

**Neuron, Volume 105**

**Supplemental Information**

**A Genetic Model of the Connectome**

**Dániel L. Barabási and Albert-László Barabási**

## Supplementary Information for A Genetic Model of the Connectome

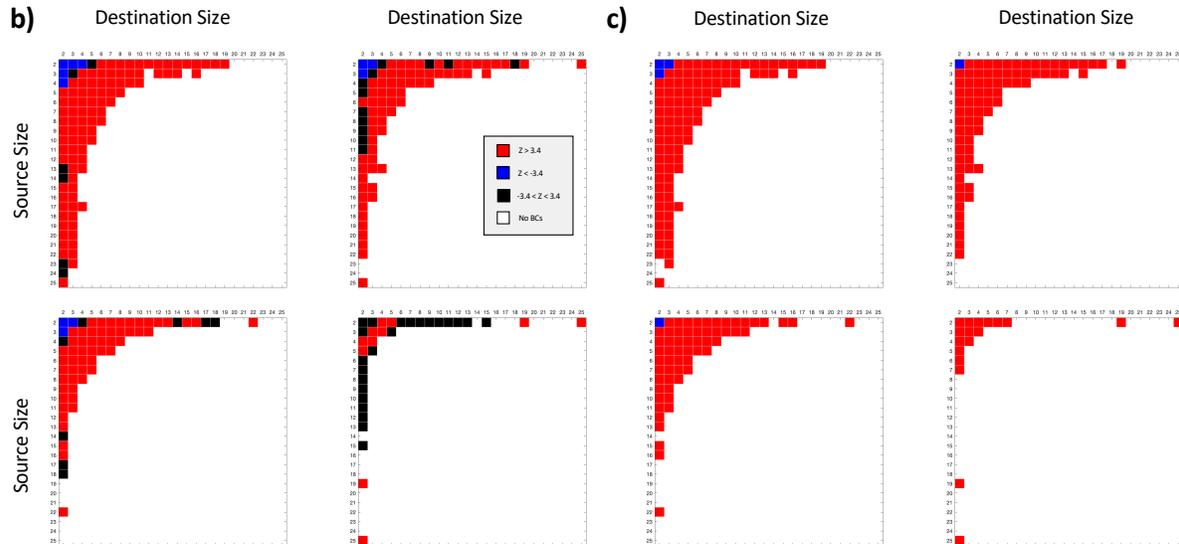
Dániel L. Barabási and Albert-László Barabási

Organism	Neurons	Synapses	TF	$b = \lceil \log_2(N) \rceil$	Data Source
<i>C. elegans</i>	302	6398	934	9	(White <i>et al.</i> , 1986; Reece-Hoyes <i>et al.</i> , 2005, 2011; Varshney <i>et al.</i> , 2011)
<i>Fruit Fly</i>	100,000	$10^7$	627	17	(Lagercrantz <i>et al.</i> , 2010; Zhang <i>et al.</i> , 2011; Zheng <i>et al.</i> , 2018)
<i>Mouse</i>	$7.09 \cdot 10^6$	$1.28 \cdot 10^{11}$	1,457	23	(Ananthanarayanan <i>et al.</i> , 2009; Zhang <i>et al.</i> , 2011)
<i>Rat</i>	$2 \cdot 10^8$	$4.48 \cdot 10^{11}$	1,371	28	(Herculano-Houzel and Lent, 2005; Ananthanarayanan <i>et al.</i> , 2009; Zhang <i>et al.</i> , 2011)
<i>Cat</i>	$7.63 \cdot 10^8$	$6.1 \cdot 10^{12}$	887	30	(Ananthanarayanan <i>et al.</i> , 2009; Zhang <i>et al.</i> , 2011)
<i>Human</i>	$8.1 \cdot 10^9$	$1.64 \cdot 10^{14}$	1,391	33	(Tang <i>et al.</i> , 2001; Azevedo <i>et al.</i> , 2009; Vaquerizas <i>et al.</i> , 2009)

**Supplementary Table 1: Neurons, synapses, and transcription factors. (Related to Results: "Encoding Neuronal Identity" and Star Methods: "Brain Sizes Across Organisms")** We compiled from the literature the number of neurons, synapses and transcription factors for various organisms. For each organism, we also show  $b = \log_2(N)$ , representing the number of TFs minimally required to offer a unique identity to all neurons in a brain. Notice that the number of TFs in each organisms exceeds  $b$ , indicating that TF combinations can reasonably offer unique cellular identity to each neuron.

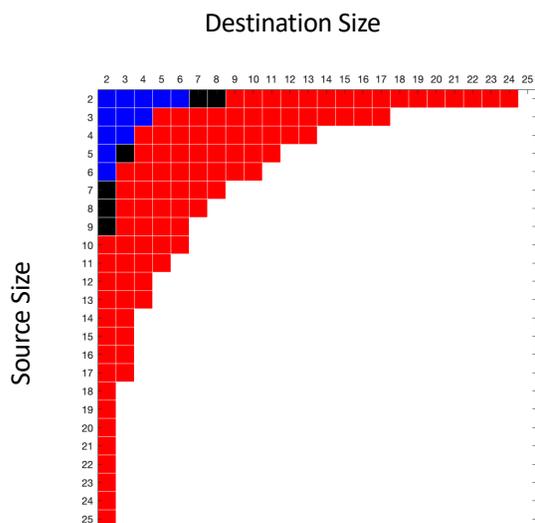
a)

		<k>	Bicliques		
			Connectome (number)	Random (number)	Z-Score
Chemical Synapses	Varshney et. al.	7.86	2,968	2,723.6±44.0	5.5573
	Cook et. al.	12.6	9,430	7,569.4±92.3	20.149
Gap Junctions	Varshney et. al.	3.70	344	314.34±8.90	3.334
	Cook et. al.	7.49	1,706	1,059.5±82.01	7.883

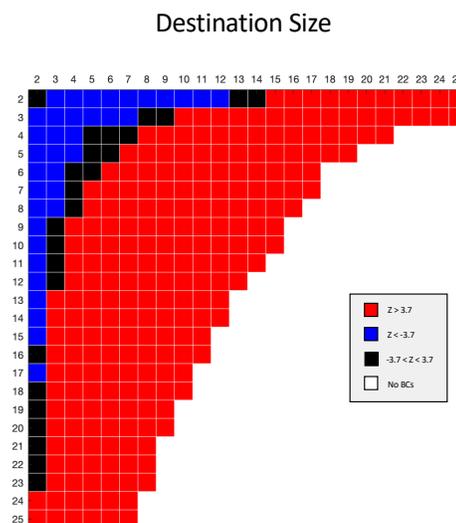


**Supplementary Figure 1: Bicliques in *C. elegans* connectome reconstructions. (Related to Star Methods: "Quantification and Statistical Analysis")** a) Biclique Numbers in Each Studied Connectome. The entries show the number of maximal bicliques found in the real (Connectome) and randomized (Random) networks, together with the overall Z-score. b) Biclique Size Distributions for Connectomes under degree preserving randomizations. Z-scores of maximal biclique sizes under degree-preserving randomization. Blue squares show maximal bicliques that are underrepresented in the real connectome compared to the random reference ( $Z < -3.4$ ) — they capture small maximal bicliques ( $2 \rightarrow n$ , or  $n \rightarrow 2$ ) that emerge frequently by chance. Red squares capture maximal bicliques that are overrepresented in the real data ( $Z > 3.4$ ). Black maximal bicliques exist but their numbers are non-significant ( $-3.4 < Z < 3.4$ ). White region corresponds to maximal bicliques that are absent in the connectome. Significance was set at  $Z = 3.4$  to correct for multiple testing for each of the biclique types, with the most stringent cutoff used for all datasets for consistency. The higher density of Cook datasets over the Varshney reconstruction is apparent in the larger maximal bicliques found, as well as the reduced number of non-significant maximal bicliques. The gap junction matrix is less sparse, however the larger maximal bicliques are more significant, as expected. c) Biclique Sizes Under Erdős-Rényi (ER) Distributions. Z-scores of maximal biclique sizes compared to ER random networks with matching density and node number. Given the low structure of ER random networks, the increased significance of maximal biclique sizes should be expected.

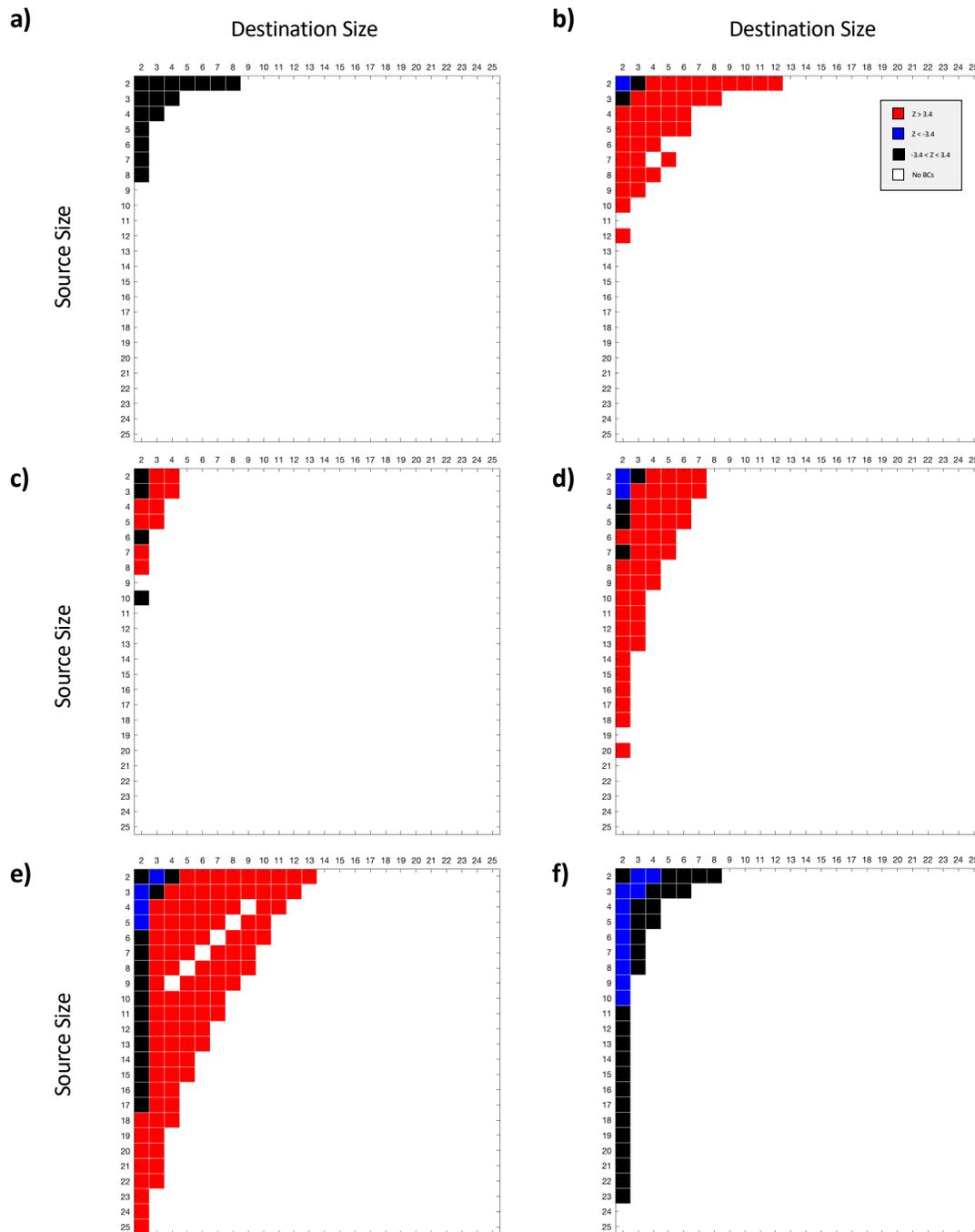
a)



b)



**Supplementary Figure 2: Bicliques in connectomes of other organisms. (Related to Star Methods: "Biclique Motifs in Other Organisms")** (a) Biclique sizes under degree-preserving randomization for *Ciona instestinalis*. Blue square is z-score less than -3.7, red square is z-score greater than 3.7, black indicates non-significant z-score ( $-3.7 < Z < 3.7$ ), while white indicates that no maximal bicliques of the given size were found in the chemical connectome. The significant z-score cutoff was set at 3.7 to account for multiple testing corrections, with a more stringent threshold used for both connectomes in this figure based on the many types of bicliques found in the Drosophila connectome. (b) Biclique sizes under degree-preserving randomization for olfactory subcircuit of Drosophila larvae.



**Supplementary Figure 3: Bicliques in network models. (Related to Results: "Bicliques in *C. elegans*" and Star Methods: "Bicliques in Network Models")** (a) Bicliques in Erdős-Rényi model. We generated a random graph of 279 nodes and 4.5% density, to match the size and average degree of the chemical connectome of *C. elegans*. As expected, we observe no significant maximal bicliques compared to degree preserved randomizations. Z-score significance cutoffs for all of Supplementary Figure 3 were set to a more stringent 3.4 standard deviations to match the thresholds in Figure 2 and Supplementary Figure 1, even though many fewer bicliques types were found in all plots of Supplementary Figure 3. (b) EDR model of *C. elegans* connectome. An exponential distance rule

network fit to the *C. elegans* connectome with  $\lambda = 10.88$ , compared to degree preserved randomizations. **(c-e)** Networks generated using the scale-free model with  $N = 279$  to match the size of the *C. elegans* connectome. The different panels correspond to different densities, generated with **(c)**  $m = 5$ , **(d)**, 8, and **(e)** 13, demonstrating the non-stationary nature of the resulting maximal bicliques. **(f)** Coreless Scale Free Network. We removed the first  $m = 13$  nodes (core) from the network profiled in (e). The resulting network has fewer large maximal bicliques, and all maximal bicliques are statistically underrepresented or nonsignificant.