

[Diffusion-Based Metrics for Biological Network Analysis](#)

In protein-protein interaction (PPI) networks, or more general protein-protein association networks, functional similarity is often inferred based on the some notion of proximity among proteins in a local neighborhood. In prior work, we have introduced diffusion state distance (DSD), a new metric based on a graph diffusion property, designed to capture more fine-grained notions of similarity from the neighborhood structure that we showed could improve the accuracy of network-based function-prediction algorithms. Boehnlein, Chin, Sinha and Liu have recently shown that a variant of the DSD metric has deep connections to Green's function, the normalized Laplacian, and the heat kernel of the graph.

Because DSD is based on random walks, changing the probabilities of the underlying random walk gives a natural way to incorporate experimental error and noise (allowing us to place confidence weights on edges), incorporate biological knowledge in terms of known biological pathways, or weight subnetwork importance based on tissue-specific expression levels, or known disease processes. Our framework provides a mathematically natural way to integrate heterogeneous network data sources for classical function prediction and disease gene prioritization problems.