**NetworkView: 3D display and analysis of protein-RNA interaction networks**

John Eargle and Zaida Luthey-Schulten*

Department of Chemistry, University of Illinois at Urbana-Champaign, 505 South Mathews Avenue Urbana, IL 61801, USA

Associate Editor: Anna Tramontano

1 INTRODUCTION

The usefulness of 2D contact network visualization in the study of protein function has already been demonstrated (Rahat et al., 2009; Sethi et al., 2009) leading to the development of software to display these networks (Doncheva et al., 2011). The NetworkView extension to the visual molecular dynamics (VMD) program (Humphrey et al., 1996) was developed to help structural biologists study allosteric and molecular signaling through network models of protein-RNA complexes in translation (Black Pyrkosz et al., 2010; Trabuco et al., 2010). Residues essential for molecular recognition and reaction mechanisms at different states of the system can be determined by using molecular dynamics (MD) simulations to calculate variations in contacts and displaying the resulting interaction networks and their properties directly onto the 3D structure (Alexander et al., 2010).

As networks are increasingly used to study the function and dynamics of biological macromolecules, the informative display of network data directly onto 3D representations of biomolecules is needed (Bhattacharyya and Vishveshwara, 2011; Csermely et al., 2010; Daily et al., 2008; Suel et al., 2002).

2 NETWORK ANALYSIS CONCEPTS

Networks are sets of nodes and edges between pairs of these nodes. Nodes here are located at specific atoms and represent amino acids (C$_\alpha$), nucleobases (N1/N9) or the sugar phosphates (P) of RNA. Edges are drawn if the average distance observed in an MD simulation (or contact distance in the X-ray structure) is less than a pre-defined cutoff. Individual edges may have associated weights or lengths based on properties such as correlated motions (see Fig. 1B), energies or physical distance. A path between two nodes is simply a set of nodes and edges connecting one node to the other, and the path length is the sum of weights for edges in the path. If two nodes lie within a connected network, there exists at least one optimal, shortest path between them, and slightly longer paths are referred to as suboptimal (see Fig. 1C).

Major communication pathways are identified by edge betweenness, the count of all pairwise optimal paths that cross a given edge (see Fig. 1F). Using the Girvan–Newman (GN) algorithm (Girvan and Newman, 2002), networks are divided into disjoint subnetworks called communities in which nodes have stronger and more connections to other nodes within the same community than they have to those outside the community. These communities, which can contain both amino acids and nucleotides, are similar to structural domains but are defined by the dynamics of the biomolecules (see Fig. 1E). A more detailed explanation of the network concepts and their associated algorithms as applied to dynamic networks is available in Sethi et al. (2009).

Once the physical network of nodes and weighted edges is generated by the program networkSetup, both optimal and suboptimal paths are calculated by subopt, and the edge betweenness and community structure are determined by gncommunities. NetworkView simplifies the viewing of the above network properties superimposed onto large biomolecular complexes.

3 IMPLEMENTATION

VMD is available for Windows, MacOS X and Linux/UNIX. NetworkView is implemented as a Tcl plugin to VMD, which allows direct access to VMD’s powerful atomselection language and wrapper procedures for the creation and display of OpenGL objects. A graphical user interface with access to the most commonly used features in NetworkView is available through the VMD menu system: Extensions/Analysis/NetworkView.

Three different types of data files can be loaded into NetworkView: network, community and suboptimal path. The
data files are ASCII text and were designed to be human readable. These files can be generated by the programs networkSetup, gncommunities and subopt. NetworkView is not technically tied to these programs, however, so it can be used to visualize and analyze network data generated by other means.

NetworkView provides an application programming interface (API) for selecting nodes and edges and then viewing the values associated with them. Documentation for the NetworkView API is available on the web at http://www.scs.illinois.edu/schulten/software/. A detailed introduction to the use of NetworkView has been provided in a tutorial with data files hosted at the same location.

4 USAGE

The necessary data files can be generated from MD trajectories or from PDB structures. To calculate dynamic networks, a trajectory file containing selected atoms is processed into a correlator from PDB structures. To calculate dynamic networks, a trajectory file containing selected atoms is processed into a correlator from PDB structures. The necessary data files can be generated from MD trajectories. These files can be generated by the programs networkSetup, gncommunities and subopt. NetworkView is not technically tied to these programs, however, so it can be used to visualize and analyze network data generated by other means.

NetworkView provides an application programming interface (API) for selecting nodes and edges and then viewing the values associated with them. Documentation for the NetworkView API is available on the web at http://www.scs.illinois.edu/schulten/software/. A detailed introduction to the use of NetworkView has been provided in a tutorial with data files hosted at the same location.

4 USAGE

The necessary data files can be generated from MD trajectories or from PDB structures. To calculate dynamic networks, a trajectory file containing selected atoms is processed into a correlation matrix made with the MD analysis program Carma (Glykos, 2006) through networkSetup. This matrix is then used to create the local-contact, weighted, adjacency matrix. If no trajectory data are available, structure networks with unweighted edges can also be generated from PDB files. To take full advantage of NetworkView’s capabilities, further processing can be done on the adjacency matrix to calculate the network’s community structure, high-betweenness paths and suboptimal paths between residues known to participate in molecular signaling.

Aside from visualization, NetworkView is also intended for quantitative analysis. Edge weights and other, user-assigned values associated with nodes and edges can be accessed through the API. For example, edge weights along a specified optimal path or at the interface between two molecules can be extracted for statistical analysis. By projecting network information onto the atoms within VMD, features related to active sites, molecular binding interfaces and areas with high sequence or structure conservation (Roberts et al., 2006) can be correlated with the calculated network properties to identify functionally important residues and structures.

5 CONCLUSION

NetworkView provides high-level, interactive views of dynamic and/or structure networks that enable the analysis and interpretation of molecular signaling. Projection of these networks onto 3D biomolecular structures is useful for rapidly identifying residues which can be targeted in mutagenesis experiments to test mechanistic hypotheses.

ACKNOWLEDGEMENTS

The authors thank VMD developers John Stone and Kirby Vandervort for helpful discussions and incorporation of NetworkView into VMD. We would also like to thank Alexis Pyrkosz, Anurag Sethi and Li Li for suggestions and help with testing.

Funding: The National Center for Supercomputing Applications (MCA03T027); the National Institutes of Health (P41-RR05969) and the National Science Foundation (MCB-0844670 and PHY-0822613).

Conflict of Interest: none declared.

REFERENCES