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

Reddy Rani Vangimalla and Jaya Sreevalsan-Nair
Graphics-Visualization-Computing Lab,
International Institute of Information Technology Bangalore, India

Problem Statement

A bipartite graph (bigraphs or 2-mode networks) consists of two disjoint sets of vertices (nodes), with only inter-set edges (links) representing many-to-many mapping between the sets.

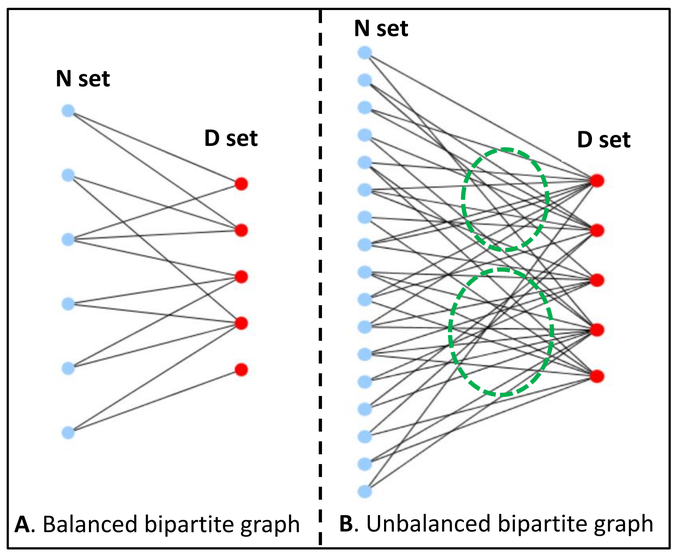
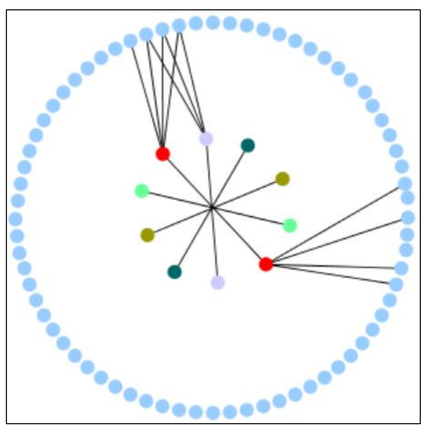


Figure 1*: Example of a balanced and unbalanced bipartite graphs*

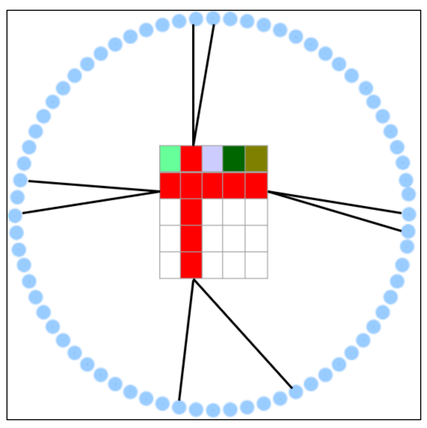
The node-link diagrams suffer from visual clutter, with disjoint node sets of cardinalities N and D, where **D << N**. Hence a novel layout is proposed.

1. Methods and Materials

RadTrix is a ***composite*** of two visualizations in a single view, i.e., matrix visualization for smaller (D) set and circular/radial graph layout for the larger (N) set in a bipartite graph. The layout follows the nesting composite visualization view design [1].



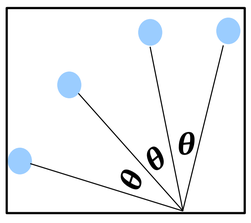
A. Two connection points



B. Four connection points

Figure 2: *Comparison of choice of layout for D nodes*.

- **Reduction in visual clutter**
  - Using quadratic spatial complexity of matrix and 4 connection points.
  - Choice of connection point for a (V_N, V_D) node pairs using shortest distance.
- **Uniform distribution of N nodes on the circumference of circle**



$\theta = 2\pi/N$, N = No of Nodes
Polar Coordinates =
$(r\cos(k\theta), r\sin(k\theta))$
Where, $r \geq 4*N$
$k \in Z$ and $0 \leq k < N$

2. Readability Tasks of RadTrix

Readability assessment tasks that are specified in [3] and are achieved with RadTrix are:

- **Task-1**: Approximate estimation of the number of nodes in the graph, referred to as "**Node Count**".
- **Task-2**: Approximate estimation of the number of links in the graph, referred to as "**Edge Count**".
- **Task-3**: The most connected node, referred to as "**Most Connected**".
- **Task-4**: A node given its label, referred to as "**Find Node**".
- **Task-5**: A link between two specified nodes, referred to as "**Find Link**".
- **Task-6:** A common neighbor between two specified nodes, referred to as "**Find Neighbor**".
- **Task-7**: A path between two nodes, referred to as "**Find Path**".

Additional Tasks RadTrix facilitates on readability are:

- **Task-A**: Analyze the distribution of association across the two sets (of genotypes and phenotypes), namely, to find:
  - ***A1***: Distribution of genes corresponding to each disease, based on the node-degree.
  - **A2**: Distribution of diseases to each gene.
- **Task-B**: Find specific genes exclusive to a specific disease.
- **Task-C**: Find specific genes that belong to at least two diseases.
- **Task-D**: Find the set of diseases a gene associates with.

3. Results: Using RadTrix for Diseasome with 73 Genes and 5 Diseases

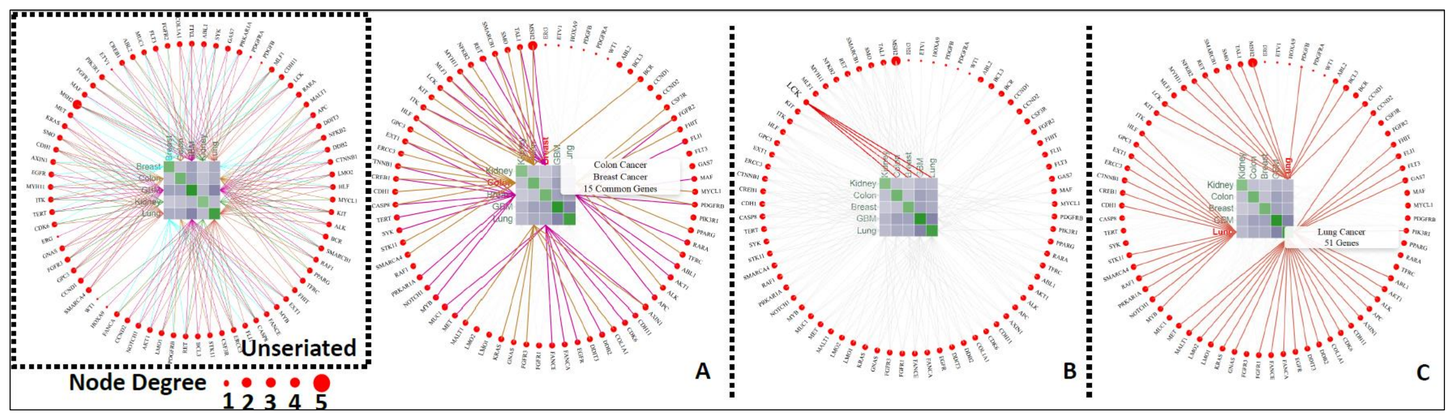


Figure 3: Visualization of the diseasome, implemented using **D3.js**. The diseasome has many-to-many mapping, where, 35, 36, 58, 29 and 51 genes correspond to Breast, Colon, GBM, Kidney, and Lung cancer profiles, respectively. (Leftmost) Overview using the unseriated version. Highlighting both matrix 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.

We have used RadTrix for a case-study of the disease-gene association network, the diseasome, constructed using the analysis of multi-omics data [2]. The user interactions such as highlight of corresponding node-links and gene nodes on mouse over show the set cardinality (Figure 3(C)), and set constituencies (Figure 3(A)) on and off the diagonal, respectively.

4. Comparison with State-of-the-Art Off-the-Shelf Toolkit

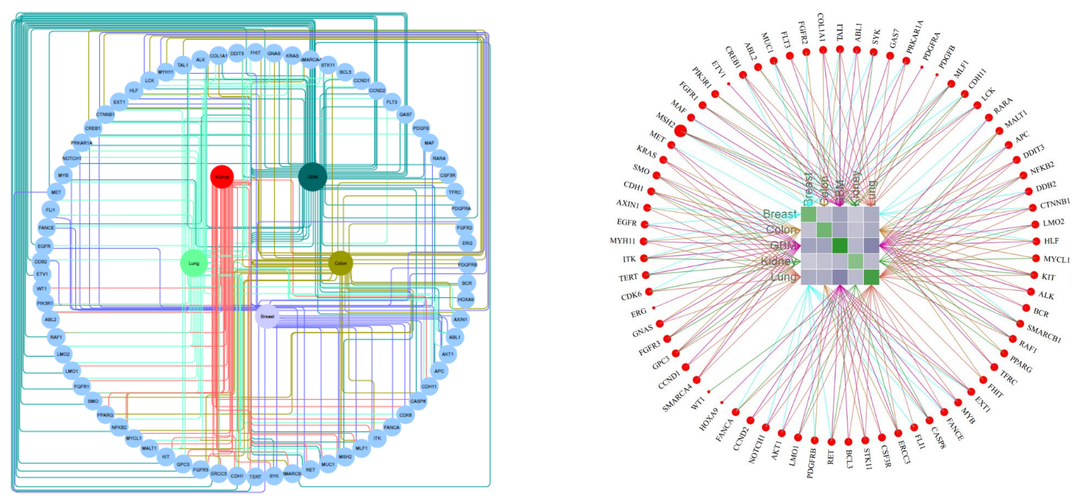


Figure 4: The Orthogonal Edge Router layout of the diseasome in our case study, generated using yFiles, in Cytoscape.

5. Conclusions

- Visual representation with RadTrix layout can be effectively used for understanding the unbalanced bipartite graphs.
- Our proposed layout has met with the design requirements arising from the challenges in the network topologies.

6. Acknowledgements

The authors would like to thank support from the Graphics Visualization-Computing Lab (GVCL) and E-health Research Center (EHRC) at the International Institute of Information Technology Bangalore. The authors are grateful for the financial support from the Visvesvaraya Ph.D. scholarship, initiated by the Ministry of Electronics & IT, Government of India.

Contact

Reddy Rani V & Jaya Sreevalsan-Nair
International Institute of Information Technology, Bangalore.
Reddyrani.Vangimalla@iiitb.org, jnair@iiitb.ac.in
https://www.iiitb.ac.in/GVCL/

References

1. Waqas Javed and Niklas Elmqvist. Exploring the design space of composite visualization. In 2012 IEEE Pacific Visualization Symposium, pages 1–8. IEEE, 2012.
2. Reddy Rani Vangimalla, Hyun-hwan Jeong, and Kyung-Ah Sohn. Integrative regression network for genomic association study. BMC Medical Genomics, 9(1):31, 2016.
3. Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Information Visualization, 4(2):114–135, 2005