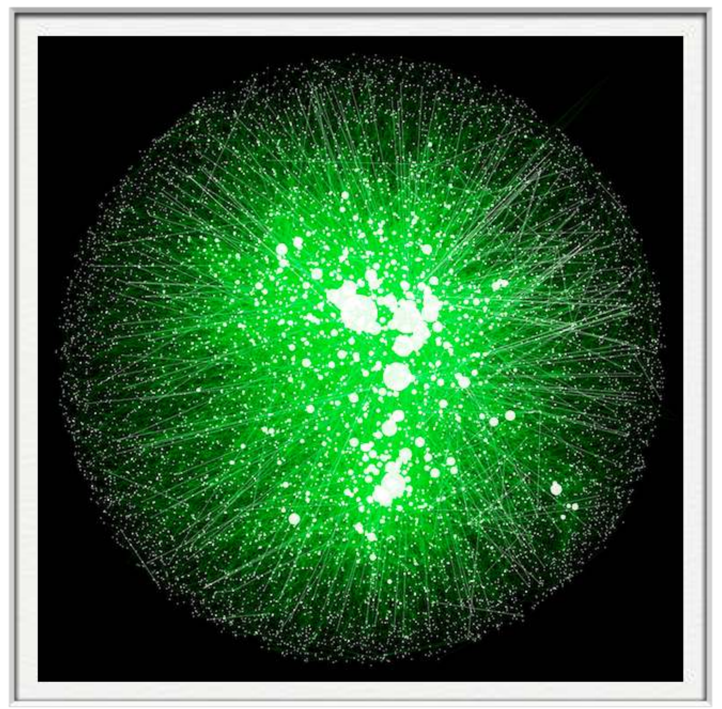
GraphPrism: Compact Visualization of Network Structure

Sanjay Kairam, Diana MacLean, Manolis Savva, Jeffrey Heer



Motivation

Traditional network visualization approaches, such as node-link diagrams, were developed for smaller graphs. When these techniques are applied to larger graphs, the results may be beautiful, but they are not always informative.

Research Question

How can we design a visualization technique which supports important network analysis tasks involving the characterization, comparison, and classification of large networks?

APPROACH

Instead of showing information about individual nodes and edges, we visualize information about the graph as a whole through the construction of ***facets***.

Each facet corresponds to a single node- or edge-specific metric. Facets are constructed by calculating distributions of the values of a given metric over increasingly large local neighborhoods (this process is explained in more detail below).

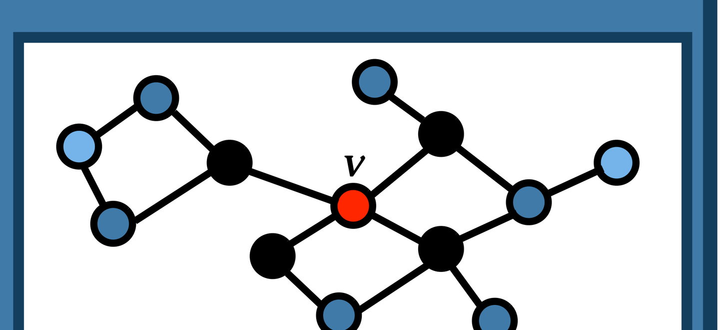
The resulting distributions are visualized as a stacked, multi-scale histogram. Facets can be inspected to determine values of specific graph statistics such as degree distribution or global transitivity or viewed holistically to observe global patterns of graph connectivity.

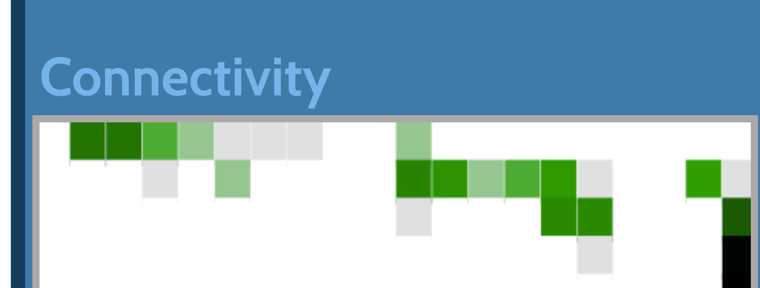
Constructing Facets

We illustrate how facets are constructed by considering the ***Connectivity*** facet. This facet represents the metric ***l*-connectivity**, or the ratio of nodes reachable by a given node v within *l* hops along the edges of the graph. Thus, the 1-connectivity (nodes reachable in one hop) is simply the degree of *v*.

Consider this example graph and the node *v* marked in red. There are 12 other nodes in the graph.

$1\text{-connectivity}(v) = 4/12 = 0.33$
$2\text{-connectivity}(v) = 10/12 = 0.83$
$3\text{-connectivity}(v) = 12/12 = 1.00$





The top row of this ***Connectivity*** facet represents the distribution of 1-connectivity values for the given graph. The second row represents 2-connectivity values, and so on.

Relevant Network Analysis Concepts

Degree	For a node *v* in a graph *G*, the number of edges incident on *v*.
Diameter	The length of the maximum shortest path between any pair of connected nodes *u*, *v* ∈ *G*.
Triad	A triplet of connected nodes. A triad is closed if edges connect all three nodes to each other.
Local Transitivity	For some *v* ∈ *G*, the fraction of triads in *G* which are closed for the subgraph defined by *v* and its immediate neighbors.
Global Transitivity	The fraction of all triads in *G* which are closed.

METRICS

The method used to create the ***Connectivity*** facet can be used to create facets for any metric which can be calculated over a node or set of nodes. In our analysis, we considered the following:

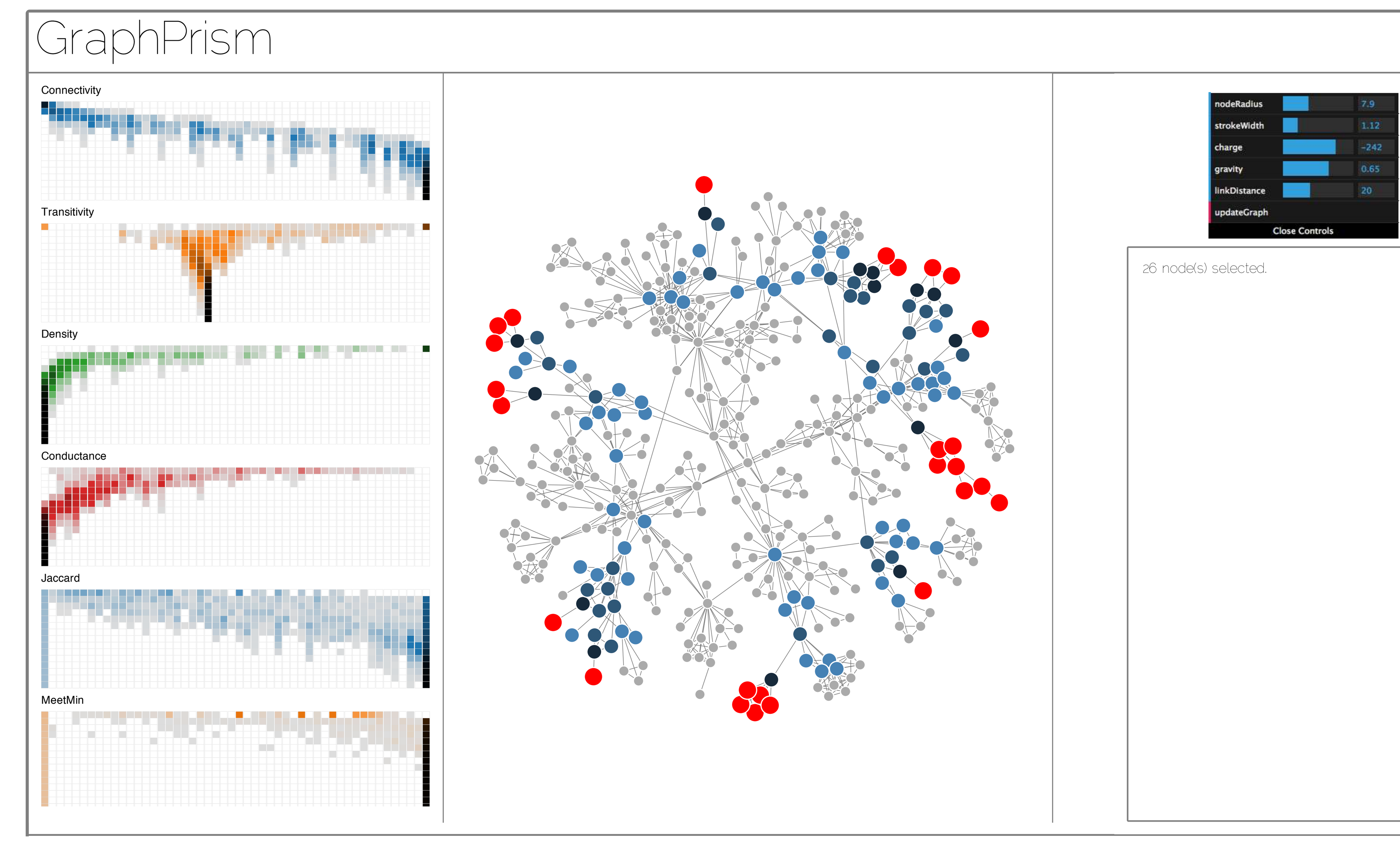
In-Connectivity: Ratio of nodes which can reach *v* within *l* hops.
Out-Connectivity: Ratio of nodes which *v* can reach within *l* hops.
Transitivity: Global transitivity of subgraph defined by the *l*-neighborhood of *v*.
Density: Edge density of subgraph defined by the *l*-neighborhood of *v*.
Conductance: Conductance of node subset defined by *l*-neighborhood of *v*.

We extend this concept to the analysis of edges by considering the *l*-neighborhoods of the two nodes (*u*, *v*) joined by a given edge. We considered the following edge-specific metrics, as well:

Jaccard Overlap: Ratio of the size of the intersection of the *l*-neighborhoods of *u* and *v* to the size of the union of the two neighborhoods.
MeetMin Overlap: Ratio of the size of the same intersection to the size of the *l*-neighborhood which is smaller.

CASE STUDY

Using data from a Network Science co-authorship network, we illustrate how GraphPrism might be used in tandem with a node-link view as part of an interactive system.



Complementing the Node-Link Diagram

Look at the first row of the ***Connectivity*** facet; you can see that node degrees follow what may be a power-law distribution.

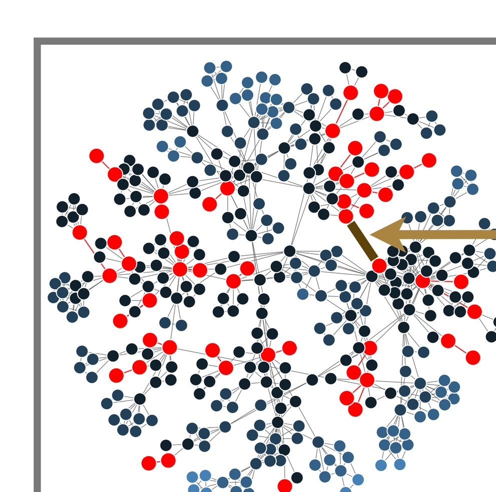
The ***Transitivity*** facet as a whole shows a relatively high global transitivity (~0.45), and the first row reveals that many nodes have local transitivity at or near 1 (resulting from the fact that papers with multiple authors create cliques in the network.

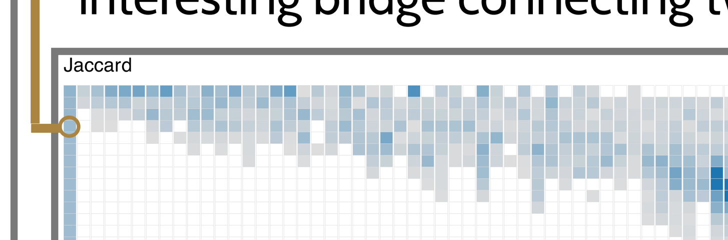
Interaction via Linked Selection

The main interaction technique is **linked selection** (brushing & linking) between GraphPrism facets and the node-link diagram.

In the picture above, we have selected the leftmost colored cell in the third row of the ***Connectivity*** facet. This cell represents the nodes which can reach the fewest nodes within 3 hops (who may represent 'outsiders' to this community). These nodes are highlighted in red; their neighbors are also highlighted in colors ranging from **black** for nodes immediately neighboring a selected node to blue for more distant neighbors.

Edges connecting nodes which are loosely connected or disconnected are called *bridges*. We can identify bridges using the ***Jaccard*** facet by selecting cells on the leftmost column (**Jaccard Overlap** = 0). The cell circled in the figure below identifies bridges which span a distance of at least 8. This helps us to identify an interesting bridge connecting two clusters (highlighted in brown).*





*Additional investigation reveals that this edge represents a paper by G. Bianconi and A. Capocci entitled "Number of Loops of Size *h* in Growing Scale-Free Networks", which interestingly creates its own sizable loop in this network.