POINTS OF VIEW

Networks

We describe graphing techniques to support exploration of networks.

Most biological phenomena arise from the complex interactions between the cell's many constituents such as proteins, DNA, RNA and small molecules. The graphical representations of networks can be useful in exploring this complex web of interactions. Choosing a suitable network visualization based on the patterns one hopes to highlight can yield meaningful insights into data.

Various techniques developed for visualizing networks will bring out different salient qualities of relational data. Two relevant features of networks are hubs and clusters. Hubs are single nodes connected to many other nodes, and clusters are sets of highly interconnected nodes. These data features characterize different classes of networks. The goal is to choose a graphing technique that is appropriate to the scale of the data and a resolution at which we care to study the networks.

Networks are known as graphs in mathematics and describe a set of pairwise relationships. A common plotting technique for such data is as 'node-link' diagrams (**Fig. 1**). In biology, these diagrams typically represent molecules as nodes and the connections between the nodes as straight or curved lines (also known as edges). A network is said to be directed if the edges are asymmetric (**Fig. 1a**) and undirected if the edges are symmetric (**Fig. 1b**,**c**). Cytoscape[1] and Gephi (http://gephi.org/) are two popular and freely available software tools for generating network diagrams.

Node-link diagrams have the distinct advantage of preserving the local detail of the network, making it easy to identify nearest neighbors for a particular node and to trace paths through the network. With these diagrams, different layouts of the same data can dramatically affect how we perceive the relationships of the data objects. For example, a circular layout with nodes sequenced by their number of connections can reveal the general connectedness of a network (**Fig. 1b**). However, layouts that simulate physical systems (for example, imagining connections as forces or springs) will often produce visible aggregates of nodes, making it easier to spot hubs and clusters (**Fig. 1c**). Node-link diagrams can be highly useful but unfortunately do not scale well. As a dataset becomes larger, the visual complexity that results from the added information density approaches an incomprehensible 'hairball'.

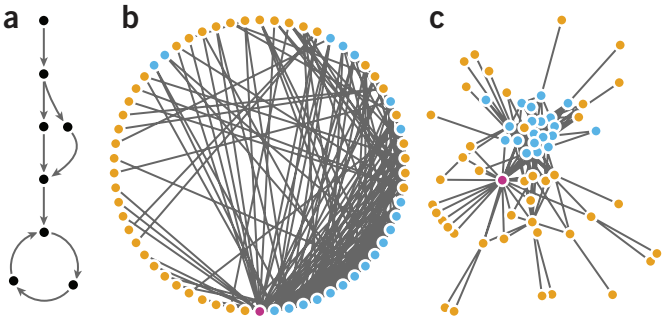

Figure 1 | Node-link diagrams. (**a**) A directed graph typical of a biological pathway. (**b**) An undirected graph with nodes arranged in a circle. (**c**) A spring-embedded layout of data from **b**.

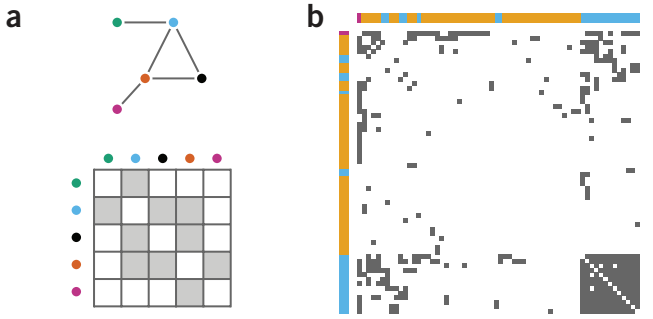

Figure 2 | Adjacency matrices. (**a**) Nodes are ordered as rows and columns; connections are indicated as filled cells. (**b**) A matrix representation of data from **Figure 1b**.

For larger undirected networks, 'adjacency matrices' are a practical solution (**Fig. 2**). In this compressed representation, every node in the network is shown as a row and a column with the order of nodes being the same on both axes. A link between two nodes is indicated by filling the two corresponding cells at the intersections of the nodes (**Fig. 2a**). In this way, adjacency matrices do not suffer from the data occlusions and edge crossings synonymous with node-link diagrams. One drawback, however, is that adjacency matrices make it difficult to understand the relationships between two nodes that are not directly connected.

To maximize the utility of adjacency matrix visualizations, reorder the nodes such that as many filled cells appear next to each other as possible. The result is that clusters are evident as marks near the diagonal and connections 'between' clusters appear as clumps away from the diagonal. Similarly, hubs are seen as rows and columns with many filled cells (**Fig. 2b**).

There may be times when both node-link diagrams and adjacency matrices are inadequate for the size of the network. In these instances, it may be useful to limit the representation to a partial network or rely on relevant statistical measures. For example, a clustering coefficient can be computed that describes the extent of interconnectivity in the neighborhood of a node.

Next month, we will examine another essential plotting technique: heatmaps.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Nils Gehlenborg & Bang Wong

1. Smoot, M. *et al. Bioinformatics* 27, 431–432 (2011).

Nils Gehlenborg is a research associate at Harvard Medical School and the Broad Institute. Bang Wong is the creative director of the Broad Institute and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.