Visual

Analytics

Graphs & Trees

- **Graph**

  Vertex/node with one or more edges connecting it to another node.

  Cyclic or acyclic

  Edge can be weighted (value) or categorized

- **Tree**

  Undirected graph where two nodes are connected by only one edge

  used for hierarchy

  Rooted or unrooted

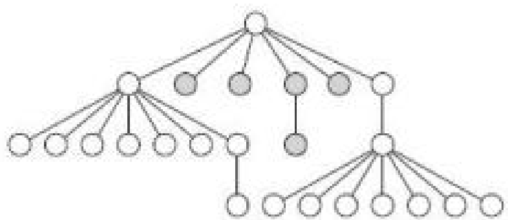
  Edge can be weighted (value) or categorized

Design Choices

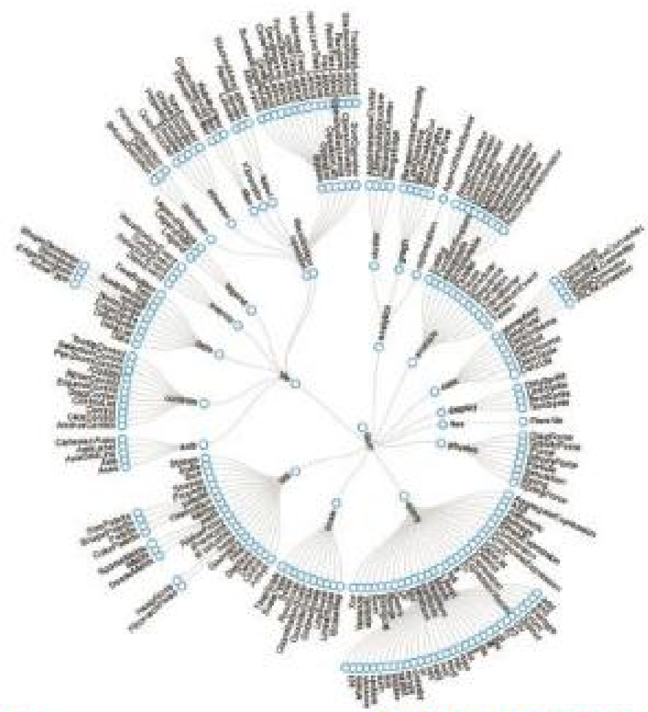
- Connectivity
  - Node-link graphs
  - Good for finding pairwise/multiway relations
  - Good for following paths through structure
  - Force-directed placement
- Containment
  - Effective at showing hierarchical structure
  - Good for finding attributes of leaf nodes
  - Treemaps, nested views
- Matrices

Node link diagrams

- The most common visual encoding idiom for tree and network data is with node–link diagrams, where nodes are drawn as point marks and the links connecting them are drawn as line marks.



(a) **Triangular vertical**



(b) **Spline radial layout**

Node link diagrams

- Networks are also very commonly represented as node–link diagrams, using connection.
- Nodes that are directly connected by a single link are perceived as having the tightest grouping, while nodes with a long path of multiple hops between them are less closely grouped.
- The number of hops within a path - the number of individual links that must be traversed to get from one node to another - is a network-oriented way to measure distances.
- The connection marks support path tracing via these discrete hops.

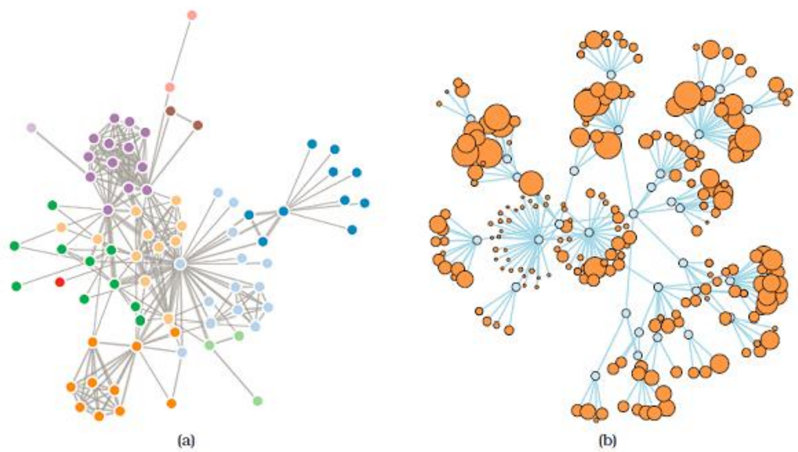
tasks that involve understanding the network topology: the direct and indirect connections between nodes in terms of the number of hops between them through the set of links.

- Examples of topology tasks include finding all possible paths from one node to another, finding the shortest path between two nodes, finding all the adjacent nodes one hop away from a target node, and finding nodes that act as a bridge between two components of the network that would otherwise be disconnected.

Ex: Force-Directed Placement

- One of the most widely used idioms for node–link network layout using connection marks is force-directed placement.
- There are many variants in the force-directed placement idiom family; in one variant, the network elements are positioned according to a simulation of physical forces where nodes push away from each other while links act like springs that draw Force-directed placement their endpoint nodes closer to each other.
- Strengths: Very easy to implement. Relatively easy to understand and explain at a conceptual level,
- Analyzing the visual encoding created by force-directed placement is somewhat subtle.
- Spatial position does not directly encode any attributes of either nodes or links; the placement algorithm uses it indirectly.
- A tightly interconnected group of nodes with many links between them will often tend to form a visual clump, so spatial proximity does indicate grouping through a strong perceptual cue.

- However, some visual clumps may simply be artifacts: nodes that have been pushed near each other because they were repelled from elsewhere, not because they are closely connected in the network. Thus, proximity is sometimes meaningful but sometimes arbitrary; this ambiguity can mislead the user.



(a) with size coding for link attributes. (b) with size coding for node attributes.

- Weekness

  One weakness of force-directed placement is that the layouts are often nondeterministic, meaning that they will look different each time the algorithm is run, rather than deterministic approaches such as a scatterplot or a bar chart that yield an identical layout each time for a specific dataset.

  A major weakness of force-directed placement is scalability, both in terms of the visual complexity of the layout and the time required to compute it.

Multilevel network idioms

- Many recent approaches to scalable network drawing are multilevel network idioms, where the original network is augmented with a derived cluster hierarchy to form a compound network.
- The Cluster hierarchy is computed by coarsening the original network into successively simpler networks that nevertheless attempt to capture the most essential aspects of the original's structure.

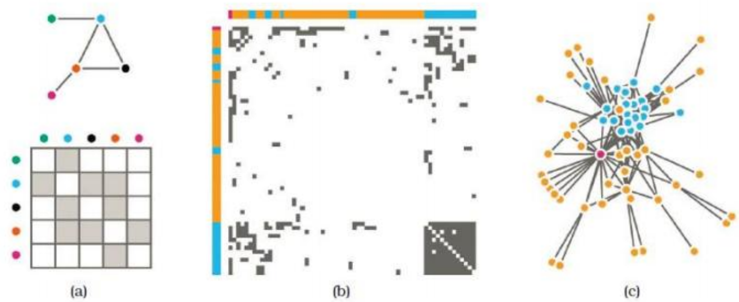
Idiom	Force-Directed Placement
What: Data	Network.
How: Encode	Point marks for nodes, connection marks for links.
Why: Tasks	Explore topology, locate paths.
Scale	Nodes: dozens/hundreds. Links: hundreds. Node/link density: $L < 4N$

Idiom	Multilevel Force-Directed Placement (sfdp)
What: Data	Network.
What: Derived	Cluster hierarchy atop original network.
What: Encode	Point marks for nodes, connection marks for links.
Why: Tasks	Explore topology, locate paths and clusters.
Scale	Nodes: 1000–10,000. Links: 1000–10,000. Node/link density: $L < 4N$.

Idiom	Adjacency Matrix View
What: Data	Network.
What: Derived	Table: network nodes as keys, link status between two nodes as values.
How: Encode	Area marks in 2D matrix alignment.
Scale	Nodes: 1000. Links: one milllion.

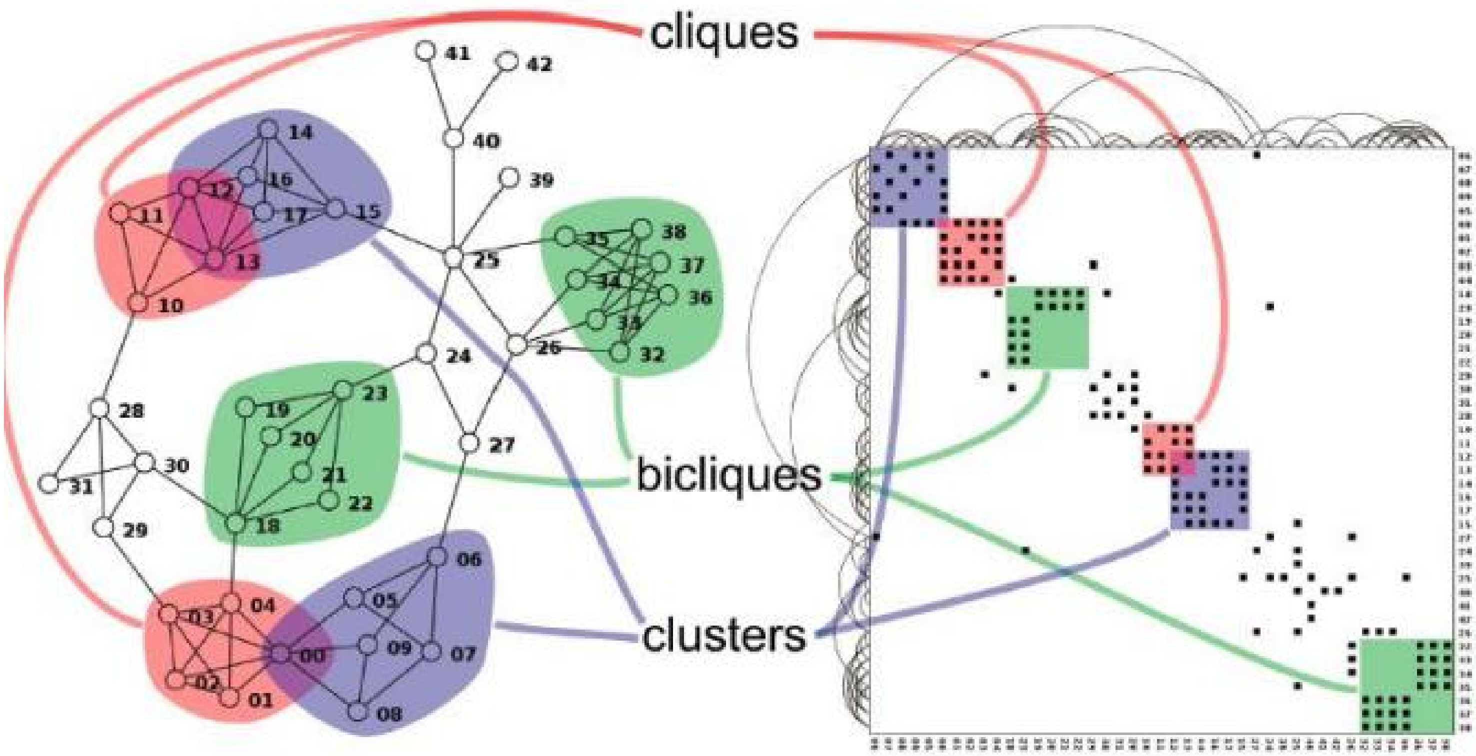
Matrix Views

- Network data can also be encoded with a matrix view by deriving a table from the original network data.
- **Example: Adjacency Matrix View**
- where all of the nodes in the network are laid out along the vertical and horizontal edges of a square region and links between two nodes are indicated by coloring an area mark in the cell in the matrix that is the intersection between their row and column.
- That is, the network is transformed into the derived dataset of a table with two key attributes that are separate full lists of every node in the network, and one value attribute for each cell records whether a link exists between the nodes that index the cell.



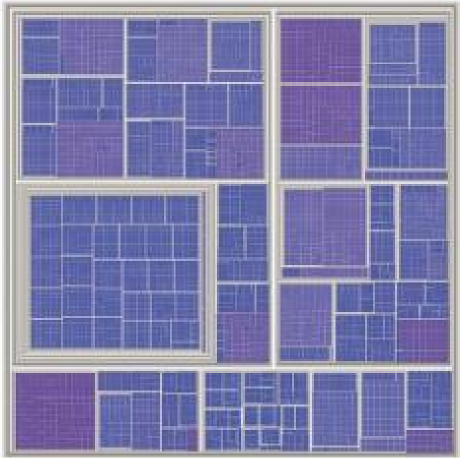
Matrix views of networks can achieve very high information density, up to a limit of one thousand nodes and one million edges, just like cluster heatmaps and all other matrix views that use small area marks.

Costs and Benefits: Connection versus Matrix



Containment: Hierarchy Marks

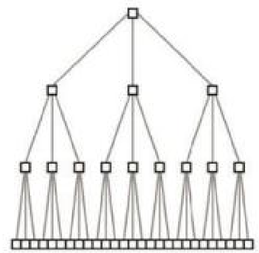
- Containment marks are very effective at showing complete information about hierarchical structure, in contrast to connection marks that only show pairwise relationships between two items at once.
- Tree Maps: The idiom of treemaps is an alternative to node–link tree drawings, where the hierarchical relationships are shown with containment rather than connection.



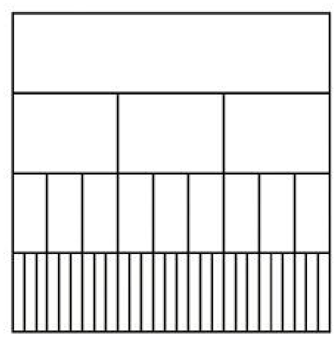
Seven different visual encoding idioms for tree data

Using different combinations of visual channels.

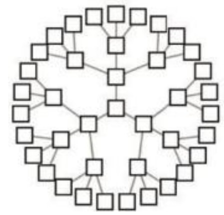
(a) Rectilinear vertical node–link, using connection to show link relationships, with vertical spatial position showing tree depth and horizontal spatial position showing sibling order.



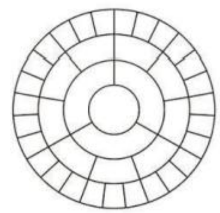
- (b) Icicle, with vertical spatial position and size showing tree depth, and horizontal spatial position showing link relationships and sibling order.



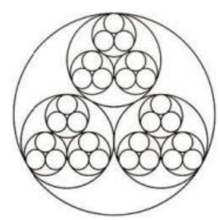
(c) Radial node–link, using connection to show link relationships, with radial depth spatial position showing tree depth and radial angular position showing sibling order.



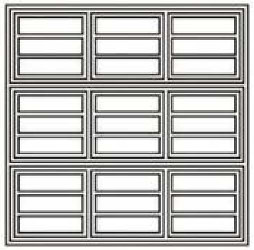
(d) Concentric circles, with radial depth spatial position and size showing tree depth and radial angular spatial position showing link relationships and sibling order.



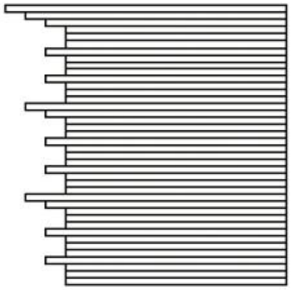
(e) Nested circles, using radial containment, with nesting level and size showing tree depth.



(f) Treemap, using rectilinear containment, with nesting level and size showing tree depth.



(g) Indented outline, with horizontal spatial position showing tree depth and link relationships and vertical spatial position showing sibling order.



Compound Network

- A compound network is a combination of a **network** and a **tree** on top of it, where the nodes in the network are the leaves of the tree.
- Thus, the interior nodes of the tree encompass multiple network nodes.

(a) shows a network (b) shows a cluster hierarchy built on top of it. (c) shows a combined view using of containment marks for the associated hierarchy and connection marks for the original network links