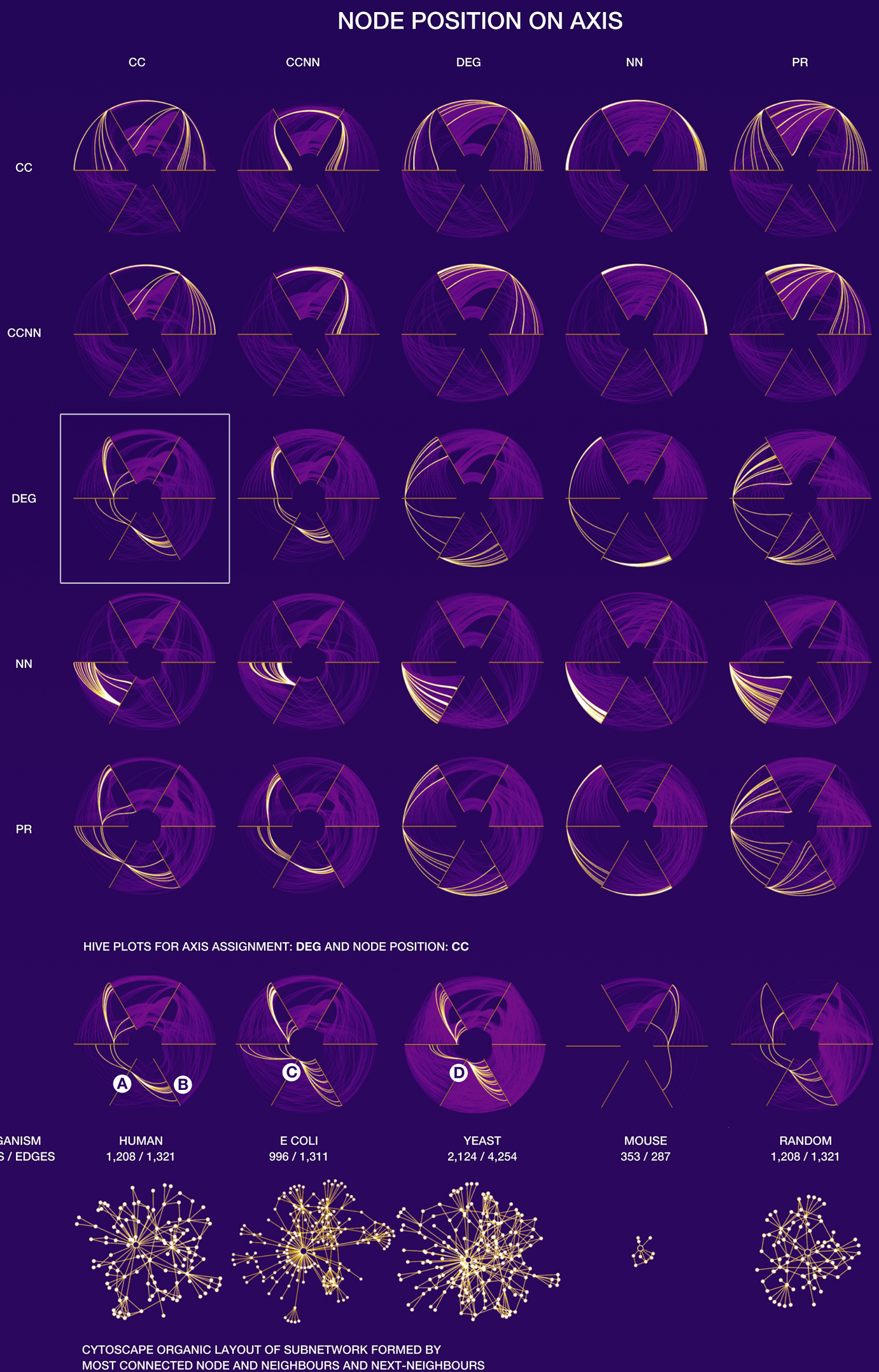
hive panel

HIVE PLOT MATRIX FOR INSIGHT INTO NETWORK STRUCTURE

M KRZYWINSKI, I BIROL, S JONES, M MARRA
CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE

HIVEPLOT.COM

BRITISH COLUMBIA CANCER RESEARCH CENTRE
100-570 W 7th Ave Vancouver BC V5Z 4S6 Canada www.bcgsc.ca



Complex networks are notoriously difficult to visualize because their structure cannot be effectively communicated in a single image [1]. We propose that this can be achieved by multiple and independent visual signatures, each based on a different combination of structural properties. Visualization schemes based on global layout rules (e.g. force-directed) are of limited use for this purpose because they cannot be tuned to be sensitive to patterns in structural attributes, cannot be directly compared, and scale poorly. We present the concept of a *hive panel*, a matrix of hive *plots* [2], each created with different layout rules to specifically interrogate different aspects of network structure.

Hive plots are a rational network visualization, whose layout is based entirely on meaningful network properties, not aesthetics. Hives of two networks are directly comparable and the degree of difference can be assessed. In a hive panel, each plot represents a view of the network based on a pair of structural or annotation node or edge properties of interest.

In the figure [3], the core interactions of the human-human protein interaction database [4] are depicted as a hive panel using the following properties: node clustering coefficient (*cc*), next-neighbor clustering coefficient (*ccnn*), node connectivity (deg), number of next-neighbors (*nn*), and page rank (*pr*). To demonstrate how the panel can focus attention, links to the most connected node are highlighted.

A single hive plot (*deg* vs *cc*) from hive panels of four organisms and a random network are shown below the human panel to demonstrate differences in connectivity and clustering coefficient. Shown also are organic layouts of the locale of the most connected node formed by its neighbors and next-nearest neighbors, the region of the network highlighted in the hive plots. Though it is not possible to confidently conclude anything from the organic layouts, the hive plots clearly communicate differences in a quantitative manner. For example, the most connected node in the human set (Ⓐ) is more cliquey (large *cc*) than E. coli (Ⓒ) and yeast (Ⓓ) and is connected to nodes which themselves are uniformly cliquey (Ⓑ). These and other patterns can be quickly identified within the panel.

[1] Stanford Large Network Dataset Collection cise.ufl.edu/research/sparse/matrices/SNAP
[2] www.hiveplot.com
[3] hiveplot.com/ismb2011/netbiosig/hivepanel-human-dip.png
[4] Xenarios, I. et al. DIP: the database of interacting proteins. Nucleic Acids Res 28, 289-291 (2000).