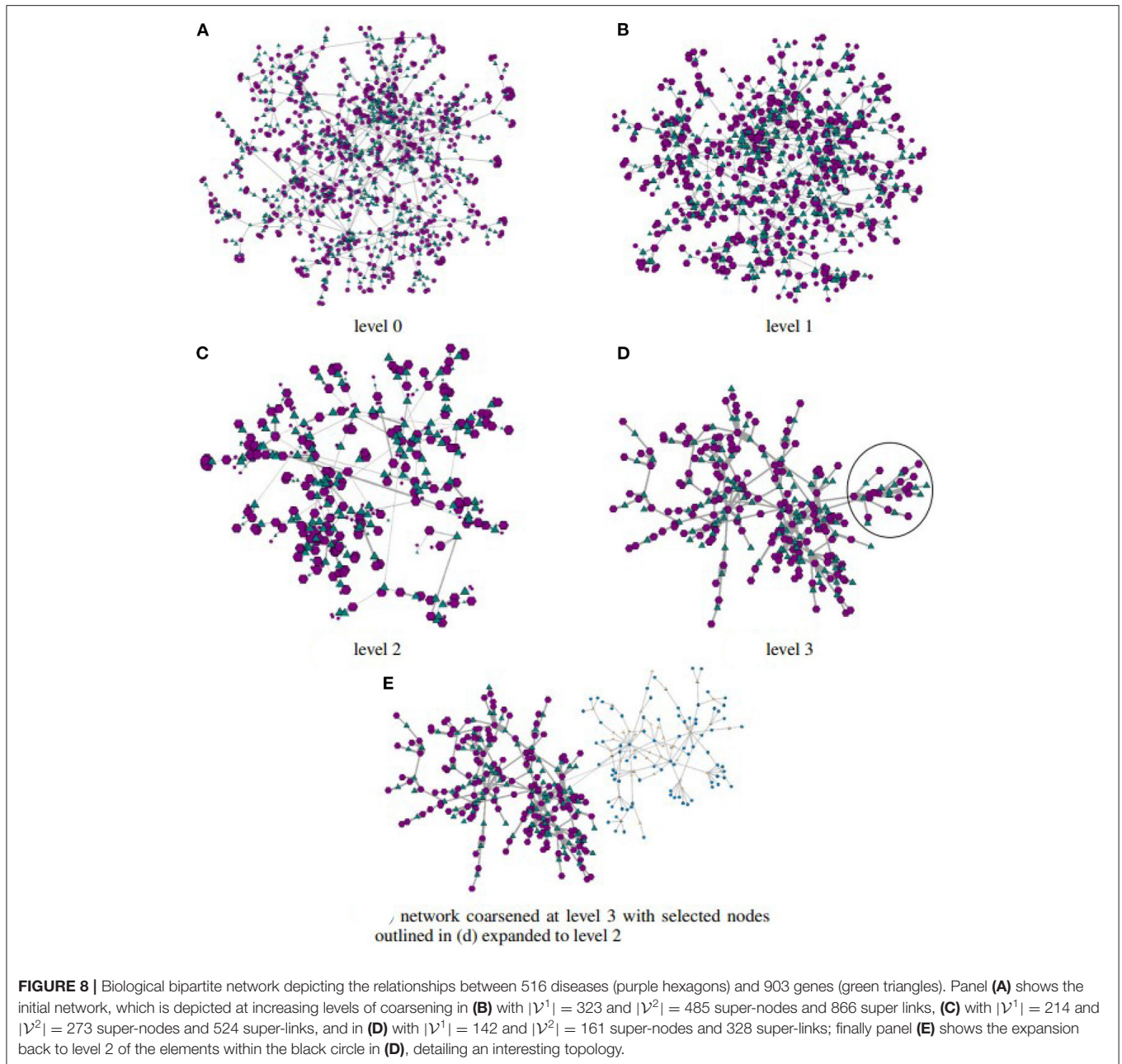
locally on-demand by “exploding” super-nodes into their predecessor nodes. In this perspective, groups of diseases become more evident, which is convenient for detailed exploration.

Highly connected super-vertices thus represent densely connected groups of genes. **Figure 9A** depicts the giant component with the gene layer coarsened at level 4. The degree of the gene nodes is mapped to the triangle size, so that larger triangles indicate higher degrees. We notice how the disease nodes appear clustered around the gene super-nodes, e.g., the two gene super-nodes outlined by the light blue and the black circles are highly connected to several disease nodes. One may expand again those regions, for more detail. **Figure 9B** illustrates the local expansion of the gene super-node (green triangle) enclosed by the black circle. The expanded region includes diseases nodes (purple hexagons) related to “Endocrine” diseases, namely: “Hypothyroidism,” “Goiter,” “Thyroid hormone resistance,” “Total iodide organization defect,” “Graves disease” and “Autoimmune thyroid disease.” Interestingly, the disease highlighted with a green circle represents a type of cancer called “Thyroid carcinoma.” In this case, the TSHR-gene bridges the diseases Hyperthyroidism

(Endocrine) and “Thyroid carcinoma” (Cancer). Notice that the “Thyroid carcinoma” is linked with another group of several cancer-related diseases (highlighted with a light blue circle), such as: “Oligodendrogloma,” “Multiple malignancy syndrome,” “Li-Fraumeni syndrome,” “Sezary syndrome,” “Lymphoma,” “Adrenocortical carcinoma,” “Dermatobrosarcoma protuberans,” “T-cell lymphoblastic leukemia” and “Li-Fraumeni syndrome.” Future investigations could analyze whether the “Thyroid carcinoma,” as a cancer-related to the thyroid, could be explored as a bridge between cancer-related diseases and other thyroid diseases through the TSHR gene.

5.4. Exploring a Scientific Collaboration Network

We include an example illustrating how the BiNetViz approach can be applied to the analysis of scientific collaborations expressed in terms of co-authorship in papers. Such analysis can be aimed at assessing, for instance, scientific influence and scope of collaborations, influential groups of researchers on topics of interest, or the overall distribution structure of groups publishing on a particular topic.

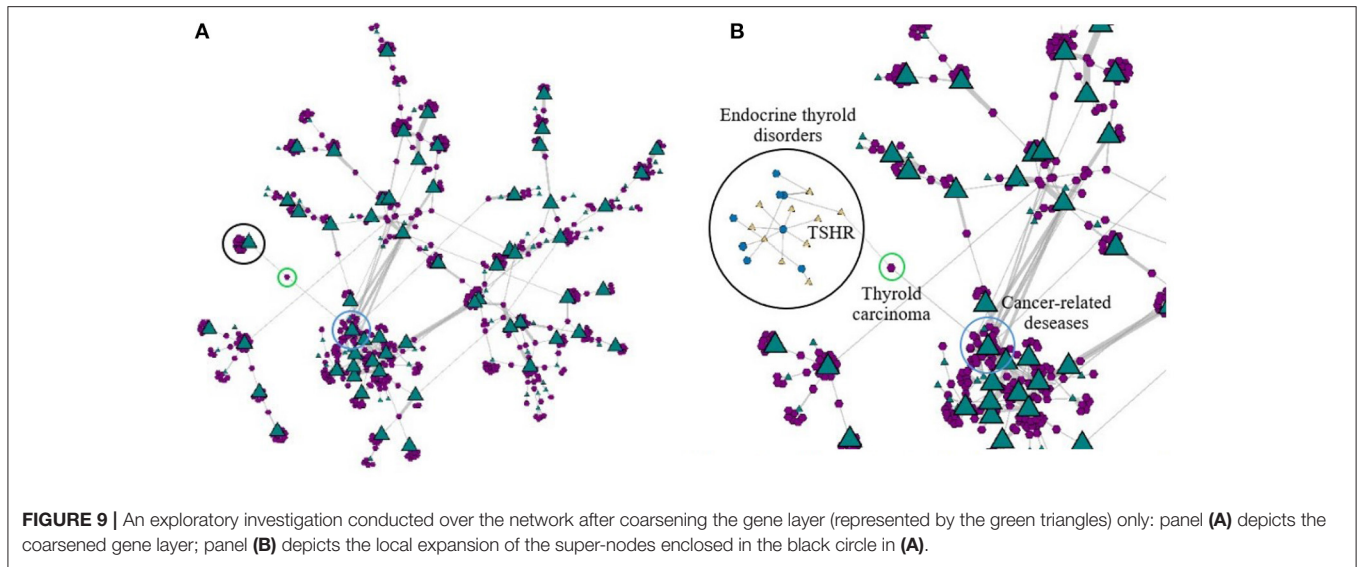


The CORONA-19 database is a freely available comprehensive collection of coronavirus literature accessible for data mining. Created as a joint effort by a coalition of renowned research organizations, the collection includes over 500,000 research papers about COVID-19, SARS-CoV-2, and similar coronaviruses, 200,000 of them with full text.⁵ We considered this database to create a bipartite network of papers and authors, with a focus on publications related with the “CoronaVac” vaccine developed by the Chinese pharmaceutical company

Sinovac Biotech. CoronaVac was one of the earliest vaccines adopted in Brazil, after an agreement for local production. We filtered the database to retrieve only papers published from 2020 for which the title or abstract included the term “CoronaVac” and created the bipartite network formed by the papers retrieved and their authors, linking the papers with their corresponding authors.

The resulting network is formed by $|\mathcal{V}^1| = 3,531$ author nodes and $|\mathcal{V}^2| = 451$ paper nodes, with 5,058 links indicating authorship. This network has been loaded in BiNetViz and it is displayed in **Figure 10A** with a node-link layout generated by the Fruchterman-Reingold algorithm. Author nodes are depicted

⁵<https://www.kaggle.com/datasets/allen-institute-for-ai/CORONA-19-research-challenge>



as purple hexagons and paper nodes are depicted as green triangles. We notice many small isolated communities toward the boundary regions, and larger communities in the central area, apparently more connected, indicating group collaborations. A three-level hierarchy (levels 1 to 3) has been created applying algorithm *Gmb* on both layers, with a target reduction factor at each level equal to 50%. The coarsened network at level 1 is formed by 1,866 super-nodes in the author layer and 335 super-nodes in the paper layer; whereas for the networks at levels 2 and 3 these numbers are, respectively, 1,119 and 563 author super-nodes and 286 and 264 paper super-nodes. **Figure 10B** depicts the coarsest network (level 3), in which the patterns of boundary isolated groups and central connected groups is again very evident. From this representation, users may easily select and expand communities of scientists and papers to obtain relevant contextual information by on-demand navigation and exploration.

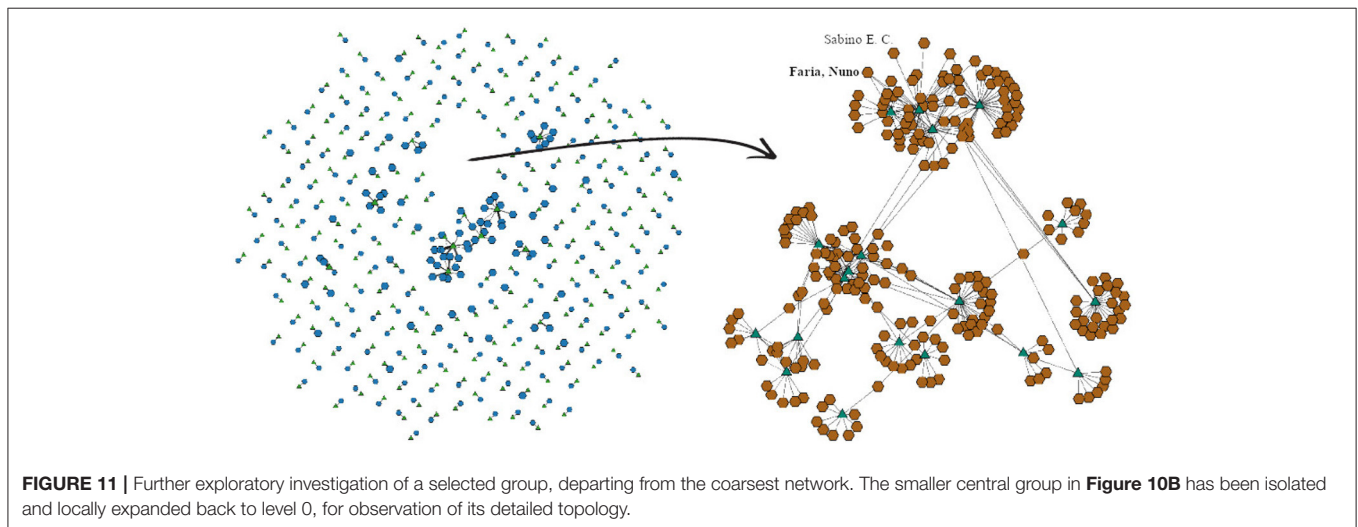
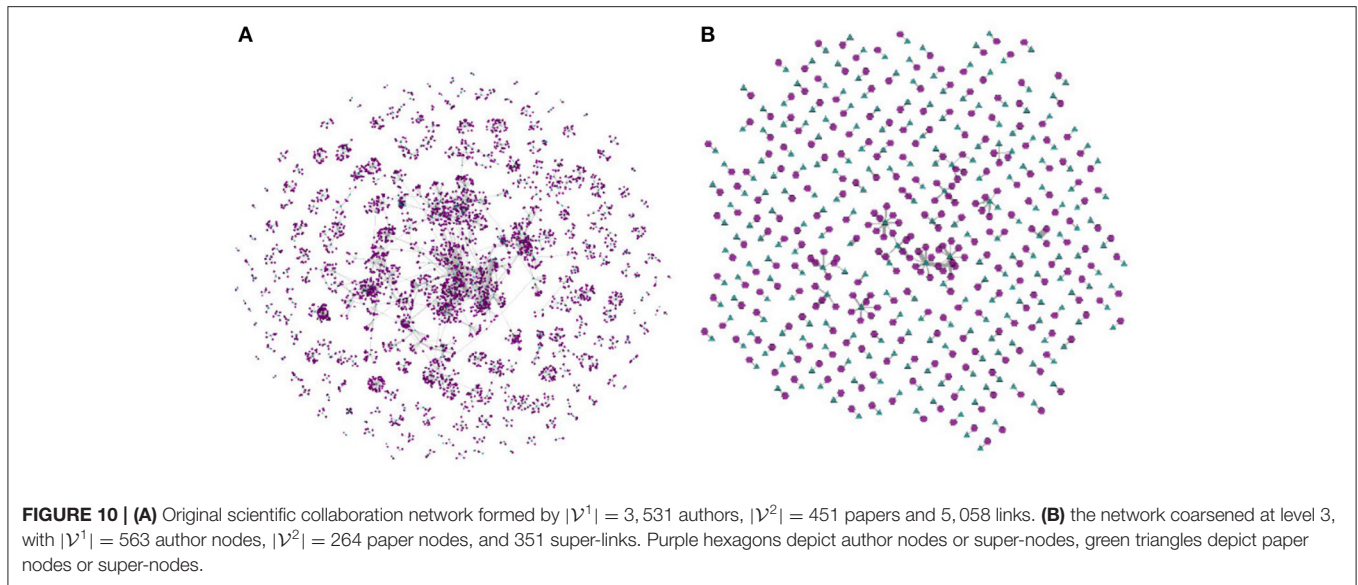
Figures 11, 12 illustrate the expansion of two specific groups back to level 0 (brown octagons correspond to author super-nodes, whereas the green triangles depict paper super-nodes). A feature of BiNetViz enables to select and crop a group of elements, and then reposition it on the screen to navigate over multiple levels-of-detail. In this case the two groups identified in **Figure 10B** have been cropped from the main view and locally expanded. The smaller group is mostly related to Brazilian researchers, while the largest group is characterized by researchers related to the “Sinovac laboratory,” responsible by the “CoronaVac vaccine.” **Figure 11** shows the expansion of the smaller group, which includes several sub-groups depicting teams led by reputed Brazilian researchers who contributed papers on the “CoronaVac” vaccine. For instance, Nuno Faria, a researcher affiliated with the University of Oxford, United Kingdom, has co-authored several studies related to the spreading of COVID-19 in Brazil, in partnership with Brazilian researchers, e.g., Li et al., 2022; Mee et al., 2022; Prete et al., 2022. This group also includes Ester C. Sabino (entry by “Sabino, E.

C.”), immunologist and researcher at the University of São Paulo (USP). Several other visible groups and their connections could also be further investigated.

Figure 12 shows the expansion to level 0 of greater and denser group, which as in the previous case has been isolated prior to expansion. Important information can be identified in this sub-network, which corresponds to the largest group of connected authors and their papers. Highlighted within the blue circle, we notice a highly connected node corresponding to researcher “Zeng, Gang,” who is a medical director at Sinovac, the pharmaceutical company responsible for the development of the CoronaVac vaccine. The brown nodes indicated within the green line correspond to a small group of Brazilian researchers only, authors of the study represented in the green triangle visible in this group (Aikawa et al., 2022). This recently published paper reports results from a study on the immunogenicity pattern induced by the CoronaVac vaccine in SARS-CoV-2 seropositive patients with autoimmune rheumatic diseases. Interestingly, Ester C. Sabino is again one of the co-authors, however, now identified as “Sabino, Ester C.”. Of course, using the solution in this application domain would require additional processing, e.g., to identify replicate author entries. Again, many groups are visible in the detailed visualization that could be further investigated for connections and other information.

6. CONCLUSIONS AND FURTHER WORK

We introduced a visualization framework for bipartite networks assisted by the multilevel method that admits a conceptual organization very consistent with the well-known visual information seeking mantra stated as *overview first, zoom and filter, then details-on-demand*. We introduce a system as a proof-of-concept on the feasibility of employing the multilevel strategy coupled with the familiar node-link views to visualize and investigate large bipartite networks. Our framework allows



departing from an overview of the major topological structures in a network and then focus on relevant elements such as individual or multiple nodes, which may be selected and rendered at a less coarsened level, with further topological details displayed. The combination of the multilevel strategy with suitable software technologies and computationally inexpensive design decisions regarding the rendering of node-link representations yields a visual interface manageable with simple interactive operations that can effectively support navigation on large-scale networks.

The proposed visualization framework may be incorporated into interactive visualization systems of large networked data sets, aimed at different application domains, with added domain-specific functionalities for data analytics. It does require further validation on practical analytical settings, as the usage scenarios of interactive knowledge discovery are inherently complex, and often domain dependent. Indeed,

approaches beyond those considered here could be devised for coarsening networks in general and bipartite networks in particular, e.g., coarsening could consider domain specific properties of the networked data. Further investigation on coarsening choices and their impact on executing exploratory visualization tasks in different application domains is required. As a downside, using this strategy does require from users some familiarity with the multilevel coarsening algorithms.

Beyond data exploration, we believe our current implementation could be tuned into a tool to guide and inform developers of novel multilevel algorithms and applications. For instance, certain multilevel strategies require selecting pivot nodes to guide the coarsening procedure, a task that could benefit from an interactive visual interface. The same applies to developers who wish to compare the outcome of multiple executions of coarsening algorithms, e.g., with different

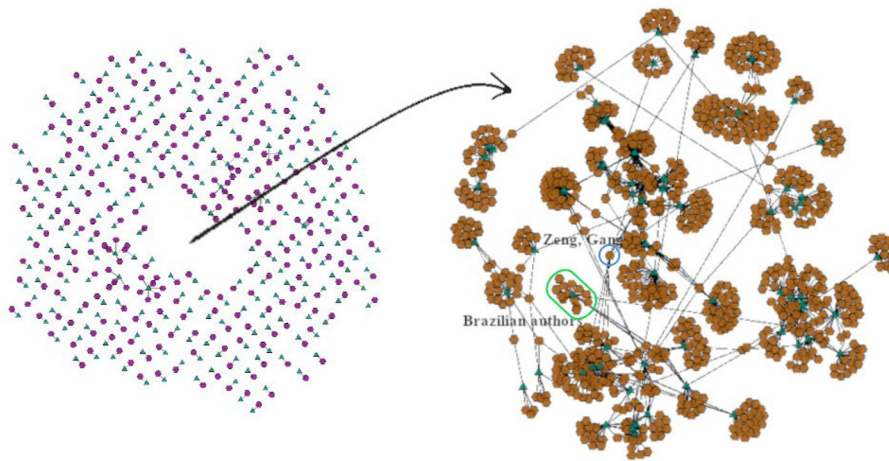


FIGURE 12 | Further exploratory investigation of a selected group, departing from the coarsest network. The bigger central group in **Figure 10B** has been isolated and locally expanded back to level 0, for observation of its detailed topology.

parameter settings. Finally, despite our current focus and interest on bipartite networks, a similar underlying rationale is clearly applicable to unipartite networks, or can be generalized to heterogeneous networks, as long as the underlying multilevel methods are provided to build a hierarchical representation.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://sparse.tamu.edu/HB/jagmesh7>, https://sparse.tamu.edu/Schulthess/N_reactome, and <https://www.pnas.org/content/suppl/2007/05/03/0701361104.DC1>.

AUTHOR CONTRIBUTIONS

AV, AA, and MO contributed the main conceptual ideas and underlying theory. AV, RF, and MO worked out the technical

details in system development. All authors listed contributed in the literature review, planning of the case studies, analysis and interpretation of results, structuring, and writing the text. All authors read and revised the original draft and the final manuscript, and approved it for publication.

FUNDING

This research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001; the Center for Artificial Intelligence (C4AI-USP), with support by the São Paulo Research Foundation (FAPESP) and by the IBM Corporation under grant FAPESP 2019/07665-4; FAPESP grants 2017/05838-3, 2018/22214-6, 2013/07375-0, 2015/50122-0, and 2019/14429-5, and the Brazilian National Research Council (CNPq) research fellowships 304040/2019-3, 301847/2017-7 and 303199/2019-9.

REFERENCES

- Abello, J., Van Ham, F., and Krishnan, N. (2006). Ask-graphview: a large scale graph visualization system. *IEEE Trans. Visual. Comput. Graph.* 12, 669–676. doi: 10.1109/TVCG.2006.120
- Aikawa, N. E., Kupa, L. V., Pasoto, S. G., Medeiros-Ribeiro, A. C., Yuki, E. F., Saad, C. G., et al. (2022). Immunogenicity and safety of two doses of the coronavac sars-cov-2 vaccine in sars-cov-2 seropositive and seronegative patients with autoimmune rheumatic diseases in Brazil: a subgroup analysis of a phase 4 prospective study. *Lancet Rheumatol.* 4, e113–e124. doi: 10.1016/S2665-9913(21)00327-1
- AlQuraishi, M., Koytiger, G., Jenney, A., MacBeath, G., and Sorger, P. K. (2014). A multiscale statistical mechanical framework integrates biophysical and genomic data to assemble cancer networks. *Nat. Genet.* 46, 1363. doi: 10.1038/ng.3138
- Archambault, D., Munzner, T., and Auber, D. (2007). Topolayout: multilevel graph layout by topological features. *IEEE Trans. Visual. Comput. Graph.* 13, 305–317. doi: 10.1109/TVCG.2007.46
- Archambault, D., Munzner, T., and Auber, D. (2008). Grouseflocks: steerable exploration of graph hierarchy space. *IEEE Trans. Visual. Comput. Graph.* 14, 900–913. doi: 10.1109/TVCG.2008.34
- Arleo, A., Didimo, W., Liotta, G., and Montecchiani, F. (2016). “A distributed multilevel force-directed algorithm,” in *International Symposium on Graph Drawing and Network Visualization* (Athens: Springer), 3–17. doi: 10.1007/978-3-319-50106-2_1
- Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56. doi: 10.1038/nrg2918
- Bartel, G., Gutwenger, C., Klein, K., and Mutzel, P. (2010). “An experimental evaluation of multilevel layout methods,” in *International Symposium on Graph Drawing* (Berlin; Heidelberg: Springer), 80–91. doi: 10.1007/978-3-642-18469-7_8
- Bartel, G., Gutwenger, C., Klein, K., and Mutzel, P. (2011). “An experimental evaluation of multilevel layout methods,” in *Graph Drawing* (Berlin; Heidelberg: Springer), 80–91.
- Batagelj, V., Brandenburg, F. J., Didimo, W., Liotta, G., Palladino, P., and Patrignani, M. (2010). Visual analysis of large graphs using (x, y)-clustering

- and hybrid visualizations. *IEEE Trans. Visual. Comput. Graph.* 17, 1587–1598. doi: 10.1109/TVCG.2010.265
- Brandt, A. (1988). “Multilevel computations: review and recent developments,” in *Multigrid methods: Theory, Applications, and Supercomputing, Proceedings of the 3rd Copper Mountain Conference on Multigrid Methods* (Rehovot), 35–62.
- Chan, G. Y., Xu, P., Dai, Z., and Ren, L. (2019). VIBR: visualizing bipartite relations at scale with the minimum description length principle. *IEEE Trans. Visual. Comput. Graph.* 25, 321–330. doi: 10.1109/TVCG.2018.2864826
- Chan, G. Y.-Y., Xu, P., Dai, Z., and Ren, L. (2018). VIBR: visualizing bipartite relations at scale with the minimum description length principle. *IEEE Trans. Visual. Comput. Graph.* 25, 321–330.
- Cintra, D., Valejo, A., Lopes, A., and Oliveira, M. (2019). “Visualization to assist interpretation of the multilevel paradigm in bipartite graphs,” in *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications – IVAPP, 133-140, 2020* (Valletta). doi: 10.5220/0008903501330140
- Di Giacomo, E., Didimo, W., Montecchiani, F., and Tappini, A. (2021). “A user study on hybrid graph visualizations,” in *Graph Drawing and Network Visualization*, eds H. C. Purchase and I. Rutter (Cham: Springer International Publishing), 21–38. doi: 10.1007/978-3-030-92931-2_2
- Dias, D. M., Mansour, M., Dias, F., Petronetto, F., Silva, C. T. and Nonato, L. S. (2017a). “A hierarchical network simplification via non-negative matrix factorization,” in *30th SIBGRAPI Conference on Graphics, Patterns and Images* (Niterói: IEEE), 119–126.
- Dias, M. D., Mansour, M. R., Dias, F., Petronetto, F., Silva, C., and Nonato, L. G. (2017b). “A hierarchical network simplification via non-negative matrix factorization,” in *30th SIBGRAPI Conference on Graphics, Patterns and Images* (SIBGRAPI) (Niterói: IEEE), 119–126. doi: 10.1109/SIBGRAPI.2017.22
- Dörk, M., Riche, N. H., Ramos, G., and Dumais, S. (2012). Pivotpaths: strolling through faceted information spaces. *IEEE Trans. Visual. Comput. Graph.* 18, 2709–2718. doi: 10.1109/TVCG.2012.252
- Faleiros, T. P., Rossi, R. G., and de Andrade Lopes, A. (2017). Optimizing the class information divergence for transductive classification of texts using propagation in bipartite graphs. *Pattern Recogn. Lett.* 87, 127–138. doi: 10.1016/j.patrec.2016.04.006
- Ferreira, V., Valejo, A., and Lopes, A. (2019). “A scalability approach based on multilevel optimization for link prediction methods,” in *Brazilian Conference on Intelligent Systems (BRACIS)* (Salvador), 365–370. doi: 10.1109/BRACIS.2019.00071
- Frishman, Y., and Tal, A. (2007). Multi-level graph layout on the GPU. *IEEE Trans. Visual. Comput. Graph.* 13, 1310–1319. doi: 10.1109/TVCG.2007.70580
- Fruchterman, T. M. J., and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Softw. Pract. Exp.* 21, 1129–1164. doi: 10.1002/spe.4380211102
- Gajer, P., and Kobourov, S. G. (2001). “Grip: graph drawing with intelligent placement,” in *Graph Drawing*, ed J. Marks (Berlin; Heidelberg: Springer), 222–228. doi: 10.1007/3-540-44541-2_21
- García-Algarra, J., Pastor, J. M., Mouronte, M. L., and Galeano, J. (2018). A structural approach to disentangle the visualization of bipartite biological networks. *Complexity* 2018, 6204947. doi: 10.1155/2018/6204947
- Ghoniem, M., Fekete, J.-D., and Castagliola, P. (2005). On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. *Inform. Visual.* 4, 114–135. doi: 10.1057/palgrave.ivs.9500092
- Ghoniem, M., Fekete, J. D., and Castagliola, P. (2004). “A comparison of the readability of graphs using node-link and matrix-based representations,” in *IEEE Symposium on Information Visualization* (Austin, TX), 17–24.
- Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabási, A.-L. (2007). The human disease network. *Proc. Natl. Acad. Sci. U.S.S.* 104, 8685–8690. doi: 10.1073/pnas.0701361104
- Grujić, J. (2008). “Movies recommendation networks as bipartite graphs,” in *International Conference on Computational Science* (Minneapolis, MN: Springer), 576–583. doi: 10.1007/978-3-540-69387-1_66
- Guillaume, J.-L., and Latapy, M. (2004). Bipartite structure of all complex networks. *Inform. Process. Lett.* 90, 215–221. doi: 10.1016/j.ipl.2004.03.007
- Guillaume, J.-L., and Latapy, M. (2006). Bipartite graphs as models of complex networks. *Phys. A Stat. Mech. Appl.* 371, 795–813. doi: 10.1016/j.physa.2006.04.047
- Hachul, S., and Jünger, M. (2004). “Drawing large graphs with a potential-field-based multilevel algorithm,” in *International Symposium on Graph Drawing* (New York, NY: Springer), 285–295. doi: 10.1007/978-3-540-31843-9_29
- Harel, D., and Koren, Y. (2000a). “A fast multi-scale method for drawing large graphs,” in *International Symposium on Graph Drawing* (Colonial Williamsburg, VA: Springer), 183–196. doi: 10.1007/3-540-44541-2_18
- Harel, D., and Koren, Y. (2000b). “A fast multi-scale method for drawing large graphs,” in *Proceedings Working Conference on Advanced Visual Interfaces* (Palermo), 282–285. doi: 10.1145/345513.345353
- Heinrich, J., Seifert, R., Burch, M., and Weiskopf, D. (2011). “Bicluster viewer: a visualization tool for analyzing gene expression data,” in *Advances in Visual Computing* (Berlin; Heidelberg: Springer), 641–652. doi: 10.1007/978-3-642-24028-7_59
- Hinge, A., Richer, G., and Auber, D. (2017). “MuGDAD: multilevel graph drawing algorithm in a distributed architecture,” in *Conference on Computer Graphics, Visualization and Computer Vision* (Lisbon), 189.
- Hoi, S. C., Liu, W., and Chang, S.-F. (2010). Semi-supervised distance metric learning for collaborative image retrieval and clustering. *ACM Trans. Multimedia Comput. Commun. Appl.* 6, 18:1–18:26. doi: 10.1145/1823746.1823752
- Holten, D. (2006). Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. *IEEE Trans. Visual. Comput. Graph.* 12, 741–748. doi: 10.1109/TVCG.2006.147
- Hu, Y. (2005). Efficient, high-quality force-directed graph drawing. *Math. J.* 10, 37–71. Available online at: <http://asus.myds.me:6543/paper/ktall/37%20-%201984%20-%20Efficient,%20High-Quality%20Force-Directed%20Graph%20Drawing.pdf>
- Hwang, T., Scotte, H., Tian, Z., Wu, B., Kocher, J.-P., Wigle, D. A., et al. (2008). Robust and efficient identification of biomarkers by classifying features on graphs. *Bioinformatics* 24, 2023–2029. doi: 10.1093/bioinformatics/btn383
- Kamada, T., and Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* 31, 7–15. doi: 10.1016/0020-0190(89)90102-6
- Karypis, G., and Kumar, V. (1998). A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.* 20, 359–392. doi: 10.1137/S1064827595287997
- Lhuillier, A., Hurter, C., and Telea, A. (2017). State of the art in edge and trail bundling techniques. *Comput. Graph. Forum* 36, 619–645. doi: 10.1111/cgf.13213
- Li, S. L., Acosta, A. L., Hill, S. C., Brady, O. J., de Almeida, M. A., Cardoso, J., et al. (2022). Mapping environmental suitability of haemagogus and *Sabethes* spp. mosquitoes to understand sylvatic transmission risk of yellow fever virus in Brazil. *PLoS Neglect. Trop. Dis.* 16, e0010019. doi: 10.1371/journal.pntd.0010019
- Mee, P., Alexander, N., Mayaud, P., Gonzalez, F., d., J. C., Abbott, S., et al. (2022). Tracking the emergence of disparities in the subnational spread of COVID-19 in Brazil using an online application for real-time data visualisation: a longitudinal analysis. *Lancet Region. Health Am.* 5, 100119. doi: 10.1016/j.lana.2021.100119
- Minatel, D., Ferreira, V., and Lopes, A. (2019). “A multilevel approach for building location-based social network by using stay points,” in *Brazilian Conference on Intelligent Systems (BRACIS)* (Salvador), 359–364. doi: 10.1109/BRACIS.2019.00070
- Minatel, D., Valejo, A., and Lopes, A. (2018). “Trajectory network assessment based on analysis of stay points cluster,” in *Brazilian Conference on Intelligent Systems (BRACIS)* (São Paulo), 564–569. doi: 10.1109/BRACIS.2018.00103
- Newman, M. E. (2001). Scientific collaboration networks. I. network construction and fundamental results. *Phys. Rev. E* 64, 016131. doi: 10.1103/PhysRevE.64.016131
- Noack, A., and Rotta, R. (2009). “Multi-level algorithms for modularity clustering,” in *International Symposium on Experimental Algorithms* (Dortmund: Springer), 257–268. doi: 10.1007/978-3-642-02011-7_24
- Nobre, C., Wootton, D., Harrison, L., and Lex, A. (2020). *Evaluating Multivariate Network Visualization Techniques Using a Validated Design and Crowdsourcing Approach*. New York, NY: Association for Computing Machinery. doi: 10.31219/osf.io/9ndb3
- Pawson, T., and Linding, R. (2008). Network medicine. *FEBS Lett.* 582, 1266–1270. doi: 10.1016/j.febslet.2008.02.011

- Perrot, A., and Auber, D. (2018). Cornac: tackling huge graph visualization with big data infrastructure. *IEEE Trans. Big Data* 6, 80–92. doi: 10.1109/TBDDATA.2018.2869165
- Pezzotti, N., Fekete, J.-D., Höllt, T., Lelieveldt, B., Eisemann, E., and Vilanova, A. (2018). Multiscale visualization and exploration of large bipartite graphs. *Comput. Graph. Forum* 37, 12–25. doi: 10.1111/cgf.13441
- Prete, C. A., Buss, L. F., Buccheri, R., Abraham, C. M., Salomon, T., Crispim, M. A., et al. (2022). Reinfection by the sars-cov-2 gamma variant in blood donors in Manaus, Brazil. *BMC Infect. Dis.* 22, 127. doi: 10.1186/s12879-022-07094-y
- Raghavan, U. N., Albert, R., and Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* 76, 036106. doi: 10.1103/PhysRevE.76.036106
- Sharan, R., Ulitsky, I., and Shamir, R. (2007). Network-based prediction of protein function. *Mol. Syst. Biol.* 3, 88. doi: 10.1038/msb4100129
- Shneiderman, B. (1996). “The eyes have it: a task by data type taxonomy for information visualizations,” in *VL96: Proceedings IEEE Symposium on Visual Languages* (Boulder, CO: IEEE CS Press), 336–343.
- Staudt, C. L., Sazonovs, A., and Meyerhenke, H. (2016). Networkit: a tool suite for large-scale complex network analysis. *Netw. Sci.* 4, 508–530. doi: 10.1017/nws.2016.20
- Steinbock, D., Groller, E., and Waldner, M. (2018). “Casual visual exploration of large bipartite graphs using hierarchical aggregation and filtering,” in *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)* (Konstanz: IEEE), 1–10. doi: 10.1109/BDVA.2018.8533894
- Sun, M., Koop, D., Zhao, J., North, C., and Ramakrishnan, N. (2019). “Interactive bicluster aggregation in bipartite graphs,” in *2019 IEEE Visualization Conference (VIS)* (Vancouver, BC: IEEE), 246–250. doi: 10.1109/VISUAL.2019.8933546
- Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. (2015). “Line: large-scale information network embedding,” in *Proceedings 24th International Conference on World Wide Web* (Florence: International World Wide Web Conferences Steering Committee), 1067–1077. doi: 10.1145/2736277.2741093
- Toosi, F. G., and Nikolov, N. S. (2016). “Vertex-neighboring multilevel force-directed graph drawing,” in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Budapest: IEEE), 002996–003001. doi: 10.1109/SMC.2016.7844696
- Valejo, A., Faleiros, T. P., Oliveira, M. C. F., and Lopes, A. (2020a). A coarsening method for bipartite networks via weight-constrained label propagation. *Knowledge Based Syst.* 195, 105678. doi: 10.1016/j.knsys.2020.105678
- Valejo, A., Ferreira, V., Fabbri, R., Oliveira, M. C. F., and Lopes, A. (2020b). A critical survey of the multilevel method in complex networks. *ACM Comput. Surv.* 53, 35. doi: 10.1145/3379347
- Valejo, A., Ferreira, V., Oliveira, M., and Lopes, A. (2017a). “Community detection in bipartite network: a modified coarsening approach,” in *International Symposium on Information Management and Big Data (SIMBig)* (Lima), 123–136. doi: 10.1007/978-3-319-90596-9_9
- Valejo, A., Lopes, A., Filho, G., Oliveira, M., and Ferreira, V. (2017b). “One-mode projection-based multilevel approach for community detection in bipartite networks,” in *International Symposium on Information Management and Big Data (SIMBig)* (Lima), 101–108.
- Valejo, A., Oliveira, M. C. F., Rocha Filho, G. P., and Andrade Lopes, A. (2018). Multilevel approach for combinatorial optimization in bipartite network. *Knowledge Based Syst.* 151, 45–61. doi: 10.1016/j.knsys.2018.03.021
- Valejo, A., Valverde-Rebaza, J., Drury, B., and Lopes, A. A. (2014a). “Multilevel refinement based on neighborhood similarity,” in *Proceedings of the 18th International Database Engineering & Applications Symposium*, 67–76. doi: 10.1145/2628194.2628227
- Valejo, A., Valverde-Rebaza, J., and Lopes, A. A. (2014b). “A multilevel approach for overlapping community detection,” in *Brazilian Conference on Intelligent Systems (BRACIS)* (São Paulo), 390–395. doi: 10.1109/BRACIS.2014.76
- Valejo, A. D. B., de Oliveira dos Santos, W., Naldi, M. C., and Zhao, L. (2021). A review and comparative analysis of coarsening algorithms on bipartite networks. *Eur. Phys. J. Spec. Top.* 230, 2801–2811. doi: 10.1140/epjs/s11734-021-00159-0
- van der Zwan, M., Codreanu, V., and Telea, A. (2016). CUBu: universal real-time bundling for large graphs. *IEEE Trans. Visual. Comput. Graph.* 22, 2550–2563. doi: 10.1109/TVCG.2016.2515611
- Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J. J., Fekete, J.-D., et al. (2011). Visual analysis of large graphs: state-of-the-art and future research challenges. *Comput. Graph. Forum* 30, 1719–1749. doi: 10.1111/j.1467-8659.2011.01898.x
- Waldner, M., Steinböck, D., and Gröller, E. (2020). Interactive exploration of large time-dependent bipartite graphs. *J. Comput. Lang.* 57, 100959. doi: 10.1016/j.cola.2020.100959
- Walshaw, C. (2000). “A multilevel algorithm for force-directed graph drawing,” in *International Symposium on Graph Drawing* (Colonial Williamsburg, VA: Springer), 171–182. doi: 10.1007/3-540-44541-2_17
- Wong, P. C., Mackey, P., Cook, K. A., Rohrer, R. M., Foote, H., and Whiting, M. A. (2009). “A multi-level middle-out cross-zooming approach for large graph analytics,” in *2009 IEEE Symposium on Visual Analytics Science and Technology* (Atlantic City, NJ: IEEE), 147–154. doi: 10.1109/VAST.2009.5333880
- Xu, P., Cao, N., Qu, H., and Stasko, J. (2016). “Interactive visual co-cluster analysis of bipartite graphs,” in *2016 IEEE Pacific Visualization Symposium (PacificVis)* (Taipei), 32–39. doi: 10.1109/PACIFICVIS.2016.7465248
- Zhao, J., Sun, M., Chen, F., and Chiu, P. (2018). Bidots: visual exploration of weighted biclusters. *IEEE Trans. Visual. Comput. Graph.* 24, 195–204. doi: 10.1109/TVCG.2017.2744458
- Zhao, J., Sun, M., Chen, F., and Chiu, P. (2019). “Missbin: visual analysis of missing links in bipartite networks,” in *2019 IEEE Visualization Conference (VIS)* (Vancouver, BC: IEEE), 71–75. doi: 10.1109/VISUAL.2019.8933639

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Valejo, Fabbri, de Andrade Lopes, Zhao and de Oliveira. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.