



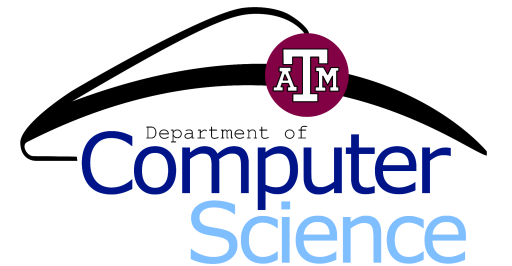
Network Connectivity Analysis on the Temporally Augmented *C. elegans* Web: A Pilot Study

YOONSUCK CHOE, BRUCE H. MCCORMICK, AND WONRYULL KOH

Brain Networks Laboratory, Department of Computer Science, Texas A&M University, College Station, Texas 77843-3112

Email: choe@tamu.edu, {mccormick,wkoh}@cs.tamu.edu; Web: <http://research.cs.tamu.edu/bnl>

Prog# 921.9



1. Introduction

Abstract

We have constructed the *C. elegans* Web (CEW), an XML web-based database of a complete connectivity structure of the nematode *C. elegans*.[†] Each neuron is assigned its own webpage; and hyperlinks between pages represent synapses, designate pre- and postsynaptic neurons, and thus wire up the network. Network analysis tools from the CAIDA tools were used to crawl the CEW to obtain the connectivity graph and visualize its network topology. The advantages of this representation are: (1) it is highly distributed and scalable, (2) hyperlinks naturally represent connection topology, and (3) analysis tools developed for computer network analysis can be used without much modification.

One important property of the nervous system not sufficiently addressed in such hyperlink network is that of conduction delay and neuronal integration time, which adds to it a temporal dimension. We have developed a method of augmenting the connectivity information with temporal information by splitting vertices and edges in the connectivity graph. Such a temporally augmented CEW can again be analyzed with standard network analysis tools.

The *C. elegans* Web serves as a prototype of our similarly constructed web-based storage and retrieval scheme, the Mouse Brain Web (MBW). Here, the MBW representations are virtual views of a common database built from pooling data from multiple mice imaged with typically 5 basic stains. Such a biological database enables us to characterize the mouse brain network by discovering its basic circuits, which model geometrically a significant part of the mouse brain network. In sum, representing neuronal connectivity data in a web-based database provides a natural and efficient way of storing and analyzing connectivity data by leveraging on existing network analysis tools.

Acknowledgments: This research was partly supported by the Department of Computer Science and the Office of the Vice President for Research at Texas A&M University. We would like to thank Ricardo Gutierrez-Osuna, Anand Aluri, J. Perry Evans, Daniel J. Corlette, Matthew W. Sanders, Jason G. Leiloglou, and Maritza Johnson for their contribution. Part of the visualization and analysis was done using tools from CAIDA (www.caida.org) and Mfinder by Uri Alon (Milo et al. 2002b).

[†] The data was kindly provided by Dr. K. Kawamura at Keio University, Japan: Oshio, K., Morita, S., Osana, Y., and Oka, K.; *C. elegans* synaptic connectivity data, Tech. Rep., CCEP, Keio Future No. 1 (1998).

Research Goals and Approach

Web-based database for storage and analysis:

- Enables the use of existing computer network analysis and visualization tools such as CAIDA (www.caida.org).
- Supports multiple views of the same dataset.

Neuronal connectivity analysis:

- Take into account temporal properties of neural circuits by augmenting the connection topology with geometrical information.
- Study scalability issues.

Constructing the *C. elegans* Web

- The *C. elegans* Web (CEW) database was constructed from the complete connectivity data of the *C. elegans* (Oshio et al. 1998) where each neuron was assigned a single web page, and the connectivity was represented as hyperlinks (href in HTML).
- A total of 316 web pages were generated and stored on a single server. The approach can easily be extended to populate multiple web servers.

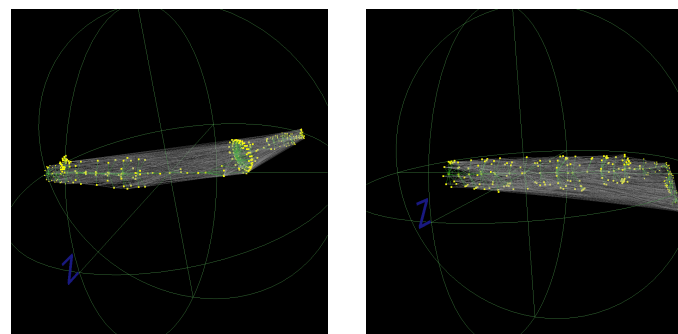
2. *C. elegans* Web

Crawling and Visualization with CAIDA Tools

The web-based storage of the *C. elegans* network enables the use of powerful computer network analysis and visualization tools such as CAIDA (www.caida.org).

- Web crawlers in CAIDA were modified to harvest connectivity information (and the spanning tree) from the CEW.
- The 3D visualization tool Walrus (from the CAIDA tools) was used to visualize the CEW, and the results compared to random networks.

Visualization of the *C. elegans* Network



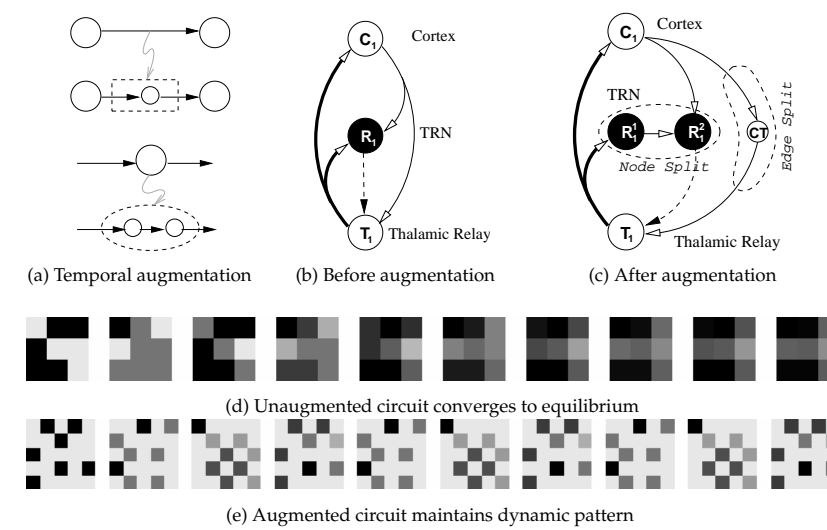
- The spanning tree (thick trunks) covering the entirety of the neurons (squares) of the *C. elegans* network is shown. Faint fibers represent connections.
- (a) A complex structure is revealed, even though the actual neuron placement is somewhat arbitrary (which depends on the particular spanning tree).
- (b) In a random network, the spanning tree and the connectivity structure was more regular, without any noticeable local structural characteristics.

3. Connectivity Analysis

Incorporating Temporal Information

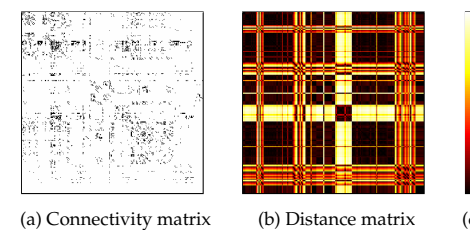
To understand how the structural connectivity determines the function of the biological systems, researchers have studied properties of the connectivity in various systems (Sporns 2002; Sporns and Tononi 2002; Kötter and Stephan 2003; Tononi et al. 1998; Barabási 2002; Milo et al. 2002b). However, the role of temporal delays has not been fully investigated. Here, we investigated how to augment connectivity matrices with temporal information.

Motivation: Temporal Augmentation

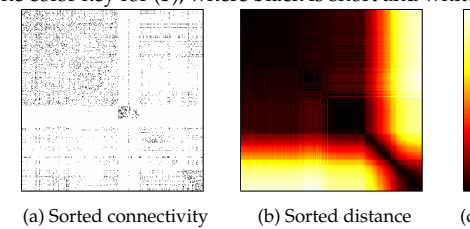


- (a–c) The plain connectivity in a connected graph can be augmented with temporal properties of neurons such as membrane time constant and axonal delay.
- (d–e) The unaugmented and augmented circuits show different higher-order connectivity characteristics (the connectivity matrix (leftmost) and its higher powers).

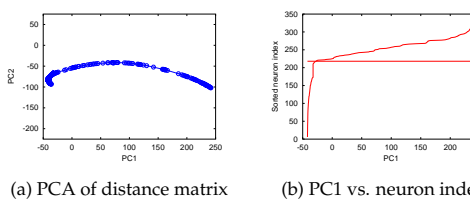
C. elegans Connectivity and Distance Matrices



- **Connectivity and distance matrix:** (a) The neural connectivity in *C. elegans* can be represented as an adjacency matrix (white: no connection; black: connection). (b) Based on the data in White et al. (1986), the (x, y) location of each neuron in the Sagittal plane was calculated, and the distance measured. (c) The color key for (b), where black is short and white is long distance.

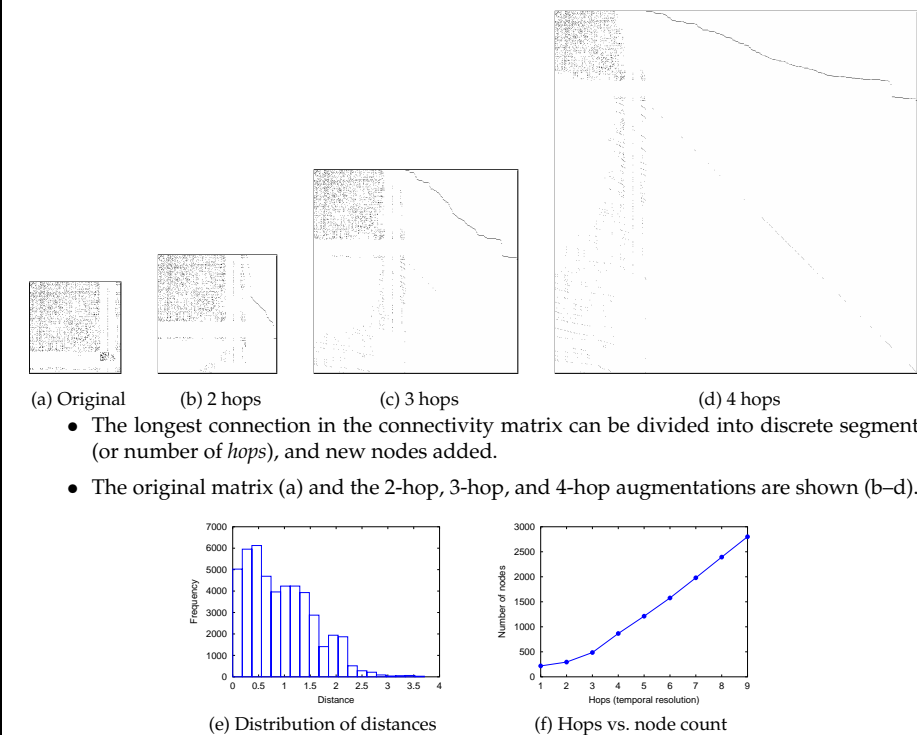


- **Neurons sorted using PCA:** For this work, we assumed that conduction delay is proportional to the distance between the neurons. We used Principal Components Analysis (PCA) to extract 218 neurons within close range in preparation of temporal augmentation. (a) The connectivity matrix after sorting the neuron index using the first principal component projection is shown. (b) The distance matrix sorted with the same order as in (a) is shown. The large black area mostly contains neurons in the head region in the *C. elegans*. (c) The color key for the distance matrix in (b).



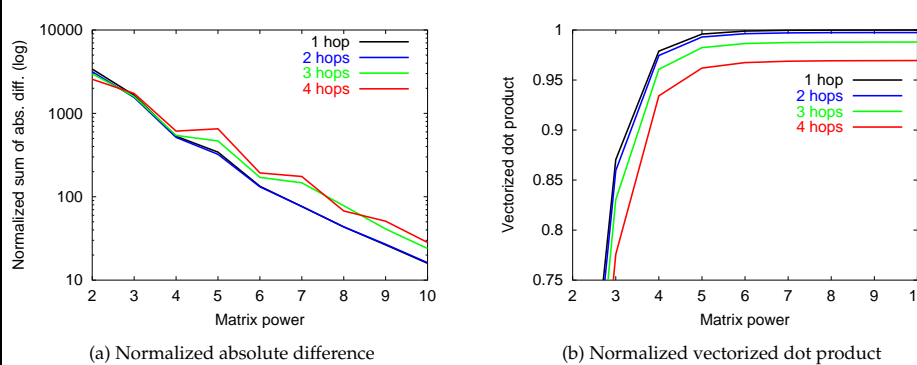
- **PCA projection:** (a) The projection of the distance data on the first- and the second principal component axes is shown. As expected, an elongated profile similar to the *C. elegans* body was obtained. (b) The sorted neuron index is plotted against the projection on the first principal component axis, which shows that most of the neurons are concentrated in the head region.

Temporally Augmenting the *C. elegans* Network



- The longest connection in the connectivity matrix can be divided into discrete segments (or number of hops), and new nodes added.
- The original matrix (a) and the 2-hop, 3-hop, and 4-hop augmentations are shown (b–d).
- (e) The distribution of distances in the original distance matrix gives a clue about the appropriate number of hops.
- (f) The resulting node count as a function of the number of hops is shown. A rough linear relationship is observed.

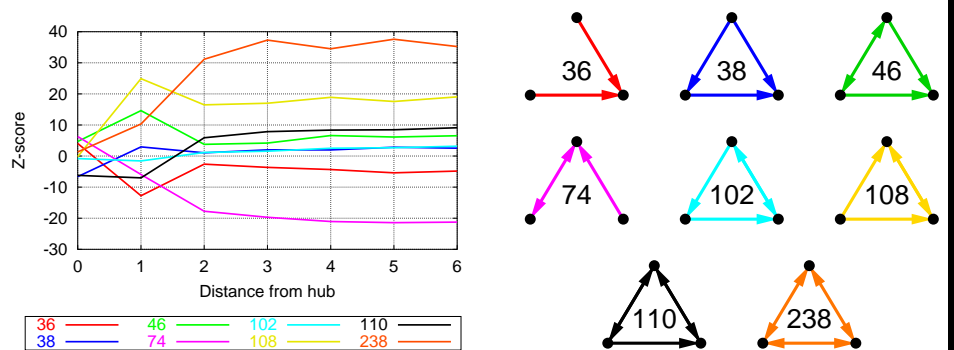
Convergence in Higher-Order Connectivity



- To compare augmented and unaugmented *C. elegans* connectivity matrices at various hops, we used two measures: (a) The sum of normalized absolute difference in successive higher powers of the augmented and unaugmented matrices; and (b) The normalized vectorized dot product of successive higher powers.

- In (a), high values correspond to greater difference, and in (b) to greater similarity.
- For the unaugmented matrix, difference between successive higher powers rapidly decreases. However, for the augmented matrices, the difference is generally greater and has an interesting fluctuation (in a).
- These pilot results suggest that connectivity analysis with temporal augmentation can potentially reveal important functional properties in biological networks.

Scalability



- The final concern, coming back to the CEW, is scalability. The work presented here is a pilot study to develop techniques that can scale up to larger networks, such as the *Mouse Brain Web* (McCormick et al. 2004). For this, we used network motif analysis as a testbed for assessing scalability (Milo et al. 2002a; Reigl et al. 2003).
- We investigated sampling algorithms to extract a subnetwork that has an equivalent motif signature as the full network. First, we obtained a subnetwork consisting of heavily connected hubs only, and then gradually included nodes within a certain radius of the hubs (hub sampling). The figure above shows the results of the motif signature relative to the radius (or distance). It shows that not all nodes are needed to obtain motifs similar to the full network (nodes with distance up to 2 were enough), thus suggesting that hub sampling may be effective in addressing scalability issues.

4. Discussion and Conclusion

In this work, the *C. elegans* Web prototype was developed to demonstrate feasibility in leveraging existing computer network analysis and visualization infrastructure, and to study scalability. Furthermore, a simple method of augmenting conventional connectivity analysis with temporal information was proposed and tested. The pilot results indicated that both directions warrant further investigation.

References

- Barabási, A.-L. (2002). *Linked*. Cambridge, MA: Perseus Publishing.
- Kötter, R., and Stephan, K. E. (2003). Network participation index: Characterizing component roles for information processing in neural networks. *Neural Networks*, 16:1261–1275.
- McCormick, B. H., Choe, Y., Koh, W., Abbott, L. C., Keyser, J., Melek, Z., Doddapaneni, P., and Mayerich, D. (2004). Construction of anatomically correct models of mouse brain networks. *Neurocomputing*, 58: 60:379–386.
- Milo, R., Sen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002a). Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002b). Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827.
- Oshio, K., Morita, S., Osana, Y., and Oka, K. (1998). *C. elegans* synaptic connectivity data. Technical Report CCEP, Keio Future No. 1, Keio University, Japan.
- Reigl, M., Alon, U., and Chklovskii, D. B. (2003). Search for computational modules in the *C. elegans* brain. Unpublished draft.
- Sporns, O. (2002). Graph theory methods for the analysis of neural connectivity patterns. In Kötter, R., editor, *Neuroscience Databases: A Practical Guide*. Boston, MA: Kluwer Publishers.
- Sporns, O., and Tononi, G. (2002). Classes of network connectivity and dynamics. *Complexity*, 7:28–38.
- Tononi, G., McIntosh, A., Russell, D., and Edelman, G. (1998). Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. *Neuroimage*, 7:133–149.
- White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B*, 314:1–340.