# **BOTTLENECK CAPACITY OF RANDOM GRAPHS FOR CONNECTOMICS**

Lav R. Varshney

Coordinated Science Laboratory Department of Electrical and Computer Engineering University of Illinois at Urbana-Champaign Urbana, IL 61801, USA

#### ABSTRACT

With developments in experimental connectomics producing wiring diagrams of many neuronal networks, there is emerging interest in theories to understand the relationship between structure and function. Efficiency of information flow in networks has been proposed as a key functional in characterizing cognition, and we have previously shown that informationtheoretic limits on information flow are predictive of behavioral speed in the nematode Caenorhabditis elegans. In particular, we defined and computed a notion called effective bottleneck capacity that emerged from a pipelining model of information flow. It was unclear, however, whether the particular C. elegans connectome had unique capacity properties or whether similar properties would hold for random networks. Here, we determine the effective bottleneck capacity for several random graph ensembles to understand the range of possible variation and compare to the C. elegans network.

*Index Terms*— connectomics, graph signal processing, information flow, random graphs

## 1. INTRODUCTION

With recent developments in experimental connectomics producing complete anatomical wiring diagrams of brain regions or whole organisms at the individual synapse level [1-4], there is emerging interest in theoretical methods to understand relationships between structure and function in the brain. Indeed, this is the basic problem in connectomics.

One approach to analysis is discovery-oriented, where one would simultaneously screen for several features of the anatomical network that may be associated with cognitive, behavioral, and psychiatric phenotype differences between nervous systems, cf. [5]. That is, one would perform a *connectome-wide association study* (CWAS). Such studies can be done on an edge-by-edge basis over all edges in a network [6], or for functional imaging data on a pixel-bypixel basis [7]. Statistical techniques for CWAS (controlling false discovery rate, etc.), however, have not been developed to simultaneously screen for several interpretable networktheoretic functionals. Backing away from such a discoveryoriented approach, here we are in a sense inspired by graph signal processing [8,9] and use a hypothesis-driven approach to study a particular network functional that may provide insight into neural function.

Neural efficiency hypotheses of intelligence argue that information flows better in the nervous systems of bright individuals [10, 11]. Separately, we have previously argued that information-theoretic capacity limits on information flow are predictive of behavioral speed in the nematode *Caenorhabditis elegans* through a combination of theoretical development [12] and connectome data analysis [2]. This hypothesis can potentially be reexamined with detailed simulation [13, 14].

We specifically considered the bottleneck capacity of the network [15, 16], a notion that emerges from a pipelining model of information flow we defined [12]. Note that due to the difference between the maximum capacity view of information flow [17, 18] and the widest path view of information flow adopted herein, bottleneck capacity is related to notions of graph diameter rather than notions of graph conductance [19]. We computed the capacity of such "circuit-switched" information flow for the *C. elegans* gap junction network<sup>1</sup> and compared to experimental measurements of behavioral speed in the worm, finding concordance.

An alternate question we ask here is whether the bottleneck capacity of the *C. elegans* network is significantly different from random graphs from ensembles that match other network functionals. Is the network non-random [20] in allowing behavior that is faster or slower than other networks? Note that this is different from asking for the capacity of random networks under the other notion of information flow [21, 22].

This paper finds the bottleneck capacity for several random graph ensembles. Our previous investigation showed functional sub-circuits have certain optimality properties for information flow [12], but now we find that the complete con-

This work was supported in part by Systems on Nanoscale Information fabriCs (SONIC), one of the six SRC STARnet Centers, sponsored by MARCO and DARPA.

 $<sup>^{1}</sup>C.$  *elegans* has two kinds of synaptic connections: chemical synapses using neurotransmitters to pass information; and gap junctions where two neurons are directly electrically connected; we focus on gap junctions. Synaptic connection between two neurons may have more than one gap junction.

nectome is much slower than graphs in related random ensembles. This suggests that functional sub-circuits are a primary organizational principle of the nervous system.

#### 2. PIPELINING MODEL OF INFORMATION FLOW AND BOTTLENECK CAPACITY

Let us think of a neuronal network as a communication network where neurons are nodes and synapses are capacitated links. Consider a network where information to be transmitted from a source node to a destination node can be split into pieces in time and sent in a pipelined fashion over (possibly) several hops using as many time slots as needed. The information, however, must go over a single route rather than being split over several routes to be recombined by the destination.

In such a model, maximizing information flow requires finding the single best route between the two nodes: the route that minimizes the weight of the maximum-weight edge in the route. In the context of the *C. elegans* gap junction network this amounts to finding paths with bottlenecks (in terms of inverse number of gap junctions) that are the largest. This can be computed by taking paths in a maximum spanning tree [16], though this may be computationally intensive.

Consider the following standard definitions of graph distance for weighted graphs.

**Definition 1.** Let G = (V, E) be a weighted graph. Then the geodesic distance between nodes  $s, t \in V$  is denoted  $d_G(s, t)$  and is the number of edges connecting s and t in the path with the smallest number of hops between them. If there is no path connecting the two nodes, then  $d_G(s, t) = \infty$ .

**Definition 2.** Let G = (V, E) be a weighted graph. Then the weighted distance between nodes  $s, t \in V$  is denoted  $d_W(s, t)$  and is the total weight of edges connecting s and t in the path with the smallest total weight between them. If there is no path connecting the two nodes, then  $d_W(s, t) = \infty$ .

If all edge weights are 1, then the weighted distance and geodesic distance coincide.

An alternate notion of distance arises from the pipelining model. We want a path between two nodes that has a small number of hops such that the weight of the maximum-weight edge is small. Then we measure path length weighted by this bottleneck weight. Under this notion of distance, the space is ultrametric rather than metric [23].

**Definition 3.** Let G = (V, E) be a weighted graph. Then the bottleneck distance between nodes  $s, t \in V$  is denoted  $d_B(s,t)$  and is the number of edges connecting s and t, scaled by the weight of the maximum-weight edge, in the path with the smallest total scaled weight between them. If there is no path connecting the two nodes, then  $d_W(s,t) = \infty$ .

**Proposition 1** ([12]). *If weights of all actual edges are* 1 *or less, geodesic distance upper bounds the bottleneck distance:* 

$$d_B(s,t) \le d_G(s,t).$$

**Proposition 2** ([12]). Weighted distance lower bounds the bottleneck distance:

$$d_B(s,t) \ge d_W(s,t).$$

Any of these distance functions define a diameter.

**Definition 4.** The graph diameter is

$$D = \max_{s,t \in V} d(s,t).$$

We also define a notion of effective diameter where node pairs that are outliers in the empirical distribution of distances do not enter into the calculation. Recall that the quantile function corresponding to cumulative distribution function (cdf)  $F(\cdot)$  is

$$Q(p) = \inf\{x \in \mathbb{R} | p \le F(x)\}$$
(1)

for a probability value 0 .

**Definition 5.** For a network of size n, let F(x) be the empirical cdf of the distances of all  $\binom{n}{2}$  distinct node pairs. Then the effective diameter is:

$$D_e = Q(0.95).$$
 (2)

This definition is more stringent than other definitions of effective diameter in the literature [24]. Of course,  $D_e \leq D$ . Moving forward, we use effective diameter rather than diameter since it characterizes when most of the information would have reached its destination. Thresholds other than 0.95 can of course be easily defined.

The notion of *bottleneck capacity* that we is proportional to the diameter under bottleneck distance. The constant of proportionality is the Shannon capacity of a single gap junction, about 1000 bits/sec as derived in [12]. We omit this constant in the sequel since we are concerned with structural characterization.

We wish to study whether the bottleneck capacity of the *C. elegans* gap junction network is more than, less than, or similar to the bottleneck capacity of random graphs that have certain other network functionals fixed.

#### 3. MAIN RESULTS

The structural properties of the gap junction connections in the *C. elegans* neuronal network have been described in detail in prior work [2]; there it is shown to be a small-world network due to large clustering coefficients. The somatic network consists of a giant component with 248 neurons, two small connected components, and several isolated neurons. Within the giant component, the average geodesic distance between two neurons is 4.52.

Here we bound the effective diameter of the giant component of the gap junction network, with respect to the bottleneck distance using the effective diameter with respect to geodesic and weighted distances. Figure 1 shows the survival function of the empirical all-pairs geodesic distance and of the empirical all-pairs weighted distance. As can be seen, the effective diameter for weighted distance is about 6 and for geodesic distance is about 7. Hence the effective diameter for bottleneck distance is between 6 and 7.

To evaluate the nonrandomness of the bottleneck capacity of the *C. elegans* network giant component, we compare it with the same quantity expected in random networks. We start with a weighted version of the Erdös-Rényi random network ensemble because it is a basic ensemble. Constructing the topology requires a single parameter, the probability of a connection between two neurons. There are 514 gap junction connections over 279 somatic neurons in *C. elegans*, and so we choose the probability of connection as 0.0133 = $2 \times 514/279/278$ . After fixing the topology, we choose the multiplicity of the connections by sampling randomly according to the *C. elegans* multiplicity distribution [2, Fig. 3(B)], which is well-modeled as a power-law with parameter 2.76. Note that in general the giant component for such a construction will be much larger than that of *C. elegans*.

Figure 2 shows the survival function of the empirical allpairs geodesic distance and of the empirical all-pairs weighted distance of one hundred random networks. A random example is highlighted. As can be observed, the effective diameter for weighted distance is about 5 and for geodesic distance is about 6. Hence the effective diameter for bottleneck distance is between 5 and 6, significantly less than that for the *C. elegans* network.

Now we consider a degree-matched weighted ensemble of random networks. In such a random network, the degree distribution matches the degree distribution of the gap junction network; the degree of a neuron is the number of neurons with which it makes a gap junction. Such a random ensemble is created using a numerical rewiring procedure to generate samples. Upon fixing the topology, the multiplicity of connections is samples as for the Erdös-Rényi ensemble. Note that in general the giant component for such a construction will be much larger than that of *C. elegans*.

Figure 3 shows the survival function of the empirical allpairs geodesic distance and of the empirical all-pairs weighted distance of one hundred random networks. A random example is highlighted. As can be observed, the effective diameter for weighted distance is about 2 and for geodesic distance is about 5. Hence the effective diameter for bottleneck distance is between 2 and 5, quite significantly less than that for the *C. elegans* network.

These results reveal a key nonrandom feature in synaptic connectivity of the *C. elegans* gap junction network, but perhaps contrary to expectation. The network has a nonrandomly *worse* bottleneck capacity compared to basic random graph ensembles. It enables globally *slower* behavioral speed than similar random networks. In contrast, we had previously found that at the micro-level of small functional



**Fig. 1**. Survival function for the empirical all-pairs distance distributions of the *C. elegans* gap junction neuronal network giant component. The weighted distance is listed in terms of inverse gap junctions. The dashed line indicates the 95th percentile used to define effective diameter.



**Fig. 2.** Survival function for the empirical all-pairs distance distributions of 100 Erdös-Rényi random network giant components; a random example is highlighted in red. The weighted distance is listed in terms of inverse gap junctions. The dashed line indicates the 95th percentile used to define effective diameter.



**Fig. 3.** Survival function for the empirical all-pairs distance distributions of 100 degree-matched random network giant components; a random example is highlighted in red. The weighted distance is listed in terms of inverse gap junctions. The dashed line indicates the 95th percentile used to define effective diameter.

sub-circuits, the *C. elegans* gap junction network has several hub-and-spoke structures [25, 26], which are actually optimal from an information flow perspective [12]. Thus, these results lend some greater nuance to efficient flow hypotheses in neuroscience.

## 4. CONCLUSION

Our previous results had shown that the behavioral speed of the nematode worm *C. elegans* is close to the lower bound obtained from optimal information flow speed, as constrained by the noise, signaling constraints, and topology of the network [12]. Here we considered the possibility of changing the network topology itself and discovered that the network has much lower bottleneck capacity than similar random graphs (whether Erdös-Rényi or degree-matched). The network does not seem to be optimized for global information flow.

On the other hand, we had also noted the prominence of hub-and-spoke functional sub-circuits in the *C. elegans* gap junction network [25, 26] and proved their optimality properties for information flow under number of gap junctions constraints [12].

In terms of neural organization, this suggests that smaller sub-circuits within the larger neuronal network are responsible for specific functional reactions, and these should have fast information flow (to quickly achieve the computational objective of that circuit, such as chemotaxis). Behavioral speed of the global network may not be biologically relevant.

With the growing prevalence of theoretical connectomics, hypotheses of neural organization according to functional sub-circuits with large information flow capacity within them should be investigated and further screened for to understand phenotype differences among individuals.

Within the field of signal processing itself, it is of interest to study the bottleneck capacity of other kinds of networks, whether random ensembles—such as Watts-Strogatz small worlds, Barabási-Albert scale-free networks, Kronecker random graphs, or random geometric graphs—or networks that arise in other application domains.

### 5. ACKNOWLEDGMENT

Thanks to Devavrat Shah, Dmitri B. Chklovskii, Sanjoy K. Mitter, and Aki Nikolaidis for inspiring discussions. Thanks to Tongtong Li for the impetus.

#### 6. REFERENCES

- J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the nervous system of the nematode *Caenorhabditis elegans*," *Proc.-R. Soc. Lond., Biol. Sci.*, vol. 314, no. 1165, pp. 1–340, Nov. 1986.
- [2] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii, "Structural properties of the *Caenorhabditis elegans* neuronal network," *PLoS Comput. Biol.*, vol. 7, no. 2, p. e1001066, Feb. 2011.
- [3] M. Helmstaedter, K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk, "Connectomic reconstruction of the inner plexiform layer in the mouse retina," *Nature*, vol. 500, no. 7461, pp. 168–174, Aug. 2013.
- [4] S. ya Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, T. Zhao, J. A. Horne, R. D. Fetter, S. Takemura, K. Blazek, L.-A. Chang, O. Ogundeyi, M. A. Saunders, V. Shapiro, C. Sigmund, G. M. Rubin, L. K. Scheffer, I. A. Meinertzhagen, and D. B. Chklovskii, "A visual motion detection circuit suggested by *Drosophila* connectomics," *Nature*, vol. 500, no. 7461, pp. 175–181, Aug. 2013.
- [5] D. J. Bumbarger, M. Riebesell, C. Rödelsperger, and R. J. Sommer, "System-wide rewiring underlies behavioral differences in predatory and bacterial-feeding nematodes," *Cell*, vol. 152, no. 1-2, pp. 109–119, Jan. 2013.
- [6] A. Zalesky, A. Fornito, and E. T. Bullmore, "Networkbased statistic: Identifying differences in brain networks," *NeuroImage*, vol. 53, no. 4, pp. 1197–1207, Dec. 2010.

- [7] Z. Shehzad, C. Kelly, P. T. Reiss, R. C. Craddock, J. W. Emerson, K. McMahon, D. A. Copland, F. X. Castellanos, and M. P. Milham, "A multivariate distance-based analytic framework for connectome-wide association studies," *NeuroImage*, vol. 93, no. 1, pp. 74–94, Jun. 2014.
- [8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [9] S. Segarra, W. Huang, and A. Ribeiro, "Diffusion and superposition distances for signals supported on networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 1, pp. 20–32, Mar. 2015.
- [10] I. J. Deary, L. Penke, and W. Johnson, "The neuroscience of human intelligence differences," *Nat. Rev. Neurosci.*, vol. 11, no. 3, pp. 201–211, Mar. 2010.
- [11] T.-W. Lee, Y.-T. Wu, Y. W.-Y. Yu, H.-C. Wu, and T.-J. Chen, "A smarter brain is associated with stronger neural interaction in healthy young females: A resting EEG coherence study," *Intelligence*, vol. 40, no. 1, pp. 38–48, Jan.-Feb. 2012.
- [12] L. R. Varshney and D. Shah, "Informational limits of neural circuits," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2011, pp. 1757–1763.
- [13] C. Rockland, "The nematode as a model complex system: A program of research," M.I.T. Laboratory for Information and Decision Systems, Working Paper WP-1865, Apr. 1989.
- [14] B. Szigeti, P. Gleeson, M. Vella, S. Khayrulin, A. Palyanov, J. Hokanson, M. Currie, M. Cantarelli, G. Idili, and S. Larson, "OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*," *Front. Comput. Neurosci.*, vol. 8, p. 137, Nov. 2014.
- [15] M. Pollack, "The maximum capacity through a network," *Oper. Res.*, vol. 8, no. 5, pp. 733–736, Sept.-Oct. 1960.
- [16] T. C. Hu, "The maximum capacity route problem," Oper. Res., vol. 9, no. 6, pp. 898–900, Nov.-Dec. 1961.
- [17] L. R. Ford, Jr. and D. R. Fulkerson, "Maximal flow through a network," *Can. J. Math.*, vol. 8, pp. 399–404, 1956.
- [18] P. Elias, A. Feinstein, and C. E. Shannon, "A note on the maximum flow through a network," *IRE Trans. Inf. Theory*, vol. IT-2, no. 4, pp. 117–119, Dec. 1956.

- [19] O. Ayaso, D. Shah, and M. A. Dahleh, "Information theoretic bounds for distributed computation over networks of point-to-point channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6020–6039, Dec. 2010.
- [20] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, "Highly nonrandom features of synaptic connectivity in local cortical circuits," *PLoS Biol.*, vol. 3, no. 3, pp. 0507–0519, Mar. 2005.
- [21] A. Ramamoorthy, J. Shi, and R. D. Wesel, "On the capacity of network coding for random networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2878–2885, Aug. 2005.
- [22] R. A. Costa and J. Barros, "On the capacity of smallworld networks," in *Proc. IEEE Inf. Theory Workshop* (*ITW*'06), Mar. 2006, pp. 302–306.
- [23] R. Rammal, G. Toulouse, and M. A. Virasoro, "Ultrametricity for physicists," *Rev. Mod. Phys.*, vol. 58, no. 3, pp. 765–788, July-Sept. 1986.
- [24] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," ACM *Trans. Knowl. Discovery Data*, vol. 1, no. 1, p. 2, Mar. 2007.
- [25] E. Z. Macosko, N. Pokala, E. H. Feinberg, S. H. Chalasani, R. A. Butcher, J. Clardy, and C. I. Bargmann, "A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*," *Nature*, vol. 458, no. 7242, pp. 1171–1176, Apr. 2009.
- [26] I. Rabinowitch, M. Chatzigeorgiou, and W. R. Schafer, "A gap junction circuit enhances processing of coincident mechanosensory inputs," *Curr. Biol.*, vol. 23, no. 11, pp. 963–967, Jun. 2013.