Eurographics Workshop on Visual Computing for Biology and Medicine (2019)
J. Byška and G. Mistelbauer (Poster Chairs)

RadTrix: A Composite Hybrid Visualization for Unbalanced Bipartite Graphs in Biological Datasets

Reddy Rani Vangimalla[1] and Jaya Sreevalsan-Nair[†1]

[1]Graphics-Visualization-Computing Lab, International Institute of Information Technology, Bangalore, Karnataka, India.

Abstract

*Bipartite graphs (bigraphs or 2-mode networks) are widely used in ecological networks, biomedical networks, biomolecular networks, protein complexes in protein-protein interaction, epidemiological networks, etc. [PKP*18]. A bipartite graph consists of two disjoint sets of vertices (nodes), with only inter-set edges (links) representing many-to-many mapping between the sets. These graphs are usually visualized using a node-link diagram where the node sets are laid out in parallel lines. However, these node-link diagrams suffer from visual clutter, which particularly worsens in the case of unbalanced sets of nodes. Thus, we are specifically interested in unbalanced bipartite graphs, which occur in biological datasets, e.g.,* ***diseasome****. We propose a novel graph visualization, RadTrix, for unbalanced bipartite graphs, consisting of disjoint sets of nodes with cardinalities N and D, where* $D \ll N$.

RadTrix is a composite visualization which uses matrix representation of the smaller set of nodes and circular/radial graph layout for the larger set in the bipartite graph. It follows the nesting composite visualization view design [JE12]. The matrix layout of nodes follows a two-way one-mode representation of D nodes. Each node on the matrix has four "connection points", on the either ends of the corresponding row as well as column in the matrix. Using the nearest connection point reduces edge length, and thereby reduces clutter. We place the N nodes on the circumference of the circle uniformly. We estimate the radius of the circle, r, as $max(D \times s, 4 \times N)$ *where s is the size of the rendered square cell in the matrix. The matrix and circular layouts of the D and N nodes, respectively, also enable use of visualization marks (matrix cells, glyphs for nodes, etc.) and channels (color, size, etc.). These marks and channels represent properties pertaining to unary and binary operators within the set of D nodes and unary operators within the N nodes.*

We have used RadTrix for a case-study of the disease-gene association network, the diseasome, constructed using the analysis of multi-omics data [VJS16]. RadTrix demonstrates the disease-gene associations, as shown in Figure 1. The user interactions such as highlight of corresponding node-links and gene nodes on mouse over show the set cardinality (Figure 1(C)), and set constituencies (Figure 1(A)) on and off the diagonal, respectively. Overall, our work shows how visual representation with RadTrix layout can be effectively used, due to its reduced edge crossings, and consequent reduced visual clutter. The future scope of our work includes extending this work for bipartite graphs in other domains, and seriating the matrices and the nodes on the circumference for minimizing edge clutter.

CCS Concepts

• ***Human-centered computing*** → *Graph drawings; Empirical studies in interaction design;*

References

[JE12] JAVED W., ELMQVIST N.: Exploring the design space of composite visualization. In *2012 IEEE Pacific Visualization Symposium* (2012), IEEE, pp. 1–8. 1

[PKP*18] PAVLOPOULOS G. A., KONTOU P. I., PAVLOPOULOU A., BOUYIOUKOS C., MARKOU E., BAGOS P. G.: Bipartite graphs in systems biology and medicine: a survey of methods and applications. *Gigascience 7*, 4 (2018), giy014. 1

[VJS16] VANGIMALLA R. R., JEONG H.-H., SOHN K.-A.: Integrative regression network for genomic association study. *BMC Medical Genomics 9*, 1 (2016), 31. doi:10.1186/s12920-016-0192-7. 1

† Senior Member, IEEE

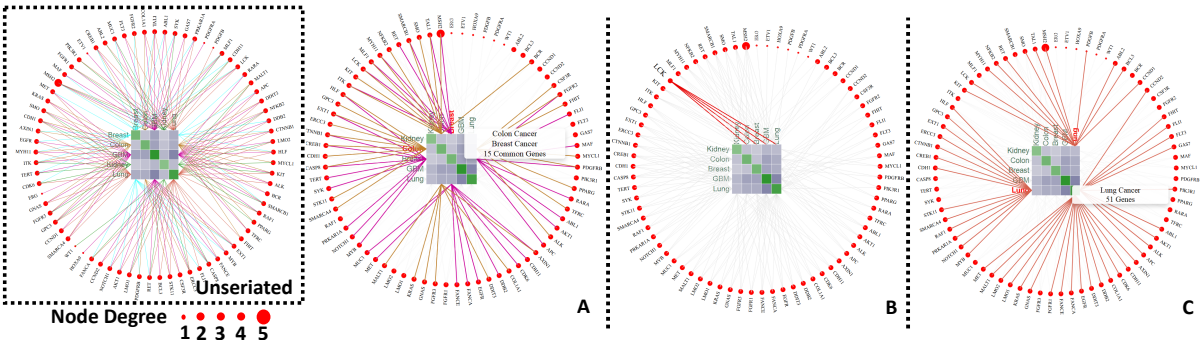

Figure 1: *Visualization of the diseasome, an unbalanced bipartite graph of 73 genes and 5 cancer profiles, using RadTrix, implemented using D3.js. The diseasome has many-to-many mapping, where, 35, 36, 58, 29 and 51 genes correspond to Breast, Colon, GBM, Kidney, and Lung cancer profiles, respectively. (Leftmost) Overview using the unseriated version. Highlighting both matrix and radial nodes, after an initial seriation for clustering, to visualize: (A) common genes between Colon and Breast cancer, (B) a gene common among multiple diseases, and (C) all genes, corresponding to Lung cancer profile.*