Network depth: identifying median and contours in complex networks.

G. Bertagnolli,[1,2] C. Agostinelli,[1] and M. De Domenico[2]

[1] *Università degli Studi di Trento, via Sommarive 14, 38123 Povo (TN), Italy*

[2] *Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy*

Node centrality is one of the most fundamental concepts of network science as it quantifies the *importance* of the units inside the system. The characterisation of the micro-scale structure of networks is also important for the understanding of global and dynamical properties of the whole system.

Despite the large number of network centrality measures available nowadays, it is still poorly understood how to identify the node which can be considered as the "centre", or the median, of a complex network. In this work we show how *statistical depth functions* provide the most natural extension of order statistics (median and quantiles) to networks. A *statistical data depth* is a measure of centrality (depth) or "outlyingness" of a sample with respect to its underlying distribution. Since it induces a centre-outward ordering of empirical observations, it enables the generalisation of non-parametric statistical tools (order statistics among others) to multivariate, or even functional, data analysis [1], [2].

Based on the first, famous and most studied *halfspace* Tukey depth, we define a new multivariate depth function, which we call Projected Tukey Depth (PTD) [3]. Thereafter, network embedding allows us to map the problem of finding the centrality of nodes in a network to that of computing the depth of points in space. We show that different metrics (shortest-path and diffusion [4] distance) on networks induce very diverse distributions of points in embedding spaces and consequently, different depth-based node rankings, medians and contours. We also provide a graph-theoretic interpretation of our network depth, based on the framework for centrality measures proposed by Borgatti and Everett in [5]. The depth based in diffusion embeddings $\mathrm{PTD}(G; D_t, t, p)$ is particularly interesting: given that the diffusion distance between two nodes i, j $D_t(i, j; G)$ is small, if there is a large transition probability from i to j along random walks of length at most t, the median node is the one that has the most heterogeneous connectivity in the network. In a social network, the median will be the an Actor, who is well-mixed among all other nodes and does not have a clear, strong belonging to a specific community. In a communication network, the deepest nodes will be those exchanging efficiently information with all other vertices.

The results obtained on the Drosophila's connectome [6] are shown in Fig. 1. The LPUs (nodes) in the depth region of order $\alpha = 0.9$ (multivariate generalisation of quantiles intervals) are biologically relevant. In particular Superpenduncular Protocerebrum (SPP/spp), Dorsomedial Protocerebrum (DMP/dmp) and Inferior Dorsofrontal Protocerebrum (IDFP/idfp) are known to be inter-modular connectors for olfactory, auditory and pre-motor respectively. Ventrolateral Protocerebrum Dorsal part VLP-D/vlp-d, for the right/left visual module have lower depth, maybe because of the high degree of left-rigth separation in the visual module. Shih et al. underline the importance of Subesophageal Ganglion (SOG) in information flow – as an alternative or parallel pathway from sensory input to motor output – and loops formation – essential for recurrent and reverberant information flow.

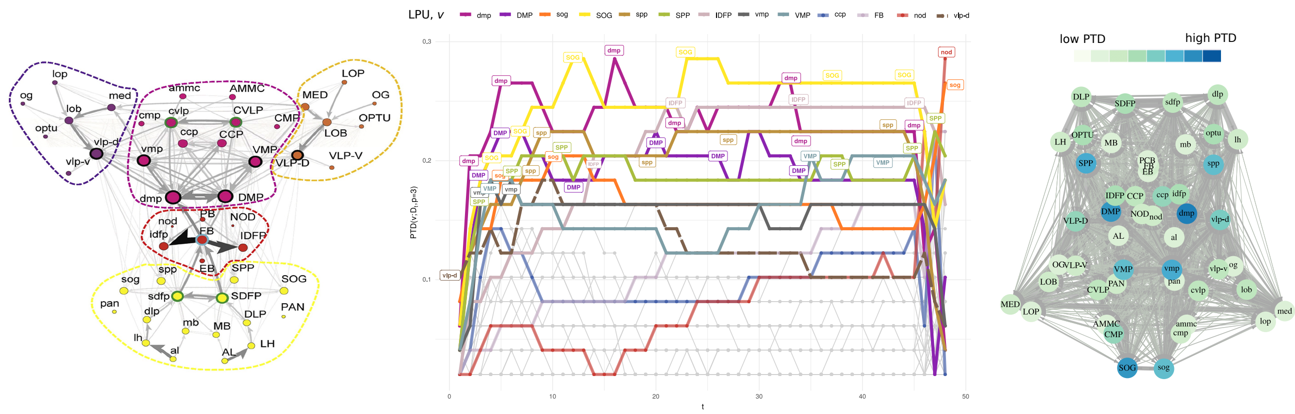

Figure 1: Drosophila's Connectome: small directed and weighted network with 49 nodes, divided in 43 functional units (LPUs) and 6 interconnecting units (LNs). 83% of all possible edges are present. Left: module representation from [6]; centre: depth patterns, labelled lines correspond to LPUs in the 90% depth region; right: the connectome with nodes' colour depending on their depth in the diffusion space with parameters (t=3, p=3).

[1] Y. Zuo and R. Serfling, Annals of Statistics **28**, 461 (2000).

[2] R. Y. Liu, J. M. Parelius, and K. Singh, Annals of Statistics **27**, 783 (1999).

[3] G. Bertagnolli, C. Agostinelli, and M. De Domenico, Journal of Complex Networks (2019), 10.1093/comnet/cnz041.

[4] M. De Domenico, Physical review letters **118**, 168301 (2017).

[5] S. P. Borgatti and M. G. Everett, Social Networks **28**, 466 (2006).

[6] C.-T. Shih and O. e. a. Sporns, Current Biology **25**, 1249 (2015).