Using Linear Visualization to Explore Large Graphs

William Longabaugh
Senior Software Engineer
Institute for Systems Biology

Exploring Graphs at Scale (EGAS) 2015
IEEE VIS 2015 Workshop
October 26, 2015
Node-Link Diagrams: Very Intuitive, BUT…

Figure from: Jacob L. Moreno, "Who Shall Survive", 1934.
What worked with hand-drawn small datasets fails with computer-driven big datasets!

Data Source: Regulon DB, Release 8.5, 11-28-2013
BioFabric: Nodes as Horizontal Lines

Data Source: Regulon DB, Release 8.5, 11-28-2013
BioFabric: *Change the Basic Geometry*

- It’s a node-link diagram, *but*…
- Nodes are represented as *horizontal lines, not as points*
BioFabric: Introduction

- Can be used for both directed and undirected networks
- Shadow Links: both endpoints get a personal copy
Super-Quick Demo


BioFabric has been around for a while…

BioFabric was published in 2012:

UML Sequence Diagrams: *Nodes as Lines* are Intuitive if Node Entity is Shown as Passing Through Time

- Figure from: http://kenhoward01.blogspot.com/2008/06/three-little-pigs-in-uml.html
History: Visibility Representations, Where Edges Do NOT Cross Node Blocks

http://pigale.sourceforge.net/images.html

Contrast BioFabric: Orthogonal Node/Edge Crossings are **Ubiquitous** and **Unremarkable**

Re: ignorable perpendicular links, see:
Huang, Hong, Eades. Pacific Visualization, 2008.

Instead, links *drawn darker*, in front, and *endpoints* are tagged to stand out

*(N.B.: The recurring wedge pattern is indicative of a clique)*

Data source: http://string-db.org/
Matrix-Based: GeneaQuilts

Rows are individuals, columns are parent/child sets

• Note how the parallel sets of links are laid out to form distinctive edge wedges for each node. Best edge arrangements are contiguous sets, ordered shortest to longest.

• Link and node colors are only used to help the user to track features and maintain context while scanning vertically or horizontally. Links are drawn darker than nodes to make them stand out.

• Alternating light pink/blue blocks highlight edges contained in each node zone.

Data from the Human Connectome Project (Van Essen and Ugurbil)  
http://biovis.net/year/2014/info/data_contest
An important optional BioFabric feature is *shadow links*.
Each edge is *duplicated*, so one copy can be tightly associated with *each* endpoint node in the dedicated *node zone* on the diagonal.
The above network is the *same* as shown previously, but with shadow links added.

Data from the Human Connectome Project (Van Essen and Ugurbil)
http://biovis.net/year/2014/info/data_contest
Comparison of “Edge Wedges”

- The links above the main diagonal are the shadow links that are duplicates of the original links drawn below the diagonal.
- Now each blue/pink node zone contains all the links incident on each node, and every node appears along the prominent diagonal.
- With shadow links enabled, the complete node neighborhoods of sets of nodes can be directly compared using the visually distinct shapes of their edge wedges.

Data from the Human Connectome Project (Van Essen and Ugurbil)  
http://biovis.net/year/2014/info/data_contest
Various Random Networks

Data generated by igraph R package (Csardi and Nepusz, 2006)

~Two hops from Node 1 gets to other side of network

Node 2 shares ~30% of its neighbors with Node 1
Clustering is a Common Network Analysis Technique

BioFabric View of Same Network

- Separation of concerns: *Intra-cluster* links are grouped *separately* from *inter-cluster* links!
Very Large Networks: Stanford Web Network

Nodes: 281,903  Edges: 2,312,497

Default layout is breadth-first search from highest degree node (edge direction ignored); search ordered by decreasing neighbor degree. Layout time is terrible for this example due to naïve implementation, but pan and zoom are acceptable.

Instead of showing *one* edge wedge per node, we can split the set of edges up into subsets (*link groups*) and display them as *multiple* edge wedges.

To compare two or more networks:

- Create the union of the networks
- Assign the edges for each network to different link groups
- Compare the networks side by side for each node
Network Comparison Using Link Groups

• Targets of four regulators in oleate versus glucose (Oaf1p, Pip2p, Oaf3p and Adr1p)
• No shadow links used in this case

“Venn Diagram” Networks From Blakley, Blakley, and Blakley (2014)

Describes a nodes-as-lines technique for an \((M, N)\) network comparison that juxtaposes the networks \((M \setminus N)\), \((M \cap N)\), and \((N \setminus M)\). Their Fig. 10:

![Graph](image)

Fig. 10. Two graphs compared in a single figure

“Venn Diagram” Networks with Link Groups

• What happens when we apply this (M | N), (M ^ N), and (N | M) approach “in the small” when assigning link groups?

• The user can visually estimate the Jaccard Index |M ^ N| / |M U N| for each region by comparing width of the center wedge |M ^ N| to full width of the node region |M U N|.

• It’s both a network visualization and a (distributed) Venn Diagram.

Data from the Human Connectome Project (Van Essen and Ugurbil)
http://biovis.net/year/2014/info/data_contest
BioFabric and Social Networks: Clusters

769 nodes and 16,656 links (2x for shadows)

Data Source: Data: Traud et al., SIAM Review 53(3), 2011
http://sociograph.blogspot.com/2011/02/visualizing-large-facebook-friendship.html
Multigraph “Hack” to Mimic Edge Weights

• Using edge thickness to represent weights fits well with overall approach.
• Arguably important to round edge “thicknesses” to the fixed grid.
• Current version of BioFabric does not support this.
• In the above example, approximate this by drawing multiple edges to represent thickness.

Data Source: http://circos.ca/presentations/articles/vis_tables1/#example2
Scaling BioFabric: Vertical Compression by Tossing Out Nodes

Lower right: vertical compression by throwing away unused nodes

Scaling BioFabric: Keep ALL Edges, but Group Many Nodes As One

Transitive reduction citation network, without modification, in BioFabric

Scaling BioFabric: Keep ALL Edges, but Group Many Nodes As One

*Keep all the links*, but combine e.g. 25 nodes into one node. This makes the *edge wedges larger* and more visible.

Network Radius: Watts and Strogatz Model

1000 nodes, 1-D lattice, 10 neighbors, p = 0 (top) and .01 (bottom)

Radius (top) = 50
Radius (bottom) = 6

Data generated by igraph R package (Csardi and Nepusz, 2006)
Some Final Points

• We can create a useful visualization, despite billions of individual line intersections.

• Standard Java2D rendering is adequate. No crippling aliasing artifacts, and the system does a good job of painting hundreds of lines passing through a single pixel.

• This might not be intuitive “out of the box”, but can be learned. Compared to an adjacency matrix, we are retaining the intuitive node and link model.

• Node layout is simply a linear ordering, as is edge layout. The standard “edge wedge” link layout approach is often sufficient.

• The added degree of freedom allows the user to create meaningful semantic groupings of edges.

• With edges, more is better!
BioFabric was developed with support from the National Cancer Institute, Award U24CA143835 (Ilya Shmulevich, PI).

Author is currently supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, Award R01HD073113.

www.BioFabric.org

Combing the Hairball blog: biofabric.blogspot.com
@wjrl59 on Twitter